ISSN 0394-1914

HYSTRIX
the Italian Journal of Mammalogy

Volume 24(1) • 2013

published by
Associazione Teriologica Italiana

HYSTRIX
the Italian Journal of Mammalogy
Volume 24(1) • 2013
Edited and published by Associazione Teriologica Italiana

Impact Factor (2012) 0.352

***HYSTRIX*, the Italian Journal of Mammalogy** is an Open Access Journal published twice per year (one volume, consisting of two issues) by Associazione Teriologica Italiana. Printed copies of the journal are sent free of charge to members of the Association who have paid the yearly subscription fee of 30 €. Single issues can be purchased by members at 35 €. All payments must be made to Associazione Teriologica Italiana onlus by bank transfer on c/c n. 54471, Cassa Rurale ed Artigiana di Cantù, Italy, banking coordinates IBAN: IT13I0843051080000000054471.

Associazione Teriologica Italiana secretariat can be contacted at segreteria.atit@gmail.com

Information about this journal can be accessed at http://www.italian-journal-of-mammalogy.it

The Editorial Office can be contacted at info@italian-journal-of-mammalogy.it

ISSN 0394-1914

HYSTRIX
the Italian Journal of Mammalogy

Volume 24(1) • 2013

Virtual Morphology and Evolutionary Morphometrics in the new millenium.

edited by
Andrea Cardini and Anna Loy

published by
Associazione Teriologica Italiana

Publication information: Hystrix, the Italian Journal of Mammalogy is published as a printed edition (ISSN 0394-1914) twice per year. A single copy of the printed edition is sent to all members of Associazione Teriologica Italiana. The electronic edition (ISSN 1825-5272), in Adobe® Acrobat® format is published "online first" on the Journal web site (http://italian-journal-of-mammalogy.it). Articles accepted for publication will be available in electronic format prior to the printed edition, for a prompt access to the latest peer-reviewed research.

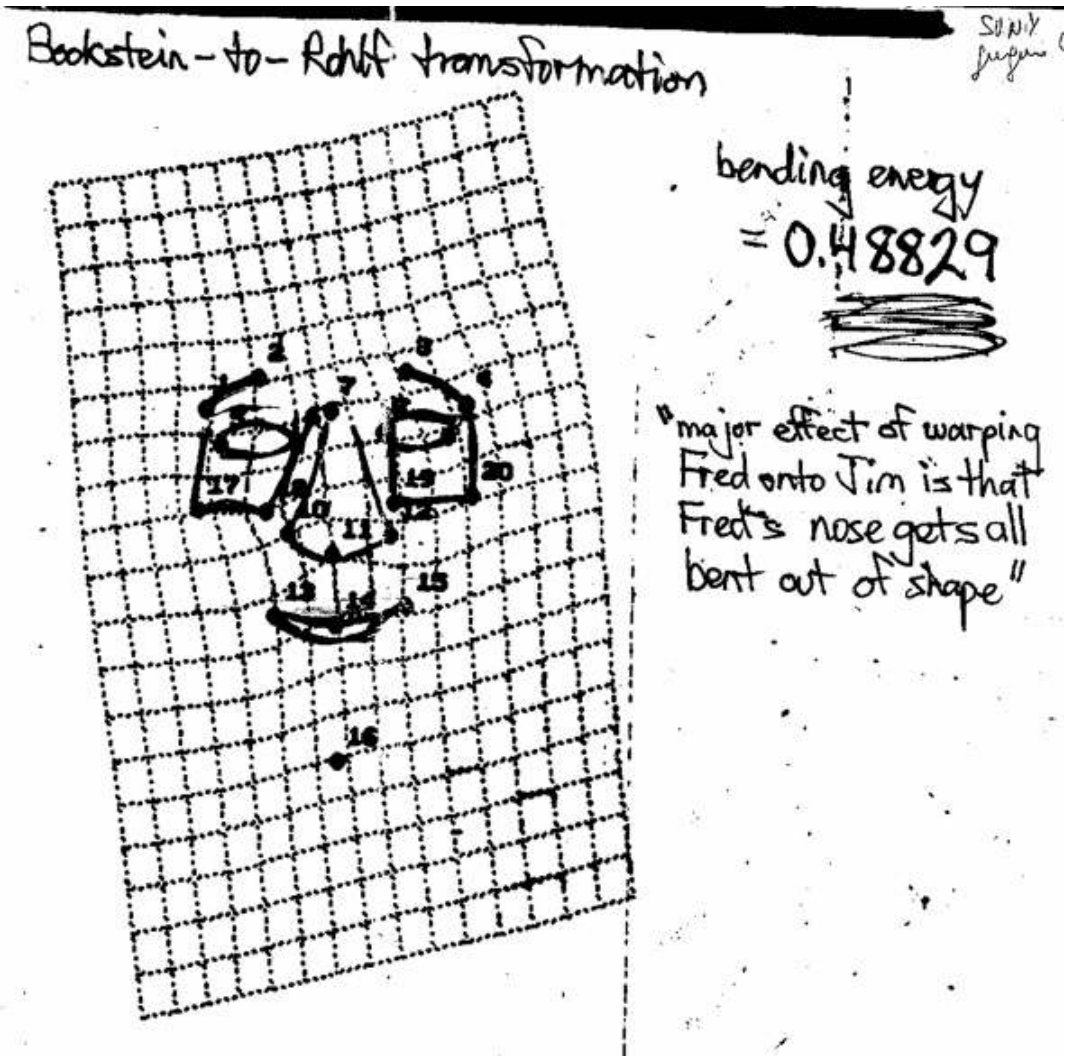
Best Paper Award

Associazione Teriologica Italiana established a Best Paper Award for young researchers. Eligible researchers are leading authors less than 35 years old, and within 7 yers from their PhD (but young researcher at an even earlier stage of their career, i.e. without a PhD, are also eligible), who have expressed interest in the award in the Communications to the Editor (step 1 of the online submission procedure; for details, see the Electronic Publication Guide; http://www.italian-journal-of-mammalogy.it/public/journals/3/authguide.pdf).

If the eligible leading researcher is not the corresponding author, the latter should express interest on the leading researcher's behalf. Criteria are innovation, excellence and impact on the scientific community (e.g., number of citations).

The award will be assegned yearly, in the second semester of the year following that of reference (i.e., Best Paper Award for 2013 will be assigned in the second semester of 2014). The Editorial Commitee is responsible to assign the award. A written motivation will be made public on the journal website.

Finito di stampare nel mese di luglio 2013 - Typeset in LaTeX
Stampa: Edizioni Belvedere, via Adige, 45 – 04100 Latina (Italia)



1990 – Stony Brook Morphometric Workshop at SUNY: first thin plate splines produced by the attendants (scanned by Anna Loy).

This and other recollections from the "early days" of GM in biology will appear throughout the book.

LATE NIGHT WITH JIM ROHLF'S

TOP TEN QUOTATIONS FROM THE 1990 MORPHOMETRIC WORKSHOP.

Fred Bookstein: Its in my book, turn to figure X.X.X.

Les Marcus: I'm not getting this across very well but I can write you a program to do it in SAS.

Neal Oden: Just do some kind of test that tells you your data is junk.

Cliff Lemen: I'd like to do some mitocrobial genetics on dogs.

Les Marcus: Plot the hell out of your variables.

Jim Rohlf: If it tells you want you want then thats the method you should use.

Maria Teresa: Leaf landmarks? (scratch your head).

Jorge Santiago-Blay: I guess I shouldn't have labelled the red oak with the green pen.

Maria Teresa: I've been eating morphometrics for 2 weeks and now, like Steve's salamanders, I can't swallow it.

Cliff Lemen: You mean you don't disagree with yourself Fred?

Steve Reilly: Fred? Could you please describe every point and every line on figurte 7.2.6?

Gene Albrecht: For most people, the methods we are learning are not practical: we are seeing that here.

Junhyong Kim: Because Fred programs in FORTRAN I had to do some retroprogramming.

Don Miles: I'm interested in the dark side of distances.

Don Straney: You can be compulsive or you can be productive.

Dennis Slice: My first announcement is that this is the first announcement I have ever made.

The 1990 preferred new morphometric method: the principal resistant-fit canonical acetate-plate generalized finite fourier spline.

Antonio Valdecasas : please no questions, only answers.

1990 – Stony Brook Morphometric Workshop at SUNY: memorable statements (scanned by Anna Loy).

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 1–5, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/8749/pdf
doi:10.4404/hystrix-24.1-8749

Commentary

On growth and form in the "computer era": from geometric to biological morphometrics

Andrea Cardini[a,b,c,*], Anna Loy[d]

[a] *Dipartimento di Scienze Chimiche e Geologiche, Università di Modena e Reggio Emilia, l.go S. Eufemia 19, 41121 Modena, Italy*
[b] *Centre for Anatomical and Human Sciences, Hull York Medical School, University of Hull, Cottingham Road, Hull, HU6 7RX, UK*
[c] *Centre for Forensic Science, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia*
[d] *Dipartimento Bioscienze e Territorio, Università del Molise, I-86090 Pesche, Italy*

Keywords:
evolutionary biology
mammals
multivariate statistics
morphometric revolution
semilandmarks
shape analysis
comparative method
phenotype

Article history:
Received: 28 January 2013
Accepted: 15 May 2013

Acknowledgements
Publishing this edited volume would have never been possible without the enthusiastic cooperation of all contributors, as well as the reviewers and the editorial staff of Hystrix. We are deeply grateful to all of them. A special thank to Damiano Preatoni (University of Insubria), for being the most supportive Technical Editor one might have wished. We are also deeply grateful to Dean Adams (Iowa State University) for his insightful comments on a previous version of this introduction to the "Yellow Book", which helped us to greatly improve it. A.C. also thanks Adriano Martinoli (University of Insubria) and the Hystrix Editorial Board for inviting him to join the board in 2012. A.C. is also greatly in debt with the "Roman and Viennese schools of morphometrics" for their crucial help in his early days as an amateur morphometrician; he also acknowledges the support of the Durham University Department of Anthropology, and the Durham International Fellowships for Research and Enterprise (DIFeREns) co-funded by Durham University and the European Union Funding.

Both A.C. and A.L. contributed equally to this paper and to the editing of the "Yellow Book".

Abstract

Almost 100 years after the publication of Thompson's seminal book "On growth and form", the study of animal morphology is becoming again central to biology. This is also thanks to the development of powerful computerized quantitative methods for statistical shape analysis, collectively known as Geometric Morphometrics (GM). GM was announced as a revolution just two decades ago. The "revolution" is now a standard tool in numerical analyses of phenotypic variation in mammals and other organisms. Hundreds of studies are published every year that take advantage of GM (e.g., more than 800 entries in Google Scholar only for 2012). We celebrate the 20th anniversary of the "revolution in morphometrics" (Rohlf and Marcus 1993, p. 129) with the publication of a "Yellow Book", a special issue of Hystrix dedicated to Evolutionary Morphometrics and Virtual Morphology. A series of 14 papers by leading morphometricians summarizes the main achievements in GM (surface methods, comparative methods in shape analysis, phenotypic trajectories quantification, modularity/integration, the use of R in morphometrics), describes its most innovative developments (ecometrics, eigensound analysis, biomechanical GM), and discusses common misunderstandings of well established methods (visualization of shape differences). Besides celebrating the success of statistical shape analysis in biology, this issue aims at introducing to GM readers unfamiliar with or intimidated by its strong numerical background. This is why, as Editors, we asked all contributors to provide concise and accurate but also clear and simple descriptions of techniques and applications. We hope that we succeeded in this aim, and wish that this Yellow Book may help to tighten the connection between biologists and statisticians for a truly "biological" GM.

It is not accidental that Hystrix, the Italian Journal of Mammalogy, strongly supported and solicited the publication of this volume on the advances in Geometric Morphometrics (GM – Adams et al. 2004 and Adams et al. this issue) in the last 20 years. Italian theriologists have had a long and strong connection with the "dream team" that led the morphometric revolution in the '90s (Rohlf and Marcus, 1993; Corti, 1993). They have also been among the most enthusiastic applicants of the new methods since the early time of the morphometric synthesis (Adams et al., this issue). It is indeed thanks to the fruitful interactions between theoreticians and biologists in Italy and elsewhere that modern morphometrics has become deeply rooted in biology. Even the somewhat old fashioned use of morphology in taxonomic assessment has been revitalized by GM: famously, a study of cranial variation in Old World moles (Rohlf et al., 1996) has become a standard reference in the field (Rohlf, 1998; Viscosi and Cardini, 2011), obtaining almost 200 citations (Google Scholar, January 2013) and becoming a best known early example of the impact that can be achieved when statisticians and morphologists join their forces.

GM is "a collection of approaches for the multivariate statistical analysis of Cartesian coordinate data, usually (but not always) limited to landmark point locations" (http://life.bio.sunysb.edu/morph/glossary/glossl.html). Greatly inspired by Thompson's "On growth and form" (1917), GM puts together geometry, multivariate morphometrics, computer science and three dimensional imaging techniques for a powerful and accurate study of organismal forms. Today, it is a leading family of methods in quantitative biology and the type of computerized image analysis which is likely to generate the main source of data and analytical tools in the emerging field of phenomics, the comprehensive study of phenotypic variation and the latest of the "–omics" after genomics, proteomics, metabolomics and all the other –omics (Houle, 2010).

Despite its modern success, however, not that many scientists seemed to have grasped the potential of the new methods in the early days of GM. Especially in Italy, very few biologists were brave enough to venture into its statistical meanders. Marco Corti and one of us (A.L.), theriologists from the University of Rome La Sapienza, strongly believed in the "revolution" and struggled to make it spread across Italy. After a first ground breaking meeting two years before in Ann Arbor (Michigan, USA), they were the only Italian participants to the second GM workshop in 1990 (Stony Brook, USA – Fig. 1). There, Marco and Anna met some of the American "giants" of the methodological development of modern morphometrics: James Rohlf, Fred Bookstein, Dennis Slice and Leslie Marcus. This encounter marked the beginning a long human and professional companionship, as well as of a series

*Corresponding author
Email address: alcardini@gmail.com (Andrea Cardini)

of memorable methodological discussions, which led to common projects, joint papers, and fundamental advancements.

Indeed, the "partnership" between morphometrics and Italian theriology has a long history (Corti et al., 2000), and one that continues today. The fifth international Theriological Congress (Marcus and Corti, 1989) hosted what was likely to have been the last workshop (Marcus and Corti, 1989) on traditional morphometrics (Marcus, 1990). Less than 10 years after, another theriological congress, the Euro-American Mammal Congress (Santiago de Compostela, Spain, 1998), was home to another morphometric workshop, but, this time, one entirely devoted to the use of GM in mammalogy (Corti et al., 2000). This workshop led to an edited volume, published in Hystrix (http://www.italian-journal-of-mammalogy.it/issue/view/264), which included seminal papers, such as Rohlf's (2000) contribution on the theory of shape spaces, and Marcus et al.'s (2000) examination of the applicability of GM to craniometric data spanning the whole range of variation of placental mammals. This pioneering research paved the way to an exponential increase in the use of GM in mammalogy and other fields of biology (Adams et al., 2004). Those early studies contributed to a better understanding of the methods and their applications, and a long series of international workshops helped to make GM not only better known to the scientific community but also more accessible to biologists. Indeed, all these meetings were characterized by a strong focus on practice and several of them were held in Italy (e.g., Ciocco, 1993; Rome, 1997-2002; Turin, 2008; Genova, Firenze and Pesche, 2010 – Fig. 1).

A pivotal role in the first decade of the "morphometric revolution" was played by James Rohlf. Among his many merits, he contributed to propel the field by 'translating' (e.g., Rohlf 1993) Bookstein's theoretical advancements, famously as brilliant as cryptic, into a form intelligible by biologists, who often lack the gift of intuition of numerical abstractions and seldom have an adequate background in mathematics. Rohlf was a "translator" for biologists also in another sense: he developed a variety of user-friendly and free GM programs, which over time became known as the `tps` Series (`tpsDig`, `tpsUtil`, `tpsRelw` etc. at: http://life.bio.sunysb.edu/morph/toc-software.html). Rohlf's software was likely the most broadly used in the '90s and is now one of the programs that enabled the biological community to take its first steps in the complex multivariate shape space first described by Kendall in the '80s (Slice 2001, and references therein). It is a fact that this increasing availability of user-friendly GM programs has been and still is a double-edge sward. It might lead users to do things that they do not understand and to treat analytical methods as a black-box. However, it is also undeniable that the success of GM in biology owes much to free software developers, and the range of programs that now goes from comprehensive easy-to-use executables, such as `MorphoJ` (Klingenberg, 2011), or the `IMP` series (Sheets and Zelditch, this issue; Zelditch et al., 2012), to powerful and flexible routines in R (Claude, 2008, this issue; Adams, Otárola-Castillo, 2013). Detailed user guides and introductory manuals (Zelditch et al., 2004, 2012; Claude, 2008) for biologists also played an important role in making GM less intimidating to the non-numerically oriented scientists.

The theoretical development of GM is far from over. New methods and innovative applications are constantly appearing in the literature. We are also reaching a much better understanding of the "old" methods, including some potential pitfalls frequently overlooked even by expert practitioners. For instance, visualizing shape variation with thin plate spline deformation grids, outline or surface rendering, and displacement vectors have become almost default options in biological GM. However, each of these methods has some largely under-appreciated limitations, but no clear and simple discussion of their advantages and disadvantages (Klingenberg, 2013a, this issue). It was also to fill some of these gaps in the literature, and to provide state of the art reviews and examples of GM in evolutionary biology that we decided to invite leading morphometricians to contribute to this special issue of Hystrix on "Evolutionary morphometrics and virtual morphology". We did choose, however, not to exclusively focus on GM, as GM belongs to a much broader family of computerized methods for quantitative morphological investigation (e.g., Evans' ecometric shape descriptors this issue).

Adams et al. open the volume with a broad overview of the field. The paper is the long awaited update of their famous "10 years from the revolution" paper (Adams et al., 2004): also "born" in Italy, as a contribution in the proceedings of the Rome 2002 morphometric workshop, Adams et al. (2004) has now been cited more than 600 times (Google Scholar, January 2013). After readers are introduced to the concept of the "Procrustes Paradigm", have discovered what is new in the field, and what they can expect to see in the near future, 13 papers, among them reviews and example studies, offer a detailed presentation of some of the main methodological advancements and "hot topics" in GM research in mammalogy and evolutionary biology.

The reviews focus on a disparate set of methodological issues. We hope they will benefit readers with their up to date information. Also, and more importantly, their aim is to provide clear discussions on technical aspects, as well as simple explanations of the pros and cons of different methods, and brief recommendations for the less experienced morphometricians. The first review (Klingenberg, 2013a, this issue) after Adams et al., however, might come as a surprise. It is on a topic which most practitioners do not find particularly hard to get, and actually it seems one of the most intuitive aspects of GM: the visualization of shape differences. Klingenberg carefully shows that this is something in fact far from obvious. Whether we use grids, vectors or other diagrams, and how we should interpret them, might represent the most overlooked and misunderstood issue in modern morphometrics. Also, as the author's acknowledgements suggest, it is an issue that may have unexpected consequences even for… airplane passengers!

As the title of this volume implies, morphometrics has been traditionally employed in evolutionary studies and this is still, very likely, the main field where GM finds its most enthusiastic users. This makes a vast number of GM analyses "comparative", and comparative data generally require a special statistical treatment. Monteiro reviews how one might take phylogeny into account in GM studies and why this is important. Polly et al. show that even an apparently straightforward principal component analysis (PCA) may be performed within a phylogenetic framework. A variety of "comparative methods" has been developed over the last three decades to address the non-independence of observations due to phylogenetic relationships. These methods are now routinely applied in ecology and ethology, and have become more popular also among morphometricians especially in the last few years. Their use, however, raises non-trivial questions and may not always be as simple as one might wish. This is partly because of the multivariate nature of shape data and therefore the need of multivariate extensions of univariate methods. Monteiro provides a broad and careful overview of the field within the context of statistical shape analysis, and makes several suggestions and recommendations including tips on the software for performing GM comparative analyses. Polly et al., in contrast, focus on a more specific and fairly recent development: the use of phylogenetic comparative methods in PCAs. They exemplify phylogenetic PCA (pPCA) using both simulated and real data (carnivoran humeri), and suggest where the main differences are compared to a standard PCA. For instance, Polly et al. demonstrate that pPCs are not uncorrelated, but do preserve original shape distances and therefore any potential phylogenetic signal in the data. Also, they may have a profound impact on specific analyses, such as tests of modularity using eigenvalues.

Modularity and integration is now a central topic in evolutionary developmental biology (evo-devo), and one that must be investigated at all levels: from genes to individual organisms and specific and supraspecific taxa. (Klingenberg, 2013b, this issue), who has been a pioneer of this type of studies using GM, summarizes the all the different methods that have been employed until now to test modularity and integration. He reviews what their application to a variety of groups of mammals has taught us, and shows also how examining symmetry vs. asymmetry might help to infer the strength of modularity and integration. Even more importantly, he opens the discussion on one of the

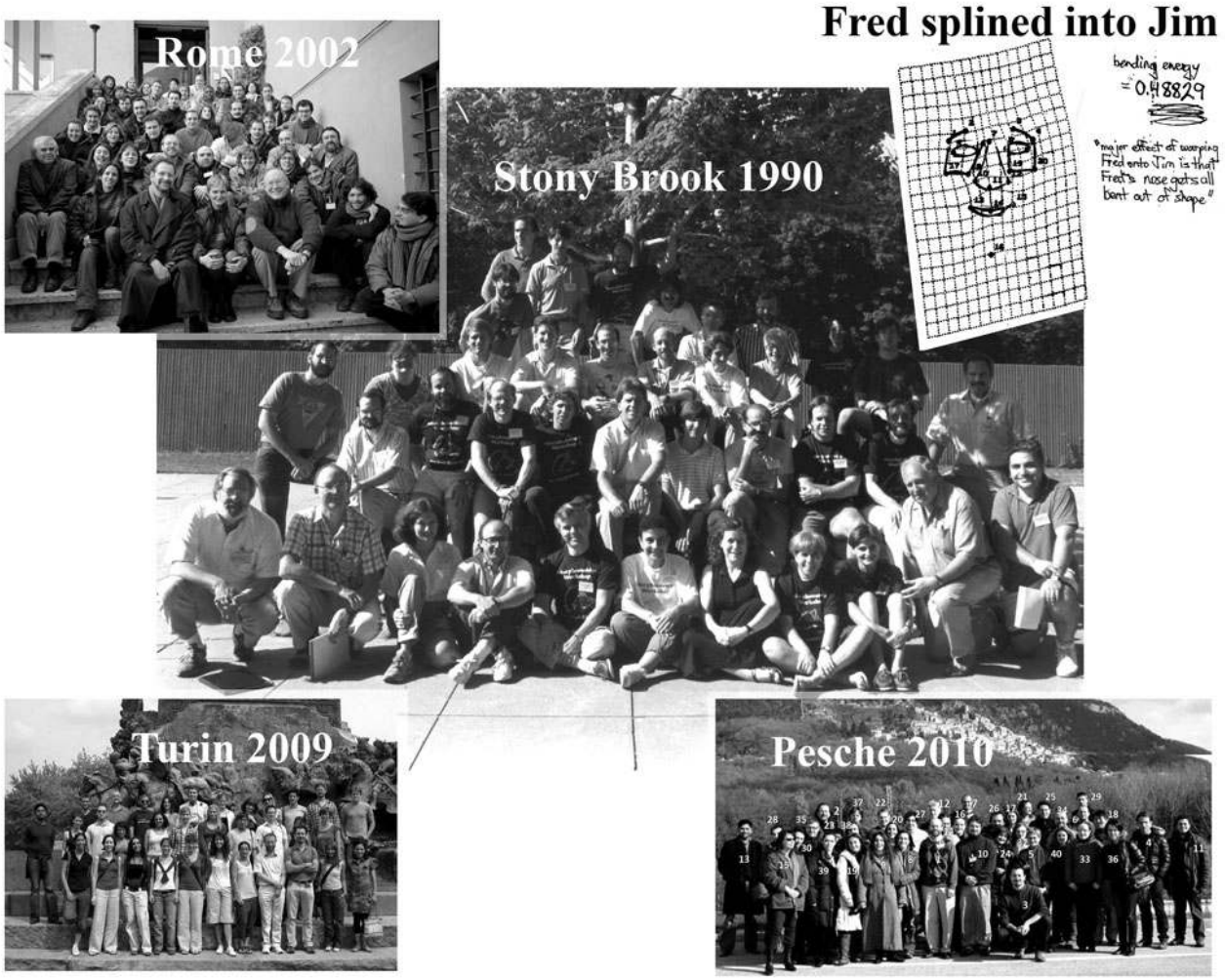


Figure 1 – Group pictures from the "historic" workshop held in Stony Brook (1990) and three of the several GM workshops held in Italy. In the top left inset, the "Bookstein to Rohlf" face-warping diagram sketched as a joke by one of the Stony Brook workshop participants.

most pervasive aspects of shape variation, with a central role as an integrating factor: size-related shape variation or allometry.

In the next contribution, Mitteroecker et al. describe how to study allometry using accurate multivariate methods, and also show how to effectively visualize size-related variation in shape. They suggest that there might be several ways for estimating and comparing allometries, including regressions and form spaces (Dryden and Mardia, 1998). Finally, they exemplify these methods on real data. By exploring how faces vary in humans in relation to factors such as body height, they demonstrate that GM is not limited to the analysis of hard tissues and, intriguingly, find that shorter men tend to be more childlike in facial features.

Measuring allometry and estimating its significance and magnitude is a first step, which is often followed by tests of differences in allometric vectors (e.g., men vs. women, humans vs. chimps etc.). Generally, whenever shape trajectories are compared, different aspects might be considered such as direction, overlap, length, start and end points. Probably the most traditional method to test trajectories across groups is a multivariate analysis of covariance (MANCOVA – e.g., Viscosi and Cardini 2011), a technique available in most statistical packages. However, MANCOVAs only compare the direction of within-group covariation and overlap between groups, and generally do so by performing parametric tests. MANCOVAs, as well as tests of vector angles, length etc., can be performed using resampling statistics. Resampling methods make fewer assumptions than parametric tests, and are generally more suitable to highly dimensional data and relatively small samples. A multiplicity of resampling techniques have been proposed in GM to test trajectories. However, nobody has yet offered a synthetic and updated overview of these methods. Sheets and Zelditch, in this issue, fill this gap with great completeness. As in their introductory book on GM for biologists (Zelditch et al., 2004, 2012), they guide the reader with clarity through a maze of statistical methods, in which a biologist could easily get lost.

Collyer and Adams carry on discussing how to study phenotypic trajectories but they take a novel and broader view. Framing scientific questions in terms of trajectories allows to test not only group differences in multivariate spaces ("static" approach) but also differences in the ways groups change positions in relation to ecological or evolutionary gradients ("dynamic" approach). With this dynamic measurement of change, "raw data' become similar to series of shapes in locomotion analysis, where locomotory trajectories are represented as forms whose size, shape and main axes of variation can be quantified and compared in a GM framework (Adams and Cerney, 2007).

Collyer and Adams' methods might be soon become another tool for investigating what set the main direction of morphological evolution in mammals, the main focus of Renaud and Auffray's contribution. Using teeth and mandibles of murine rodents, they provide inspirational insight into how shape analysis might help us to link micro- to macroevolution. By testing lines of least resistance, that might channel but potentially facilitate change, Renaud and Auffray produce results that can be help to design future experimental studies and explore the processes behind the main phenotypic patterns they measure in mice.

Murine rodents are a most common study group in GM research, and also the subject of the paper by Claude. His aim, however, is different:

Claude, the author of Morphometrics with R (2008), examines the discriminatory power of different morphometric descriptors and analytical methods, when closely related species are compared. By doing this, he shows how to carry on these analyses in R, possibly the most widely used statistical environment in science. Claude convincingly makes the point that, despite a steep initial learning curve, investing time to learn how to perform GM analyses in this free open source software may be worth the effort, and rapidly lead to flexibility and a range of potential applications hardly matched by any other program.

The next three contributions are very heterogeneous. All of them represent innovative extensions of GM methods. Renaud and Auffray, as well as Claude, have provided examples of analyses of outlines using Fourier methods. These methods, which pre-date the "Procrustes Paradigm", have been fruitfully employed for at least three decades to measure outlines. Outlines, and more generally curves and surfaces, tend to lack landmarks with a straightforward and precise correspondence across individuals. In the '90s, however, Bookstein and colleagues (references in Gunz and Mitteroecker this issue) showed how curves and surfaces can be also measured using Cartesian coordinates of points within a Procrustean framework. They suggested elegant algorithmic manipulations to improve the mathematical correspondence of a special type of points, called semilandmarks, employed to "discretize" variation in forms without landmarks. Although there is no biological model behind these algorithms, and answering the question of the biological correspondence of semilandmarks is far from trivial (Klingenberg, 2008; Oxnard and O'Higgins, 2011), semilandmark methods have allowed morphometricians to combine landmarks and points on outlines or surfaces in a single analysis. Semilandmark methods have thus opened a huge range of analytical possibilities that few morphometricians could review better than Gunz et al. These authors belong to the "Viennese school" of morphometrics, which largely developed semilandmark and surface analyses and massively applied them in anthropology. In their review, they explain methods and applications in two and three dimensions, and discuss how semilandmarks can become part of a larger set of tools used for estimating missing landmarks in incomplete specimens, as in the recently born discipline of virtual reconstruction of fossils (Gunz et al. 2009, and referenes therein). Semilandmarks represent a fundamental step forward in shape analysis, and one that is rapidly gaining popularity thanks to a user-friendly implementation in 2D by Rohlf (http://life.bio.sunysb.edu/morph/morphmet/tpsrelww32.exe). More recently, free scripts and packages for 3D analyses have also been released (see documentation and files at: https://sourceforge.net/projects/morpho-rpackage, http://CRAN.R-project.org/package=geomorph, http://hdl.handle.net/2022/14613). As these new possibilities arise, the nicer visualization and apparent increase in information might tempt users to add semilandmarks without a proper consideration of whether they really matter for a specific study aim. More (points in this case) may not always be better, and, as it happened in the past with the interpretation of the biology (or lack of it) behind partial warps (Rohlf, 1998), time will help to better understand how powerful mathematical methods for the GM analysis curves and surfaces relate to underlying biology (Klingenberg, 2008; Oxnard and O'Higgins, 2011; Viscosi and Cardini, 2011).

In the next two studies, MacLeod et al., and O'Higgins and Milne, show that GM is not only extending its analytical toolkit, but also is taking directions which few might have predicted. MacLeod et al. argue that sounds can be measured as 3D shapes. By doing so, in their eigensound study, they manage to discriminate bat species with a high cross-validated accuracy ($\geq 80\%$). This is better than using traditional acoustic analyses and suggests a potential for innovative GM applications, well beyond the traditional range of morphological investigation. Indeed, this is also what O'Higgins and Milne demonstrate in a rather different context: the functional and biomechanical study of armadillos' femurs. The authors use functional simulations and techniques from mechanical engineering to measure the biomechanical performance of the femur. This first type of analysis produces deformations in a computerized 3D model of the bone. The resulting 'deformed' virtual bones become new shapes, and these can be compared and visualized using GM. This new tool might open "new avenues of investigation of skeletal form and function in evolutionary biology" (O'Higgins and Milne, this issue), although, as for semilandmarks and eigensound methods, there might be a lot that we need to learn on validity and accuracy, and the correspondence between these numeric representations and their underlying biology (Adams et al. this issue).

Last but not least, Evans stimulates the reader to keep her/his mind open and not to be led by a preferred or fashionable set of methods. He urges to go beyond the boundaries of 'standard GM' and explore the usefulness of other shape descriptors to investigate specific questions on function, ecology and the interactions between organisms and their environment. Using examples from extinct mammals, Evans points out that hard tissues and especially teeth (the most frequently preserved component of the mammalian fossil record) have much to say on how species evolved and adapted to the environment. He reviews a variety of shape statistics, which do not encode shape itself, such as shear ratios, measures of wear, relief indexes, fractals and other measurements of sharpness, complexity, height and curvature of cusps. Thus, after Gunz et al. (this issue) have shown us how to numerically build virtual fossils, Evans demonstrates that a combination of biological insight and sophisticated quantitative analyses of forms can help to reconstruct palaeo-environments of extinct forms.

"Virtual morphology and evolutionary morphometrics" celebrates the 20th anniversary since the "revolution in morphometrics" (Rohlf and Marcus, 1993), and the enthusiasm with which Italian theriologists joined this revolution. This special issue of Hystrix will be probably known as the "Yellow Book", following a long tradition of naming GM books after the colour of their covers. It is our wish that the "Yellow Book" will make the enthusiasm for GM even more contagious, and contribute to increase the number of biologists that, like us, believe that times are mature for a more profound integration of biological knowledge and numerical methods in the transition from geometric to biological morphometrics (Oxnard and O'Higgins, 2011).

Some of the readers will be fascinated by these methods, but also intimidated by the complex theory of multivariate statistical shape analysis. As biologists, we have shared and still share these feelings. However, we have also learnt that the morphometric community is a great source of friendly and precious advice. The readiness with which the "Giants" of GM accepted our invitation to contribute to this volume went beyond our expectations. We do hope that the "Yellow Book" will provide beginners and less experienced morphometricians with easy-to-follow introductions to the most modern methods in morphometrics. However, this volume will not and cannot answer all questions. As James Rohlf said in a memorable workshop in Rome ten years ago, beginners should not be shy and, when they cannot find clear explanations in the literature, they should look for help in MORPHMET (the discussion list of morphometricians: http://morphometrics.org/morphmet.html), by directly contacting colleagues or by attending one of the many GM courses and workshops, which are now run regularly in many countries (http://life.bio.sunysb.edu/morph/notices.html).

Finally, this volume gives us a chance to remember a few friends, who are are no longer with us: Marco Corti, Robin Hennessy, Les Marcus and Santiago Reig. They were all passionate scientists, keen pioneers of "biological GM", loving friends and great mentors. Santiago's memories of Marco and the early days of GM, a few years before Santiago's death, are testimony to both the great enthusiasm and the friendship we and many others experienced in the morphometric community.

MORPHMET
Re: Marco Corti: 1950-2007
Tue, 13 Feb 2007

"It was a big shock and extremely sad news to hear that Marco is not with us anymore. For those that witnessed the enthusiastic beginning of the "GM revolution", Marco was always the reference-guy for many of us that were on the most zoological side of the crowd. Although definitely a tool-user, Marco was among the advanced and smart users, much more confident on math and stats than the rest of us, and also truly proficient on SAS, a great deal of this learned directly from Les

Marcus. That is why he often played a role as the connecting liaison between developers and clients at the various morphometrics symposia and workshops during the 90'. If you didn't get something during the lecture, you still had the chance of learning it after-hours from Marco, in plain words and using friendly examples... If he admitted that he didn't understand something very well, then forget about it, there'll be little chances that you'll get it... but hey, knowing that Marco was also lost in that something, would make you feel less frustrated about it. Marco managed to be the first one (together with Anna Loy, most of the times) to put into practice and publish papers using the latest morphometric techniques applied to the kind of evolutionary problems of mammals that we all had in mind. In that way he was a pioneer, clearing the path and serving as an inspiration for many mammalogists. His papers were a model to follow because he showed us how to put into practice methods that seemed so complicated and unreachable for many of us that we would have never dared to use, had it not been for his papers. And above all, Marco will always be remembered as the happy man, the happy man who was always smiling and willing to share his discoveries, to show everyone else around him, the exciting and friendly side of morphometrics. We'll miss a generous friend and a most stimulating colleague."

Santiago Reig (1958-2011), Gregorio Maranon Hospital, Universidad Complutense de Madrid, Medical Imaging Lab

References

Adams D.C., Cerney M.M., 2007. Quantifying biomechanical motion using Procrustes motion analysis. Journal of Biomechanics. 40: 437–444.

Adams D.C., Otárola-Castillo E., 2013. `geomorph`: an R package for the collection and analysis of geometric morphometric shape data. Methods in Ecology and Evolution 4: 393–399. http://cran.r-project.org/web/packages/geomorph/index.html

Adams D.C., Rohlf F.J., Slice D.E., 2004. Geometric morphometrics: ten years of progress following the "revolution". Italian Journal of Zoology 71: 5–16.

Adams D.C., Rohlf F.J., Slice D.E., 2013. A field comes of age: geometric morphometrics in the 21st century. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6283

Claude J., 2008. Morphometrics with R. Springer Verlag.

Claude J., 2013. Log-shape ratios, Procrustes superimposition, elliptic Fourier analysis: three worked examples in R. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6316

Collyer M.L., Adams D.C, 2013. Phenotypic trajectory analysis: comparison of shape change patterns in evolution and ecology. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6298

Corti M., 1993. Geometric morphometrics: An extension of the revolution. Trends in Ecology & Evolution. 8: 302–303.

Corti M., Marcus L.F., Hingst-Zaher E., 2000 Introduction to the symposium: geometric morphometrics in mammalogy. Hystrix 10(2): 3–7. doi:10.4404/hystrix-11.1-4133

Dryden I.L., Mardia K.V., 1998. Statistical shape analysis. John Wiley & Sons, New York.

Evans A.R., 2013. Shape descriptors as ecometrics in dental ecology. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6363

Gunz P., Mitteroecker P., Neubauer S., Weber G.W., Bookstein F.L., 2009. Principles for the virtual reconstruction of hominin crania. Journal of Human Evolution. 57: 48–62.

Gunz P., Mitteroecker P., 2013. Semilandmarks: a method for quantifying curves and surfaces. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6292

Houle D., 2010. Colloquium Paper: Numbering the hairs on our heads: The shared challenge and promise of phenomics. Proceedings of the National Academy of Sciences. 107: 1793–1799.

Klingenberg C.P., 2008. Novelty and "Homology-free" Morphometrics: What's in a Name? Evol Biol. 35: 186–190.

Klingenberg C.P., 2011. `MorphoJ`: an integrated software package for geometric morphometrics. Molecular Ecology Resources. 11: 353–357.

Klingenberg C.P., 2013a. Visualizations in geometric morphometrics: how to read and how to make graphs showing shape changes. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-7691

Klingenberg C.P., 2013b. Cranial integration and modularity: insights into evolution and development from morphometric data. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6367

MacLeod N., Krieger J., Jones K.E., 2013. Geometric Morphometric Approaches to Acoustic Signal Analysis in Mammalian Biology. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6299

Marcus L.F., 1990. Traditional morphometrics. Proceedings of the Michigan morphometrics workshop. 2: 77–122.

Marcus L.F., Corti M., 1989. Data analysis in systematics. A manual for workshop W3, at Fifth International Theriological Congress, Rome.

Marcus L.F., Hingst-Zaher E., Zaher H., 2000. Application of landmark morphometrics to skulls representing the orders of living mammals. Hystrix 11(1): 27–47. doi:10.4404/hystrix-11.1-4135

Mitteroecker P., Gunz P., Windhager S., Schaefer K., 2013. A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6369

Monteiro L.R., 2013. Morphometrics and the comparative method: studying the evolution of biological shape. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6282

O'Higgins P., Milne N., 2013. Applying geometric morphometrics to compare changes in size and shape arising from finite elements analyses. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6284

Oxnard C., O'Higgins P., 2011. Biology Clearly Needs Morphometrics. Does Morphometrics Need Biology? Biological Theory. 4: 84– 97.

Polly P.D., Lawing A.M., Fabre A.-C., Goswami A., 2013. Phylogenetic Principal Components Analysis and Geometric Morphometrics. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6383

Renaud S., Auffray J.-C., 2013. The direction of main phenotypic variance as a channel to morphological evolution: case studies in murine rodents. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6296

Rohlf F.J., 1993. Relative warp analysis and an example of its application to mosquito wings. In: Marcus L.F., Bello E., Garcia-Valdecasas A. (Eds.) Contributions to Morphometrics. Museo Nacional de Ciencias Naturales (CSIC) Vol. 8. Madrid, Spain. 131–159.

Rohlf F.J., 1998. On Applications of Geometric Morphometrics to Studies of Ontogeny and Phylogeny. Systematic Biology. 47: 147–158.

Rohlf F.J., 2000. On the use of shape spaces to compare morphometric methods. Hystrix 11(1): 9–25. doi:10.4404/hystrix-11.1-4134

Rohlf F.J., Loy A., Corti M., 1996. Morphometric Analysis of Old World Talpidae (Mammalia, Insectivora) Using Partial-Warp Scores. Systematic Biology. 45: 344–362.

Rohlf F.J., Marcus L.F., 1993. A revolution morphometrics. Trends in Ecology & Evolution 8: 129–132.

Sheets H.D., Zelditch M.L., 2013. Studying ontogenetic trajectories using resampling methods and landmark data. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6332

Slice D.E., 2001. Landmark Coordinates Aligned by Procrustes Analysis Do Not Lie in Kendall's Shape Space. Syst. Biol. 50: 141–149.

Thompson D.W., 1917. On growth and form. Cambridge Univ. Press, Cambridge.

Viscosi V., Cardini A., 2011. Leaf Morphology, Taxonomy and Geometric Morphometrics: A Simplified Protocol for Beginners. PLoS ONE. 6(10): e25630. doi:10.1371/journal.pone.0025630

Zelditch M.L., Swiderski D.L., Sheets H.D., 2004. Geometric Morphometrics for Biologists: a primer. Academic Press.

Zelditch M.L., Swiderski D.L., Sheets H.D., 2012. Geometric Morphometrics for Biologists: a primer (Second Edition). Academic Press.

1990 – Stony Brook Morphometric Workshop at SUNY: Gavin Naylor (with a tube as hat) and Steve Reilly tubing in Long Island during a relaxing break of the workshop (picture by Anna Loy).

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 7–14, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/6283/pdf
doi:10.4404/hystrix-24.1-6283

Research Article

A field comes of age: geometric morphometrics in the 21st century

Dean C. Adams[a,*], F. James Rohlf[b], Dennis E. Slice[c]

[a] *Department of Ecology, Evolution, and Organismal Biology, and Department of Statistics, Iowa State University, Ames IA, 50011, USA*
[b] *Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, 11794-5245, USA*
[c] *Department of Scientific Computing, Florida State University, Tallahassee, FL 32306, USA*

Keywords:
shape analysis
Procrustes analysis
statistics
morphology

Article history:
Received: 29 May 2012
Accepted: 21 October 2012

Acknowledgements
We thank M. Collyer and A. Kaliontzopoulou for constructive comments and discussion, and benefitted greatly from discussions with F. Bookstein. This work was supported in part by NSF grant DEB-1118884 (to DCA) and CDC contract 254-2007-M-21314 (to DES).

Abstract

Twenty years ago, Rohlf and Marcus proclaimed that a "revolution in morphometrics" was underway, where classic analyses based on sets of linear distances were being supplanted by geometric approaches making use of the coordinates of anatomical landmarks. Since that time the field of geometric morphometrics has matured into a rich and cohesive discipline for the study of shape variation and covariation. The development of the field is identified with the *Procrustes paradigm*, a methodological approach to shape analysis arising from the intersection of the statistical shape theory and analytical procedures for obtaining shape variables from landmark data. In this review we describe the Procrustes paradigm and the current methodological toolkit of geometric morphometrics. We highlight some of the theoretical advances that have occurred over the past ten years since our prior review (Adams et al., 2004), what types of anatomical structures are amenable to these approaches, and how they extend the reach of geometric morphometrics to more specialized applications for addressing particular biological hypotheses. We end with a discussion of some possible areas that are fertile ground for future development in the field.

Introduction

> The study of form may be descriptive merely, or it may become analytical. We begin by describing the shape of an object in the simple words of common speech: we end by defining it in the precise language of mathematics; and the one method tends to follow the other in strict scientific order and historical continuity.
>
> D'Arcy Wentworth Thompson (1915)

For centuries, naturalists have marvelled at the diversity and complexity of life on earth. From the simple observation that organisms differ in both their anatomical attributes and in their use of these traits, scholars have sought to describe morphological and anatomical differences among taxa and explain how these differences have evolved (Darwin, 1859). Indeed, this long-standing fascination with biological form has shaped our current perspectives on many biological topics, including our notions of taxonomic discontinuities, our methods of classification, and our hypotheses of the structure-function relationship. As was accurately prophesized by Thompson in 1915, the study of form has developed into a rigorous quantitative discipline; the field of morphometrics. For much of the 20th century morphometric analyses were accomplished by applying univariate and multivariate statistics to sets of measured traits that included linear distances, ratios, and angles (i.e., traditional morphometrics *sensu*: Blackith and Reyment 1971; Marcus 1990; Reyment 1991; for a unique history of the field see: Reyment 1996). However, over time it became apparent that certain shortcomings limited the biological interpretations that were possible with these methods. For example, graphical depictions of shape and shape changes cannot always be generated from the results of these approaches, as the geometric relationships among variables were usually not preserved in the measurements taken (Strauss and Bookstein 1982; for discussion see: Rohlf and Marcus 1993; Adams et al. 2004). Thus, another approach to the study of shape was required.

In the 1980's, alternative approaches were proposed and developed that represented a radical shift in the way the shapes of anatomical structures were quantified and analysed (but see Boas 1905; Galton 1907; Sneath 1967 for earlier work). These alternatives captured the geometry of the morphological structures of interest and retained this information throughout the analyses. This new approach was called *geometric morphometrics* (Corti, 1993). Not surprisingly, the conversion from traditional to geometric morphometrics could be viewed as a "revolution" (Rohlf and Marcus, 1993). This revolution embodied both a shift in methodology, and in the conceptual and statistical underpinnings of the field. Geometric morphometric approaches generally utilize fundamentally different types of data to quantify shape; landmark coordinates, outline curves, and surfaces. In its early years the field developed at a rapid pace, as methodological advances for shape quantification intertwined with a rich statistical theory for shape analysis (the morphometric "synthesis", *sensu* Bookstein 1996). As a result of this paradigm shift, landmark-based geometric morphometrics (GM) provides a powerful technique in the quantitative biologists' repertoire for the study of shape variation and the identification of its causes. Not surprisingly, these methods are increasingly used to quantify anatomical shapes in a wide range of scientific disciplines.

Nearly a decade ago, we reviewed the field of geometric morphometrics and described the important advantages that these approaches have relative to alternative methods of shape analysis (Adams et al. 2004; for other reviews see: O'Higgins 2000; Slice 2005, 2007; Mitteroecker and Gunz 2009). We also summarized some of the methodological advances that took place in the ten years following the "revolution", and provided an account of the state-of-the field at that time. Much has happened in geometric morphometrics in the years since that review, mandating a new appraisal of this now mature discipline. In this review, we describe some of the theoretical developments that have ensued during the past decade. Our review focuses on advances in landmark-based morphometric approaches, though we recognize that developments in related areas have also occurred (e.g., MacLeod 2008; McPeek et al.

*Corresponding author
Email address: dcadams@iastate.edu (Dean C. Adams)

2008; Shen et al. 2009). Our aim is to provide readers with a succinct summary of the current "toolkit" of geometric morphometrics, and an appreciation for the types of applications and extensions that can be used to address specific biological hypotheses. We also provide our perspective of what the future may hold in store in the coming years.

Geometric morphometrics and the "Procrustes paradigm"

Geometric morphometrics is the statistical analysis of shape variation and its covariation with other variables (Bookstein, 1991). These methods quantify variation in the shape of anatomical objects using the Cartesian coordinates of anatomical landmarks, after the effects of non-shape variation have been mathematically held constant. Geometric morphometric studies are accomplished through what can be called the *Procrustes paradigm*; an approach to shape analysis that emerged from the unification of a rigorous statistical theory for shape (Kendall, 1981, 1984, 1985) with analytical procedures for superimposing landmark configurations to obtain shape variables (Gower, 1975; Bookstein, 1986; Rohlf and Slice, 1990; Rohlf, 1999b). In a typical morphometric analysis, the Procrustes paradigm is implemented as a series of operations (Fig. 1).

First, from each specimen, a set of two- or three-dimensional landmark coordinates is obtained, which record the relative positions anatomically-definable locations. These landmarks can be considered "fixed points", as they define the locations of particular anatomical traits representing discrete biological attributes (see below for a discussion of semilandmarks). Next, a generalized Procrustes analysis (GPA: Gower 1975; Rohlf and Slice 1990) is used to superimpose the configurations of landmarks in all specimens to a common coordinate system, and to generate a set of shape variables. This least-squares procedure translates all specimens to the origin, scales them to unit centroid size, and rotates them to minimize the total sums-of-squares deviations of the landmark coordinates from all specimens to the average configuration. After superimposition, the aligned *Procrustes shape coordinates* describe the location of each specimen in a curved space related to Kendall's shape space (Rohlf, 1999b; Slice, 2001). These are typically projected orthogonally into a linear tangent space yielding *Kendall's tangent space coordinates* (Dryden and Mardia, 1993, 1998; Rohlf, 1999b; Kent and Mardia, 2001), on which multivariate analyses of shape variation are then conducted[1]

The third step of a morphometric study is to test biological hypotheses using multivariate statistical methods. For instance, Hotelling's T^2 or multivariate analysis of variance (MANOVA) can be used to test for shape differences among groups, while multivariate regression or partial least squares (PLS: Rohlf and Corti 2000) can be used to help identify patterns of covariation between shape and other continuous variables. In addition, methods that partition shape variation in particular ways can be utilized to address more specialized biological hypotheses. For instance, shape variation due to directional and fluctuating asymmetry can be quantified and analysed to test hypotheses of symmetry in a sample (Klingenberg and McIntyre, 1998; Mardia et al., 2000; Kent and Mardia, 2001; Klingenberg et al., 2002; Schaefer et al., 2006). Finally, graphical methods are used to visualize patterns of shape variation and facilitate descriptions of shape changes. Here, ordination methods such as principal components analysis (PCA) generate scatterplots representing the dispersion of shapes in tangent space, while thin-plate spline transformation grids can provide a visual description of the shape differences between objects. The latter display shape changes and differences in a manner similar to D'Arcy Thompson's transformation grids (Thompson, 1917) where one object is transformed (or "warped") into another using the thin-plate spline (Bookstein, 1989, 1991). Importantly, transformation grids may be generated for actual specimens in the data set or for estimated specimens, such as group means or predicted specimens along a regression line. Thus, by comparing transformation grids, differences in shape between objects and trends in shape change in specific directions in Kendall's tangent space can be depicted and anatomically described. This combination of rigorous statistical analysis with visualizing shape changes represents one of the more powerful aspects of the Procrustes paradigm.

[1]Note that alternative mathematical approximations of tangent space are possible (see e.g., Weber and Bookstein 2011). These should give very similar results when shape variation is small.

Advances over the past decade

The dawn of the 21st century brought with it the maturation phase of geometric morphometrics and the emergence of the Procrustes paradigm as the standard methodological approach for analysing shape characterized by landmark data. This was due in large part to the realization that Procrustes-based approaches outperformed alternative methods from a statistical perspective (e.g., Rohlf 1999a, 2000, 2003). However, despite the pre-eminence of this approach to shape analysis, much theoretical progress continues in the discipline. Particularly active has been the development of more specialized applications to address particular biological problems and hypotheses. In this section we highlight several of these theoretical advances, and some ways in which they enhance the morphometrician's toolkit.

Use of three-dimensional data

One major change in geometric morphometrics from a decade ago has been the rapid increase in the use of three-dimensional data. Interestingly, there are generally no mathematical limitations for handling data in three dimensions. Indeed the algorithms commonly used for superimposition, projection, and statistical analysis are all generalized to accommodate data of any dimensionality. Instead, the restriction to two-dimensional data has been decidedly more practical. Until recently, acquiring three-dimensional data required specialized equipment and was prohibitively expensive, and the use of many three-dimensional devices was limited to specimens in a restricted size range. However, over the past decade a number of lower-cost options have become available, including surface scanners and other data-collection devices. Because these devices are now more accessible to the research community, many more geometric morphometric studies are conducted using three-dimensional landmark data. This shift is perhaps most pervasive in the field of anthropology, where the specimen sizes are particularly well-matched to the optimal size ranges of three-dimensional data acquisition tools (Slice, 2007).

Semilandmarks and missing landmarks

Geometric morphometric analyses are typically performed on landmark coordinates describing specific anatomical locations (i.e., "fixed"

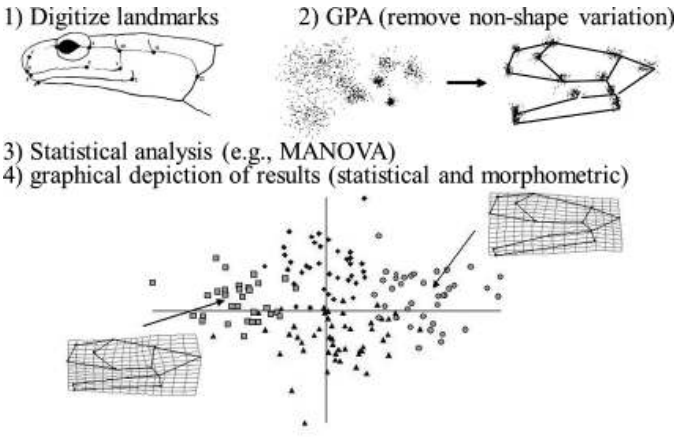


Figure 1 – Graphical depiction outlining the steps of the Procrustes paradigm for landmark-based geometric morphometrics. 1) Digitize raw data (landmarks recorded on head of a *Plethodon* salamander), 2) Generalized Procrustes analysis to remove non-shape variation (landmarks of 156 specimens before and after GPA), 3) statistical analysis (e.g., MANOVA), and 4) graphical depiction of results (ordination plot of specimens with thin-plate spline transformations for the mean specimens of two groups). Data from Adams and Rohlf (2000).

anatomical points), yet the shape of other anatomical features may also be of interest. For example, *semilandmarks* can be used to capture the shape of boundary curves, which can then be included with a set of fixed landmarks in a Procrustes-based shape analysis (Bookstein, 1997; Bookstein et al., 1999). With this approach, a series of locations along the curve are digitized, and an additional step may be incorporated in the Procrustes algorithm which slides these points along vectors tangent to the curve until their positions align as closely as possible with the semilandmarks on the corresponding curve of a reference specimen (for related approaches see: Hammond et al. 2004; McCane and Kean 2011). This algebra for semilandmarks has now been extended to three dimensions, so that the shapes of both curves and surfaces can be quantified (Gunz et al., 2005). Here, semilandmarks on curves are usually slid along their tangent vectors, and semilandmarks on surfaces are slid within their tangent planes, until their positions minimize the shape difference between specimens. This reduces the effect of the arbitrary initial placement of the semilandmarks, and can be accomplished either by minimizing bending energy or Procrustes distance between the reference and the target specimen (Bookstein et al., 1999; Gunz et al., 2005; Rohlf, 2010). It should be emphasized that care must be taken when placing semilandmarks relative to other structures (Oxnard and O'Higgins, 2011). Additionally, these methods tend to work best when used on relatively smooth curves and surfaces, or when many points are included. Nevertheless, by using both landmarks and semilandmarks, information from points, curves, and surfaces may be combined for a more comprehensive quantification and analysis of biological shape variation (Fig. 2).

Methods have also been developed to account for missing landmarks. Because morphometric analyses require that all specimens have the same set of landmarks, incomplete specimens must either be eliminated from the analysis, or missing landmarks must be eliminated from the dataset. Clearly, neither solution is ideal: a far preferable alternative is to estimate the locations of missing landmarks in some way, so that partial specimens can be included. For structures exhibiting bilateral symmetry, one approach is to estimate missing landmark locations by reflecting their corresponding landmarks across the mid-line of the structure (Claude, 2008). Unfortunately, this method can only be implemented for symmetrical objects. Inferring the locations of missing landmarks can also be accomplished through statistical estimation approaches based on regression or expectation maximization (e.g., Neeser et al. 2009; Couette and White 2010). A third alternative extends the logic of semilandmarks to the estimation of missing landmarks (Bookstein et al., 1999; Gunz et al., 2009). Here the locations of missing landmarks are determined using the thin-plate spline, such that they minimize the shape difference between the incomplete specimen and the reference, thereby exerting minimal influence on the resulting patterns of shape variation (the method is presently based on bending energy, but an approach that minimizes Procrustes distances could also be envisioned). Together, these methods greatly expand the utility of the Procrustes paradigm for fields where missing landmarks are common, such as palaeontology and anthropology (see e.g., Weber and Bookstein 2011).

Analysis of symmetry

Biologists have long been interested in identifying deviations in symmetry in populations of organisms, including: fluctuating asymmetry (the random departures from perfect symmetry across individuals) and directional asymmetry (biased departures from perfect symmetry across individuals). Over a decade ago, the methods commonly employed on traditional measurements to analyse patterns of asymmetry were generalized for landmark-based shape data (Klingenberg and McIntyre, 1998; Klingenberg et al., 2002). Subsequently, a mathematical decomposition of shape variation into its symmetry components (for bilaterally symmetric shapes) was derived (Mardia et al., 2000; Kent and Mardia, 2001), and recent extensions of the approach have been proposed for other types of symmetry (Savriama and Klingenberg, 2011). As a consequence of these methodological advances, the analysis of symmetry, and the identification of a lack of symmetry, is now frequently examined in geometric morphometric studies.

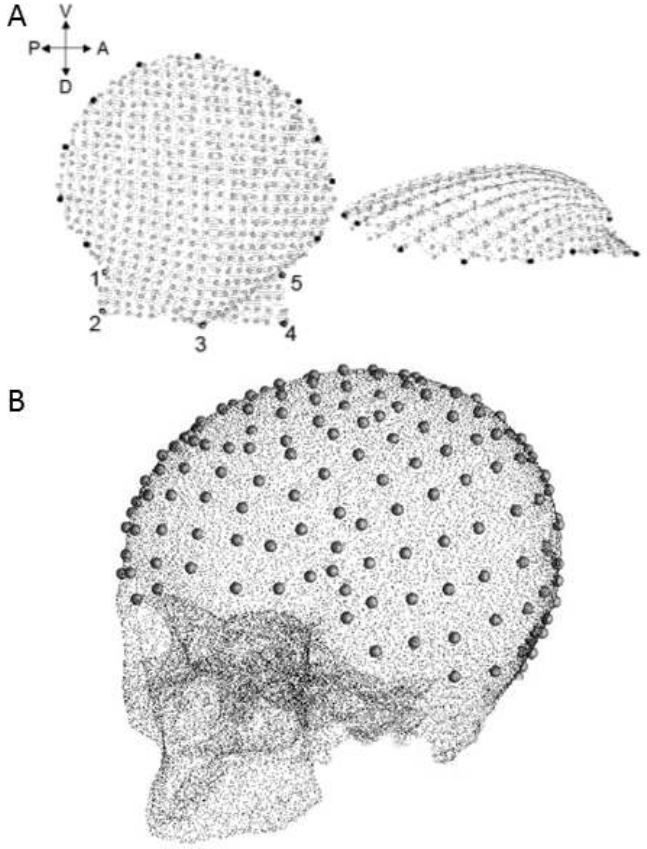


Figure 2 – A) A set of landmarks defining fixed anatomical points, the boundary curve, and the surface of a scallop shell (from Serb et al. 2011). B) Sliding semilandmarks quantifying the surface of a skull (reproduced from Adams et al. 2004; after Mitteroecker 2001; MS thesis, Univ. Vienna).

Visualizing allometry

A nearly ubiquitous property of organisms is that individuals of different sizes also have different shapes. The association of size and shape is investigated through the analysis of allometry. Allometric studies have a long history in morphometrics (Huxley, 1932; Jolicoeur, 1963; Cock, 1966; Klingenberg, 1996; Sidlauskas et al., 2011), yet a recurring challenge has been to generate graphical summaries that adequately illustrate these patterns. This problem is particularly acute for geometric morphometric data, where allometric trajectories describe the multivariate relationship between shape and size. Several recent methods address this issue and help to facilitate biological interpretation of allometric and ontogenetic trends (Fig. 3). For instance, a principal components analysis of the matrix of tangent space coordinates augmented with the vector of log centroid size reveals the major direction of variation in this size-shape space (*sensu* Mitteroecker et al. 2004), which often is due to allometry. When multiple groups are present the common allometric component (CAC: Mitteroecker et al. 2004) can be calculated, which is an estimate of the common within-group trend. Shape residuals from the CAC can then be plotted against size to identify differences among allometric (or ontogenetic) trajectories. A related approach is to estimate shape scores from the regression of shape on size, and plot these against size (Drake and Klingenberg, 2008). It should be noted that for a single group, this approach is mathematically identical to the CAC, as the average within-group trend (i.e. the CAC) is simply the trend for a single group. In a third approach, an allometric trajectory is represented by estimating predicted values from a regression of shape on size, and a stylized graphic is obtained by plotting the first principal component of the predicted values versus size (Adams and Nistri, 2010). Finally, transformation grids representing the shapes of small and large specimens can be included with the

scatterplots to provide a graphical depiction of how shape changes as a function of size. Together, these methods provide complementary visualizations of the allometric and ontogenetic patterns frequently present in morphometric data.

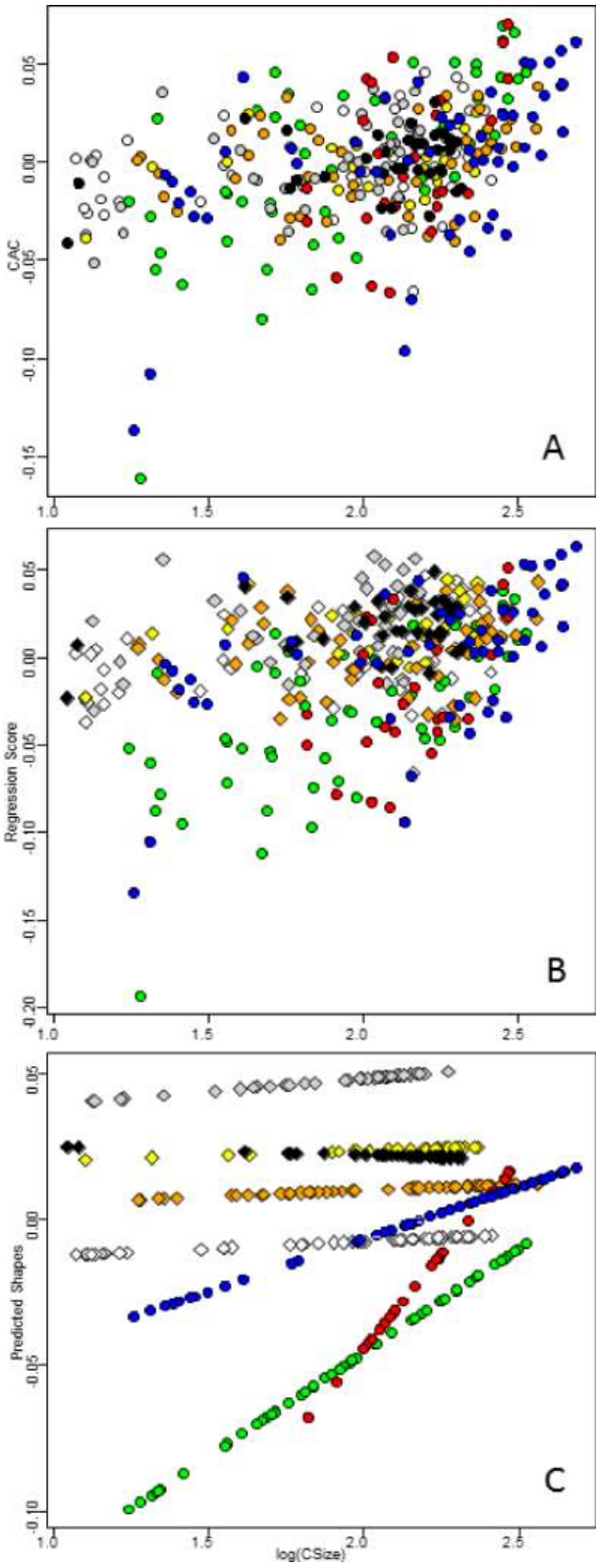


Figure 3 – Examples of various visualizations of shape allometry for eight species of Italian plethodontid salamanders (data from Adams and Nistri 2010). A) The common allometric component (CAC) versus log centroid size; B) Regression scores versus log centroid size; C) PCI of predicted values (from a regression of shape on size) versus log centroid size.

Quantifying phenotypic trajectories and motion paths

The sequence of changes in shape due to allometry corresponds to a path (or trajectory) of shape changes in tangent space. This trajectory reveals how shape changes as a function of size. However, patterns of shape change can also be generated from processes other than allometry; thus a general approach for analysing phenotypic change trajectories is required. Recently, one approach was proposed that examines various attributes of phenotypic change trajectories and uses these for statistical comparison (Adams and Collyer, 2007; Collyer and Adams, 2007; Adams and Collyer, 2009; Collyer and Adams, this issue). Termed *phenotypic trajectory analysis*, this approach quantifies the size, orientation, and shape of phenotypic paths (trajectories) and examines these attributes to determine whether sets of trajectories are similar or different (beyond any differences they may have in their location in tangent space). Importantly, the method can quantify any trajectory of shape change; such as those representing allometric or ontogenetic growth trajectories, temporal sequences, evolutionary shape changes, shape changes resulting from ecological shifts, or shape changes observed in studies of phenotypic plasticity (see Adams and Collyer 2009). Recent applications have identified parallel evolution of shape changes resulting from competitive interactions (Adams, 2010), and patterns of ontogenetic convergence among closely related taxa (Adams and Nistri, 2010; Piras et al., 2010).

A related approach has been proposed for the study of motion paths (Adams and Cerney 2007; for related ideas see: Slice 1999; O'Higgins et al. 2002). Here a motion is represented by a sequence of shapes in tangent space that corresponds to differences in the relative position of their anatomical parts. This sequence forms a trajectory in shape space whose attributes can be quantified and used to identify similarities or differences among motion trajectories (Adams and Cerney, 2007). This method is a special case of the phenotypic trajectory analysis described above. Finally, a complementary procedure quantifies motion paths as the change in relative position of a single landmark during a movement, and sets of these motion paths are then quantitatively compared using Procrustes analysis (Decker et al., 2007).

Applications to quantitative genetics

Another area where considerable development occurred over the past decade is the utilization of shape data in quantitative genetics. For instance, a number of studies have used quantitative trait loci to assess the genetic underpinnings of shape variation (see e.g., Klingenberg and Leamy 2001; Klingenberg et al. 2004; Burgio et al. 2009; Boell et al. 2011). In addition, several methods for linking the algebra of quantitative genetics with that of geometric morphometrics have been proposed. One method is based on Procrustes distance, allowing the estimation of heritability and other quantitative genetics attributes in a univariate framework (Monteiro et al., 2002). An alternative uses the multivariate generalization of the breeder's equation (Klingenberg and Leamy, 2001), retaining the full multivariate nature of the geometric morphometric shape data throughout the analysis and allowing implementations based on the "animal model" (*sensu* Lynch and Walsh 1998). Considerable discussion highlighted the advantages and disadvantages of these two alternatives (e.g., Klingenberg 2003; Monteiro et al. 2003; Klingenberg and Monteiro 2005; see also Myers et al. 2006). Ultimately, for accurate estimates of magnitudes and directions of shape heritability the multivariate approach is preferred (Klingenberg and Monteiro, 2005). Recent applications of this approach can be found in (e.g., Gómez et al. 2009; Klingenberg et al. 2010; Adams 2011; Martínez-Abadías et al. 2012).

Integration and modularity

It has long been observed that some phenotypic traits are highly correlated while other sets of traits display less correlation. Such observations lead Olson and Miller to propose the concepts of morphological integration and modularity (Olson and Miller, 1951, 1958). Morphological integration describes the nature of correlated characters and the cohesion among traits that result from developmental, evolutionary, and functional processes (Klingenberg, 2008; Mitteroecker, 2009).

Several methods have been proposed for empirically identifying modular or integrated components in morphometric data, and these approaches address related, but subtly distinct biological hypotheses. For instance, an exploratory way to identify integration among traits is based on conditional independence (Magwene, 2001, 2009). Here no *a priori* hypothesis of modularity is required; rather, the partial correlations among traits are used to hypothesize combinations of traits that may describe modular structure. However, a number of shortcomings limit the utility of this approach, particularly as the number of traits per module increases (Mitteroecker and Bookstein, 2007, 2009).

When prior evidence exists to posit a hypothesis of modularity, the amount of integration between modules can be assessed through partial least squares (Bookstein et al., 2003). This approach identifies the degree of covariation between sets of traits, and when used with geometric morphometric data, is called *singular warps analysis* (Bookstein et al., 2003; Mitteroecker and Bookstein, 2008). An alternative approach tests the hypothesis that there is modularity according to a given partition of landmarks against the null hypothesis that there is no modularity. This method expresses the degree of covariation between modules relative to variation within modules using Escoufier's RV coefficient (Klingenberg, 2009). This approach can be used to determine whether an *a priori* partitioning of landmarks exhibits covariation expected under the hypothesis of modularity, or to identify putative modules that have minimal covariation between them. Finally, one can test alternative modularity hypotheses by comparing the observed pattern of covariation among traits to patterns of covariation expected under a particular modular hypothesis (Marquez, 2008; Parsons et al., 2012). Here a goodness of fit statistic between the two covariance matrices is obtained and assessed via Monte Carlo methods. Together, these techniques provide new and exciting avenues for examining patterns of integration and modularity in morphometric data.

The next ten years: a possible future

The past twenty five years have seen a surge of methodological developments in geometric morphometrics. What began as an alternative approach to shape quantification has developed into a rigorous discipline that combines a rich statistical theory with shape data from points, curves and surfaces, to test an ever-increasing breadth of biological hypotheses. It is therefore not surprising that geometric morphometric methods are being utilized more than ever before to describe and investigate patterns of shape variation and covariation in many areas of biological research. If history is any guide, the next ten years will be equally exciting, and theoretical developments in geometric morphometrics will continue at a rapid pace. As we did in our review a decade ago (Adams et al., 2004), here we highlight a few areas that we predict are ripe for future development.

Morphometrics and phylogenetics

It is with a sense of irony that we include this topical area under future developments, as a decade ago we commented that the intersection of morphometrics and phylogenetics needs to be further explored (Adams et al., 2004). In some sense this area may remain a perennial candidate for future development, because despite considerable effort, combining these two disciplines in a cohesive manner has remained stubbornly elusive. The power of geometric morphometrics to quantify patterns of shape variation is undeniable; thus it is natural to consider how morphometric variables may be used to estimate phylogenetic trees. However, most attempts to do so have relied on methods based on cladistic parsimony, and unfortunately, there is a fundamental disconnect between these approaches (which require independent, discrete variables) and geometric morphometric shape variables (which are continuous, multivariate, and inter-dependent). Discussions of earlier attempts at bridging this divide can be found in (Fink and Zelditch 1995; Adams and Rosenberg 1998; Rohlf 1998; MacLeod 2002; Swiderski et al. 2002 see also Adams et al. 2011). In fact, even a recent approach for estimating phylogenies from shape data (Gonzalez-Jose et al., 2008) suffers these same shortcomings, as this approach is ultimately based on a rank-based representation of shape differences among species along individual shape axes (for discussion see: Adams et al. 2011; also Klingenberg and Gidaszewski 2010). Other recent work has focused on obtaining a phylogenetically-informed superimposition (see e.g., Goloboff and Catalano 2011; Catalano and Goloboff 2012). Importantly, while these methods use morphometric shape data as input, they fall outside the realm of the Procrustes paradigm, as the landmark-by-landmark parsimony approaches they employ do not conform with shape distances as defined by Kendall's shape space and tangent space (Klingenberg and Gidaszewski, 2010; Adams et al., 2011). Finally, several recent approaches use maximum likelihood methods to obtain estimates of phylogenetic relationship from shape data (Caumul and Polly, 2005; Cardini and Elton, 2008). However, current implementations of this approach assume that the characters evolve independently and with equal variances following a Brownian motion model of evolution; assumptions that are not met with geometric morphometric shape data (see Adams et al. 2011). Nevertheless, because likelihood methods can incorporate continuous multivariate data, we feel that they hold the most promise for this application, and recommend that future work in this area be focused in that direction.

Morphometrics and phylogenetics can also be combined by utilizing an existing phylogeny to address hypotheses of shape change through evolutionary time. For instance, one can visualize predicted patterns of shape evolution by estimating ancestral shapes and projecting both the hypothesized ancestors and the phylogeny into tangent space (e.g., Rohlf 2002; Klingenberg and Gidaszewski 2010). Methods have also been developed for estimating phylogenetic signal from morphometric data (Klingenberg and Gidaszewski, 2010), though more work is required in this area. A few studies have looked at the tempo and mode of macroevolutionary changes in shape (e.g., Monteiro and Nogueira 2011). However, more work is needed to integrate morphometric data, which is both multivariate and multi-dimensional (*sensu* Klingenberg and Gidaszewski 2010), with recent evolutionary methods for comparing models of phenotypic evolution on phylogenies (e.g., Butler et al. 2000; Blomberg et al. 2003; Butler and King 2004), and for estimating rates of phenotypic evolution on phylogenies (e.g., O'Meara et al. 2006; Revell and Harmon 2008; see also Bookstein 2012b for recent Brownian motion models for shape evolution).

Estimating landmark covariances

Another area that represents a hold-over from our previous review is estimating landmark covariance structure. For many reasons, morphologists are interested in evaluating the relative variability within and among landmarks. However, the superimposition procedure itself alters the resulting patterns of landmark covariance, which makes direct interpretation of the resulting covariance matrix challenging (Rohlf and Slice, 1990; Walker, 2000; Rohlf, 2003). While it is important to recognize that this shortcoming does not affect statistical tests of shape differences performed in Kendall's tangent space, the issue still needs to be resolved. One approach to the problem is to perform simulations that can identify the extent to which observed patterns of landmark covariance reflect what is expected under a particular model of shape change. Alternatively, it may be possible to extend recent approaches that evaluate the relative influence of individual landmarks on allometric trends ("jackknife-GPA" *sensu*: van der Linde and Houle 2009), or methods for assessing the relative contribution of each landmark to overall patterns of shape variation (Albert et al., 2003), for the estimation of landmark covariance structure. More work is needed on this issue.

Morphometrics and biomechanics

One area we believe will be fruitful in generating future methodological developments is the intersection of geometric morphometrics and biomechanics. In biomechanics, finite element analysis (FEA) is often used to estimate the stresses and strains that anatomical structures endure when subjected to forces or loadings (e.g., Dumont et al. 2005). Like geometric morphometrics, FEA begins by capturing the geometry of an anatomical structure based on the coordinates of (many) point locations. Therefore it is logical to suppose that the two approaches can

be combined in some meaningful way (for discussion see: O'Higgins et al. 2011; Weber et al. 2011; Parr et al. 2012). Indeed, a number of approaches for combining FEA and GM in the same analysis have recently been proposed. One method uses the thin-plate spline to deform an existing FE model from one specimen into another (e.g., Pierce et al. 2008; Stayton 2009; Rivera and Stayton 2011). Here, geometric morphometric methods estimate the smooth deformation mapping between specimens (based on the locations of their landmarks), and the deformation function is subsequently used to generate a predicted FE model to evaluate stresses and strains for the target specimen. An alternative method combines GM and FEA in a different manner. Here, the thin-plate spline is used to visualize the structural deformations implied by the FE model itself (e.g., Cox et al. 2011; Groning et al. 2011; O'Higgins et al. 2011).

While these approaches represent exciting new developments and hold great promise for future studies in functional and geometric morphometry, a number of serious challenges remain. First, little has been done to validate whether the FE models predicted from morphometric mappings represent what would be obtained from an FEA performed directly on the target specimen. Second, when a single FE model is used to predict models for multiple specimens (e.g., Rivera and Stayton 2011; Parr et al. 2012), distinct FE models are obtained for each target specimen. However, all the information about stresses and strains is based on the geometric and densiometric properties of the original specimen: no new stress information is obtained. As such, this "filtering" of the FE model through the thin-plate spline does not result in novel biomechanical predictions for each specimen, but simply re-expresses the same biomechanical information in some non-linear, but predictable, manner. Further, at present it is not clear how representations of shape differences as displayed by the thin-plate spline relate to the stresses and strains described by FEA. Additionally, even when the densiometric properties of both the reference and target specimens are assumed to be identical (which is unlikely to be the case), predicted FE models generated from thin-plate spline mappings do not provide the correct deformations in terms of the physics of the actual deformations (Bookstein, 2013). Finally, no mathematical theory exists for quantitatively relating difference in shape to differences in stresses and strains that would occur on those specimens (see Weber et al. 2011; Bookstein 2013). Thus, we currently do not know whether the relationship between FE models and shape differences (Procrustes distance) is a one-to-one or many-to-one mapping. As alluded to above, one problem is that the thin-plate spline does not take into account the material properties of the structure, and thus some biomechanical properties estimated by the FE model cannot be directly related to shape deformations. Clearly, additional theoretical work is required to resolve these issues, and to provide a methodological framework for future studies in this area.

Resolution of modularity approaches

The past several years has seen tremendous growth in studies of integration and modularity as applied to morphometric data. In fact, three distinct analytical approaches are currently in use: one based on partial least squares (Bookstein et al., 2003; Mitteroecker and Bookstein, 2008), one based on Escoufier's RV coefficient (Klingenberg, 2009), and one based on comparing observed and expected covariance matrices (Marquez, 2008; Parsons et al., 2012). Further, two of these methods (PLS and the RV coefficient) are mathematically related: they both examine the cross-covariance matrix between sets of traits; only they use this information in different ways. Despite these advances however, a number of issues remain to be resolved in this area. One issue is that the present methods test against an unreasonable null hypotheses of modular structure. That is, present methods test all combinations of contiguous landmarks to see whether their internal level of covariation is higher than that to landmarks not in a putative module. However, these combinations generally ignore the distances between landmarks, yet landmarks cannot usually be sampled uniformly across a structure. A simple alternative to the existence of modules is a model where the level of covariance between landmarks is inversely proportional to their distance. In some datasets, the same modules are obtained by maximizing within-module covariance or by simply clustering landmarks based on interlandmark distances (F.J. Rohlf, unpublished). Thus, in these instances, there is no need to invoke a hypothesis of modular structure, as the observed covariance patterns are simply the result of the relative distances between landmarks. Further, at present it is not clear which approach is most effective at identifying non-spatially determined patterns of integration and modularity, and under what circumstances. Therefore, what is needed is a thorough comparison of the statistical properties of these approaches under known conditions, to determine when each performs relatively well and when each performs more poorly. This would give some much-needed clarity to the field, and provide end-users with clear recommendations on which integration and modularity methods should be utilized, when, and why. We note that this suggestion echoes the approach taken to resolve an earlier issue; namely, which landmark-based morphometric method should be used. Here, statistical simulations determined that Procrustes approaches outperformed alternative methods (e.g., Rohlf 1999a, 2000, 2003).

Software development

As with all quantitative fields, new methods will only be used if researchers have access to software for their implementation. Fortunately, the morphometric community is replete with theorists who also generate software, and thus numerous packages are available. Nearly all morphometrics packages include the standard components of the Procrustes paradigm: including Procrustes superimposition, basic statistical analyses of shape data (PCA, MANOVA, regression), and visualization through thin-plate spline transformation grids. A partial list of these packages includes the `TPS`-series (e.g., `TPSRelw`: Rohlf 2010), `IMP` (Sheets, 2003), the `EVAN` Toolbox (EVAN Society 2012), `MorphoJ` (Klingenberg, 2011), `Morpheus et al.` (Slice, 1998), and several libraries for R (e.g., routines in Claude 2008, and the libraries "`shapes`": Dryden 2013, and "`geomorph`": Adams amd Otárola-Castillo 2013). More specialized analyses, such as mapping shapes on phylogenies, asymmetry analyses, and modularity approaches, are found in fewer software packages, as are methods for incorporating three-dimensional surface semilandmarks in GPA and for estimating missing landmarks using deformation procedures. In the future, we predict that new methodological developments will be incorporated into the software packages mentioned here, and that new software options will become available to the end-user.

Conclusions

It is not an overstatement to say that the past twenty five years in morphometrics has been exciting. We have witnessed a fundamental shift in how morphology is quantified as the field migrated towards the use of landmark coordinates on anatomical points, curves, and surfaces. From the union of statistical shape theory and methods for obtaining shape variables a standard approach to shape analysis has emerged (the Procrustes paradigm), and new extensions based on Procrustes methods continue to be developed to address specific biological hypotheses. Patterns of shape changes are now quantified and compared, and graphical representations of shapes are generated to facilitate biological interpretation of these statistical trends. From all of these developments it is clear that D'Arcy Thompson's dream of investigating biological form in a fully quantitative manner has now been realized. We look forward to the exciting new developments that will emerge in the years to come.

References

Adams D.C., 2010. Parallel evolution of character displacement driven by competitive selection in terrestrial salamanders. BMC Evol. Biol. 10: 1–10.

Adams D.C., 2011. Quantitative genetics and evolution of head shape in *Plethodon* salamanders. Evol. Biol. 38: 278–286.

Adams D.C., Cardini A., Monteiro L.R., O'Higgins P., Rohlf. F.J., 2011. Morphometrics and phylogenetics: principal components of shape from cranial modules are neither appropriate nor effective cladistic characters. J. Hum. Evol. 60: 240–243.

Adams D.C., Cerney M.M., 2007. Quantifying biomechanical motion using Procrustes motion analysis. J. Biomech. 40: 437–444.

Adams D.C., Collyer M.L., 2007. The analysis of character divergence along environmental gradients and other covariates. Evolution 61: 510–515.

Adams D.C., Collyer M.L., 2009. A general framework for the analysis of phenotypic trajectories in evolutionary studies. Evolution 63: 1143–1154.

Adams D.C., Nistri A., 2010. Ontogenetic convergence and evolution of foot morphology in European cave salamanders (Family: Plethodontidae). BMC Evol. Biol. 10: 1–10.

Adams D.C., Otárola-Castillo E., 2013. `geomorph`: an R package for the collection and analysis of geometric morphometric shape data. Methods in Ecology and Evolution 4: 393–399. http://cran.r-project.org/package=geomorph

Adams D.C., Rohlf F.J., 2000. Ecological character displacement in *Plethodon*: biomechanical differences found from a geometric morphometric study. Proceedings of the National Academy of Sciences, USA 97: 4106–4111.

Adams D.C., Rohlf F.J., Slice D.E. 2004. Geometric morphometrics: ten years of progress following the "revolution". It. J. Zool. 71: 5–16.

Adams D.C., Rosenberg M.S., 1998. Partial warps, phylogeny, and ontongeny: a comment on Fink and Zelditch (1995). Syst. Biol. 47:168–173.

Albert M.H., Le H., Small C.G., 2003. Assessing landmark influence on shape variation. Biometrika 90: 669–678.

Blackith R.E., Reyment R.A., 1971. Multivariate Morphometrics. Academic Press, New York.

Blomberg S.P., Garland T., Ives A.R., 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57: 717–745.

Boas F., 1905. The Horizontal Plane of the Skull and the General Problem of the Comparison of Variable Forms. Science 21: 862–863.

Boell L., Gregorová S., Forejt J., Tautz D., 2011. A comparative assessment of mandible shape in a consomic strain panel of the house mouse (*Mus musculus*) - implications for epistasis and evolvability of quantitative traits. BMC Evol. Biol. 11(309): 1–12.

Bookstein F.L., 1986. Size and shape spaces for landmark data in two dimensions. (With discussion and rejoinder). Statistical Science 1: 181–242.

Bookstein F.L., 1989. Principal warps: thin-plate splines and the decomposition of deformations. Institute of Electrical and Electronics Engineers, Transactions on Pattern Analysis and Machine Intelligence 11: 567–585.

Bookstein F.L., 1991. Morphometric tools for landmark data: Geometry and Biology. Cambridge Univ. Press, New York.

Bookstein F.L., 1996. Biometrics, biomathematics and the morphometric synthesis. Bulletin of Mathematical Biology 58: 313–365.

Bookstein F.L., 1997. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Medical Image Analysis 1: 225–243.

Bookstein F.L., 2013. Allometry for the Twenty-First century. Biol. Theory 7(1): 10–25. doi:10.1007/s13752-012-0064-0

Bookstein F.L., 2012b. Random walk as a null model for high-dimensional morphometrics of fossil series: geometrical considerations. Paleobiology 39(1): 52–74.

Bookstein F.L., Gunz P., Mitteroecker P., Prossinger H., Schaefer K., Seidler H., 2003. Cranial integration in *Homo*: singular warps analysis of the midsagittal plane in ontogeny and evolution. J. Hum. Evol. 44: 167–187.

Bookstein F.L., Schafer K., Prossinger H., Seidler H., Fieder M., Stringer G., Weber G.W., Arsuaga J.-L., Slice D.E., Rohlf F.J., Recheis W., Mariam A.J., Marcus L.F., 1999. Comparing frontal cranial profiles in archaic and modern *Homo* by morphometric analysis. Anat. Rec. (New Anat.) 257: 217–224.

Burgio G., Baylac M., Heyer E., Montagutelli X., 2009. Genetic analysis of skull shape variation and morphological integration in the mouse using interspecific recombinant congenic strains between C57BL/6 and mice of the *Mus spretus* species. Evolution 63: 2668–2686.

Butler M.A., King A.A., 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164: 683–695.

Butler M.A., Schoener T.W., Losos J.B., 2000. The relationship between sexual size dimorphism and habitat use in Greater Antillean *Anolis* lizards. Evolution 54: 259–272.

Cardini A., Elton S., 2008. Does the skull carry a phylogenetic signal? evolution and modularity in the guenons. Biol. J. Linn. Soc. 93: 813–834.

Catalano S.A., Goloboff P.A., 2012. Simultaneously mapping and superimposing landmark configurations with parsimony as optimality criterion. Syst. Biol. 61(3): 392–400.

Caumul R., Polly D., 2005. Phylogenetic and environmental components of morphological variation: skull, mandible, and molar shape in marmots (*Marmota*, Rodentia). Evolution 59: 2460–2472.

Claude J., 2008. Morphometrics with R. Springer, New York.

Cock A.G., 1966. Genetical aspects of metrical growth and form in animals. Quart. Rev. Biol. 41: 131–190.

Collyer M.L., Adams D.C., 2007. Analysis of two-state multivariate phenotypic change in ecological studies. Ecology 88: 683–692.

Collyer M.L., Adams D.C., 2013. Phenotypic trajectory analysis: comparison of shape change patterns in evolution and ecology. Hystrix 24(1) (Online First). doi:10.4404/hystrix-24.1-6298

Corti M., 1993. Geometric morphometrics: An extension of the revolution. Trends Ecol. Evol. 8: 302–303.

Couette S., White J., 2010. 3D geometric morphometrics and missing data. Can extant taxa give clues for the analysis of fossil primates? Comptes Rendus Palevol 9: 423–433.

Cox P.G., Fagan M.J., Rayfield E.J., Jeffery N., 2011. Finite element modelling of squirrel, guinea pig and rat skulls: using geometric morphometrics to assess sensitivity. J. Anat. 219: 696–709.

Darwin C., 1859. On the origin of species by means of natural selection, or the preservation of favored races in the struggle for life. Murray, London.

Decker L., Berge C., Renous S., Penin X., 2007. An alternative approach to normalization and evaluation for gait patterns: Procrustes analysis applied to the cyclograms of sprinters and middle-distance runners. J. Biomech. 40: 2078–2087.

Drake A.G., Klingenberg C.P., 2008. The pace of morphological change: Historical transformation of skull shape in St Bernard dogs. Proceedings of the Royal Society B, Biological Sciences 275: 71–76.

Dryden I.L., 2012. `shapes`: Statistical shape analysis. R package version 1.1-8. http://cran.r-project.org/package=shapes

Dryden I.L., Mardia K.V., 1993. Multivariate shape analysis. Sankhya 55: 460–480.

Dryden I.L., Mardia K.V., 1998. Statistical shape analysis. John Wiley & Sons, New York.

Dumont E.R., Piccirillo J., Grosse I.R., 2005. Finite element analysis of biting behavior and bone stress in the facial skeletons of bats. Anat. Rec. 293: 319–330.

Fink W.L., Zelditch M.L., 1995. Phylogenetic analysis of ontogenetic shape transformations: a reassessment of the piranha genus *Pygocentrus* (Teleostei). Syst. Biol. 44: 344–361.

Galton F., 1907. Classification of Portraits. Nature 76: 617–618.

Goloboff P.A., Catalano S.A., 2011. Phylogenetic morphometrics (II): algorithms for landmark optimization. Cladistics 27: 42–51.

Gómez J.M., Abdelaziz M., Muñoz-Pajares J., Perfectti F., 2009. Heritability and genetic correlation of corolla shape and size in *Erysimum mediohispanicum*. Evolution 63: 1820–1831.

Gonzalez-Jose R., Escapa I., Neves W.A., Cuneo R., Pucciarelli H.M., 2008. Cladistic analysis of continuous modularized traits provides phylogenetic signals in Homo evolution. Nature 453: 775–779.

Gower J.C., 1975. Generalized Procrustes analysis. Psychometrika 40: 33–51.

Groning F., Fagan M.J., O'Higgins P., 2011. The effects of the periodontal ligament on mandibular stiffness: a study combining finite element analysis and geometric morphometrics. J. Biomech. 44: 1304–1312.

Gunz P., Mitteroecker P., Bookstein F.L., 2005. Semilandmarks in three dimensions. In: Slice D.E. (Ed.). Modern morphometrics in physical anthropology. Klewer Academic/Plenum, New York, pp. 73–98.

Gunz P., Mitteroecker P., Neubauer S., Weber G.W., Bookstein F.L., 2009. Principles for the virtual reconstruction of hominin crania. J. Hum. Evol. 57: 48–62.

Hammond P., Hutton T.J., Allanson J.E., Campbell L.E., Hennekam R.C.M., Holden S., Patton M.A., Shaw A., Temple I.K., Trotter M., Murphy K.C., Winter R.M.. 2004. 3D analysis of facial morphology. Am. J. Med. Genet. 126A: 339–348.

Huxley J.S., 1932. Problems of relative growth. Methuen, London. Reprinted in 1993 by Johns Hopkins University Press, Baltimore.

Jolicoeur P., 1963. The multivariate generalization of the allometry equation. Biometrics 19: 497–499.

Kendall D.G., 1981. The statistics of shape. In: Barnett V. (Ed). Interpreting multivariate data. Wiley, New York, pp. 75–80.

Kendall D.G., 1984. Shape-manifolds, Procrustean metrics and complex projective spaces. Bulletin of the London Mathematical Society 16: 81–121.

Kendall D.G., 1985. Exact distributions for shapes of random triangles in convex sets. Adv. Appl. Prob. 17: 308–329.

Kent J.T., Mardia K.V., 2001. Shape, Procrustes tangent projections and bilateral symmetry. Biometrika 88: 469–485.

Klingenberg C.P., 2003. Quantitative genetics of geometric shape: heritability and the pitfalls of the univariate approach. Evolution 57: 191–195.

Klingenberg C.P., 2008. Morphological integration and developmental modularity. Annu. Rev. Ecol. Evol. Syst. 39: 115–132.

Klingenberg C.P., 2009. Morphometric integration and modularity in configurations of landmarks: tools for evaluating a priori hypotheses. Evol. Develop. 11: 405–421.

Klingenberg C.P., 2011. MorphoJ: an integrated software package for geometric morphometrics. Molec. Ecol. Res. 11: 353–357.

Klingenberg C.P., Barluenga M., Meyer A., 2002. Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. Evolution 56: 1909–1920.

Klingenberg C.P., Debat V., Roff D.A., 2010. Quantitative genetics of shape in cricket wings: developmental integration in a functional structure. Evolution 64: 2935–2951.

Klingenberg C.P., Gidaszewski N.A., 2010. Testing and quantifying phylogenetic signals and homoplasy in morphometric data. Syst. Biol. 59: 245–261.

Klingenberg C.P., Leamy L.J., 2001. Quantitative genetics of geometric shape in the mouse mandible. Evolution 55: 2342–2352.

Klingenberg C.P., Leamy L.J., Cheverud J.M., 2004. Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible. Genetics 166: 1909–1921.

Klingenberg C.P., McIntyre G.S., 1998. Geometric morphometrics of developmental instability: Analyzing patterns of fluctuating asymmetry with Procrustes methods. Evolution 52: 363–1375.

Klingenberg C.P., Monteiro L.R., 2005. Distances and directions in multidimensional shape spaces: implications for morphometric applications. Syst. Biol. 54: 678–688.

Klingenberg K.P., 1996. Multivariate allometry. In: Marcus L.F., Corti M., Loy A., Naylor G.J.P., Slice D.E. (Eds). Advances in Morphometrics. Plenum Press, New York, pp. 23–49.

Lynch M., Walsh B., 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland.

MacLeod N., 2002. Phylogenetic signals in morphometric data. In: MacLeod N., Forey P.L. (Eds.). Morphology, shape and phylogeny. Taylor and Francis, London, pp. 100–138.

MacLeod N., 2008. Understanding morphology in systematic contexts: 3D specimen ordination and 3D specimen recognition. In: Wheeler Q. (Ed.). The New Taxonomy. CRC Press, Taylor & Francis Group, London, pp. 143–210.

Magwene P.M., 2001. New tools for studying integration and modularity. Evolution 55: 1734–1745.

Magwene P.M., 2009. Statistical methods for studying modularity: a reply to Mitteroecker and Bookstein. Syst. Biol. 58: 146–149.

Marcus L.F., 1990. Traditional morphometrics. In: Rohlf F.J., Bookstein F.L. (Eds.). Proceedings of the Michigan morphometrics workshop. University of Michigan Museum of Zoology, Ann Arbor, pp. 77–122.

Mardia K.V., Bookstein F.L., Moreton I.J., 2000. Statistical assessment of bilateral symmetry of shapes. Biometrika 87: 285–300.

Marquez E.J., 2008. A statistical framework for testing modularity in multidimensional data. Evolution 62: 2688–2708.

Martínez-Abadías N., Esparza M., Sjøvold T., González-José R., Santos M., Hernández M., Klingenberg C.P., 2012. Pervasive genetic integration directs the evolution of human skull shape. Evolution 66: 1010–1023.

McCane B., Kean M.R., 2011. Integration of parts in the facial skeleton and cervical vertebrae. Am. J. Orthodont. Dento. Orthoped. 139: e13–e30.

McPeek M.A., Shen L., Torrey J.Z., Farid H., 2008. The tempo and mode of three-dimensional morphological evolution in male reproductive structures. Am. Nat. 171: E158–E178.

Mitteroecker P., 2009. The developmental basis of variational modularity: insights from quantitative genetics, morphometrics, and developmental biology. Evol. Biol. 36: 377–385.

Mitteroecker P., Bookstein F.L., 2007. The conceptual and statistical relationship between modularity and morphological integration. Syst. Biol. 56: 818–836.
Mitteroecker P., Bookstein F.L., 2008. The evolutionary role of modularity and integration in the hominoid cranium. Evolution 62: 943–958.
Mitteroecker P., Bookstein F.L., 2009. Examining modularity via partial correlations: a rejoinder to a comment by Paul Magwene. Syst. Biol. 58: 346–348.
Mitteroecker P., Gunz P., 2009. Advances in geometric morphometrics. Evol. Biol. 36: 235–247.
Mitteroecker P., Gunz P., Bernhard M., Schaefer K., Bookstein F.L., 2004. Comparison of cranial ontogenetic trajectories among great apes and humans. J. Hum. Evol. 46: 679–698.
Monteiro L.R., Diniz-Filho J.A.F., Reis S.F.D., Araujo E.D., 2002. Geometric estimates of heritability in biological shape. Evolution 56: 563–572.
Monteiro L.R., Diniz-Filho J.A.F., Reis S.F.D., Araujo E.D., 2003. Shape distances in general linear models: Are they really at odds with the goals of morphometrics? A reply to Klingenberg. Evolution 57: 196–199.
Monteiro L.R., Nogueira M.R., 2011. Evolutionary patterns and processes in the radiation of phyllostomid bats. BMC Evol. Biol. 11(137): 1–23.
Myers E.M., Janzen F.J., Adams D.C., Tucker J.K., 2006. Quantitative genetics of plastron shape in slider turtels (*Trachemys scripta*). Evolution 60: 563–572.
Neeser R., Ackermann R.R., Gain J., 2009. Comparing the accuracy and precision of three techniques used for estimating missing landmarks when reconstructing fossil hominin crania. Am. J. Phys. Anthropol. 140: 1–18.
O'Higgins P., 2000. The study of morphological variation in the hominid fossil record: biology, landmarks and geometry. J. Anat. 197: 103–120.
O'Higgins P., Cobb S.N., Fitton L.C., Groning F., Phillips R., Liu J., Fagan M.J., 2011. Combining geometric morphometrics and functional simulation: an emerging toolkit for virtual functional analyses. J. Anat. 218: 3–15.
O'Higgins P., Jones N., Ghattaura A., Hutton T., Carr M.. 2002. Geometric morphometric approaches to the study of soft tissue growth and expression in the human face. Am. J. Phys. Anthropol. Supp. 34: 119.
O'Meara B.C., Ane C., Sanderson M.J., Wainwright P.C., 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60: 922–933.
Olson E.C., Miller R.L., 1951. A mathematical model applied to a study of the evolution of species. Evolution 5: 325–338.
Olson E.C., Miller R.L., 1958. Morphological Integration. University of Chicago Press, Chicago.
Oxnard C., O'Higgins P., 2011. Biology Clearly Needs Morphometrics. Does Morphometrics Need Biology? Biol. Theory 4: 84–97.
Parr W.C.H., Wroe S., Chamoli U., Richards H.S., McCurry M.R., Clausen P.D., McHenry C., 2012. Toward integration of geometric morphometrics and computational biomechanics: New methods for 3D virtual reconstruction and quantitative analysis of Finite Element Models. J. Theor. Biol. 301: 1–14.
Parsons K.J., Marquez E., Albertson R.C., 2012. Constraint and opportunity: the genetic basis and evolution of modularity in the cichlid mandible. Am. Nat. 179: 64–78.
Pierce S.E., Angielczyk K.D., Rayfield E.J., 2008. Patterns of morphospace occupation and mechanical performance in extant crocodilian skulls: a combined geometric morphometric and finite element modeling approach. J. Morph. 269: 840–864.
Piras P., Colangelo P., Adams D.C., Buscalioni A., Cubo J., Kotsakis T., Meloro C., Raia P., 2010. The Gavialis-Tomistoma debate: the contribution of skull ontogenetic allometry and growth trajectories to the study of crocodylian relationships. Evol. Develop. 12: 568–579.
Revell L.J., Harmon L.J., 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. Evol. Ecol. Res. 10: 311–331.
Reyment R.A., 1991. Multidimensional paleobiology. Pergamon Press, New York.
Reyment R.A., 1996. An idiosyncratic history of early morphometrics. In: Marcus L.F., Corti M., Loy A., Naylor G.J.P., Slice D.E. (Eds.). Advances in Morphometrics. Plenum Press, New York, pp. 15–22.
Rivera G., Stayton C.T., 2011. Finite element modeling of shell shape in the freshwater turtle *Pseudemys concinna* reveals a trade-off between mechanical strength and hydrodynamic efficiency. J. Morph. 272: 1192–1203.
Rohlf F.J., 1998. On applications of geometric morphometrics to studies of ontogeny and phylogeny. Syst. Biol. 47: 147–158.
Rohlf F.J., 1999a. On the use of shape spaces to compare morphometric methods. Hystrix 11: 1–17.
Rohlf F.J., 1999b. Shape statistics: Procrustes superimpositions and tangent spaces. Journal of Classification 16: 197–223.
Rohlf F.J., 2000. Statistical power comparisons among alternative morphometric methods. Am. J. Phys. Anthropol. Supp. 111: 463–478.
Rohlf F.J., 2002. Geometric morphometrics and phylogeny. In: Macleod N., Forey P. (Eds.). Morphology, shape, and phylogeny. Taylor & Francis, London, pp. 175–193.
Rohlf F.J., 2003. Bias and error in estimates of mean shape in morphometrics. J. Hum. Evol. 44: 665–683.
Rohlf F.J., 2010. tpsRelw: Relative warps analysis. Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, NY.
Rohlf F.J., Corti. M., 2000. The use of partial least-squares to study covariation in shape. Syst. Biol. 49: 740–753.
Rohlf F.J., Marcus L.F., 1993. A revolution in morphometrics. Trends Ecol. Evol. 8: 129–132.
Rohlf F.J., Slice D.E., 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39: 40–59.
Savriama Y., Klingenberg C.P., 2011. Beyond bilateral symmetry: geometric morphometric methods for any type of symmetry. BMC Evol. Biol. 11(280): 1–20.
Schaefer K., Lauc T., Mitteroecker P., Gunz P., Bookstein F.L., 2006. Dental Arch Asymmetry in an Isolated Adriatic Community. Am. J. Phys. Anthropol. 129: 132–142.
Serb J.M., Alejandrino A., Otárola-Castillo E., Adams D.C., 2011. Morphological convergence of shell shape in distantly related scallop species (Mollusca: Pectinidae). Zool. J. Linn. Soc. 163: 571–584.
Sheets H.D., 2003. IMP-Integrated Morphometrics Package, version 7, Department of Physics, Canisius College, Buffalo, New York. Available from http://www.canisius/edu/sheets/morphsoft.html.
Shen L., Farid H., McPeek M.A., 2009. Modeling 3-dimensional morphological structures using spherical harmonics. Evolution 63: 1003–1016.
Sidlauskas B.L., Mol J.H., Vari R.P., 2011. Dealing with allometry in linear and geometric morphometrics: a taxonomic case study in the Leporinus cylindriformis group (Characiformes: Anostomidae) with description of a new species from Suriname. Zool. J. Linn. Soc. 162: 103–130.
Slice D.E., 1998. Morpheus et al.: software for morphometric research. Department of Ecology and Evolution, State University of New York, Stony Brook, New York.
Slice D.E., 1999. Geometric motion analysis. Am. J. Phys. Anthropol. Supp. 28: 253–254.
Slice D.E., 2001. Landmark coordinates aligned by Procrustes analysis do not lie in Kendall's shape space. Syst. Biol. 50: 141–149.
Slice D.E., 2005. Modern morphometrics. In: Slice D.E. (Ed.). Modern morphometrics in physical anthropology. Klewer Press, New York, pp. 1–45.
Slice D.E., 2007. Geometric Morphometrics. Ann. Rev. Anthropol. 36: 261–281.
Sneath P.H.A., 1967. Trend-surface analysis of transformation grids. J. Zool. Lond. 151: 65–122.
Society E., 2012. EVAN Toolbox. build 1626, Vienna. Available from www.evan-society.org.
Stayton C.T., 2009. Applications of thin-plate spline transformations to finite element models, or, how to turn a bog turtle into a spotted turtle to analyze both. Evolution 63: 1348–1355.
Strauss R.G., Bookstein F.L., 1982. The truss: body form reconstructions in morphometrics. Syst. Zool. 31: 113–135.
Swiderski D.L., Zelditch M.L., Fink W.L., 2002. Comparability, morphometrics and phylogenetic systematics. In: MacLeod N., Forey P.L. (Eds.). Morphology, shape and phylogeny. Taylor & Francis, London, pp. 67–99.
Thompson D.W., 1915. Morphology and mathematics. Trans. Roy. Soc. Edinburgh 50: 857–895.
Thompson D.W., 1917. On Growth and Form. Cambridge, London.
van der Linde K., Houle D., 2009. Inferring the nature of allometry from geometric data. Evol. Biol. 36: 311–322.
Walker J.A., 2000. The ability of geometric morphometric methods to estimate a known covariance matrix. Syst. Biol. 49: 686–696.
Weber G.W., Bookstein F.L., 2011. Virtual anthropology: a guide to a new interdisciplinary field. Springer Verlag.
Weber G.W., Bookstein F.L., Strait D.S., 2011. Virtual anthropology meets biomechanics. J. Biomech. 44: 1429–1432.

Associate Editor: A. Loy

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 15–24, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/7691/pdf
doi:10.4404/hystrix-24.1-7691

Research Article

Visualizations in geometric morphometrics: how to read and how to make graphs showing shape changes

Christian Peter KLINGENBERG[a,*]

[a] *Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, United Kingdom*

Keywords:
geometric morphometrics
landmarks
Procrustes superimposition
thin-plate spline
transformation grids
warping

Article history:
Received: 17 September 2012
Accepted: 16 January 2013

Acknowledgements
I thank Paul O'Higgins for extensive and contentious discussion of the problems with different visualization methods while flying from Paris to Santiago de Chile and back in October 2011. This paper is very much the outcome of those discussions, with arguments refined and sharpened as a result of the exchange. I hope the several hundred other passengers on those flights have forgiven us for keeping them awake by arguing, in hushed but agitated tone and for hours on end, about something that must have appeared to them as utterly arcane and incomprehensible.

This is also an opportunity to thank Larry Leamy for many discussions on geometric morphometrics, back in the late nineties and early noughties, which considerably helped me to develop the ideas that I have finally spelled out in this paper. It was during these discussions that we coined the expression "lollipop graph", but I do not recall whether Larry or I first came up with it.

I am grateful to Andrea Cardini for inviting me to contribute to this special issue of Hystrix. I also would like to thank him, Anna Loy, Paul O'Higgins and the members of my lab for providing insightful comments on an earlier version of this paper.

Abstract

An important aspect of geometric morphometrics, since its beginnings, has been the visualization of shape changes. A range of methods has been developed with advances in the theory of statistical shape analysis and new possibilities in computer graphics. Most approaches are based either on relative shifts of landmark positions in starting and target shapes after superimposition or on D'Arcy Thompson's idea of transformation grids. Both approaches are in wide use in current morphometrics, and both have their distinctive advantages and shortcomings. This paper discusses the assumptions and some caveats of both approaches. The paper also offers some recommendations for authors of geometric morphometric studies.

Introduction

The visualization of shape changes is at the very core of geometric morphometrics. Indeed, one of the key advantages of geometric morphometrics is that shape differences can be visualized directly as illustrations or computer animations. Accordingly, at the time of the "revolution" in morphometrics, when geometric morphometrics was established as a discipline, the ease of visualization was used as an important argument in favour of geometric morphometrics by comparison with "traditional" morphometrics (Rohlf and Marcus, 1993). Since then, the success of geometric morphometrics is substantially due to the fact that the various methods of visualization can communicate even complex morphological changes much more effectively than the tables of coefficients that result from traditional morphometric analyses. Above all, these visualizations provide information on morphological changes in their immediate anatomical context. Visualizing shape changes remains an important tool for understanding morphological variation, as geometric morphometrics is used to address an increasingly varied range of questions about evolution and development of organisms (Klingenberg, 2010).

A number of different kinds of visualizations for shapes and shape changes have been widely used in geometric morphometrics. They are based mainly on two principles: visualization of shape change by showing the relative displacements of corresponding landmarks in different shapes or by showing the deformation of a regular grid, an outline or a surface that is interpolated from the shape change. These two approaches can also be used in combination. Advances in computer graphics have made it easier to produce appealing illustrations of shape changes in two or three dimensions with both these approaches, and computer animation holds further potential.

All methods for visualising shape changes are based on particular lines of reasoning that makers and viewers of such graphics need to understand and follow, and which occasionally give rise to misunderstandings and controversy. To avoid those problems, it is also important to consider the language in which shape changes and their visualizations are described in the morphometric literature. Despite their crucial importance for geometric morphometrics, the principles and implicit assumptions of the methods for visualizing shape changes have not been discussed in any detail.

This paper explores the conceptual basis and implicit assumptions of visualization through landmark displacements and through graphs based on deformation, such as transformation grids and warped 3D surfaces. I show that both types of visualization have problems that authors and readers of morphometric studies need to keep in mind when making and interpreting visualizations of shape changes. My aim is not to recommend or to condemn a particular approach, but rather to compare and contrast the different logical and visual bases of the approaches. I hope that this will help morphometricians, on the one hand, to interpret

*Corresponding author
Email address: cpk@manchester.ac.uk (Christian Peter KLINGENBERG)

published illustrations and descriptions of shape changes correctly and, on the other hand, to produce visualizations and to describe them in a manner that will communicate morphometric results to readers without misunderstandings.

Shape and shape changes: two key concepts in geometric morphometrics

Shape is defined technically as all the geometric features of an object except for its size, position and orientation (e.g. Dryden and Mardia 1998). This definition may seem abstract and complicated, but in fact it refers to exactly the sort of information we are using when we interpret what appears on a photograph. Imagine you are looking at a picture of a friend. You easily recognize your friend in the picture even though the picture is much smaller than your friend and despite the fact that you may be looking at the picture far away from your friend or from where the picture was taken. The fact that size and position of the picture are "wrong" does not affect your ability to recognize the person in the image (and, in the first place, to see that it is a person). You might even be able to interpret the photograph if you hold it upside down or at some other angle: orientation also is not essential for our ability to recognize people in images. The main source of information for interpreting images is the shape of the objects they contain, but we tend to discount information on size, position and orientation. So, whenever we are looking at a picture, we are intuitively and unconsciously applying the technical definition of shape.

Because of this definition of shape, we can display the shapes of objects without worrying about their size, position and orientation for the purposes of visualization in morphometric analyses. This is true both for diagrams showing individual shapes and visualizations that display a combination of two or more shapes to show the differences between them – but of course, the resulting diagrams may look very different. Because size, position and orientation are not part of shape, they can be varied freely without affecting the shape information. It does not matter how big a visualization of a shape appears on the page or screen, where exactly it appears and what its orientation is. Therefore, size and location can be chosen so that diagrams fit together conveniently and meet the requirements of a journal format or screen size. Orientation is best chosen to follow anatomical convention for the structure and organisms under study (e.g. the convention that the dorsal side is up and the ventral side is down).

Kendall's shape space is a key component of geometric morphometrics (Bookstein, 1996a; Small, 1996; Dryden and Mardia, 1998; Kendall et al., 1999). Each possible shape (for a given number of landmarks and dimensionality) corresponds to a single point in the shape space and every point in shape space corresponds to a particular shape. Because shape spaces are curved, non-Euclidean spaces, it is advantageous to use a local approximation by a linear, Euclidean tangent space that touches the shape space at the location of the average shape in the sample. This approximation is the same as the one that is used when the curved surface of the Earth is approximated by a flat map of a particular region. For limited ranges of shape variation, as they are usually used in biological studies even at large taxonomic scale, this approximation tends to be very good (e.g. Marcus et al. 2000). Just as in the shape space, each point in the shape tangent space corresponds to a particular shape, and each shape that is sufficiently close to the average shape in the sample (the shape at the tangent point) corresponds to a point in the shape tangent space. Because of this one-to-one correspondence, it is possible to reconstruct the physical shape of an object corresponding to each point in the shape space or tangent space. Therefore, it is possible to go back and forth between the abstract results of statistical analyses in the shape tangent space and the actual shapes of objects. This back-and-forth is the key to visualization in geometric morphometrics. Visualization is possible not just for the particular specimens included in a morphometric study, but for any point in the shape space or shape tangent space (to be precise: any point in the tangent space that is a projection of a point in shape space).

Shape changes are equally fundamental in geometric morphometrics as the concept of shape itself, and it is shape changes that are normally the subject of visualization. Shape changes are among the results of many statistical analyses, such as principal components, regression, partial least squares and others. The concept of shape change is closely related to the concept of shape difference. A difference between two shapes implies that the two shapes are not the same, but it is not directed – there is no designated start or end. In contrast, a shape change involves a directed difference from a starting shape to a target shape. An example of a shape difference is sex dimorphism, which may be studied as the shape change from male to female or from female to male (Gidaszewski et al., 2009; Astúa, 2010). Examples of shape changes are ontogenetic changes associated with growth and development, where there is a clear directionality from younger to older organisms, or historical changes from earlier to later time (Drake and Klingenberg, 2008; Weisensee and Jantz, 2011). Whereas a shape corresponds to a single point in shape space or shape tangent space, a shape change is the movement from the point representing the starting point to the point representing the target shape. This means that it is a vector that has a direction and a magnitude (or length; Klingenberg and Monteiro 2005).

In the context of geometric morphometrics, shape changes are part of the results of most multivariate analyses. Many multivariate procedures, including principal component or partial least squares analysis, provide a system of new coordinate axes that have particular properties (maximum variance, maximum covariance with other features, etc.). Because these axes are in the shape tangent space, their directions can be interpreted as features of shape variation – together with a magnitude and a sign ("up" or "down" direction along the axis), each axis can specify a shape change (Dryden and Mardia, 1998; Klingenberg and McIntyre, 1998; Klingenberg and Zaklan, 2000; Klingenberg et al., 2003b; Drake and Klingenberg, 2010). If the analysis is conducted in the space of Procrustes coordinates (projected to tangent space), the coefficients from the multivariate analysis (eigenvector for principal components, singular vector for partial least squares, etc.) can be directly used, with an appropriate choice of scaling, for visualizing the shape change. Similarly, multivariate regression of shape on some other variable yields a vector of regression coefficients that indicates the change of shape per unit of change in the independent variable (e.g., the shape change for an increase of size by one unit or per unit of time, etc.; Monteiro 1999; Drake and Klingenberg 2008; Rodríguez-Mendoza et al. 2011; Weisensee and Jantz 2011). For regression, the scaling of the shape change is defined because the regression analysis provides the amount of shape change expected per unit of change in the predictor variable. Other multivariate analyses yield results that can be visualized as shape changes in similar ways.

Shape changes always need to be visualized in conjunction with a shape. In order to interpret the change in shape, we need to interpret the relative displacements of landmarks in the context of their overall arrangement. This is partly due to our perception, which requires a shape as the context for making sense of a shape change. Yet, human perception is not the only reason for this. Whereas, in principle, it is straightforward to think of "transplanting" the same shape change vector from its original context to any other point in the same shape space, this does not necessarily make sense. Even though analyses of *Drosophila* wings and mouse mandibles may both use the same shape space for 15 landmarks in two dimensions (e.g. Klingenberg et al. 2003b; Breuker et al. 2006; Klingenberg 2009), it does not mean that a shape change taken from one of these systems and applied to the other has any biological meaning. Thus, shape changes are only interpretable in the context of the structure for which they were found and in conjunction with the shape of that structure. In addition, there is a slightly more subtle limitation: shape changes need to be expressed in the same coordinate system as the shapes from which they are computed, so that the alignment of the coordinate system of landmarks with the anatomical axes of the structure is consistently the same. This is usually not a problem if all shape changes are derived from the same Procrustes superimposition and visualizations are produced using starting and target shapes with the the same Procrustes alignment – this ensures that the shape change is aligned correctly in relation to the starting shape used

in the visualization (only once the visualization has been produced, it can be freely rotated, translated and scaled and as a whole).

Shape changes can vary in their magnitudes and directions. Because some biological phenomena are associated with relatively small shape variation, even biologically important shape differences can be subtle. Examples include deformations of skulls under mechanical loading (O'Higgins et al., 2011; O'Higgins and Milne, this issue), effects of quantitative trait loci (Klingenberg et al., 2001; Workman et al., 2002; Klingenberg et al., 2004), or directional asymmetry (Klingenberg et al., 1998, 2010), but the same applies to many studies of shape variation within populations because there is often relatively little variation at this level. To make such small shape changes better visible to viewers, visualizations often show them with exaggerated magnitudes. Choosing by how much to exaggerate a shape change is a compromise: the exaggerated shape change should be big enough to be clearly visible, but not so big that it results in major distortions. Of course, viewers need to be alerted, for instance in a figure legend, that the shape change has been exaggerated.

There are several ways of visualizing shape changes, which all have their advantages and shortcomings. Therefore, they need to be used and interpreted with appropriate caution. In the rest of this paper, I will discuss different options for visualization. I hope this discussion will be useful both for investigators who are facing choices of how to present results from their geometric morphometric analyses and for readers of morphometric studies who need to know how to interpret such graphs.

Landmark displacements and Procrustes superimposition

Because shape is defined as all the geometric features of an object except its size, position and orientation, one way to analyze the variation of shape in a data set is to remove the variation of size, position and orientation in some way. The remaining variation then concerns shape only. For data consisting of landmark coordinates, the variation of shape, position and orientation is most often removed with a Procrustes superimposition that achieves a best fit of landmark configurations according to a least-squares criterion (Sneath, 1967; Gower, 1975; Rohlf and Slice, 1990; Goodall, 1991; Dryden and Mardia, 1998). The least-squares Procrustes superimposition is of fundamental importance in geometric morphometrics because it is providing the link to the theory of Kendall's shape spaces (Small, 1996; Dryden and Mardia, 1998; Kendall et al., 1999), which provides a sound mathematical foundation for empirical studies of shape variation.

The generalized Procrustes superimposition, which is used for extracting shape information from samples of multiple landmark configurations, is an iterative procedure that fits each configuration to the mean shape in the sample as closely as possible. Variation in size is removed by scaling each configuration so that is has centroid size 1.0 (centroid size is a measure of size that quantifies the spread of landmarks around their centroid, or centre of gravity). Variation in position is removed by shifting the configurations so that they share the same centre of gravity. The remainder of the procedure deals with rotations to find an optimal orientation for each configuration. The procedure then aligns all configurations in the dataset to one particular configuration, for instance the first one, using least-squares Procrustes fitting so that the sum of squared distances between corresponding landmarks of each configuration and the common target configuration is minimal. After this step, an average shape is computed by averaging landmark positions (and rescaling to ensure that the centroid size of the average is exactly 1.0). In the next iteration, this average shape is used as the new target configuration and every configuration is fitted to it. A new average is then calculated, and the procedure is repeated until the average no longer changes (changes are usually negligible after two or three rounds).

As a result of this procedure, every configuration in the sample is optimally aligned to the average configuration. Also, unless the shape variation in the sample is unusually large, every configuration is nearly optimally aligned to every other configuration in the sample. Because the configurations are aligned so that size, position and orientation are kept constant according to the criterion for the least-squares fit, the remaining variation in landmark positions is due to variation of shape. Accordingly, the relative displacements of the landmarks from the mean configuration to any particular configuration, or the relative displacements from one shape to another shape nearby in shape space, can be used to assess the corresponding shape difference. These relative displacements provide a visualization of the shape change by showing how the landmarks are rearranged against each other when the non-shape components of variation, size position and orientation, are held constant with the Procrustes superimposition.

Note that I have referred consistently to relative displacements of landmarks. Because all the landmarks are included in the Procrustes superimposition and jointly determine the alignment of each configuration in relation to the mean shape, variation in the position of each landmark after superimposition is relative to the positions of all other landmarks. This interpretation of shape changes as relative shifts of landmarks against one another is central for all types of visualizations based on Procrustes fits.

Scatter of Procrustes-superimposed samples

At the beginning of most shape analyses, investigators want to get an overview of their data and the variation they contain. In geometric morphometrics, this is complicated by the fact that the raw coordinate data contain variation in the position and orientation of objects that is of no biological relevance (it simply reflects the variable positioning of specimens relative to the camera or other equipment used for obtaining images or landmark coordinates). The simplest way of viewing the data without these extraneous components of variation is to keep them constant by plotting the scatter of landmark locations after a Procrustes superimposition (Fig. 1). Similar representations are also available for three-dimensional data. This type of presentation of data is rather popular, both in textbooks (Dryden and Mardia, 1998; Monteiro and dos Reis, 1999; Claude, 2008; Weber and Bookstein, 2011; Zelditch et al., 2012) and in research papers (e.g. Bookstein et al. 1999; Robinson et al. 2001; Dworkin and Gibson 2006; Bruner et al. 2010; Webster 2011). It

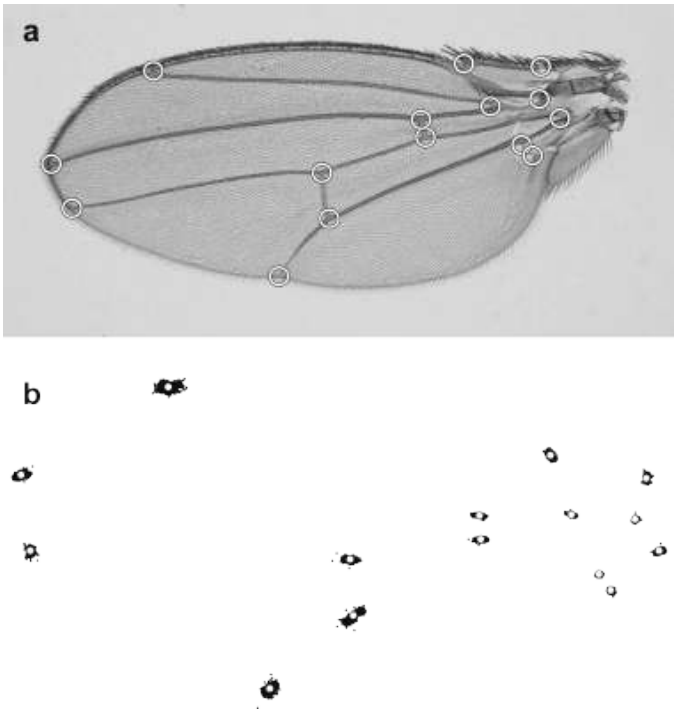


Figure 1 – Procrustes fit for the example of *Drosophila* wings. (a) The set of 15 landmarks used in the example (modified from Breuker et al. 2006). The dataset includes 834 wings of male and female *Drosophila melanogaster*, either wild type (Oregon-R strain) or heterozygous for the *spalt-major*^{1} mutation. (b) The landmark configurations of all wings in the dataset after Procrustes superimposition. For each landmark, the white circle indicates the location of the landmark for the average shape and the black dots indicate the locations for individual wings.

is therefore useful to discuss briefly what graphs of this kind can show and what they inevitably hide.

The example in Fig. 1b shows the scatter of landmark positions around the average shape after Procrustes superimposition of a sample of fly wings. It appears that the spread of positions for some of the landmarks in the distal part of the wing is considerably greater than for some of the landmarks at the wing base (but there are differences among the landmarks even within these regions). There is some justification for this impression, but interpreting the difference between the amounts of variation at different landmarks is complicated by the fact that variation present at any one landmark depends on all other landmarks. Both the amount of variation at other landmarks and their spacing in relation to each other can affect the variability of the position of a given landmark (readers can verify this by conducting a thought experiment or computer simulation in which they drastically increase variation at one landmark or move its average location far from the others and examine the consequence for the Procrustes superimposition). Therefore, the amounts of variation of positions of individual landmarks are not attributable to those landmarks on their own, but they result from the superimposition of the entire configurations.

The scatters of variation of several landmarks show a clear directionality, which, for some landmarks, lines up with anatomical features such as the wing veins in the *Drosophila* example (Fig. 1). This directionality indicates that shape variation is organized in specific ways so that it is concentrated mainly in certain directions, whereas other directions have less variation. In other words, shape variation appears to be integrated (e.g. Klingenberg 2008a). Although some patterns are clearly apparent, the scatter of landmark positions after the Procrustes fit shows only a part of the covariance structure in the data. What can be seen is the covariation between the x and y coordinates at each landmark. By contrast, the patterns of covariation among landmarks do not appear at all in the plot. To assess and display those patterns of integration, fully multivariate methods of analysis are required.

Plots or three-dimensional displays of landmark positions after Procrustes superimposition are often used as preliminary scans for differences among groups of observations in the data, for instance different species or age groups. Because the Procrustes fit does not take into account a possible group structure in the data, there is no guarantee that group differences will be visible in the plot even if they do exist. If such differences are very large, they may be visible, but even then this is by no means certain. For any purpose beyond a preliminary look at the data, fully multivariate analyses in the shape tangent space should be used. These analyses take into account the complete structure of the data, including the covariation among landmarks. For examining group structure in the data, for instance, there are analyses that specifically focus on differences between groups, such as discriminant analysis or canonical variate analysis (e.g. Rohlf et al. 1996; Duarte et al. 2000; Klingenberg et al. 2003a; Weinberg et al. 2009; Weisensee and Jantz 2011; Florio et al. 2012; Klingenberg et al. 2012). Analyses in the shape tangent space also produce graphical outputs that characterize group differences in optimal ways, because they can use all the information about shape variation, even if it does not appear in a display of Procrustes-aligned landmark positions.

Plots of landmark positions after Procrustes superimposition are useful for a first, informal assessment of the data. They cannot show the covariation among landmarks and therefore hide a fundamental aspect of the variation in morphometric data. Therefore, they are not suitable for a thorough and complete examination of the variation in the data. Even preliminary analyses such as the search for outliers in the data should use the complete information about variation in the landmark data and thus should be conducted in shape tangent space, not just by inspecting displays of superimposed landmark configurations. My recommendation is not to use this kind of graphs for formal presentation or publication.

Why shape changes should not be attributed to particular landmarks

The situation where variation is extremely localized to a small region of the landmark configuration, or even a single landmark, relative to the remaining landmarks is widely known as the "Pinocchio effect" and has been extensively discussed in the morphometrics literature (e.g. Siegel and Benson 1982; Chapman 1990; Rohlf and Slice 1990; Walker 2000). These discussions have pointed out that the least-squares Procrustes superimposition tends to spread variation from landmarks with greater variation to landmarks with less variation. This has been illustrated with examples and simulations where variation is concentrated in a single landmark versus all the others, which are constant in relation to reach other. Chapman (1990) provides the fictitious example of Pinocchio's nose, where the tip of the nose varies relative to all other landmarks of the head, and he demonstrates the example of a starfish with a single amputated and regenerating arm that is contrasted with the intact starfish before amputation, so that the tip of the amputated arm varies relative to all other landmarks of the starfish. These examples are used to argue that the superimposition by the least-squares Procrustes superimposition is in some way incorrect because the variation is "spread around" from the one variable landmark to all others. For instance,Chapman (1990, p. 260 f.) directly contrasted the Procrustes superimposition to a resistant-fit superimposition that he considered to be correct.

Such examples have considerable intuitive appeal and appear immediately convincing. We might ask, however, how we know that it is the tip of Pinocchio's nose that is moving anteriorly and not the remainder of his head moving posteriorly. Of course, the actual change is occurring in the tissue of the nose between the landmarks, not in the landmarks themselves. Therefore, the question which landmarks are responsible for a shape change misses the point: the tip of the nose moving forward (with the remainder of the head remaining in the same position) or the remainder of the head moving backward (with the tip of the nose remaining constant) are both describing the consequences of the lengthening of the nose. Both descriptions of the change are equivalent, and there is no obvious criterion for choosing one over the other. Curiously, however, we find that the description with just the tip of the nose moving seems quite natural, whereas the description with the remainder of the head moving away from the tip of the nose sounds distinctly odd. Apparently, we intuitively apply some criterion of parsimony when evaluating changes in the arrangements of landmarks, which makes us prefer descriptions that involve changes in as few landmarks as possible.

This bias in our perception makes us favour one way to characterize the change, as one landmark moving alone, over the equivalent one, as all other landmarks moving the opposite way, even though both are equally accurate and both descriptions equally miss the point that the change actually originates between the landmarks. Because this intuitive bias appears to be quite strong, we need to be aware of it and consider how it affects the ways in which we visualize and interpret shape changes.

Above all, we should resist the temptation to attribute shape changes to particular landmarks. Shape changes involve shifts in the positions of landmarks relative to other landmarks. This relative nature of landmark movements introduces a certain ambiguity in how a given shape change can be visualized. For any shape change, there are multiple ways of visualizing it (Fig. 2).

Because it is the same shape change, all these visualizations are equivalent, even though they may appear different. What differs between them is how the starting and target configurations are scaled, translated and rotated relative to each other. The differences are in the size, position and orientation of the two configurations, and therefore do not affect their shapes. Fig. 2 provides an example of this: the three panels show three ways in which the same two shapes can be compared. The landmark configurations after a Procrustes fit show differences in the positions of all landmarks (Fig. 2a). For the specific shapes used in this example, it is possible to change the sizes and positions so that the two configurations coincide in four of the five landmarks (Fig. 2b).

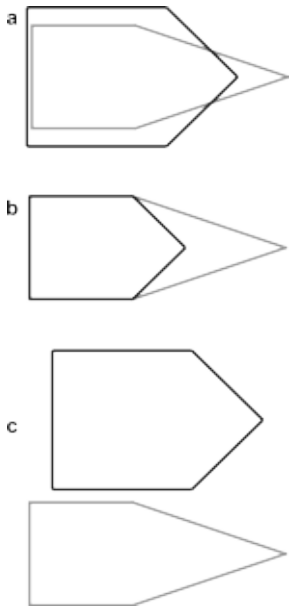


Figure 2 – Three equivalent ways of comparing two shapes. (a) Procrustes fit of the two configurations. (b) Superimposition so that four points match between the two configurations. (c) The two configurations drawn next to each other, in arbitrary positions. Note that the difference between the three ways of presenting the same shape change is in the size and position of the configurations (the orientation happens to be the same in all these three examples, but might also vary).

The Procrustes superimposition is optimal in the sense that it minimizes the sum of squared distances between corresponding landmarks (and also for various theoretical reasons), whereas the second type of superimposition minimizes the number of landmarks that differ in their positions (note that it is not always possible to obtain a superimposition for which more than two landmarks coincide in their positions). Finally, the configurations can be drawn side by side (Fig. 2c). This is in some ways the least problematic of the possible comparisons of shapes, but if the shape difference is subtle, it may be hard to see the difference.

Recall that all three displays are equivalent because the same two shapes are compared in each of them, so the shape change is the same. Yet, the absolute displacements of individual landmarks are quite different. What is constant are the relative displacements: in all three comparisons, it is clear that the landmark on the right protrudes further relative to the other four landmarks in the grey shape than in the black shape. But note that this change cannot be attributed to this landmark by itself, because the change might just as well be assigned to the other four landmarks, or to some change of all five landmarks in the configuration. When interpreting visualizations of shape changes, it is important for viewers to keep in mind this inherently relative nature of landmark displacements.

Arrows, lollipops and other visualizations based on landmark displacements

Various types of graphs visualize shape changes with landmark displacements after some kind of superimposition, most often the Procrustes superimposition. The simplest of these graphs indicate shifts of landmarks positions by lines or arrows, for instance in the "lollipop" graph, where the landmark positions of the starting shape are indicated by dots (the "candy" of the lollipop) and the shifts of landmarks to the target shape are represented as lines (the lollipop stick; Fig. 3a). Visualizations with arrows or lollipops are suitable in two or three dimensions (but they are somewhat cumbersome in the 3D context). Visualizations of this type were used from the beginning of geometric morphometrics (e.g., Siegel and Benson 1982; Chapman 1990; Rohlf and Marcus 1993), but more recently have been replaced mostly by more sophisticated visualization techniques that include more anatomical details.

Visualizations that contain only the shifts of landmark positions (Fig. 3a) are difficult to read because they do not provide any information that can help the viewer to relate those shifts of landmarks to the underlying anatomical structure. To provide that additional information and to make the graphs appear as more realistic representations of the structures under study, different ways are available to indicate the morphological context of the landmarks. One option is to use a pair of superimposed wireframe graphs that connect the landmarks with straight lines for the starting and target shapes (Fig. 3b; e.g. Siegel and Benson 1982; Baylac and Penin 1998; Marcus 1998; Frost et al. 2003; Willmore et al. 2005; Drake and Klingenberg 2008; Weisensee and Jantz 2011) or by superimposed outline drawings that are warped using the thin-plate spline according to the information from the landmarks (see below; Fig. 3c).

All these types of graphs show shifts of the landmarks from the starting to the target shape. It is important for viewers to keep in mind that these shifts, at any landmark, are relative to the shifts at all the other landmarks. No matter what graphical means are used to indicate anatomical context, this relative nature of landmark movements is shared by all visualizations based on superimpositions. Also, viewers need to remember that different registrations of the same starting and target shapes might result in displays with fundamentally different appearance.

Displacements of surfaces in three dimensions

Over the past decade, several methods for analysing the shape of entire surfaces have been proposed that are based on superimposing the entire surface directly. In this context, shape changes are often visualized with a representation of the surface that is coloured according to a heat map that represents the distance of the surfaces at each point – for instance, "cold" colours for areas where the target shape recedes within the start-

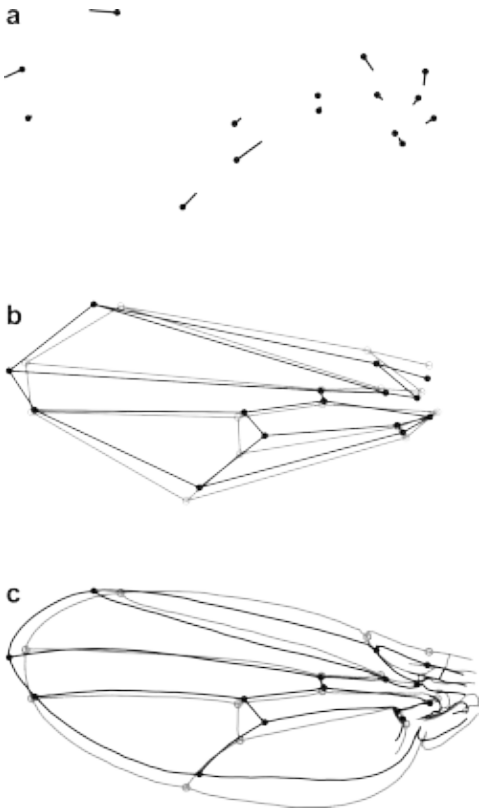


Figure 3 – Three ways of visualizing the same shape change that are based on the Procrustes superimposition. (a) A lollipop diagram, in which the positions of landmarks in the starting shape are shown as dots and the shifts of landmarks to the target shape are indicated by lines. (b) A wireframe graph with two wireframes: a gray one for the starting shape and a black one for the target shape. (c) A graph with a warped outline drawing (gray) for the starting shape and another one (black) for the target shape. The thin-plate spline was used, separately for the starting and target shapes, to warp the outline drawing to match the landmark configurations.

ing shape, and "warm" colours where the target shape bulges out of the starting shape (e.g., Zollikofer and Ponce de Léon 2002; Hammond et al. 2005; Kristensen et al. 2008; Claes et al. 2011). The displacements of surfaces are computed as the displacement of each point on one surface in relation to the nearest point on the other surface. These point-to-point relations can be interpreted as the equivalent of displacements between landmarks in approaches that use discrete landmark points.

In common with the visualization methods based on landmark displacements, the displacements of points on the surfaces, and thus the visualizations of those distances via heat-map colouration, depend critically on the specific superimposition that is used for aligning the surfaces. Different superimposition methods may produce markedly different results. This is particularly relevant for surfaces, because no standard method for aligning surfaces exists that is used universally. Above all, the "hot" and "cold" spots on the surface that correspond to the regions with greatest positive and negative displacements (one or the other surface is outside) are not local features, but depend on the entire surfaces. For instance, whether a part of the nose appears more prominent in one surface that the other not only depends on the nose in the two surfaces, but also on the remainder of the face or head that is included in the analysis. Even though there are no landmark displacements in this sort of study, the difficulties with interpreting this type of visualization are the same as with graphs based on landmark shifts.

Transformations

Visualizations based on the transformation grids of D'Arcy Thompson have played a special role in geometric morphometrics. The method was originally presented as part of an argument for the importance of mathematical and physical ideas for understanding animal forms (Thompson, 1961; Arthur, 2006). Even though the physical analogies have not have a substantial influence on contemporary understanding of growth and evolution, Thompson's appealing diagrams have captured the imagination of many biologists through most of the twentieth century, and efforts to develop a rigorously quantitative approach for making such diagrams contributed significantly to the development of geometric morphometrics (Sneath, 1967; Bookstein, 1978, 1989; Rohlf, 1993).

The idea is that changes of biological forms can be characterized by examining the transformations that have to be applied to the coordinate space in order to change one form into another. Imagine that you draw one form on a rubber sheet and stretch and compress it in different directions until the form drawn on the rubber sheet matches the other form exactly. Then the transformation that has been applied to the rubber sheet can be used to characterize the difference between the two forms. To make this transformation visible directly, a rectangular grid can be drawn on the rubber sheet before it is transformed. When the rubber sheet is stretched and compressed so that the two forms match, the originally rectangular grid is distorted and shows the nature of the transformation. In practice, of course, no rubber sheet is used, but the lines of the grid are fitted with the aid of morphological landmarks that are recognisable on both forms. Thompson fitted the lines of the grid by hand, but now this can be done computationally, using some interpolation method that guarantees that the relation of the grid lines to the landmark positions is correct. The available methods are based on different mathematical or statistical considerations (e.g. Dryden and Mardia 1998, chapter 10), but none of them has any biological basis. The method that is currently by far the most widely used is the thin-plate spline interpolation (Bookstein, 1989; Dryden and Mardia, 1998).

The thin-plate spline

The thin-plate spline is an interpolation technique that was brought into morphometrics as a flexible and mathematically rigorous implementation of D'Arcy Thompson's transformation grids (Bookstein, 1989). The thin-plate spline is a technique that guarantees that the corresponding points of the starting and target form appear precisely in corresponding positions in relation to the untransformed and transformed grids (something that is not guaranteed if the grids are drawn by hand or with some other computational approaches) and it provides the smoothest possible transformation for any pair of starting and target forms (in the sense of minimizing second derivatives). Much has been written in the morphometrics literature about the metaphor of an infinitely thin and infinitely large metal plate and about the related notion of bending energy, the measure of localized deformation. The key point for morphometric application is the property of the smoothest possible transformation, which also emerges from the consideration of deforming a metal plate because the metal plate will resist abrupt bending (a greater bending energy would be required). The thin-plate spline method has a number of desirable properties. For instance, unlike some earlier methods such as trend surface analysis (Sneath, 1967), the thin-plate spline always produces transformation grids where the changes diminish towards the margins, outside the region occupied by landmarks.

The thin-plate spline provides a one-to-one mapping not only between the landmarks of the starting and target forms, but also between every point of the plane or three-dimensional space in which the starting form is embedded and the plane or space of the target shape. It is this interpolation to the whole plane or space that provides the opportunity for using the thin-plate spline to draw transformation grids, as the mapping can be done for the points on the grid lines in the plane or space of the starting shape to obtain the transformed grids in the plane or space of the target shape.

In addition to its use as a tool for visualization, the thin-plate spline has also been used as a method to decompose shape changes into different geometric components: a component of uniform (or affine) shape change that is the same throughout the entire configuration and a non-uniform component of shape change that is localized to various degrees (Bookstein, 1989, 1991, 1996c; Rohlf and Bookstein, 2003). The only justification for this distinction was the suggestion that biomechanical effects on shape should be uniform (Bookstein, 1991, 1996b,c; Rohlf and Bookstein, 2003), but this idea has been discredited by numerous studies showing that biomechanical forces are highly non-uniform (e.g. Pierce et al. 2008; Fortuny et al. 2011; O'Higgins et al. 2011), which is unsurprising given the localized nature of joints, muscle insertions and similar structures involved in biomechanical function. The non-uniform component can be further broken down into partial warps, which are geometrically separate components of a shape change. Together, the uniform component and all partial warps provide a coordinate system for analysing shape variation, which has been widely used in geometric morphometrics (Rohlf, 1993). The complete set of partial warps and uniform components forms a coordinate system in shape tangent space that differs from the coordinate system of projected Procrustes coordinates only by a rotation (Rohlf, 1993, 1999). The results of multivariate analyses of the same data based on the two coordinate systems are therefore equivalent. Because of the considerable conceptual and computational complexity of the partial warps and uniform component, it is preferable to use Procrustes tangent coordinates as the basis for morphometric analyses, and to limit the use of the thin-plate spline as a tool for visualization.

Transformation grids

Transformation grids produced with the thin-plate spline (e.g. Fig. 4) are in such widespread use in geometric morphometrics that they have become one of its most familiar tools, and they are covered in detail in textbooks of geometric morphometrics (Bookstein, 1991; Monteiro and dos Reis, 1999; Zelditch et al., 2012) and statistical shape analysis (Dryden and Mardia, 1998). The immediate visual appeal of transformation grids reinforces this familiarity. But despite this familiarity, or perhaps because of it, there has not been very much discussion about transformation grids as a visualization tool. Here, I will present some caveats that should be taken into account when using and interpreting transformation grids.

A first and most fundamental point is to keep in mind that the grids are based on an analogy, but do not represent actual biological phenomena. The analogy of the rubber sheet is an interesting and intuitive way to think about biological shape change, but there is usually no direct equivalent to such a smooth elastic deformation in the biological

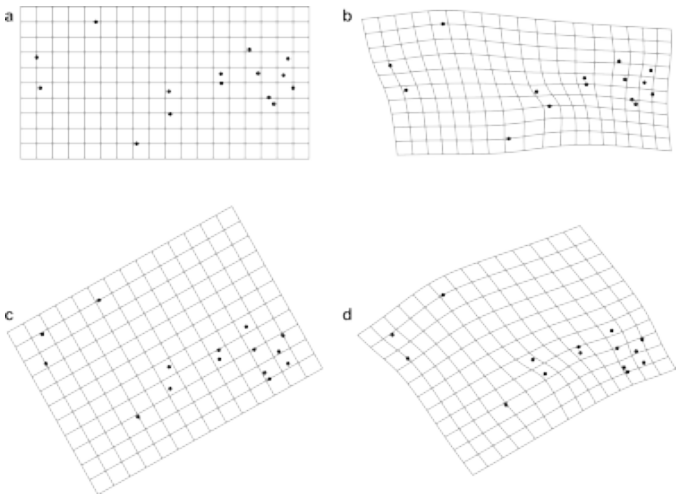


Figure 4 – Visualizations using the thin-plate spline, using the same shape change as in Fig. 3 (see also Fig. 1). (a) The starting shape, with grid lines aligned so that the horizontal grid lines approximately follow the anterior–posterior compartment boundary of the wing. (b) The target shape, aligned as in (a). (c) The starting shape, with the grid in a different alignment relative to the wing. (d) The target shape, with the grid aligned as in (c). Note that example of a shape change is not an extreme case of the effect of a change in how the grid is aligned to the structure.

processes that produce biological shapes and shape variation. Accordingly, the transformation grids do not depict a biological reality, but are an imagined aid for visualization. Unlike other visual aids such as arrows or lollipops, which indicate relative shifts of landmark positions, transformation grids render the space between the landmarks, precisely where no data are available. Literally, for that reason, transformation grids are pure fiction!

It is debatable whether the median (sagittal) plane is an exception to this line of reasoning. One might argue that the median plane is not just imagined, but reflects an anatomical reality. It is tangibly embodied in structures such as the nasal septum, and the midline also plays an important role in development. Accordingly, one might argue with some justification that the median plane is a real developmental and anatomical entity.

Other anatomical lines and planes, however, arbitrarily cut through various structures. These planes may be useful for establishing a frame of reference for alignment (e.g. the Frankfurt plane in human anatomy), but in general they have no biological significance beyond that. Viewers should keep this in mind when interpreting visualizations of shape changes using transformation grids. Some software packages allow users to specify arbitrary planes in visualizations of three-dimensional that are then warped with the thin-plate spline to visualize some shape change (e.g. O'Higgins and Jones 1998). In visualizations of shape changes that contain such warped surfaces, it is usually difficult to determine what their spatial positions and anatomical relations are, and it is therefore doubtful whether they contribute to the viewer's understanding of the shape change. Even if such warped surfaces provide the viewer with a better feeling of the transformation as interpolated by the thin-plate spline, it is an open question whether that reflects the actual shape change.

The algorithms used for computing transformation grids, such as the thin-plate spline, have no biological basis but are based exclusively on geometric or statistical criteria. The interpolation between landmarks is therefore problematic and cannot be expected to be biologically realistic. In the immediate vicinity of landmarks, the behaviour of the grids is driven by the nearby landmarks and therefore by the actual biological data. In regions that are devoid of landmarks, however, the influence of the biological information diminishes with increasing distance from the landmarks. Therefore, particularly in regions that are at some distance from the nearest landmarks, transformation grids need to be interpreted with caution.

Depending on the alignment of the rectangular grid relative to the anatomical features in the starting form (cf. Fig. 4a versus Fig. 4c), the visual appearance of transformation grids can differ even if the shape change is the same (Fig. 4b versus Fig. 4d). Also, the visual impression of a shape change depends on whether grid lines pass through the regions with the most accentuated localized deformations. Therefore, transformation grids that are shifted relative to the landmark configurations or that differ in the intervals between grid lines (i.e. the density of the grids) can make the same shape change look quite different. These effects are not always obvious to the viewer, particularly if the starting shape with the rectangular grid is omitted to save space, as it is often done in publications (imagine Fig. 4b,d without the comparison to Fig. 4a,c). Depending on the software that is used, these properties of the visualizations can be changed by the user. In general, it is a good idea to choose an alignment that has a clear anatomical meaning (e.g. the anterior–posterior direction or median line parallel to the grid lines). In some cases, this requires an active choice by the user, because the default options will not result in an anatomically meaningful alignment. For instance, if landmark configurations from half-skulls are aligned according to the major axes of the mean configuration, the grid lines usually will be at oblique angles to the median axis or plane, which is not an anatomically sensible arrangement. Software packages for geometric morphometrics have options for users to choose anatomically meaningful alignments and other important properties of visualizations (e.g. the number of horizontal and vertical grid lines), but users need to make active choices because the default options often are not the best choice for particular study situations.

Transformation grids such as those in Fig. 4, with just the landmark positions and the grids, can be difficult to interpret because they do not provide anatomical information. This became particularly evident for me when I discovered that several transformation grids in a paper by a well-established group of investigators, published about a decade ago in a high-profile journal, are upside-down in relation to the other figures that explain the landmarks and other aspects of the study. The labels of the diagrams are printed correctly, which indicates that the figure in question was not just flipped upside-down by the printers, but the mistake happened to the authors and passed scrutiny by the reviewers, editors, and readers of the paper (the paper in question has been cited several dozen times, but it and its authors shall remain unnamed here). There has been no correction to the paper, nor was the mistake pointed out in a forum such as the Morphmet E-mail list, so that it seems the mistake has mostly gone unnoticed. If a fairly crude mistake such as upside-down transformation grids can go unnoticed even by the authors of a study, chances are that many readers will routinely overlook many of the more subtle anatomical features depicted in transformation grids.

Warped outlines or surfaces

The pitfalls of "bare" transformation grids can be avoided by providing additional information so that the viewer can easily see the anatomical context of the landmark configuration. This context can be provided by a drawing of the structure. Such drawings were an integral part of visualizations with transformation grids from the start, as D'Arcy Thompson (1961) always included a drawing of the structure with the transformation grid (the grids were derived from those drawings). The visualizations with bare grids only emerged later, when the grids were obtained by computational methods.

Another option is to use the techniques for warping transformation grids but to apply them to the drawing only, without the grid. Because the thin-plate spline interpolation works for every point of the image plane in the surroundings of the landmarks, it can be used not only to warp a rectangular grid, but just as well a drawing of the structure under study (Fig. 3c). The drawings can be simple outlines (e.g. Klingenberg and Leamy 2001) or quite detailed drawings (e.g. Rodríguez-Mendoza et al. 2011). Similarly, in three-dimensional applications, a scanned surface model can be warped (Fig. 5; e.g. Wiley et al. 2005).

For these visualizations with warped drawings or surface models, some of the same caveats apply as for transformation grids, but others do not. Because the information about shape change resides entirely in the landmarks, the interpolation between landmarks must be interpreted with caution, particularly in regions that are relatively far

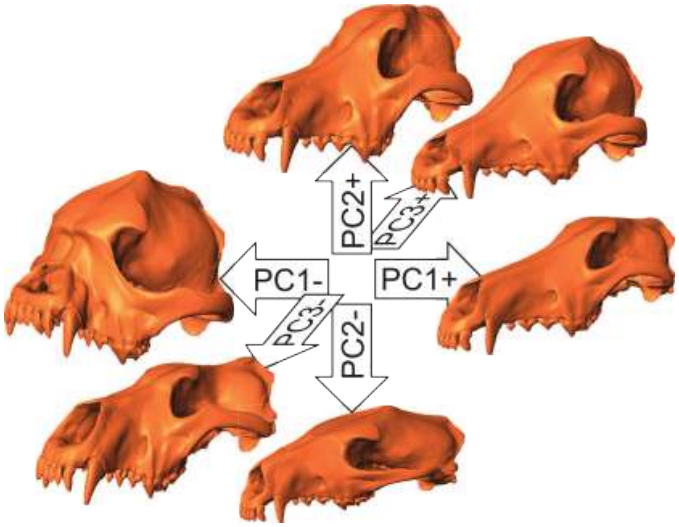


Figure 5 – Visualizations of shape changes with warped surfaces. A scanned surface of a wolf skull was warped with the thin-plate spline to the shapes near the extremes of the scores for the first three principal components of an analysis of domestic dogs and wild carnivorans (from Drake and Klingenberg 2010, Fig. 3D). This example shows shape changes at a fairly large scale, demonstrating that this method of visualization is fairly robust. Some of the changes are strongly reminiscent of actual extreme forms in the analysis (e.g. the contrast between dogs with short and elongate skulls), whereas other aspects are not realistic (e.g. the canine teeth, because no landmarks were digitized at the tips of the canines).

from the nearest landmarks (e.g. the posterior margin near the base of the wing in Fig. 3c). Because the drawing can be chosen to show all the anatomically relevant information, many of the ambiguities of the visualizations with "bare" transformation grids do not apply to warped drawings or surfaces. The interpretation of shape changes is therefore easier than for transformation grids. The main problem for authors is to choose an appropriate level of detail for the drawing that corresponds reasonably well to the coverage of the structure with landmarks. Also, investigators should avoid using drawings or surfaces that extend far beyond the landmarks (e.g. a drawing of a bird skull including the beak if no landmarks have been digitized on the beak).

In three dimensions, presentation of warped surfaces has become easier with advances in technology and standard file formats used for distribution. Even simple, "flat" images of warped surfaces can depict three-dimensional shape changes in an accessible way (e.g. Fig. 5), but new possibilities for distributing such visualizations are emerging that go beyond flat images. For instance, it is now possible to include entire three-dimensional surfaces within figures in published papers, so that readers, after downloading the respective PDF file from a journal web site, can rotate the structure under control of their computer's mouse (or similar control device) to view the warped surface from any direction (e.g. Drake 2011).

Discontinuities

A limitation of the transformation approach is that it fails in some circumstances, when the assumption of continuous deformation is violated. This assumption of continuity means that there is a one-to-one correspondence between every point in one of the structures under consideration and some particular point in every other structure under study. This correspondence applies both to the landmarks and to points between the landmarks (the difference is that the correspondence is not known for the latter, but is inferred through interpolation with an algorithm such as the thin-plate spline). Transformations can easily deal with groups of landmarks converging towards each other, as long as there is at least a small distance left between them, or with sets of landmarks diverging from each other or rotating relative to the remaining landmarks in the configurations. The transformation approach fails if continuous deformations are not sufficient to characterize a shape change – if multiple landmarks shift to the same point or, in reverse, if landmarks diverge from the same point and thus generate a "hole" in the transformation grid between them, or if portions of the landmark configuration "flip over" relative to the remainder so that the transformation grid folds over. All these problems disrupt the one-to-one correspondence of points in the planes (or 3D spaces) of the shapes under comparison, which is a fundamental assumption of the transformation approach.

Such changes are fairly widespread and cannot just be dismissed as rare exceptions (Klingenberg, 2008b; Oxnard and O'Higgins, 2009; Gómez-Robles et al., 2011). Examples are the disappearance of certain structures, such as the loss of the nasal bone in the dugong, structures that can bifurcate so that a single part in one organism corresponds to two parts in another, as it occurs for tooth cusps, or growth processes such as the closing of the fontanelle of many mammals, where the relations between adjacent cranial bones change and thus drastically affect the landmarks on those bones and on the sutures between them. Other examples of changes that violate the assumption of continuous change involve sets of landmarks that are defined in different ways: for instance, landmarks defined by the bones of the skull versus landmarks defined by the locations of muscle insertion. Because muscle insertions can shift relative to the structures of the skull, there can be drastic shifts in the relative arrangements of the respective sets of landmarks, which produce discontinuities that cannot be represented sensibly by the transformation approach (Oxnard and O'Higgins, 2009). Again, this is not a unique example, but similar problems can arise whenever different sets of landmarks can be displaced relative to each other (another example might be landmarks on butterfly wings that are defined by wing veins versus colour patterns).

A closely related situation is the "switching" of positions among nearby landmarks. Zelditch et al. (2012, p. 30f.) mention this as a possible problem in morphometric studies,giving an example of a cranial foramen that can appear on either side of a suture. The consequence is an abrupt localized deformation in the region of the landmarks in question. Accordingly, deformation grids and related visualizations will show marked distortions in that region, for instance grids that are folding over, which violate the assumption of the transformation approach that there is an unambiguous one-to-one correspondence between each point in the two- or three dimensional space in which each specimen is recorded. This invalidates the interpolation step that is central to the transformation approach. If the landmarks that switch position are sufficiently close to each other, the resulting shape change may not be very large (as measured by Procrustes distance between the shapes).

All these cases pose no serious problem for geometric morphometric methods, provided that investigators use definitions of landmarks that take the situation into account. Landmarks can be displaced relative to each other in any way without problems for the Procrustes superimposition and the various multivariate methods used in geometric morphometrics. Problems only arise for the transformation approach as a tool for visualization of shape changes, because all these examples involve discontinuous changes that transformations cannot properly represent. Therefore, these problems can be circumvented by simply choosing another type of visualization.

Discussion

Although visualization of shape changes has been a central task of geometric morphometrics since its inception, there has been surprisingly little discussion of the approaches used for visualization. That different authors strongly favour different visualization methods is not just a matter of personal taste, but it indicates that these methods come with contrasting strengths and weaknesses. There are no "right" or "wrong" methods among the main approaches used for visualizations. Each method, however, comes with its own underlying ideas and conventions that have been established through repeated use and of which the viewer needs to be aware to interpret the resulting graphs correctly. Because these ideas and conventions tend to be so familiar and intuitive to every morphometrician, at least for the type of visualization that he or she prefers, they are rarely thought about or discussed explicitly. This paper has attempted to make the ideas and conventions explicit

and compare them between the main approaches for visualization that are used in geometric morphometrics.

Making and interpreting visualizations are parts of a communication process. Authors and viewers of illustrations that visualize shape changes need to understand the visualization methods in order to communicate effectively and avoid misunderstandings. Here I sum up the main conclusions of this paper and make some recommendations.

Shifts of superimposed landmarks must not be interpreted one-by-one

In visualizations of shape changes based on shifts of superimposed landmarks, a shift is shown at every landmark, but it is important that these shifts are relative to all other landmarks. It is particularly important to keep in mind this relative nature of landmark shifts when describing shape changes and reading such descriptions – there are multiple ways of characterizing the same shape change. For instance, whether the tip of Pinocchio's nose moves anteriorly (relative to the rest of the head) or whether the rest of the head moves posteriorly (relative to the tip of the nose) does not make a difference to the shape change. Viewers need to keep in mind that the same shape change can produce different-looking patterns of landmark shifts, depending on the superimposition. Accordingly, equivalent visualizations of the same shape change might look quite different from each other and show the greatest shifts for different landmarks. Also, the same shape change might yield descriptions that sound very different because they involve different sets of landmarks (e.g. Pinocchio's nose versus the rest of the head) and thus might appear to support different biological interpretations.

A possible alternative to graphs of superimposed landmark configurations, which authors should consider seriously, is to display the starting and target shapes side by side. Exaggerating the shape change makes it clearly visible (finding the right factor for exaggeration is a matter of trial and error – the shape change should be clearly visible but not lead to gross distortions). This method avoids the ambiguity of the superimposition graphs but is usually very effective, particularly when used in combination with wireframes and similar visualization tools (e.g. O'Higgins and Jones 1998; Klingenberg et al. 2012).

Transformation grids are fiction

Despite their immediate visual appeal, it is important to keep in mind that transformation grids are merely a mathematical construct that provides a means for visualizing shape changes but do not represent a biological reality. Quite literally, these grids are fiction.

Nevertheless, with the appropriate caveats, transformation grids are a very effective tool for visualizing shape changes. The key point for viewers is to interpret them critically. In particular, the grids are likely to be unreliable guides to change in regions that are relatively far from the nearest landmarks – in these regions, the transformation is mostly resulting from the warping algorithm and not informed by biological data.

Authors should provide images of starting shapes with untransformed grids, so that viewers can understand the anatomical context of the grids (Fig. 4). For transformation grids that are used as part of three-dimensional visualizations, it is critical that authors provide detailed explanations of the position and anatomical relations of the plane used for the rectangular grid.

Using deformed outline drawings or surface models avoids some of the aspects that make transformation grids artificial, but it shares all the problems concerning the fact that the warping criteria are biologically arbitrary. Therefore, as with the grids, such visualizations need to be interpreted cautiously in regions that are relatively far from landmarks.

Despite these caveats, this method of visualization is probably the best one that is currently available because of its straightforward visual appeal and the direct biological relevance of all the elements in the graphs. To avoid the problems of superimposition outlined above, the warped outlines or surfaces are best shown side by side (Fig. 5).

Visualization as communication

In geometric morphometrics, visualization of shape changes is a key element in the exploration of data, formulating and testing of hypotheses and reporting of results to others. Just as scientific writing is best viewed as a form of communication between authors and readers, visualization is best considered as part of the same communication process. Accordingly, authors should aim to produce visualizations that are clear and easy to interpret, so that the viewers' task of reconstructing the meaning of the shape changes is straightforward and the risk of misunderstandings is minimal. In turn, viewers should be aware of the conventions and assumptions that are inherent to the various types of visualizations.

To make this communication effective, both authors and viewers of the visualizations should try to understand each other's perspective. Just as in technical writing (e.g. Gopen 2004), visualization is more effective if the author takes into account the viewers' expectations and general conventions. If an unusual type of visualization is necessary, authors should pay particular attention to provide sufficient explanations as text in publications or verbally for oral presentations. In particular, it is important to provide the information what the starting and target shapes are and whether the shape change is exaggerated (and if so, by how much). Explanations of shape changes should explicitly and consistently point out the relative nature of landmark shifts, and authors should not assume that readers will remember that landmark shifts are relative when reading a description. In turn, viewers should examine visualizations of shape changes carefully and also examine the accompanying explanations. Sometimes, an author's statement is not following these recommendations and does not really mean what it is literally saying, for instance, if the statement "landmark X is shifted dorsally" is meant as an abbreviation for something like "landmark X is shifted dorsally relative to other landmarks in the region" (but the exact meaning depends on the particular context). Readers need to anticipate and recognize the use of such shorthand or careless wording and grasp the correct interpretation that is behind them. In other words, authors should strive to be helpful to the viewers and readers of their visualizations and associate explanations, whereas readers should make the effort of examining the author's reasoning and logic.

The goal of visualizations is to communicate complex shape changes as a means for discovering and disseminating patterns of morphological variation in their full anatomical context. Sophisticated visualization techniques are not ends in themselves, but are means to support authors in sharing biological insights from their morphometric analyses with viewers. Used in this manner, the current methods for visualization of shape changes provide powerful means for communicating complex results in an intuitive and appealing manner, and future advances can continue to make important contributions to the development of geometric morphometrics.

References

Arthur W., 2006. D'Arcy Thompson and the theory of transformations. Nat. Rev. Genet. 7: 401–406.

Astúa D., 2010. Cranial sexual dimorphism in New World marsupials and a test of Rensch's rule in Didelphidae. J. Mammal. 91: 1011–1024.

Baylac M., Penin X., 1998. Wing static allometry in *Drosophila simulans* males (Diptera, Drosophilidae) and its relationships with developmental compartments. Acta Zool. Acad. Sci. Hung. 44: 97–112.

Bookstein F.L., 1978. The measurement of biological shape and shape change. Springer-Verlag, Berlin.

Bookstein F.L., 1989. Principal warps: thin-plate splines and the decomposition of deformations. IEEE Trans. Pattern Anal. Mach. Intell. 11: 567–585.

Bookstein F.L., 1991. Morphometric tools for landmark data: geometry and biology. Cambridge University Press, Cambridge.

Bookstein F.L., 1996a. Biometrics, biomathematics and the morphometric synthesis. Bull. Math. Biol. 58: 313–365.

Bookstein F.L., 1996b. Combining the tools of geometric morphometrics. In: Marcus L.F., Corti M., Loy A., Naylor G.J.P., Slice D.E. (Eds.) Advances in morphometrics. Plenum Press, New York. 131–151.

Bookstein F.L., 1996c. Standard formula for the uniform shape component in landmark data. In: Marcus L.F., Corti M., Loy A., Naylor G.J.P., Slice D.E. (Eds.) Advances in morphometrics. Plenum Press, New York. 153–168.

Bookstein F.L., Schäfer K., Prossinger H., Seidler H., Fieder M., Stringer C.B., Weber G.W., Arsuaga J.-L., Slice D.E., Rohlf F.J., Recheis W., Mariam A.J., Marcus L.F., 1999. Comparing frontal cranial profiles in archaic and modern *Homo* by morphometric analysis. Anatomical Record (New Anatomist) 257: 217–224.

Breuker C.J., Patterson J.S., Klingenberg C.P., 2006. A single basis for developmental buffering of *Drosophila* wing shape. PLoS ONE 1(1): e7. doi:10.1371/journal.pone.0000007
Bruner E., Martin-Loeches M., Colom R., 2010. Human midsagittal brain shape variation: patterns, allometry and integration. J. Anat. 216: 589–599.
Chapman R.E., 1990. Conventional Procrustes approaches. In: Rohlf F.J., Bookstein F.L. (Eds.) Proceedings of the Michigan morphometrics workshop. University of Michigan Museum of Zoology, Ann Arbor, MI. 251–267.
Claes P., Walters M., Vandermeulen D., Clement J.G., 2011. Spatially-dense 3D facial asymmetry assessment in both typical and disordered growth. J. Anat. 219: 444–455.
Claude J., 2008. Morphometrics with R. Springer, New York.
Drake A.G., 2011. Dispelling dog dogma: an investigation of heterochrony in dogs using 3D geometric morphometric analysis of skull shape. Evol. Dev. 13: 204–213.
Drake A.G., Klingenberg C.P., 2008. The pace of morphological change: historical transformation of skull shape in St. Bernard dogs. Proc. R. Soc. Lond. B Biol. Sci. 275: 71–76.
Drake A.G., Klingenberg C.P., 2010. Large-scale diversification of skull shape in domestic dogs: disparity and modularity. Am. Nat. 175: 289–301.
Dryden I.L., Mardia K.V., 1998. Statistical shape analysis. Wiley, Chichester.
Duarte L.C., Monteiro L.R., Von Zuben F.J., Dos Reis S.F., 2000. Variation in mandible shape in *Trichomys apereoides* (Mammalia: Rodentia): geometric analysis of a complex morphological structure. Syst. Biol. 49: 563–578.
Dworkin I., Gibson G., 2006. Epidermal growth factor receptor and transforming growth factor-β signaling contributes to variation for wing shape in *Drosophila melanogaster*. Genetics 173: 1417–1431.
Florio A.M., Ingram C.M., Rakotondravony H.A., Louis E.E., Jr., Raxworthy C.J., 2012. Detecting cryptic speciation in the widespread and morphologically conservative carpet chameleon (*Furcifer lateralis*) of Madagascar. J. Evol. Biol. 25: 1399–1414.
Fortuny J., Marcé Nogué J., De Esteban-Trivigno S., Gil L., Galobart À., 2011. Temnospondyli bite club: ecomorphological patterns of the most diverse group of early tetrapods. J. Evol. Biol. 24: 2040–2054.
Frost S.R., Marcus L.F., Bookstein F.L., Reddy D.P., Delson E., 2003. Cranial allometry, phylogeography, and systematics of large-bodied papionins (Primates: Cercopithecinae) inferred from geometric morphometric analysis of landmark data. Anat. Rec. 275A: 1048–1072.
Gidaszewski N.A., Baylac M., Klingenberg C.P., 2009. Evolution of sexual dimorphism of wing shape in the *Drosophila melanogaster* subgroup. BMC Evol. Biol. 9: 110.
Gómez-Robles A., Olejniczak A.J., Martinón-Torres M., Prado-Simón L., Bermúdez de Castro J.M., 2011. Evolutionary novelties and losses in geometric morphometrics: a practical approach through hominin molar morphology. Evolution 65: 1772–1790.
Goodall C.R., 1991. Procrustes methods in the statistical analysis of shape. J. R. Statist. Soc. B 53: 285–339.
Gopen G.D., 2004. The sense of structure: writing from the reader's perspective. Pearson Longman, New York.
Gower J.C., 1975. Generalized Procruses analysis. Psychometrika 40: 33–51.
Hammond P., Hutton T.J., Allanson J.E., Buxton B., Campbell L.E., Clayton-Smith J., Donnai D., Karmiloff-Smith A., Metcalfe K., Murphy K.C., Patton M.A., Pober B., Prescott K., Scambler P., Shaw A., Smith A.C.M., Stevens A.F., Temple I.K., Hennekam R.C.M., Tassabehji M., 2005. Discriminating power of localized three-dimensional facial morphology. Am. J. Hum. Genet. 77: 999–1010.
Kendall D.G., Barden D., Carne T.K., Le H., 1999. Shape and shape theory. Wiley, Chichester.
Klingenberg C.P., 2008a. Morphological integration and developmental modularity. Annu. Rev. Ecol. Evol. Syst. 39: 115–132.
Klingenberg C.P., 2008b. Novelty and "homology-free" morphometrics: What's in a name? Evol. Biol. 35: 186–190.
Klingenberg C.P., 2009. Morphometric integration and modularity in configurations of landmarks: tools for evaluating a-priori hypotheses. Evol. Dev. 11: 405–421.
Klingenberg C.P., 2010. Evolution and development of shape: integrating quantitative approaches. Nat. Rev. Genet. 11: 623–635.
Klingenberg C.P., Barluenga M., Meyer A., 2003a. Body shape variation in cichlid fishes of the *Amphilophus citrinellus* species complex. Biol. J. Linn. Soc. 80: 397–408.
Klingenberg C.P., Duttke S., Whelan S., Kim M., 2012. Developmental plasticity, morphological variation and evolvability: a multilevel analysis of morphometric integration in the shape of compound leaves. J. Evol. Biol. 25: 115–129.
Klingenberg C.P., Leamy L.J., 2001. Quantitative genetics of geometric shape in the mouse mandible. Evolution 55: 2342–2352.
Klingenberg C.P., Leamy L.J., Cheverud J.M., 2004. Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible. Genetics 166: 1909–1921.
Klingenberg C.P., Leamy L.J., Routman E.J., Cheverud J.M., 2001. Genetic architecture of mandible shape in mice: effects of quantitative trait loci analyzed by geometric morphometrics. Genetics 157: 785–802.
Klingenberg C.P., McIntyre G.S., 1998. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. Evolution 52: 1363–1375.
Klingenberg C.P., McIntyre G.S., Zaklan S.D., 1998. Left-right asymmetry of fly wings and the evolution of body axes. Proceedings of the Royal Society of London B, Biological Sciences 265(1402): 1255–1259.
Klingenberg C.P., Mebus K., Auffray J.-C., 2003b. Developmental integration in a complex morphological structure: how distinct are the modules in the mouse mandible? Evol. Dev. 5: 522–531.
Klingenberg C.P., Monteiro L.R., 2005. Distances and directions in multidimensional shape spaces: implications for morphometric applications. Syst. Biol. 54: 678–688.
Klingenberg C.P., Wetherill L.F., Rogers J.L., Moore E.S., Ward R.E., Autti-Rämö I., Fagerlund Å., Jacobson S.W., Robinson L.K., Hoyme H.E., Mattson S.N., Li T.K., Riley E.P., Foroud T., CIFASD, 2010. Prenatal alcohol exposure alters the patterns of facial asymmetry. Alcohol 44: 649–657.
Klingenberg C.P., Zaklan S.D., 2000. Morphological integration between developmental compartments in the *Drosophila* wing. Evolution 54: 1273–1285.
Kristensen E., Parsons T.E., Hallgrímsson B., Boyd S.K., 2008. A novel 3-D image-based morphological method for phenotypic analysis. IEEE Trans. Biomed. Eng. 55: 2826–2831.
Marcus L.F., 1998. Variation in selected skeletal elements of the fossil remains of *Myotragus balearicus*, a Pleistocene bovid from Mallorca. Acta Zool. Acad. Sci. Hung. 44: 113–137.
Marcus L.F., Hingst-Zaher E., Zaher H., 2000. Application of landmark morphometrics to skulls representing the orders of living mammals. Hystrix 11(1): 27–47. doi:10.4404/hystrix-11.1-4135
Monteiro L.R., 1999. Multivariate regression models and geometric morphometrics: the search for causal factors in the analysis of shape. Syst. Biol. 48: 192–199.
Monteiro L.R., dos Reis S.F., 1999. Princípios de morfometria geométrica. Holos, Ribeirão Preto.
O'Higgins P., Cobb S.N., Fitton L.C., Gröning F., Phillips R., Liu J., Fagan M.J., 2011. Combining geometric morphometrics and functional simulation: an emerging toolkit for virtual functional analyses. J. Anat. 218: 3–15.
O'Higgins P., Jones N., 1998. Facial growth in *Cercocebus torquatus*: an application of three-dimensional geometric morphometric techniques to the study of morphological variation. J. Anat. 193: 251–272.
O'Higgins P., Milne N., 2013. Applying geometric morphometrics to compare changes in size and shape arising from finite elements analyses. Hystrix 24(1) (Online First). doi:10.4404/hystrix-24.1-6284
Oxnard C.E., O'Higgins P., 2009. Biology clearly needs morphometrics. Does morphometrics need biology? Biol. Theor. 4: 84–97.
Pierce S.E., Angielczyk K.D., Rayfield E.J., 2008. Patterns of morphospace occupation and mechanical performance in extant crocodilian skulls: a combined geometric morphometric and finite element modeling approach. J. Morphol. 269: 840–864.
Robinson D.L., Blackwell P.G., Stillman E.C., Brook A.H., 2001. Planar Procrustes analysis of tooth shape. Arch. Oral Biol. 46: 191–199.
Rodríguez-Mendoza R., Muñoz M., Saborido-Rey F., 2011. Ontogenetic allometry of the bluemouth, *Helicolenus dactylopterus dactylopterus* (Teleostei: Scorpaenidae), in the Northeast Atlantic and Mediterranean based on geometric morphometrics. Hydrobiologia 670: 5–22.
Rohlf F.J., 1993. Relative warp analysis and an example of its application to mosquito wings. In: Marcus L.F., Bello E., García-Valdecasas A. (Eds.) Contributions to morphometrics. Museo Nacional de Ciencias Naturales, Madrid. 131–159.
Rohlf F.J., 1999. Shape statistics: Procrustes superimpositions and tangent spaces. J. Classif. 16: 197–223.
Rohlf F.J., Bookstein F.L., 2003. Computing the uniform component of shape variation. Syst. Biol. 52: 66–69.
Rohlf F.J., Loy A., Corti M., 1996. Morphometric analysis of Old World Talpidae (Mammalia, Insectivora) using partial-warp scores. Syst. Biol. 45: 344–362.
Rohlf F.J., Marcus L.F., 1993. A revolution in morphometrics. Trends Ecol. Evol. 8: 129–132.
Rohlf F.J., Slice D.E., 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39: 40–59.
Siegel A.F., Benson R.H., 1982. A robust comparison of biological shapes. Biometrics 38: 341–350.
Small C.G., 1996. The statistical theory of shape. Springer-Verlag, New York.
Sneath P.H.A., 1967. Trend-surface analysis of transformatrion grids. Journal of Zoology 151: 65–122.
Thompson D.A.W., 1961. On growth and form. Cambridge University Press, Cambridge.
Walker J.A., 2000. Ability of geometric morphometric methods to estimate a known covariance matrix. Syst. Biol. 49: 686–696.
Weber G.W., Bookstein F.L., 2011. Virtual anthropology: a guide to a new interdisciplinary field. Springer, Vienna.
Webster M., 2011. The structure of cranidial shape variation in three early ptychoparioid trilobite species from the Dyeran-Delamarean (traditional "lower-middle" Cambrian) boundary interval of Nevada, U.S.A. J. Paleontol. 85: 179–225.
Weinberg S.M., Andreasen N.C., Nopoulos P., 2009. Three-dimensional morphometric analysis of brain shape in nonsyndromic orofacial clefting. J. Anat. 214: 926–936.
Weisensee K.E., Jantz R.L., 2011. Secular change in craniofacial morphology of the Portuguese using geometric morphometrics. Am. J. Phys. Anthropol. 145: 548–559.
Wiley D.F., Amenta N., Alcantara D.A., Ghosh D., Kil Y.J., Delson E., Harcourt-Smith W., Rohlf F.J., St. John K., Hamann B., 2005. Evolutionary morphing. Proceedings of the IEEE Visualization 2005 (VIS'05): 431–438.
Willmore K.E., Klingenberg C.P., Hallgrímsson B., 2005. The relationship between fluctuating asymmetry and environmental variance in rhesus macaque skulls. Evolution 59: 898–909.
Workman M.S., Leamy L.J., Routman E.J., Cheverud J.M., 2002. Analysis of quantitative trait locus effects on the size and shape of mandibular molars in mice. Genetics 160: 1573–1586.
Zelditch M.L., Swiderski D.L., Sheets H.D., 2012. Geometric morphometrics for biologists: a primer. Elsevier, Amsterdam.
Zollikofer C.P.E., Ponce de León M.S., 2002. Visualizing patterns of craniofacial shape variation in *Homo sapiens*. Proc. R. Soc. Lond. B Biol. Sci. 269: 801–807.

Associate Editor: A. Cardini

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 25–32, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/6282/pdf
doi:10.4404/hystrix-24.1-6282

Research Article

Morphometrics and the comparative method: studying the evolution of biological shape

Leandro Rabello Monteiro[a,*]

[a] *Laboratório de Ciências Ambientais, CBB, Universidade Estadual do Norte Fluminense, Av. Alberto Lamego 2000, Pq. Califórnia, Campos dos Goytacazes, RJ, cep 28013-620, Brasil*

Keywords:
Adaptation
morphometrics
comparative methods
phylogenetic analysis
multivariate analysis

Article history:
Received: 27 May 2012
Accepted: 9 July 2012

Acknowledgements
The author would like to thank M.R. Nogueira, A. Cardini, J. Claude, and A. Loy for important comments and improvement suggestions on previous versions of the manuscript. Work by LRM is funded by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (bolsa de produtividade em pesquisa, processo 304118/2010-9) and Fundação Estadual de Amparo à Pesquisa do Rio de Janeiro (FAPERJ) (Cientista do Nosso Estado, processo E-26/103.044/2011).

Abstract

Phylogenetic comparative methods are one of the most important parts of the morphometric toolkit for studies of morphological evolution. The assessment of repeated independent events of evolution of phenotypic and associated ecological-functional traits is still a starting point for the study of adaptation, but modern comparative approaches go beyond correlative methods, allowing for the modeling of evolutionary scenarios and analyses of trait evolution patterns. The evidence for adaptive change due to ecological diversification is stronger (even if still circumstantial) if models that predict increases in diversification rate fit the data well and the morphological changes are associated with ecological and functional changes. A large body of literature is dedicated to methodological and theoretical aspects of comparative methods, but in the context of univariate traits. On the other hand, biological shape is a complex trait, and morphometric data is essentially multivariate. Whereas most comparative methods allow for direct multivariate extensions, dimension reduction is an almost certain requirement due to the high dimensionality of morphometric data sets and the large number of evolutionary parameters that need to be estimated by comparative methods. Objective methods with considerable statistical support to determine data dimensionality exist, but the applied literature usually relies on subjective criteria to assess how many shape dimensions should be retained. The most appropriate calculation and interpretations of principal components, the most popular dimension reduction method, are also topics that should be considered more carefully in applications. The maturity of comparative methods and the development of model-based approaches linking macroevolutionary patterns and microevolutionary processes provide an exciting perspective for the study of morphological evolution.

Introduction

The study of interspecific comparative data sets has a long history in evolutionary biology. In this context, the recurrent association between phenotypic and ecological traits in different lineages is often used as evidence, even if circumstantial, of adaptation (Harvey and Pagel, 1991; Losos, 2011a), defined non-historically as the phenotypes that present higher fitness than other phenotypes in a given environment (Martins, 2000; Lewens, 2007). The strength of the evidence is related to the number of times that a given combination of phenotype and environment arises independently during the evolution of a lineage (Harmon et al., 2005). Felsenstein (1985) called attention to the fact that trait values for species at the tips of a phylogenetic tree are not necessarily independent pieces of evidence in statistical analysis, because part of the trait's evolutionary history is shared by common descent. Felsenstein pointed out issues with previous attempts to deal with the statistical problem of non-independence and proposed an elegant method to deal with the problem, known as phylogenetic independent contrasts.

A period of effervescence followed, when a number of methods were proposed for the statistical analysis of phylogenetic non-independent data (Cheverud et al., 1985; Grafen, 1989; Garland et al., 1993; Hansen, 1997; Martins and Hansen, 1997; Diniz-Filho et al., 1998). Despite a sharp increase in popularity, the interpretation of results from comparative analyses, the assumptions being made, and the appropriate combination of methods were not always straightforward. Many authors discussed aspects of application, interpretation or whether comparative methods were useful at all (Westoby et al., 1995; Björklund, 1997; Losos, 1999; Freckleton et al., 2002; Freckleton, 2009). According to the jargon commonly used in the early period, the comparative methods would sometimes transform the original variables "correcting" for phylogenetic non-independence (phylogenetic independent contrasts) or decompose variation into "phylogenetic" (autocorrelation) or "heritable" (mixed models) components versus specific or residual components. Another possibility was to incorporate phylogenetic non-independence in the residuals using a generalised linear model (phylogenetic generalised least squares – PGLS). The analogies between statistical results from the different models with evolutionary processes started to appear in the literature, such as the decomposition of variation into a phylogenetic (reflecting "historical constraints") and a specific component (reflecting "adaptation") via autocorrelation (Cheverud et al., 1985). In these early contributions, phylogenetic non-independence was interpreted as evidence of phylogenetic constraint (a concept similar to the modern definition of phylogenetic inertia – see below), and only correlations among specific values were accepted as evidence for the action of selection, both interpretations no longer considered valid (Hansen and Orzack, 2005). The multitude of methods, applications and interpretations generated some conflicts, particularly when different authors used the same concept names with different definitions (as in phylogenetic inertia – see below). A number of reviews related statistical properties, grouping methods and pointing out similarities, differences and limitations (Martins, 2000; Rohlf, 2001; Freckleton et al., 2002; Martins et al., 2002; Hansen and Orzack, 2005). A book summarised some of the early approaches (Harvey and Pagel, 1991), but the methodological and conceptual developments have been so massive during the last decade that new books on the subject (both entry level and advanced) are urgently needed again.

* Corresponding author
Email address: lrmont@uenf.br (Leandro Rabello Monteiro)

doi:10.4404/hystrix-24.1-6282

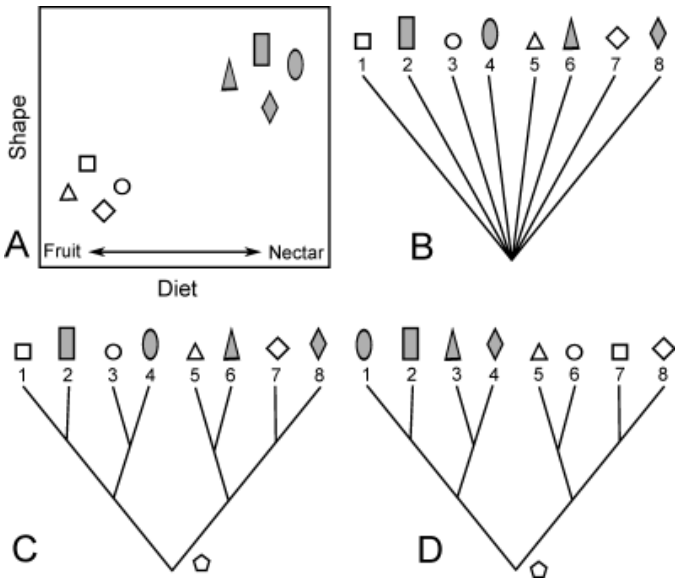


Figure 1 – Phylogenetic structure of a hypothetical correlational study, associating morphology and diet in a lineage of bats. A) Pattern of association between shape and diet, where one aspect of shape (elongation) is associated with dietary differences (relative contribution of fruit versus nectar, represented as shades filling the geometric figures). The axes relate to groups of multiple variables, but are represented in a bivariate scatterplot for simplicity. B) Star phylogeny pattern of covariance, where each OTU represents an independent piece of evidence. C) Phylogenetic structure where the evolutionary transition in diet and shape was observed four times independently. The assumed shape and diet of the ancestor (Frugivory, not elongated) is represented near the root of the tree. D) Phylogenetic structure where the evolutionary transition in diet and shape was observed only once.

Statistical approaches in comparative methods

The statistical problem tackled by comparative methods is represented in Fig. 1. When attempting to correlate two groups of variables (shape and diet) with values measured on OTUs (tips of the phylogeny – extant species), one might come to a result as in Fig. 1A, showing an association between shape (the elongation of the figures) and diet (relative importance of fruit and nectar). The correct interpretation of this pattern and its significance as evidence that natural selection might be responsible for the observed association depends on the phylogenetic covariance structure assumed for the residuals of this model. Common statistical methods (i.e. non-phylogenetic) assume that all observations are independent as in Fig. 1B. However, if the phylogenetic structure in Fig. 1C is assumed, the diet transition between frugivory and nectarivory was observed four times independently (the postulated ancestor was a frugivore), leading to a decrease in the number of degrees of freedom. The phylogenetic structure in Fig. 1D is what Felsenstein (1985) called a "worst case" phylogeny, where the diet transition was observed only once, and the four nectarivore species are not independent observations. The strength of circumstantial evidence for adaptation is greater if we assume the phylogeny in Fig. 1C than the one in Fig. 1D, but that is a statistical consequence of the reduced number of degrees of freedom. Even if we assume the phylogeny in Fig. 1D as true, it does not mean that the shape differences are "caused" or "constrained" by phylogeny.

Translating this problem into a working statistical model requires the definition of three components: hypotheses, sampling assumptions (random, independent?) and statistical assumptions (distribution of response variables) (McPherson, 1990). Of greater concern here are the sampling assumptions, as phylogenetic non-independence is the motivation behind the use of phylogenetic comparative methods. Considering the phylogeny in Fig. 1C as the actual sampling structure, the method of phylogenetic independent contrasts transforms the original values of shape and diet in a way that the resulting observations (contrasts) are standardised differences among tips or estimated nodes descending from the same immediate common ancestor (Felsenstein, 1985). Correlations of contrasts for different traits were expected to arise independently of similarities due to common ancestry. A statistically equivalent approach would be to use a generalised linear model that incorporates the structure of phylogenetic covariance among observations into the error term (Martins and Hansen, 1997; Rohlf, 2001), as can be done for non-independent spatial or time series data. In this case, the expected phylogenetic covariance for any pair of species would be proportional to the sum of branch lengths leading from the root to their last common ancestor. Using branch lengths to directly estimate phylogenetic covariances assumes that the response variables evolved according to a Brownian motion model (BM), where differences among species are proportional to the branch lengths leading to their most recent ancestor. This assumption, however, can be flexible, and other evolutionary models can be used by transforming branch lengths or species covariances to reflect alternative models, such as stabilising selection or allowing for varying rates of character evolution (Martins and Hansen, 1997; Freckleton et al., 2002; Blomberg at al., 2003).

Most applications of comparative methods assume that the species means used are representative and that within-species variation is negligible when compared to among-species variation (Garamszegi and Moller, 2010). Whereas this assumption might hold true for a number of studies, it has been shown that ignoring intraspecific variability can lead to increased type I error rates when intraspecific variances are large and sample sizes are small (Harmon and Losos, 2005). One concern for multivariate morphometric studies is the downward bias shown to influence estimates of correlation and regression coefficients when measurement error is large (Ives et al., 2007). Specific modifications of well known methods (phylogenetic independent contrasts, PGLS) were proposed to take measurement error (and within-species biological variability) into account when estimating parameters (Ives et al., 2007; Felsenstein, 2008), and the potential for these methods in morphometrics is great, both providing less biased estimates of covariances and correlations or in using maximum likelihood to compare observed and hypothetical covariance matrices or levels of variation (intraspecific, interspecific) (Felsenstein, 2008).

Early attempts to use quantitative genetics theory to discriminate between evolutionary processes came from the expectation of among-population covariances under genetic drift proposed by Lande (Lande, 1979, 1980) and the associated tests, comparing among and within population covariance matrices (Lofsvold, 1988; Ackermann and Cheverud, 2002), and more recently even incorporating phylogenetic non-independence (Revell, 2007). Modern comparative approaches use model selection to infer best fitting evolutionary models or scenarios from expected OTU means instead of among-group covariances (Butler and King, 2004; Hansen et al., 2008) and to estimate evolutionary parameters reflecting historical changes of evolutionary rates of trait diversification (O'Meara et al., 2006; Revell at al., 2012). Going back to the example discussed, one might then ask, given the data in Fig. 1A and the tree in Fig. 1D, whether the likelihood of the pattern of differences observed is higher under a Brownian motion model (which encompasses both neutral evolution or random fluctuations of adaptive optima over time) or under a stabilising selection model with two optima (for eating fruit or nectar), that function as attractors in the model and can be interpreted as fixed adaptive peaks. In this model-based approach, the focus moves from the number of independent branches showing the evolutionary transitions of interest, to the maintenance of adaptive optima under stabilising selection (Hansen, 1997). Because the evolutionary mechanisms are modeled directly, the methods also provide a way to estimate parameters relating postulated processes to observed interspecific patterns in the evolution of quantitative traits, such as the strength of selection towards an adaptive optimum for a character (Hansen and Orzack, 2005), or evolutionary divergence rates (a measure of phenotypic variation in the phylogeny) that might change along different branches (O'Meara et al., 2006; Revell at al., 2012).

Conceptual issues with comparative methods

The maturity of comparative methods have also provided clearer definitions for concepts that have been confused and misinterpreted for a long time, such as phylogenetic signal (or effect) and phylogenetic inertia, considered synonyms by some authors (Losos, 1999). Phylogen-

etic signal is a pattern of statistical non-independence, where phenotypic similarity is associated with phylogenetic relatedness (Revell et al., 2008; Klingenberg and Gidaszewski, 2010), whereas phylogenetic inertia is the tendency of a trait to resist a current adaptive force (Blomberg and Garland, 2002; Hansen and Orzack, 2005). Phylogenetic inertia can be caused by constraints in development or variation (canalisation, trade-offs due to correlations with lower fitness traits), and is itself recognised as one of the causes behind phylogenetic effects (Harvey and Pagel, 1991), along with adaptive explanations. The confusion between the concepts of phylogenetic inertia and signal led to a misconception of phylogenies as sources of variation in statistical models (Losos, 2011b), and the wrong interpretation of phylogenetic signal as evidence of "constraint". Phylogenetic inertia and adaptation are not mutually exclusive explanations for interspecific differences, nevertheless, they can only be properly assessed in comparisons where their effects on each other are jointly controlled for (Hansen and Orzack, 2005; Hansen et al., 2008). The phylogenetic generalised least squares (PGLS) models provide a useful framework for such analyses. Adaptation can be tested and estimated by the main effects in a PGLS model and inertia can be tested and estimated from the phylogenetic signal in the residual (or error) term of the same PGLS model (Hansen et al., 2008).

Several measures and tests of phylogenetic signal strength have been proposed (Blomberg and Garland, 2002; Freckleton et al., 2002; Blomberg at al., 2003; Klingenberg and Gidaszewski, 2010; Diniz-Filho et al., 2012). Interpretations relating specific values of these phylogenetic signal statistics to evolutionary processes have been proposed. For example, Blomberg's K statistic is expected to be one if trait evolution behaves as expected by a Brownian motion model, whereas it is less than one if species are less similar than expected by phylogenetic relatedness (too many convergences?) and greater than one if species are more similar than expected by neutral evolution (stabilising selection?). However, the link between evolutionary patterns and processes has been challenged by simulation studies (Revell et al., 2008), showing that different models can lead to similar patterns of species similarity within a phylogeny. The model-based approach discussed before provides a more robust basis for evolutionary process inference in the context of comparative data. Measures of phylogenetic signal have also been proposed as a means to determine whether it is necessary to use a comparative method or not, implying that traits with low or non-significant phylogenetic signals would not require the phylogenetic non-independence to be included in the model (Losos, 1999; Klingenberg and Gidaszewski, 2010; Losos, 2011b). This approach is no longer recommended, because the statistical models make assumptions about the distribution and independence of the error (differences between observed and predicted), not the raw data (Revell, 2010). Testing for phylogenetic signal on the traits directly to decide whether to use a phylogenetic comparative method is an error equivalent to testing for normality and homoscedasticity in raw data, before a linear model is fitted (Hansen and Orzack, 2005). The strength of phylogenetic signal can be jointly estimated with the statistical model to determine the most appropriate form of the covariance error matrix (assumptions regarding the evolutionary model), using a number of alternative measures of phylogenetic signal (Martins and Hansen, 1997; Blomberg at al., 2003). As a result, if the error structure of the response variable does not show strong phylogenetic signal, the phylogenetic covariance matrix will approach the identity matrix assumed when errors are expected to be independent. Because the comparative methods are flexible regarding the evolutionary model and the strength of phylogenetic signal, it is advisable to always include phylogenetic information in the statistical models.

An important issue that is frequently overlooked is the fact that studies showing correlations between ecological and morphological changes are by no means considered direct evidence of causation (Martins, 2000). A number of reasons (alternative to direct causation) can be invoked to explain correlations (Losos, 2011a), for example, selection on body size (associated with dietary differences) has probably led to correlated allometric shape differences in the skull of new world monkeys (Marroig and Cheverud, 2005). When many traits are correlated with the same fitness differences, it can be complicated to discern which ones are under selection just from correlation results. In spite of the problems, correlation studies are an important source of patterns that require further investigation by in depth studies measuring the strength of selection or the experimental link between performance, morphology, and ecology (Losos et al., 1997; Winter and von Helversen, 2003; Langerhans and DeWitt, 2004; Langerhans et al., 2004; Nogueira et al., 2009; Losos, 2011a). The problem of inferring evolutionary origin from correlation is by no means exclusive to phylogenetic comparative analyses, but a more general issue in evolutionary biology (Martins, 2000), and the subject of a deep philosophical debate (Sober, 1993; Lewens, 2007). The difficulty of inferring causation from correlation is also firmly rooted in Popper's principle of falsificationism, where it is much easier to falsify a hypothesis than to prove it (Paipneau, 2003). A single negative finding is sufficient to disprove a hypothesis, whereas no number of supporting findings will be considered a conclusive proof.

Comparative methods and morphometric data

Shape is defined as all the geometrical information that remains after location, scale and rotational effects are filtered out from an object (Dryden and Mardia, 1998), and the study of shape is essentially a multivariate undertaking. The simplest morphometric data structure is a triangle of landmarks (Bookstein, 1991), requiring two variables to describe its shape variation. As a result, morphometric studies of comparative data are presented with the additional difficulty of either extending the statistical models to multiple response variables when feasible, or using a method to reduce dimensionality to one or just a few variables. A review of published papers that presented results of morphometric analyses of comparative data provides a summary of the diversity of approaches as discussed below. In this review, I did not discriminate among geometric morphometric studies using landmark coordinates or "traditional" morphometric studies using distances measured among landmarks as data, because the relevant multivariate methods are the same for both kinds of data.

Most comparative methods are readily extensible to multivariate data, but comparative studies seldom use shape variables (as Procrustes aligned coordinates or partial warp plus uniform component scores) directly in the statistical models. They reduce the dimensionality of the data set first (using their principal components also called relative warps). It is also important to note that shape is not necessarily the response variable, and the position of shape variables in the models will depend on the hypotheses being tested. Although comparative studies mostly associate shape variation with ecological or functional variables, morphological diversification within lineages can be associated with amounts of speciation (Adams et al., 2009), or the evolution of specific morphologies can be associated with shifts in cladogenesis (Fitzjohn, 2010). One possibility for correlational studies is to use PGLS models to fit regressions between matrices of shape variables and functional and ecological variables (Rüber and Adams, 2001; Clabaut et al., 2007; Meloro et al., 2008; Raia et al., 2010). Phylogenetic independent contrasts can also be calculated for each shape variable before associating them with contrasts for ecological variables, either by multivariate regression (Figueirido et al., 2010) or partial least squares (PLS) (Klingenberg and Ekau, 1996). This approach should be statistically equivalent to PGLS, as long as the among-taxa covariances are exactly proportional to the branch lengths (assuming a Brownian motion evolutionary model) in the phylogenetic tree (Rohlf, 2001). Another possibility is to correlate morphometric distances with ecological and phylogenetic distances using a matrix correlation method (Harmon et al., 2005; Young et al., 2007; Astua, 2009; Monteiro and Nogueira, 2010). This distance matrix-based approach is less informative than other comparative methods (PGLS, model-based approaches), and is considered a less powerful option (Peres-Neto and Jackson, 2001). The multivariate models discussed above have been mostly used for significance testing, whereas the visualisation of shape variation patterns has been almost exclusively dependent on principal components analysis (as in the phylomorphospace discussed below).

The sets of shape variables can be considerably large, and the most common approach by far is the reduction of shape variables (where the observations are the species means) to a set of principal components (PCs) and fitting comparative statistical models using one or a few PCs. Dimension reduction by principal components analysis (PCA) is a technique commonly used in multivariate analysis to reduce multidimensional data sets to a small number of interpretable axes that retain a maximum amount of variation (Jolliffe, 2002). These shape PCs can be associated (using comparative methods) with ecological variables (or PCs of these) that might explain the patterns of variation among species, but they need to be carefully interpreted, because although they correspond to axes of major variation in shape space, they are not designed to maximise correlation or covariance with any set of ecological variables (as partial least squares would). This is important when interpreting non-significant results, because not finding a clear association between shape PCs and ecological variables does not mean a significant association does not exist for a different linear combination of shape variables and does not exclude multivariate significance (the ecological variables are just not associated with the main axes of interspecific variation).

Studies that reduced the original set of variables into principal components for comparative analysis have used them as univariate variables (one PC at a time) or smaller multivariate data sets to calculate regressions or correlations of phylogenetic independent contrasts (Bergmann and Irschick, 2011; Brusatte et al., 2012), fitting PGLS regression models (Nogueira et al., 2009), and evaluating alternative evolutionary scenarios with the model-based approach (Bergmann et al., 2009; van Buskirk, 2009; Collar et al., 2009). Meloro (2012) used a slightly different approach to evaluate the association between shape and functional variables in the mandible of carnivores, employing partial least squares (PLS) to extract pairs of linear combinations within both sets of variables that explain most cross-covariance between sets (Rohlf and Corti, 2000). The PLS axes were then "validated" by PGLS regression of the shape (as dependent) on diet and size. In this case, the PLS will not construct a linear combination of shape variables that maximises the variation among species, but that should be associated with the variables of interest. One possible problem with this approach, as pointed out by Revell (2009), is that the phylogenetic non-independence needs to be incorporated in the estimates of covariances themselves, not just in *a posteriori* statistical tests (further discussed below).

The principal components are more appropriately used as a multivariate set in comparative analyses (Losos, 1990; Monteiro and Nogueira, 2011). If any subset of PCs will be used as variables (as opposed to the set explaining 100% of total variation), a criterion is needed to determine how many PCs should be retained. A popular criterion in morphometrics is the broken-stick, where the percentage of variance explained by each principal component is compared with a null distribution of expected percentages when the total variation is randomly distributed among principal components (Harmon et al., 2005; Morgan, 2009; Brusatte et al., 2012). However, the vast majority of studies use subjective criteria, such as the number of PCs that sums up more than 70 or 80% of total variation (considered reasonable amounts), or retaining the PCs that account for more than 10 or 5% by themselves. On the other hand, well defined criteria to assess dimensionality of a data set do exist and extensive simulation studies have pointed to a number of well-performing stopping rules for PCs (Peres-Neto et al., 2005). One consensus method in the statistical literature for best performance in recovering subjacent dimensionality is Horn's parallel analysis (Franklin et al., 1995; Dinno, 2009), where a number of data sets with random uncorrelated variables are generated and the distribution of random eigenvalues compared with the observed values (see the R package `paran` – http://CRAN.R-project.org/package=paran – for an implementation of the method. A simplified R code for covariance matrices is also available from the author). Using a well defined criterion to determine the number of overdispersed PCs (the ones that account for more variance than expected in random uncorrelated samples) should prove more informative in determining shape space dimensionality than subjective criteria based on absolute amounts of variance explained. One caveat of parallel analysis is that it is unable to determine dimensionality of spaces larger than $\frac{p}{2.5}$, where p is the number of variables (A. Dinno, *pers. comm.*), which is not a relevant issue for most morphometric applications. A more relevant issue is that because parallel analysis attempts to estimate the "true" number of latent dimensions using statistical inference, a sample size/variables ratio larger than 3 is recommended (although the effect of high dimensional low sample size data is still to be determined with simulations).

The estimation of principal components from interspecific data is usually performed without taking the phylogenetic non-independence into account. Revell (2009) pointed out that when the observations are phylogenetically related, this procedure is not optimal, and proposed a method to estimate principal components that take phylogenetic non-independence into account. The procedure uses the among-taxa $(n \times n)$ phylogenetic covariance matrix $\mathbf{C}$ (the same used in PGLS, for example) to allow for the estimation of the eigenstructure. The elements of the $\mathbf{C}$ matrix can be flexible, depending on the evolutionary model assumed. If a Brownian motion model of evolution is assumed, the elements of $\mathbf{C}$ will be γt_{ij}, where t_{ij} is the distance along the phylogenetic tree between the root and the last common ancestor for species pair ij, and γ is a parameter related to the magnitude of trait variation (Martins and Hansen, 1997), also referred to in the literature as the "rate of evolution" along a branch (Eastman et al., 2011). A more flexible model that allows for stabilising selection constraints would have the elements of $\mathbf{C}$ estimated as $\gamma \exp[-\alpha t_{ij}]$, where α is interpreted as the strength of selection towards an adaptive optimum (Martins and Hansen, 1997). Alternatively, the off-diagonal elements of $\mathbf{C}$ can be multiplied by λ $(0 \leq \lambda \leq 1)$, estimated from the data, to take into account varying strengths of phylogenetic signal (Freckleton et al., 2002). Starting with a data matrix $\mathbf{X}$ (n taxa and p shape variables), one first estimates a vector $\mathbf{a}(p \times 1)$ of phylogenetic means, corresponding to the estimated ancestral shape (for morphometric data) at the root:

$$\mathbf{a} = [(\mathbf{1}^\top \mathbf{C}^{-1} \mathbf{1})^{-1} \mathbf{1}^\top \mathbf{C}^{-1} \mathbf{X}]^\top, \quad (1)$$

where $\mathbf{1}$ is a $n \times 1$ vector of ones. The evolutionary variance-covariance matrix among variables is calculated as:

$$\mathbf{R} = (n-1)^{-1} (\mathbf{X} - \mathbf{1}\mathbf{a}^\top)^\top \mathbf{C}^{-1} (\mathbf{X} - \mathbf{1}\mathbf{a}^\top). \quad (2)$$

The covariance matrix $\mathbf{R}$ is also called the evolutionary rate matrix, because it estimates a matrix of Brownian rate parameters (Revell and Harmon, 2008; Revell and Collar, 2009). The covariances in the off-diagonal indicate associated differences relative to an estimated ancestral shape inversely weighted by the phylogenetic covariances in $\mathbf{C}$ (the influence of neighbouring branches is removed by negative weights). As in PGLS models, multiplication by the inverse of $\mathbf{C}$ ensures that the correct error structure (phylogenetic non-independence) is used in estimating the covariances. The evolutionary covariance matrix above is equivalent to a covariance matrix calculated from phylogenetic independent contrasts, if $\mathbf{C}$ assumes a Brownian motion evolutionary model (Revell, 2009). The PCA of the evolutionary variance-covariance matrix can be performed with the spectral decomposition $\mathbf{R} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$, and the PC scores are computed in the original space as

$$\mathbf{S} = (\mathbf{X} - \mathbf{1}\mathbf{a}^\top)\mathbf{V}. \quad (3)$$

These phylogenetic PCs will indicate directions of maximum evolutionary rates in shape space because the shape changes are standardised by branch lengths. Revell (2009) strongly emphasised that this procedure does not produce "phylogenetic corrected" PC scores. It just estimates the eigenstructure properly (as judged by simulations of multivariate data evolving under a Brownian motion model), taking the phylogenetic non-independence into account. If phylogenetic non-independence was present in the original data, the phylogenetic PC scores will be non-independent as well. The assumption of evolutionary model is included in the derivation of matrix $\mathbf{C}$ elements (phylogenetic covariances), which can be changed to reflect different underlying evolutionary models (Martins and Hansen, 1997; Freckleton et al., 2002; Blomberg at al., 2003). The R package `phytools` (Revell, 2012)

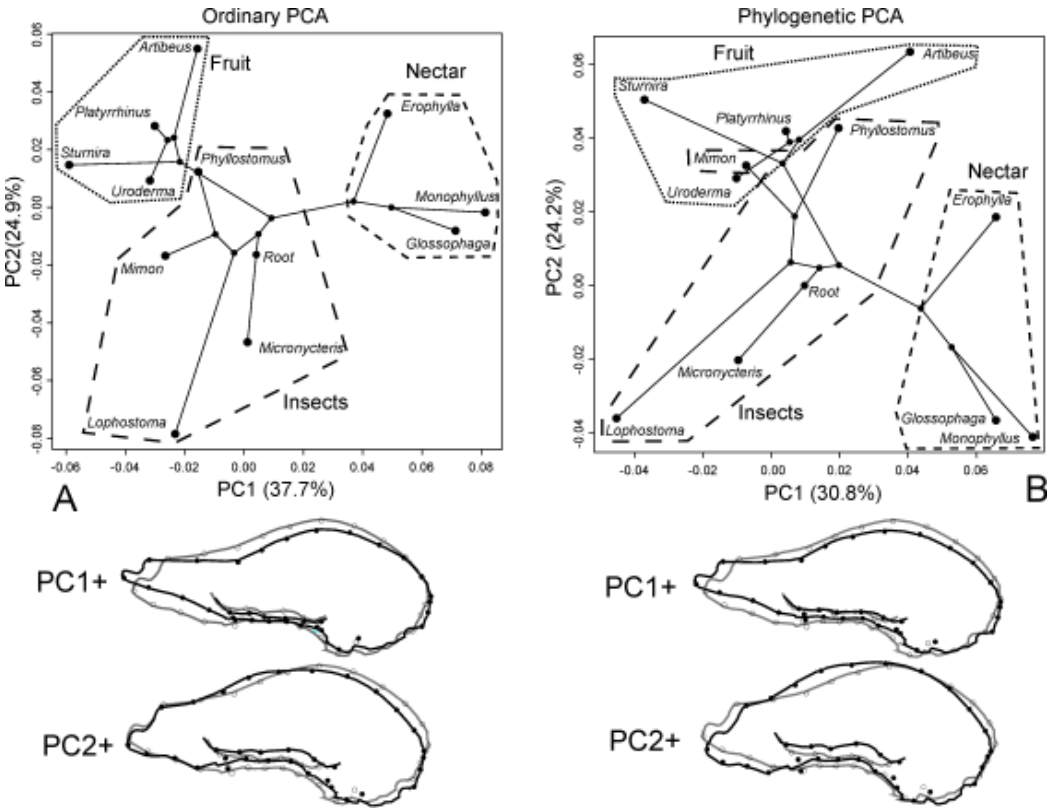


Figure 2 – Phylomorphospaces for principal components of shape, ordinating 11 species of phyllostomid bats based on lateral skull landmarks (Nogueira et al., 2009). The upper panel depicts a scatterplot for PC scores, whereas the lower panels depict the shape change along the positive direction for each axis as warped outlines (average shape as gray line, positive deviation as black line). Species with similar diets are enclosed by polygons with different line patterns. The main diet item is listed near the respective groups. The phylogenetic tree is mapped on the shape subspace. The large circles correspond to the tips and the smaller circles correspond to the nodes (estimated ancestors). A) Scatterplot for the first two PCs from ordinary PCA. B) Scatterplot for the first two phylogenetic PCs (based on the evolutionary rate matrix, see text for explanation).

provides a function to estimate phylogenetic principal components and allows for an estimate of $\mathbf{C}$ by a Brownian motion model or by the lambda transformation (which will be equivalent to Brownian motion if $\lambda = 1$). For the visualisation of shape changes associated with phylogenetic PCs, the eigenvectors can be plotted as shape changes from a mean starting shape (Claude, 2008).

This simple procedure can be easily extended to other types of multivariate analysis, such as Partial Least Squares (PLS) or Canonical Correlation Analysis (CCA). Revell and Harrison (2008) proposed the extension of the method to Canonical Correlation Analysis (CCA), where the data matrices for different sets were transformed before the analysis, using the phylogenetic covariance matrix $\mathbf{C}$ and the phylogenetic mean vector $\mathbf{a}$. The method is then carried out using regular algorithms. For PLS, the same transformed matrices for each block of variables can be used to calculate the cross-covariances in submatrix $\mathbf{R}_{12}$, and the singular value decomposition to estimate pairs of vectors explaining most covariance (Rohlf and Corti, 2000). Significance testing for singular values with permutations (for a null hypothesis of no association between blocks of variables) might still be carried out in the analyses. As in the phylogenetic PCA, the interpretation needs to take into account that the transformed data are not the shapes themselves, but shape differences from the estimated root standardised by branch lengths (similar to what one would get with phylogenetic independent contrasts). An alternative method based on maximum likelihood for comparing hypotheses concerning cross-covariances has also been proposed by Felsenstein (2008), incorporating the influence of the within-species phenotypic covariances on the among-species covariances.

A visualisation method for comparative morphometric data is mapping a phylogeny on the shape subspace composed by the major axes of variation among species (PCs). The position of tree nodes is estimated by ancestral character reconstruction and the branches are drawn as lines connecting the species and their immediate ancestors. This procedure was first proposed by Rohlf (2002), and became popular in applied studies (Nicola et al., 2003; Clabaut et al., 2007; Sidlauskas, 2008; Figueirido et al., 2010; Klingenberg and Gidaszewski, 2010; Monteiro and Nogueira, 2011), receiving the name phylomorphospace (Sidlauskas, 2008). Ancestral character reconstruction can be performed by maximum likelihood (under a Brownian motion evolutionary model) or squared-change parsimony, depending on software, but the approaches are mathematically equivalent (Sidlauskas, 2008). A least-squares approach as used in the calculation phylogenetic independent contrasts is also possible (Felsenstein, 1985). The estimation of ancestral values is controversial and probably inaccurate (Cunningham et al., 1998; Losos, 1999; Webster and Purvis, 2002), unless the real evolutionary process was similar to a Brownian motion (Martins, 1999), and should be used mostly for illustration, so the phylogenetic tree can be mapped on the ordination scatterplot. This visualisation is particularly interesting to detect possible convergences among species in different branches, to get evidence on accelerated morphological evolution, or to analyse morphospace occupation by different lineages (Sidlauskas, 2008), as long as the number of relevant PCs are determined by an objective method and the shape subspace is neither ignoring relevant PCs nor including irrelevant ones (i.e. PCs that do not explain more variation than expected at random).

An example of application of the phylomorphospace and the phylogenetic PCA is shown in Fig. 2. The data set correspond to landmarks placed on a lateral view of the skull of 11 species of phyllostomid bats, studied in Nogueira et al. (2009). The panels on the left (Fig. 2A) show an ordinary PCA with a phylogenetic tree mapped on the ordination of the first two PCs (phylomorphospace). The branch leading to the nectarivore species is aligned with the first PC, whereas the divergence between frugivores and insectivores is aligned with the second PC. The phylomorphospace plot shows that each branch diverged morphologically towards different regions of the PC shape subspace. Looking at the ordination in Fig. 2A, one might feel tempted to attribute the closeness of *Phyllostomus* with frugivore species to convergence. In fact, the species of *Phyllostomus* used (*P. hastatus*) has a mixed diet where

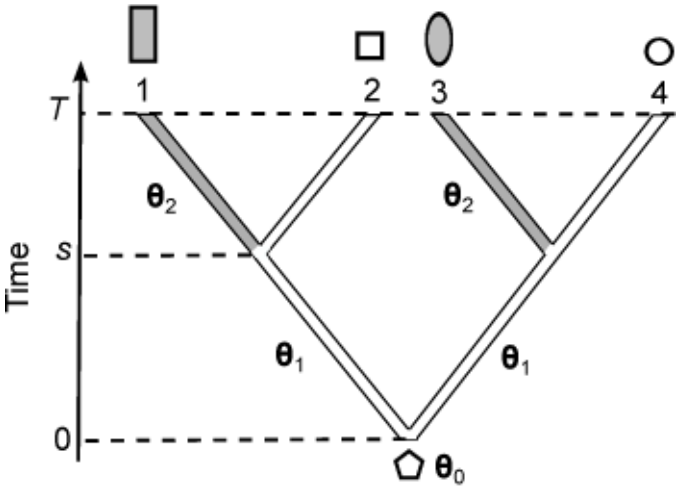


Figure 3 – Hypothetical phylogeny with four species showing a change in selective regime after a speciation event (marked by *s*). The ancestral phenotype is depicted as θ_0, whereas different selective optima are depicted as θ_1 and θ_2. The total sum of branch lengths leading from root to tip is given by T.

the contribution of fruit is almost as large as the one from insects. However, Horn's parallel analysis suggested that 3 PCs were over-dispersed (had larger observed eigenvalues than expected from simulations), and the third PC is mainly a contrast between the larger insectivores – *Mimon* and *Phyllostomus* and the remaining species (not shown in the example). Therefore, their branch might not be considered convergent if we consider all relevant shape dimensions. The phylogenetic PCA (Fig. 2B) is shown on the right panel. The ordination and the shape changes are similar to the ordinary PCA, in fact, the angles between the corresponding vectors are more similar than expected from a distribution of angles between uncorrelated uniform vectors (Claude, 2008). However, the qualitative interpretation of the ordination pattern could be different, and morphological divergence among closely related species is given greater weight in the phylogenetic PCs.

The phylogenetic PCs are a rigid rotation of the original shape space (and as a matter of fact of the ordinary PCs). Any statistical analysis that is based on the scores of the entire set of PCs will achieve the same quantitative results regardless of using phylogenetic of ordinary PCs (Polly et al. 2013, this volume). However, when a subset of the first PCs are used or displayed, one will find differences in the ordination and due to slight changes in rotation. For example, although the shape change associated with nectarivory is clearly associated with PC1 of the ordinary PCA, in the phylogenetic PCA the same shape change would be a combination of the positive changes along PC1 and negative changes along PC2. This finding also visually highlights one important recommendation: to consider the entire multivariate set of over-dispersed PCs when fitting a comparative statistical model is more informative and appropriate than analysing each PC separately as univariate variables. For the phylogenetic PCA, Horn's parallel analysis indicated 4 over-dispersed PCs, and the divergence of *Phyllostomus* from the frugivore branch is relegated to the 4th PC, but the general qualitative conclusions are similar between methods.

Model-based approaches and dimensionality

Moving beyond the phenotype-function-ecology correlations, comparative methods allow for the estimation of evolutionary parameters for trait diversification and measure the agreement between models of evolutionary processes and data (Butler and King, 2004; Freckleton and Harvey, 2006; Hansen et al., 2008). Model-based comparative analyses of multivariate data are still mostly confined to univariate analyses of PCs (Bergmann et al., 2009; van Buskirk, 2009; Collar et al., 2009; Harmon et al., 2010), possibly because both of software package limitations and restrictions of sample size imposed by the number of parameters needed in more complex models with multivariate data. However, due to the potential pitfalls with univariate analyses of PCs, a multivariate model-fitting procedure should be favoured with morphometric data, as currently available in the R package `OUCH` (King and Butler, 2009).

Monteiro and Nogueira (2011) fitted multivariate models based on Brownian motion and several postulated adaptive landscapes based on dietary differences among phyllostomid bat species, to a data set corresponding to five PCs obtained from mandible shape variables. Considering that p is the number of variables in a multivariate set, a Brownian motion model (the simplest evolutionary model possible) requires the estimation of an average vector with p elements and a square variance-covariance matrix (Sigma squared matrix) with $p \times (p+1)/2$ elements (the lower triangular matrix, including the diagonal) that measures the strength of genetic drift (Butler and King, 2004). Stochastic models for the evolution of a multivariate phenotype consistent with stabilising selection around adaptive optima, such as the Ornstein-Uhlenbeck (O-U) model (King and Butler, 2009), assume the form of the differential equation

$$d\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta}(t) - \mathbf{x}(t))dt + \mathbf{S}dB(t), \quad (4)$$

where $\mathbf{A}$ and $\mathbf{S}$ are $p \times p$ square symmetric matrices of parameters measuring the strength of stabilising selection and random drift, respectively. The vector $\boldsymbol{\theta}(t)$ is the optimum phenotype corresponding to a particular selection regime and $B(t)$ is the standard Wiener (Brownian motion) process. Fig. 3 shows a hypothetical phylogeny with four species where branch colours indicate selection regime, based on example in Butler and King (2004). The evolution of the multivariate phenotypes can be seem as weighted sums of selective optima in each branch leading to each specific extant species. In this context, the expected phenotype for species 1 depends on the ancestral phenotype (θ_0) and the estimated optima (θ_1, θ_2) for each branch leading to its current position. According to the multivariate Ornstein-Uhlenbeck model, the expected mean trait vector for species 1 is

$$\begin{aligned} E[\mathbf{x}_1(T)] =& \boldsymbol{\theta}_0 \mathbf{Q} e^{-DT} \mathbf{Q}^{-1} \\ &+ \boldsymbol{\theta}_1 \mathbf{Q} e^{-D(T-s)} \mathbf{Q}^{-1} (\mathbf{I} - \mathbf{Q} e^{-D(s)} \mathbf{Q}^{-1}) \\ &+ \boldsymbol{\theta}_2 (\mathbf{I} - \mathbf{Q} e^{-D(T-s)} \mathbf{Q}^{-1}), \end{aligned} \quad (5)$$

where the thetas are $1 \times p$ vectors of optima, and the multiplying matrices are equivalent to matrix exponentials as $e^{\mathbf{A}} = \mathbf{Q} e^{D} \mathbf{Q}^{-1}$, for $\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^{-1}$, where $\mathbf{Q}$ and $\mathbf{D}$ are the eigenvectors and eigenvalues of $\mathbf{A}$, respectively, and e^{D} is a diagonal matrix with elements e^{λ}, exponential functions of the eigenvalues of $\mathbf{A}$. The random walk around the estimated optima is determined by the covariance matrix Sigma ($\mathbf{S}$), and along with the selection strength matrix $\mathbf{A}$, is used to estimate the covariance matrix for the expectations. The thetas and expected values for each species in the tree can be calculated via generalised least squares and the matrices $\mathbf{A}$ and $\mathbf{S}$ can be optimised during GLS iterations. The procedure described above is implemented in package `OUCH` for R (King and Butler, 2009). Each model requires the estimation of $2 \times p \times (p+1)/2 + p \times n_{\theta}$ parameters (the number of degrees of freedom – DOF for the model), where n_{θ} = the number of adaptive optima. The number of parameters that can be estimated is limited by the number of taxa available, and multivariate data sets will be particularly demanding on sample sizes. For example, an O-U model with 5 shape variables and three adaptive optima will require the estimation of 45 parameters. As a result, the confidence on estimation will be reduced and such models can be severely penalised by information criteria used in model comparisons (Butler and King, 2004), as the Akaike Information Criterion will usually be in the general form $AIC = -2 \times \text{logLikelihood} + 2 \times DOF$, and models with smaller AIC fit better the data.

The assumptions of these multivariate models can be better understood if we visualise them as a restricted random walk on a multidimensional fitness landscape. The variances and covariances determining the Normal distribution of the random walk will be the same for all selective regimes and branches (single $\mathbf{S}$ matrix), and the strength of the selective pull will also be the same for all optima (single $\mathbf{A}$ matrix). These assumptions might be unrealistic, and further advances have been proposed as expansions of the original O-U model with different

variances and selective strengths for each adaptive optimum (Beaulieu et al., 2012), demanding separate $\mathbf{A}$ and $\mathbf{S}$ matrices for each selective regime and greatly increasing the number of parameters estimated. The model-based comparative approaches will usually require the reduction of dimensionality via principal components, but the limitations and suggestions discussed above regarding interpretation and choice of number of components should be kept in mind when using this solution, and the dimensionality of the shape space carefully considered to avoid excluding relevant dimensions.

The model-based approaches can use the information from macroevolutionary patterns of morphological diversification among phylogeny tips (OTUs) to make inferences about the underlying evolutionary processes. Recent improvements in model-fitting, such as Bayesian estimation of model parameters seem to be more accurate in determining evolutionary mechanisms generating data (at least from simulations) (Eastman et al., 2011). Instead of looking at the correlation of phenotypic and ecological variables, these models detect changes in rates of morphological diversification to detect bursts or variation in tempo and mode of the evolution of continuous characters among lineages as evidence of past adaptive radiations, periods of continuous gradual change or periods of stasis (Harmon et al., 2003; O'Meara et al., 2006; Harmon et al., 2010; Venditti et al., 2011; Thomas and Freckleton, 2012). Because most of these models have large numbers of parameters to be estimated, the need for dimensionality reduction and the sample size requirements are again a concern for morphometric data sets. It is also important to realise that the estimation of evolutionary model parameters and inferences about the correlation of trait blocks are complementary approaches (Paradis, 2012), providing mutual support for the understanding of morphological evolution. A claim of evidence for an adaptive radiation due to ecological diversification is made stronger if models that predict increases in diversification rate fit the data well and the morphological changes are associated with ecological and functional changes.

Conclusions

The development of morphometric methods went through an exponential phase just before the end of the 20th century with the geometric morphometric revolution (Adams et al., 2004) and started stabilising during the last 10 years. As the methods matured, the most appropriate methodological combinations, theoretical implications and concepts have become more clear, for instance, the superiority of Procrustes superimposition and associated spaces in statistical shape analysis, as compared to many alternative methods considered before (Rohlf, 2000). It is now relatively easier for a beginning researcher to find relevant advice or guidance around the literature and jargon. The comparative methods are going through a period of exponential production, where a large number of methods and alternative models are proposed, and the detailed statistical properties and relevance of different methods are being clarified through simulations and theoretical contributions (Freckleton et al., 2002; Martins et al., 2002; Hansen and Orzack, 2005; Rohlf, 2006; Revell et al., 2008). There are good introductory texts focused on practical aspects (Butler et al., 2008; Paradis, 2012), but a larger theoretical book summarising recent developments is in demand by an ever increasing community of users. One difficulty of the early days was the availability of software packages to perform the analyses, the formatting differences of all types of data (phylogenies, traits), and software bugs. The widespread use of the R environment and the large number of packages for phylogenetic and comparative analyses (Paradis, 2012) available in this system (31 packages implementing comparative method functions, checked in 31 July 2012) has greatly improved access to almost all published methods. The most comprehensive package `APE` (Paradis et al., 2004) provides not only the most common methods (contrasts, PGLS), but also phylogenetic tree edition and visualisation tools also used by most other phylogenetic packages. See also the continuously updated package descriptions on the phylogenetics CRAN task view maintained by Brian O'Meara (http://cran.r-project.org/web/views/Phylogenetics.html). The different functions and packages show considerable compatibility among each other and the authors of most new methods being proposed provide R packages and functions. A good source for help and information is the email list for the Special Interest Group `R-sig-phylo` (https://stat.ethz.ch/mailman/listinfo/r-sig-phylo), where most authors and package developers make announcements, answer questions and provide assistance with the use of comparative methods. An alternative software that can perform comparative analyses in two-dimensional (previously aligned) morphometric data is Mesquite (Maddison and Maddison, 2011), but R provides a more flexible and complete statistical environment.

The complexity of morphometric data is increasing with the greater availability of 3D data collection devices, and the models in comparative analysis are also becoming more complex in number of parameters. As a consequence, the assessment of appropriate shape space dimensionality should be a matter of great concern. It is an exciting time to be a morphometrician working with comparative data.

References

Ackermann R.R., Cheverud J.M., 2002. Discerning evolutionary processes in patterns of tamarin (genus *Saguinus*) craniofacial variation. Am. J. Phys. Anthropol. 117: 260–271.

Adams D.C., Berns C.M., Kozak K.H., Wiens J.J., 2009. Are rates of species diversification correlated with rates of morphological evolution?. Proc. R. Soc. B. 276: 2729–2738.

Adams D.C., Rohlf F.J., Slice D.E., 2004. Geometric morphometrics: ten years of progress following the "revolution". Ital. J. Zool. 71: 5–16.

Astua D., 2009. Evolution of scapula size and shape in Didelphid marsupials (Didelphimorphia: Didelphidae). Evolution 63: 2438–2456.

Beaulieu J.M., Jhwueng D., Boettiger C., O'Meara B.C., 2012. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. Evolution 66: 2369–2383.

Bergmann P.J., Irschick D.J., 2011. Vertebral evolution and the diversification of squamate reptiles. Evolution 66: 1044–1058.

Bergmann P.J., Meyers J.J., Irschick D.J., 2009. Directional evolution of stockiness coevolves with ecology and locomotion in lizards. Evolution 63: 215–227.

Björklund M., 1997. Are "comparative methods" always necessary? Oikos 80: 607–612.

Blomberg S.P., Garland T., 2002. Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. J. Evol. Biol. 15: 899–910.

Blomberg S.P., Garland T., Ives A.R., 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57: 717–745.

Bookstein F.L., 1991. Morphometric tools for landmark data: geometry and biology. Cambridge University Press, Cambridge, UK.

Brusatte S.L., Sakamoto M., Montanari S., Harcourt-Smith E.H., 2012. The evolution of cranial form and function in theropod dinosaurs: insights from geometric morphometrics. J. Evol. Biol. 25: 365–377.

van Buskirk J., 2009. Getting in shape: adaptation and phylogenetic inertia in morphology of Australian anuran larvae. J. Evol. Biol. 22: 1326–1337.

Butler M.A., King A., 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164: 683–695.

Butler M.A., O'Meara B.C., Pienaar J., 2008. Comparative methods and data analysis in R. Available from http://www2.hawaii.edu/~mbutler/Rquickstart/Rcomparative.pdf.

Cheverud J.M., Dow M.M., Leutenegger W., 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight in primates. Evolution 39: 1335–1351.

Clabaut C., Bunje P.M.E., Salzburger W., Meyer A., 2007. Geometric morphometric analyses provide evidence for the adaptive character of the Tanganyikan cichlid fish radiations. Evolution 61: 560–578.

Claude J., 2008. Morphometrics with R. Springer, New York.

Collar D.C., O'Meara B.C., Wainwright P.C., Near T.J., 2009. Piscivory limits diversification of feeding morphology in centrarchid fishes. Evolution 63: 1557–1573.

Cunningham C.W., Omland K.E., Oakley T.H., 1998. Reconstructing ancestral character states: a critical reappraisal. Trends Ecol. Evol. 13: 361–366.

Diniz-Filho J.A.F., Rangel T.F., Santos T., Bini L.M., 2012. Exploring patterns of interspecific variation in quantitative traits using sequential phylogenetic eigenvector regressions. Evolution 66: 1079–1090.

Diniz-Filho J.A.F., Sant'Ana C.E.R., Bini L., 1998. An eigenvector method for estimating phylogenetic inertia. Evolution 52: 1247–1262.

Dinno A., 2009. Exploring the sensitivity of Horn's parallel analysis to the distributional form of random data. Multivariate Behav. Res. 44: 362–388.

Dryden I.L., Mardia K.V., 1998. Statistical shape analysis. John Wiley & Sons, New York.

Eastman J.M., Alfaro M.E., Joyce P., Hipp A.L., Harmon L.J., 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65: 3578–3589.

Felsenstein J., 1985. Phylogenies and the comparative method. Am. Nat. 125: 1–15.

Felsenstein J., 2008. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. Am. Nat. 171: 713–725.

Figueirido B., Serrano-Alarcon F.J., Slater G.J., Palmqvist P., 2010. Shape at the crossroads: homoplasy and history in the evolution of the carnivoran skull towards herbivory. J. Evol. Biol. 23: 2579–2594.

Fitzjohn R.G., 2010. Quantitative traits and diversification. Syst. Biol. 59: 619–633.

Franklin S.B., Gibson D.J., Robertson P.A., Pohlmann J.T., Fralish J.S., 1995. Parallel analysis: a method for determining significant principal components. J. Veg. Sci. 6: 99–106.

Freckleton R.P., 2009. The seven deadly sins of comparative analysis. J. Evol. Biol. 22: 1367–1375.

Freckleton R.P., Harvey P.H., 2006. Detecting non-Brownian trait evolution in adaptive radiations. PLOS Biology 4: 2104–2111.

Freckleton R.P., Harvey P.H., Pagel M., 2002. Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat. 160: 712–726.

Garamszegi L.Z., Moller A.P., 2010. Effects of sample size and intraspecific variation in phylogenetic comparative studies: a meta-analytic review. Biol. Rev. 85: 797–805.
Garland T., Dickerman A.W., Janis C.M., Jones J.A. 1993. Phylogenetic analysis of covariance by computer simulation. Syst. Biol. 42: 265–292.
Grafen A., 1989. The phylogenetic regression. Philos. Trans. R. Soc. Lond. B 326: 119–157.
Hansen T.F., 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51: 1341–1351.
Hansen T.F., Orzack S.H., 2005. Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. Evolution 59: 2063–2072.
Hansen T.F., Pienaar J., Orzack S.H., 2008. A comparative method for studying adaptation to a randomly evolving environment. Evolution 62: 1965–1977.
Harmon L.J., Kolbe J.J., Cheverud J.M., Losos J.B., 2005. Convergence and the multidimensional niche. Evolution 59: 409–421.
Harmon L.J., Losos J.B., 2005. The effect of intraspecific sample size on type I and type II error rates in comparative studies. Evolution 59: 2705–2710.
Harmon L.J., Losos J.L., Davies T.J., Gillespie R.G., Gittleman J.L., Jennings W.B., Kozak K.H., McPeek M.A., Moreno-Roark F., Near T.J., Purvis A., Ricklefs R.E., Schluter D., Schulte J.A., Seehausen O., Sidlauskas B.L., Torres-Carvajal O., Weir J.T., Mooers A.O., 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64: 2385–2396.
Harmon L.J., Schulte J.A., Larson A., Losos J.B., 2003. Tempo and mode of evolutionary radiation in iguanian lizards. Science 301: 961–964.
Harvey P.H., Pagel M., 1991. The comparative method in evolutionary biology. Oxford Univ. Press, Oxford, UK.
Ives A.R., Midford P.E., Garland T., 2007. Within-species variation and measurement error in phylogenetic comparative methods. Syst. Biol. 56: 252–270.
Jolliffe I.T., 2002. Principal component analysis. Springer-Verlag, New York.
King A.A., Butler M.A., 2009. ouch: Ornstein-Uhlenbeck models for phylogenetic comparative hypotheses (R package). Available from: http://ouch.r-forge.r-project.org.
Klingenberg C.P., 2011. MorphoJ: an integrated software package for geometric morphometric morphometrics. Mol. Ecol. Resour. 11: 353–357.
Klingenberg C.P., Ekau W., 1996. A combined morphometric and phylogenetic analysis of an ecomorphological trend: pelagization in Antarctic fishes (Perciformes: Nototheniidae). Biol. J. Linn. Soc. 59: 143–177.
Klingenberg C.P., Gidaszewski N.A., 2010. Testing and quantifying phylogenetic signals and homoplasy in morphometric data. Syst. Biol. 59: 245–261.
Lande R., 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33: 402–416.
Lande R., 1980. Genetic variation and phenotypic evolution during allopatric speciation. Am. Nat. 116: 463–479.
Langerhans R.B., DeWitt T.J. 2004. Shared and unique features of evolutionary diversification. Am. Nat. 164: 335–349.
Langerhans R.B., Layman C.A., Shokrollahi A.M., DeWitt T.J., 2004. Predator-driven phenotypic diversification in *Gambusia affinis*. Evolution 58: 2305–2318.
Lewens T., 2007. Adaptation. In: Hull D.L., Ruse M. (Eds.). The Cambridge companion to the philosophy of biology. Cambridge University Press, Cambridge, UK. 1–21.
Lofsvold D., 1988. Quantitative genetics of morphological differentiation in *Peromyscus*. II. Analysis of selection and drift. Evolution 42: 54–67.
Losos J.B., 1990. Ecomorphology, performance capability, and scaling of West Indian *Anolis* lizards: an evolutionary analysis. Ecol. Monogr. 60: 369–388.
Losos J.B., 1999. Uncertainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods. Anim. Behav. 58: 1319–1324.
Losos J.B., 2011a. Convergence, adaptation, and constraint. Evolution 65: 1827–1840.
Losos J.B., 2011b. Seeing the forest for the trees: the limitations of phylogenies in comparative biology. Am. Nat. 177: 709–727.
Losos J.B., Warheit K.I., Schoener T.W., 1997. Adaptive differentiation following experimental island colonization in *Anolis* lizards. Nature 387: 70–73.
Maddison W. P., Maddison D., 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75. Available from: http://mesquiteproject.org.
Marroig G., Cheverud J.M., 2005. Size as a line of least evolutionary resistance: diet and adaptive morphological radiation in new world monkeys. Evolution 59: 1128–1142.
Martins E.P., 1999. Estimation of ancestral states of continuous characters: a computer simulation study. Syst. Biol. 48: 642–650.
Martins E.P., 2000. Adaptation and the comparative method. Trends Ecol. Evol. 15: 296–299.
Martins E.P., Diniz-Filho J.A.F., Housworth E.A., 2002. Adaptive constraints and the phylogenetic comparative method: a computer simulation test. Evolution 56: 1–13.
Martins E.P., Hansen T.F., 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149: 646–667.
McPherson G., 1990. Statistics in scientific investigation: its basis, application, and interpretation. Springer-Verlag, New York.
Meloro C., 2012. Mandibular shape correlates of tooth fracture in extant Carnivora: implications to inferring feeding behaviour of Pleistocene predators. Biol. J. Linn. Soc. 106: 70–80.
Meloro C., Raia P., Piras P., Barbera C., O'Higgins P., 2008. The shape of the mandibular corpus in large fissiped carnivores: allometry, function and phylogeny. Zool. J. Linn. Soc. 154: 832–845.
Monteiro L.R., Nogueira M.R., 2010. Adaptive radiations, ecological specialization, and the evolutionary integration of complex morphological structures. Evolution 64: 724–744.
Monteiro L.R., Nogueira M.R., 2011. Evolutionary patterns and processes in the radiation of phyllostomid bats. BMC Evol. Biol. 11: 137.
Morgan C.C., 2009. Geometric morphometrics of the scapula of South American caviomorph rodents (Rodentia: Hystricognathi): form, function and phylogeny. Mamm. Biol. 74: 497–506.
Nicola P.A., Monteiro L.R., Pessoa L.M., von Zuben F.J., Rohlf F.J., Reis S.F., 2003. Congruence of hierarchical, localized variation in cranial shape and molecular phylogenetic structure in spiny rats, genus *Trinomys* (Rodentia: Echimyidae). Biol. J. Linn. Soc. 80: 385–396.
Nogueira M.R., Peracchi A.L., Monteiro L.R., 2009. Morphological correlates of bite force and diet in the skull and mandible of phyllostomid bats. Funct. Ecol. 23: 715–723.
O'Meara B.C., Ané C., Sanderson M., Wainwright P.C., 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60: 922–933.
Papineau D., 2003. Philosophy of science. In: Bunnin N., Tsui-James E.P., (Eds.) The Blackwell Companion to Philosophy. Blackwell Publishing, Malden, MA. 286–318.
Paradis E., 2012. Analysis of phylogenetics and evolution with R. Springer-Verlag, New York.
Paradis E., Claude J., Strimmer K., 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289–290.
Peres-Neto P.R., Jackson D.A., 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia 129: 169–178.
Peres-Neto P.R., Jackson D.A., Somers K.M., 2005. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Comput. Stat. Data Anal. 49: 974–997.
Raia P., Carotenuto F., Meloro C., Piras P., Pushinka D., 2010. The shape of contention: adaptation, history, and contingency in ungulate mandibles. Evolution 64: 1489–1503.
Revell L.J., 2007. Testing the genetic constraint hypothesis in a phylogenetic context: a simulation study. Evolution 61: 2720–2727.
Revell L.J., 2009. Size-correction and principal components for interspecific comparative studies. Evolution 63: 3258–3268.
Revell L.J., 2010. Phylogenetic signal and linear regression on species data. Meth. Ecol. Evol. 1: 319–329.
Revell L.J., 2012. `phytools`: an R package for phylogenetic comparative biology (and other things). Meth. Ecol. Evol. 3: 217–223.
Revell L.J., Collar D.C., 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. Evolution 63: 1090–1100.
Revell L.J., Harmon L.J., 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. Evol. Ecol. Res. 10: 311–331.
Revell L.J., Harmon L.J., Collar D.C., 2008. Phylogenetic signal, evolutionary process, and rate. Syst. Biol. 57: 591–601.
Revell L.J., Harrison A.S., 2008. PCCA: a program for phylogenetic canonical correlation analysis. Bioinformatics 24: 1018–1020.
Revell L.J., Mahler D.L., Peres-Neto P.R., Redelings B.D., 2012. A new phylogenetic method for identifying exceptional phenotypic diversification. Evolution 66: 135–146.
Polly P.D., Lawing A.M., Fabre A.-C., Goswami A., 2013. Phylogenetic Principal Components Analysis and Geometric Morphometrics. Hystrix 24(1) (Online first) doi:l0.4404/hystrix-24.l-6383
Rohlf F.J., 2002. Geometric morphometrics in phylogeny. In: Forey P., Macleod N. (Eds.). Morphology, shape and phylogenetics. Taylor & Francis, London 175–193.
Rohlf F.J., 2000. On the use of shape spaces to compare morphometric methods. Hystrix 11: 9–25.
Rohlf F.J., 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution 55: 2143–2160.
Rohlf F.J., 2006. A comment on phylogenetic correction. Evolution 60: 1509–1515.
Rohlf F.J., Corti M., 2000. Use of two-block partial least-squares to study covariation in shape. Syst. Biol. 49: 740–753.
Rüber L., Adams D.C., 2001. Evolutionary convergence of body shape and trophic morphology in cichlids from Lake Tanganyika. J. Evol. Biol. 14: 325–332.
Sidlauskas B.L., 2008. Continuous and arrested morphological diversification in sister clades of characiform fishes: a phylomorphospace approach. Evolution 62: 3135–3156.
Sober E., 1993. The nature of selection: evolutionary theory in philosophical focus. University of Chicago Press, Chicago, IL.
Thomas G.G., Freckleton R.P., 2012. MOTMOT: models of trait macroevolution on trees. Meth. Ecol. Evol. 3: 145–151.
Venditti C., Meade A., Pagel M., 2011. Multiple routes to mammalian diversity. Nature 479: 393–396.
Webster A.J., Purvis A., 2002. Testing the accuracy of methods for reconstructing ancestral states of continuous characters. Proc. R. Soc. Lond. B 269: 143–149.
Westoby M., Leishman M.R., Lord J.M., 1995. On misinterpreting the "phylogenetic correction". J. Ecol. 83: 531–534.
Winter Y., von Helversen O., 2003. Operational tongue length in phyllostomid nectar-feeding bats. J. Mammal. 84: 886–896.
Young R.L., Haselkorn T.S., Bayaev A.V., 2007. Functional equivalence of morphologies enables morphological and ecological diversity. Evolution 61: 2480–2492.

Associate Editor: A. Cardini

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 33–41, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/6383/pdf
doi:10.4404/hystrix-24.1-6383

Research Article

Phylogenetic Principal Components Analysis and Geometric Morphometrics

P. David Polly[a,*], A. Michelle Lawing[b], Anne-Claire Fabre[c,d], Anjali Goswami[c]

[a] *Departments of Geological Sciences, Biology, and Anthropology, Indiana University, 1001 E. 10th Street, Bloomington, IN 47405, USA*
[b] *Departments of Geological Sciences and Biology, Indiana University, 1001 E. 10th Street, Bloomington, IN 47405, USA*
[c] *Departments of Genetics, Evolution & Environment and Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK*
[d] *Muséum National d'Histoire Naturelle, UMR 7207 - Centre de Recherche sur la Paléobiodiversité & les Paléoenvironnements, 8 rue Buffon - CP 38, F-75231 Paris cedex 05, France*

Keywords:
phylogenetic principal component analysis
geometric morphometrics
comparative phylogenetic methods
shape analysis

Article history:
Received: 28 July 2012
Accepted: 19 September 2012

Acknowledgements
The authors thank Andrea Cardini and Anna Loy for inviting them to contribute to this volume. Leandro Monteiro, Andrea Cardini, and an anonymous reviewer made suggestions that helped improve the manuscript. Jim Rohlf, Joe Felsenstein, and Emilia Martins discussed some of the equations that were used in the Mathematica code used to perform many of the analyses presented in this paper. We thank Jacques Cuisin, Géraldine Véron and Julie Villemain for access to specimens from the Laboratoire Mammifères et Oiseaux, MNHN, Paris, and Loïc Costeur for the loan of material from the Naturhistorisches Museum, Basel. Raphael Cornette and Stéphane Peigné provided extensive help and support during data collection and template construction for the humerus datsaset. A-CF would like to thank the UMS CNRS/MNHN 2700, "Outils et Méthodes de la Systématique Intégrative",for access to the Plate-forme de Morphométrie and the Ecole doctorale interdisciplinaire Européene Frontières du vivant ED474 program doctoral Liliane Bettencourt for support during data collection. Mathematica code to perform most of the analyses presented here is available at http://mypage.iu.edu/~pdpolly/Software.html.

Abstract

Phylogenetic Principal Components Analysis (pPCA) is a recently proposed method for ordinating multivariate data in a way that takes into account the phylogenetic non-independence among species means. We review this method in terms of geometric morphometric shape analysis and compare its properties to ordinary principal components analysis (PCA). We find that pPCA produces a shape space that preserves the Procrustes distances between objects, that allows shape models to be constructed, and that produces scores that can be used as shape variables for most purposes. Unlike ordinary PCA scores, however, the scores on pPC axes are correlated with one another and their variances do not correspond to the eigenvalues of the phylogenetically corrected axes. The pPC axes are oriented by the non-phylogenetic component of shape variation, but the positioning of the scores in the space retains phylogenetic covariance making the visual information presented in plots a hybrid of non-phylogenetic and phylogenetic. Presuming that all pPCA scores are used as shape variables, there is no difference between them and PCA scores for the construction of distance-based trees (such as UPGMA), for morphological disparity, or for ordinary multivariate statistical analyses (so long as the algorithms are suitable for correlated variables). pPCA scores yield different trait-based trees (such as maximum likelihood trees for continuous traits) because the scores are correlated and because the pPC axes differ from PC axes. pPCA eigenvalues represent the residual shape variance once the phylogenetic covariance has been removed (though there are scaling issues), and as such they provide information on covariance that is independent of phylogeny. Tests for modularity on pPCA eigenvalues will therefore yield different results than ordinary PCA eigenvalues. pPCA can be considered another tool in the kit of geometric morphometrics, but one whose properties are more difficult to interpret than ordinary PCA.

Introduction

Principal Component Analysis (PCA) is an important step in geometric morphometrics, both in its own right as a tool to understand overall patterns of shape variation and as a means for producing mathematically uncorrelated shape variables to use in subsequent analyses (Bookstein, 1997a; Dryden and Mardia, 1998). Even though PCA shape variables are mathematically uncorrelated, they may be phylogenetically correlated when shape is sampled in populations or species that are phylogenetically structured. Techniques for analyzing phylogenetically structured data exist, such as mapping shape data onto phylogenies or removing phylogenetic correlations from data, but confusion remains about what these techniques do and when they should be applied. The confusion is compounded by the term "phylogenetic comparative methods", which is applied both to techniques that highlight phylogenetic changes and techniques that remove the effects of phylogeny. Studies of evolution, adaptation, and systematics benefit from techniques that incorporate phylogeny, while studies of correlations that originate from physical processes or other non-adaptive factors benefit from techniques that remove phylogeny.

Phylogenetic principal components analysis (pPCA) is a method recently proposed for controlling for phylogenetic covariance to produce a PCA-like ordination (Revell, 2009). The major axes of pPCA shape space are not the major axes of shape variation, as in ordinary PCA, but rather the major axes of the non-phylogenetic residual variation once phylogenetic covariation has been removed. Note that pPCA is an ordination that attempts to correct for shared phylogenetic history in constructing the axes, which is very different from methods that simply project phylogenetic trees into morphospace and which are referred to by some as "phylogenetic principal components analysis" (e.g., Rohlf 2002; Polly 2008; Klingenberg and Gidaszewski 2010). pPCA can ostensibly be used to study shape variation that arises from underlying processes that are common to all taxa, such as allometric scaling in quadrupedal animals between body mass and limb structure. Because pPCA has rapidly gained popularity in evolutionary studies (e.g., Bergmann and Berk 2012; Kohlsdorf and Navas 2012), we feel it is timely to review its properties in the context of geometric morphometrics. As anticipated by Revell (2009), we find that phylogenetic correlation is not removed by pPCA, and we also find the pPCA scores are correlated between axes and that the variance of scores on pPC axes does not necessarily decrease with sequentially higher axes.

To demonstrate these findings, we critically compare pPCA with ordinary PCA in the context of geometric morphometrics. As a preface, we review PCA, including the essential properties of eigenvectors, eigenvalues, and scores, to provide a clear point of comparison for how pPCA differs. We show with simulated and real examples that pPCA scores have several underappreciated properties, most importantly that they are a rigid rotation of PCA scores and thus conserve between-

*Corresponding author
Email address: pdpolly@indiana.edu (P. David Polly)

specimen distances when measured in the full multivariate space and that the scores on different pPCA axes can be highly correlated with one another. We evaluate how these properties impact some of the standard multivariate analyses used in geometric morphometrics.

A review of principal components analysis

Before discussing pPCA, it is useful to review the technical properties and uses of ordinary PCA in geometric morphometrics. PCA is an ordination technique that is frequently used in multivariate morphometrics. In geometric morphometrics, PCA is one of two preferred methods for producing shape variables that are uncorrelated with one another and whose dimensionality is appropriate for further statistical analysis, the other method being partial warps analysis (Rohlf, 1993; Dryden and Mardia, 1998; Zelditch et al., 2004). Relative Warps Analysis (Rohlf, 1993) is a PCA of partial warp scores and is identical to ordinary shape PCA when the uniform component of shape is included and when the principal warps are weighted equally. PCA space is therefore the standard "shape space" used to show similarity and difference in shape among objects, and it is often used to model evolutionary and developmental shape transformations (e.g., Mitteroecker et al. 2004; Polly 2004; Adams and Collyer 2009). Principal component axes (PCs) function as shape variables, the first of which represents the major axis of variation among the objects. Successive PCs are orthogonal, or at right angles, to the first PC and to each other along the successively greatest axes of variation among the objects. Because the first two or three axes often represent most of the variation in the data set, plots of objects on these axes show most of the overall similarity and difference in shape. Many variations of PCA exist, and readers are referred to Jollife (2002) for a comprehensive description. Here we discuss properties of the simplest form of PCA where the ordination of objects is mean centered and based on the covariance matrix of the variables. The features of PCA that are most relevant to our discussion are the covariance matrix ($\mathbf{P}$), the eigenvectors ($\mathbf{U}$) and eigenvalues ($\mathbf{\Sigma}$) of $\mathbf{P}$, and the scores ($\mathbf{S}$) of the objects on the eigenvectors.

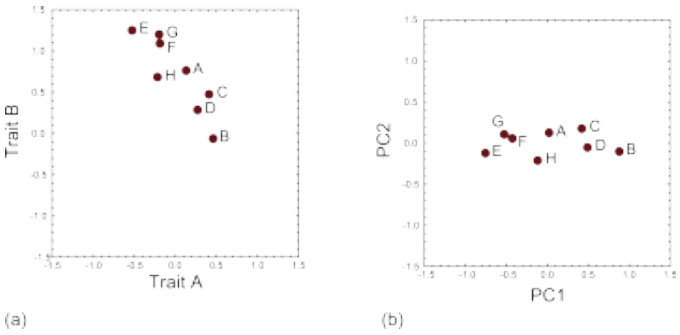


Figure 1 – Example showing the conservation of variance in PCA using a covariance matrix. (a) Data plotted in trait space, where the two clades are separated by both traits. (b) Data plotted in principal components space, where PC 1 separates the two clades. Note that for two traits, there are only two principal components so no other dimensions exist in the trait or PC spaces than those shown.

$\mathbf{P}$ is a symmetrical $m \times m$ matrix of the covariances between the m traits in the off-diagonal elements and the variances in the diagonal elements. In geometric morphometrics, the traits that go into this matrix are the mean-centered landmark coordinates of the objects after Procrustes superimposition. The trace of $\mathbf{P}$, which is the sum of its diagonal elements, is self-evidently equal to the sum of the variances of each of the superimposed landmark coordinates. For later reference, note that $\mathbf{P}$ is calculated as:

$$\mathbf{P} = (n-1)^{-1}(\mathbf{X} - mean[\mathbf{X}])^{\top} \cdot (\mathbf{X} - mean[\mathbf{X}]) \qquad (1)$$

where n is the number of taxa and $\mathbf{X}$ is a matrix of trait values. Note that the rank of this matrix is less than m, the number of traits, because $\mathbf{X}$ is a matrix of Procrustes superimposed coordinates where size, translation, and rotation have been removed, thus removing four degrees of freedom from two-dimensional landmark data $(2m-4)$ and seven for three-dimensional data $(3m-7)$ Rohlf and Slice (1990); Bookstein (1997a). The Procrustes coordinates are thus said to be "collinear" and $\mathbf{P}$ is said to be "singular" because of the loss of these degrees of freedom. The Procrustes coordinates are thus non-independent, violating the assumptions of many statistical analyses. The relevance to PCA is that the loss of degrees of freedom results in fewer PC axes than there are landmark coordinates and that PCA scores provide uncorrelated shape variables that can be used in place of the correlated Procrustes coordinates.

PC axes are defined by the eigenvectors of $\mathbf{P}$ and the variance of the objects on each of the axes is given by the eigenvalues. The term "objects" refers to the individual specimens or taxa that are being analyzed or, more technically, the variables used to measure the objects. In geometric morphometrics, the objects are represented by constellations of Procrustes superimposed points. The elements of the eigenvectors are the cosines of the angles (in radians) of each vector from each of the original variables (the Procrustes residuals in the case of geometric morphometrics). Because $\mathbf{P}$ is singular, $\mathbf{U}$ and $\mathbf{\Sigma}$ are usually calculated using the singular-value decomposition algorithm (SVD) where $\mathbf{P} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^{\top}$ (e.g., Dryden and Mardia 1998). In some implementations of SVD the matrix $\mathbf{\Sigma}$ is a diagonal matrix of singular values, which are the square-roots of the eigenvalues, whereas in others $\mathbf{\Sigma}$ is returned as the eigenvalues themselves (our notation follows the latter). The scores, $\mathbf{S}$, are the values of the objects on the PC axes, or their coordinates in shape space. $\mathbf{S}$ is calculated by projecting the Procrustes residuals into the principal component space by multiplying the Procrustes residuals by the eigenvectors:

$$\mathbf{S} = (\mathbf{X} - mean[\mathbf{X}]) \cdot \mathbf{U} \qquad (2)$$

where $\mathbf{X}$ is the matrix of the Procrustes superimposed coordinates, $mean[\mathbf{X}]$ is the consensus shape or mean of the Procrustes coordinates. Because $\mathbf{U}$ contains the cosines of the angles between the original coordinates and the eigenvectors, Equation 2 describes the translation of the original data to the mean and rotation to its major axes of variation. Importantly for our later discussion, the eigenvectors can be thought of as describing the orientation of the objects in PC space and the eigenvalues can be thought of as their scaling in that space.

For the geometry of shape to be preserved in the PCA shape space, the variance and proportionality of the x, y, and z landmark coordinates must be maintained, as must the distances between the objects in the space (Rohlf, 1993; Dryden and Mardia, 1998). PCA based on a covariance matrix has such properties: the variances of the scores on the PC axes equal the corresponding eigenvalues, and so the sum of the variances of the scores equals the sum of the eigenvalues. Furthermore, because the scores are the result of a rigid rotation, the total variance of the original data set is preserved in the scores. In other words, the sum of the variances of the scores equals not only the sum of the eigenvalues, but also the sum of the variances of the original Procrustes coordinates and the sum of the diagonal of $\mathbf{P}$. Recall that a variance is the sum of squared deviations of objects from their mean divided by the number of objects. A Euclidean distance is the square root of the sum of squared differences of one object to another. The conservation of variance in PCA therefore also conserves the distances between objects in shape space.

Table 1 – Comparison of variances in original traits with the diagonal elements in their covariance matrix ($\mathbf{P}$), the eigenvalues of $\mathbf{P}$, and the variance of the PCA scores based on $\mathbf{P}$. The original trait variances and distances are preserved when the PCA is based on the covariance matrix. Fig. 1 shows ordinations associated with these data.

Variance of traits	
Trait A	0.123
Trait B	0.217
Total	0.340

Eigenvalues of $\mathbf{P}$	
Eigenvalue 1	0.320
Eigenvalue 2	0.020
Total	0.340

Trace of $\mathbf{P}$	
Trait A	0.123
Trait B	0.217
Total	0.340

Variance scores of $\mathbf{P}$	
PC 1	0.320
PC 2	0.020
Total	0.340

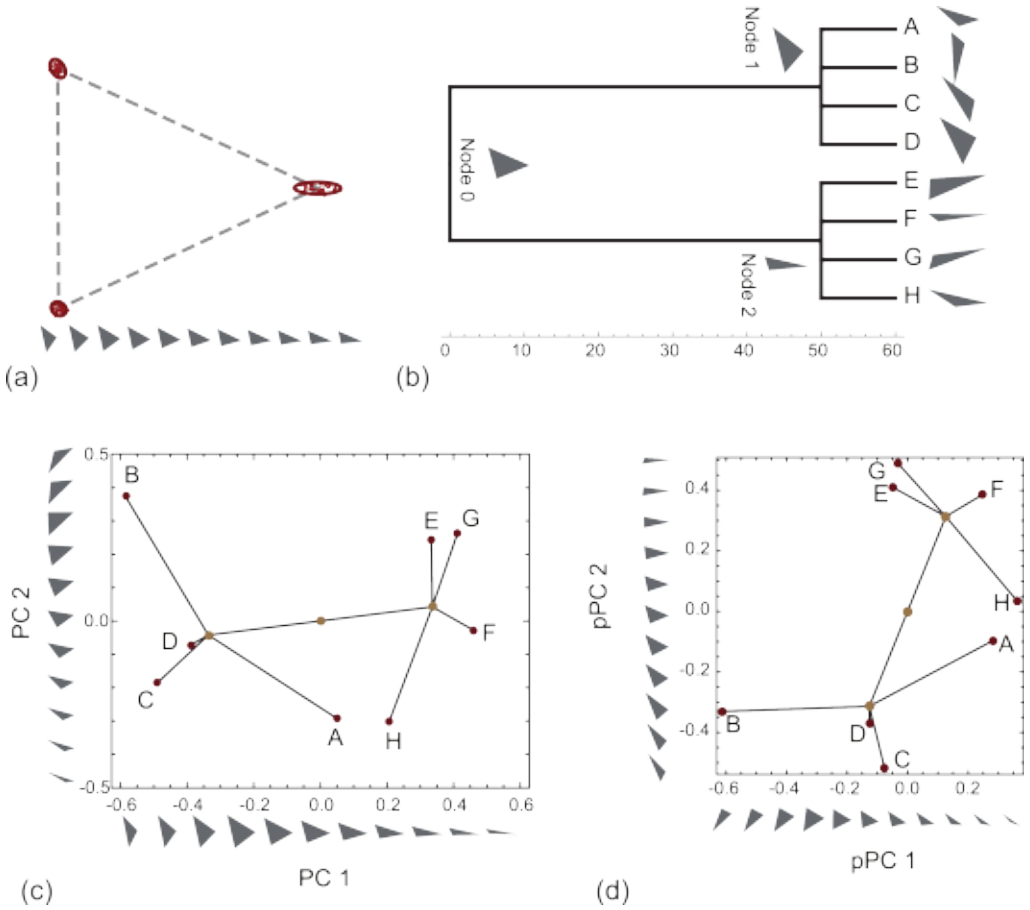


Figure 2 – A comparison of standard PCA and phylogenetic PCA using simulated evolution of eight triangles. (a) The apices of the triangles were simulated with covariances shown here. The grey triangles show the shape variation along the first eigenvector of the corresponding covariance matrix. (b) Triangles were simulated on the tree shown here using Brownian motion and the covariance matrix derived from a. After generating the eight tip shapes (shown in grey), ancestral node shapes were reconstructed on the tree (also shown in grey). (c) Ordinary PCA of the eight tip taxa from b. The phylogenetic tree has been projected into the PC space by locating the points in the space that correspond to the reconstructed ancestral node shapes. The shape variation described by each PC axis is shown as a series of small grey triangles along the two margins. (d) Phylogenetic PCA of the same data using the same conventions as in c. Note that shape space for triangles has only two dimensions ($2K - 4 = 6 - 4 = 2$) so the axes shown in c and d are the full shape space.

The example in Fig. 1 demonstrates the conservation of variance and inter-specimen distances in PCA. Randomly generated bivariate data for eight objects are plotted in trait space in Fig. 1a and in principal components space in Fig. 1b. One can see that the distances between the objects are preserved in the PC space, having been translated and rotated relative to the trait space. One can also see that in the PC space the objects are oriented along their major axis of variation. The sum of the variances of Traits A and B is the total variance in the data set, which is preserved in the sum of the variance of the scores on PC 1 and PC 2 and in the eigenvalues (Tab. 1). The mean distance between objects in the original trait space and in PC space is 0.730 units in both cases. Note that the PC space is mean centered and trait space is not.

Phylogenetic structure impacts PCA ordination, which is easily seen in the simulated example in Fig. 2. We simulated the evolution of triangles on a phylogenetic tree consisting of eight tip taxa in two distantly separated clades. We used triangles in this example because, after Procrustes superimposition, which removes 4 degrees of freedom for scaling, translation, and rotation, the shape space for triangles has only two dimensions allowing the full morphospace to be represented by just two principal components (Rohlf, 1999). We evolved the triangles on the tree (Fig. 2b) using Brownian motion and an arbitrarily defined population covariance (Fig. 2a). The major axis of this "generating" covariance is shown as a series of grey triangles in Fig. 2a. The eight simulated triangles and ancestral shapes reconstructed from them using a Brownian motion model (Martins and Hansen, 1997; Rohlf, 2001) are shown in Fig. 2b. The principal components of the triangles are shown in Fig. 2c, with the tree projected into the resulting shape space based on the node reconstructions (Rohlf, 2002; Polly, 2008). The shape variation described by each of the two PCs is shown as a series of grey triangles along the margins of the plot. The two clades form separate clusters in the plot, separated along PC 1, which is unsurprising because the branches separating the two clades are long and the greatest evolutionary differences are expected to accumulate along the longest branches. In fact, it is typical with data containing two or more clades that the first PC separates one group from the rest, the second PC separates another group, the third PC yet another, and etc. (in cases where there are more than two groups and two PCs). Also unsurprisingly, the shape variation associated with the first PC is similar to the generating covariance because the evolution along the two long branches separating the clades was simulated using that covariance structure. The structure of PC 1 is thus highly phylogenetic, but the shape variation along the axis is closely related to the underlying covariance used to simulate the data.

The code used to simulate these data, to perform the ancestral reconstructions, and to project the tree into shape space, as well as a complete description of the algorithms used, is available in the Geometric Morphometrics for Mathematica (v. 9.0) and the Phylogenetics for Mathematica (v. 2.1) packages (Polly, 2012a,b).

Properties of phylogenetic principal components analysis

Phylogenetic principal components analysis (pPCA) is similar to PCA, except that the covariance matrix is inversely weighted by phylogeny and the space is centered on the estimated phenotype of the root node of the tree instead of the mean of the tips (Revell, 2009). A key component of pPCA is the phylogenetic covariance matrix ($\mathbf{C}$), which is a symmetrical $n \times n$ matrix, where n is the number of tips on the tree, with off-diagonal elements containing the branch length shared by taxa and the diagonal elements containing the total branch length between

each tip and the root of the tree (Martins and Hansen, 1997; Rohlf, 2001; Revell, 2009). Under a Brownian motion model of evolution, this matrix describes the expected phenotypic variance and covariance among tip taxa due to common descent (Martins and Hansen, 1997; Felsenstein, 2003). Branch lengths can be given in any units, but, as elaborated below, the choice of units affects the eigenvalues of $\mathbf{C}$. In terms of geometric morphometrics, $\mathbf{C}$ describes the expected similarity in shape due to recency of common ancestry.

Following Revell (2009), the first step in pPCA is to estimate the ancestral node values of the traits on the tree:

$$a = [(\mathbf{1}^\top \cdot \mathbf{C}^{-1} \cdot \mathbf{1})^{-1} \cdot \mathbf{1}^\top \cdot \mathbf{C}^{-1} \cdot \mathbf{X}]^\top \quad (3)$$

which gives a vector of estimated ancestral values for the n traits ($\mathbf{X}$) at the root node of the tree, where $\mathbf{1}$ is a vector of ones whose length is equal to $\mathbf{C}$ and $\mathbf{X}$ is a matrix of Procrustes superimposed coordinates. This method for estimating ancestral node values is the same as the generalized linear model method (Martins and Hansen, 1997; Rohlf, 2001), and the ancestral node estimates are identical to maximum likelihood (Schluter et al., 1997) and squared-change parsimony (Maddison, 1991) estimates when the traits have evolved under a Brownian motion model. As Revell pointed out, methods that make different assumptions about evolution can be used in pPCA if so desired. The root node reconstruction is used to center the pPCA. Rohlf (1998) warned of distortions that could arise from the non-Euclidean curvature of shape space when values other than the arithmetic mean are used to center ordinations. In practice, the risk of such distortion is minimal because shape variation in biological data sets is usually small (Rohlf, 2003) and because the ancestral node reconstruction is merely a weighted mean and not radically different from the arithmetic mean, so we will not consider this issue further.

The next step in pPCA is to estimate the evolutionary covariance matrix for the traits ($\mathbf{P_P}$) (Revell, 2009). This matrix is similar to the ordinary trait covariance matrix ($\mathbf{P}$) except that taxa are weighted by their shared ancestry and traits are centered on the ancestral node values instead of their mean:

$$\mathbf{P_P} = (n-1)^{-1} \cdot (\mathbf{X} - a^\top)^\top \cdot \mathbf{C}^{-1} \cdot (\mathbf{X} - a^\top) \quad (4)$$

where n is the sample size and 1 is a scalar. Note that the calculation of $\mathbf{P_P}$ is identical to the calculation of $\mathbf{P}$ (Equation 1) except that the root ancestor is substituted for the mean and the inverse of $\mathbf{C}$ is used to weight the calculation. Because the elements of $\mathbf{C}$ are shared branch lengths, those taxa that share the longest branch lengths have the highest phylogenetic covariances are down-weighted most heavily by $\mathbf{C}$'s inverse. As with PCA, eigenvalues ($\mathbf{\Sigma_P}$) and eigenvectors ($\mathbf{U_P}$) are extracted from $\mathbf{P_P}$ using singular value decomposition. Note neither the sum of the diagonal elements of $\mathbf{P_P}$ nor the sum of its eigenvalues equal the sum of the variances in the original traits because of the weighting by the inverse of $\mathbf{C}$.

The final step of pPCA is to project the tip taxa into the space defined by the eigenvectors of $\mathbf{P_P}$ (Revell, 2009). The pPCA scores are calculated as

$$\mathbf{S_P} = (\mathbf{X} - a) \cdot \mathbf{U_P}. \quad (5)$$

Note the similarity and difference between Equations 2 and 5. Instead of being mean centered as in Equation 2, the traits $\mathbf{X}$ are centered on the root node a, and instead of being rotated to the eigenvectors of the covariance matrix $\mathbf{P}$ they are rotated to the eigenvectors of the phylogenetically weighted covariance matrix $\mathbf{P_P}$. Equation 5 is thus a rigid rotation of $\mathbf{X}$ around a just as Equation 2 is a rigid rotation of $\mathbf{X}$ around $mean[\mathbf{X}]$. The phylogenetic scores $\mathbf{S_P}$ are not rescaled by the eigenvalues $\mathbf{\Sigma_P}$, nor can they be without also changing their relative positions within the shape space. Revell (2009) rightly notes that the projection of the tip data $\mathbf{X}$ into the space defined by the eigenvectors makes pPCA different than an ordinary PCA of phylogenetically independent contrasts (Ackerly and Donoghue, 1998). The goal of pPCA is to ordinate the n tips rather than the $n-1$ contrasts, even though the goal of both is to provide a phylogenetically corrected ordination.

The PCA and pPCA ordinations are compared in Fig. 22c-d using the simulated triangle data set. In this example the two ordinations are visibly different (indeed, these particular simulated data were used because of their especially strong difference). One difference is the orientation of the objects within the space. PC 1 separates the two clades in the PCA plot because the greatest axis of phenotypic variation runs between the clades, but pPC 1 does not substantially separate the two groups because of the inverse weighting of $\mathbf{C}$ reduces the influence of shared differences of taxa, an effect that is most pronounced in the longest branches. Note, however, that clades 1 and 2 are just as distinct in the pPCA as in the PCA because the shape data themselves are not adjusted before projecting them onto the pPCA axes. All of the shape variation is therefore represented in pPCA because the calculation of scores is a rigid rotation of the original, ensuring that all the shape information is retained in both analyses. The plots differ, minorly, in where their axes are centered. In PCA the center of the axes is at the arithmetic mean of the tips (which happens to coincidentally be near Node 0 in this example), but in pPCA the center is precisely at the position of the basal node of the tree. The code for doing the pPCA ordination and projecting the phylogenetic tree into it is available in the Geometric Morphometrics for Mathematica package (version 9.0) (Polly, 2012a).

Table 2 – Trait variances in pPCA are not conserved. The sum of the variances of the tip traits does not equal the variances in the evolutionary covariance matrix $\mathbf{P_P}$ because of the adjustment for phylogenetic relationships. The sum of the variances of the pPCA axes, the eigenvalues, equals the variance in $\mathbf{P_P}$ but not the variance of the original traits. When the tips are projected into the pPCA space the sum of their variances is equal to the variance in the original data, but not the sum of the pPCA eigenvalues. Note that the variance of the scores on pPC 2 is greater their variance on pPC 1. The data summarized here are shown in Fig. 2.

Variance of Procrustes coordinates	
Apex 1x	0.055
Apex 1y	0.049
Apex 2x	0.016
Apex 2y	0.055
Apex 3x	0.047
Apex 3y	0.029
Total	0.251

Eigenvalues of $\mathbf{P}$	
Eigenvalue 1	0.182
Eigenvalue 2	0.069
Total	0.251

Eigenvalues of $\mathbf{P_P}$	
Eigenvalue 1	0.008
Eigenvalue 2	0.004
Total	0.011

Variance of scores of $\mathbf{P}$	
PC 1	0.182
PC 2	0.069
Total	0.251

Variance of scores of $\mathbf{P_P}$	
pPC 1	0.096
pPC 2	0.155
Total	0.251

An important property of pPCA is that, even though the pPC axes are orthogonal and they are numbered in descending order of their eigenvalues, the shape scores on the pPC axes are correlated and the proportion of the shape variance explained by them is not in descending order. In this example, the correlation between the pPC 1 and pPC 2 scores is 0.39 and the variance of the scores on pPC 2 is greater than on pPC 1 (Tab. 2). These unusual properties arise from the fact that phylogenetically adjusted trait data are used to construct the pPCA eigenvectors but unadjusted trait data are projected into them. We discuss these issues in detail in the following sections.

pPCA produces phylogenetically corrected eigenvalues and eigenvectors

The phylogenetic correction in pPCA adjusts the covariance matrix to remove the expected phylogenetic correlation among the traits (Revell, 2009), which it does very well. The effect of this adjustment is as though the tip data were adjusted to be no more divergent or correlated than if they had been drawn from a star phylogeny. The eigenvectors and eigenvalues of pPC 1 thus the major axis of non-phylogenetic variation, pPC 2 is the orthogonal major axis to the residual non-phylogenetic variation, and so on. The more phylogenetic structure there is among the tips, the more PCA and pPCA will differ. If the tips are drawn from a star phylogeny, the two ordinations will be

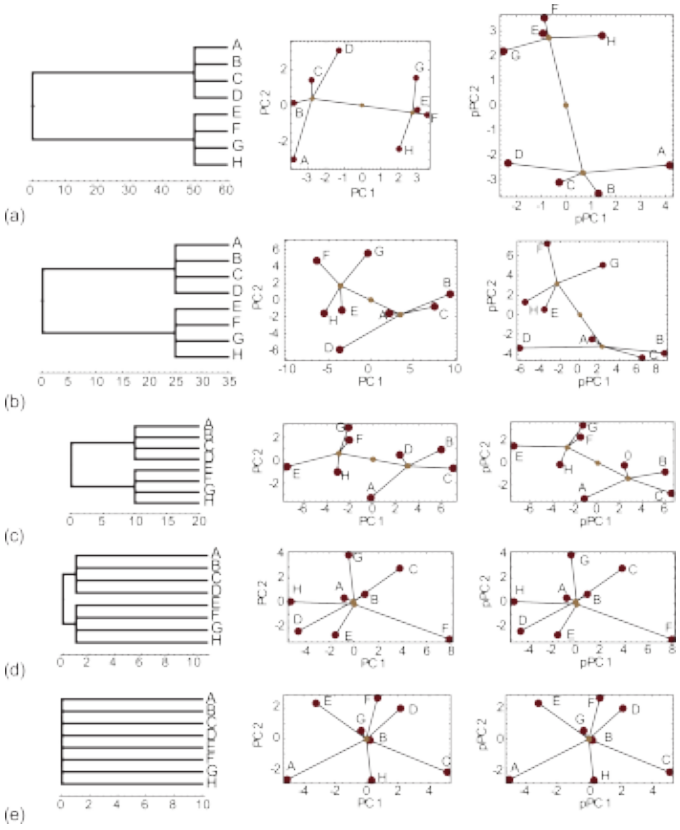


Figure 3 – Results of five simulations showing how phylogenetic structure affects PCA and pPCA ordinations. The evolution of two traits were simulated on the trees in the left column using a Brownian motion model. The PCA ordination of each simulated data set is shown in the center column and the pPCA ordination in the right column. When clades are separated by long branches (a) then the clades tend to be more different than their members and PC 1 tends to separate groups. pPCA adjusts for the phylogenetic separation and so pPC 1 differs substantially from PC 1. When the clades have shorter branches separating them than they have between members of a clade, then the between group differences tend to be small (b-d) and PC 1 is driven by non-phylogenetic variation and so does not differ substantially from pPC 1. As the tree topology approaches a star phylogeny (b-d) the PCA and pPCA ordinations become identical.

identical; and the farther the phylogeny is from a star, the more the two ordinations will differ (Fig. 3).

Note that the magnitudes of the pPCA eigenvalues depend on the units used for branch lengths and thus cannot be viewed as a simple proportion of the original shape variance. In our example, the sum of the eigenvalues is 0.011, but if the branch lengths are multiplied by 10 (e.g., if they were scaled in hundreds of thousands of years instead of millions of years) the sum of the eigenvalues declines to 0.001. In order for the pPCA eigenvalues to be proportional to the original shape variance, the branch lengths would have to be converted to phenotypic variance units so they would be on the same scale as the data (Rohlf, 2001; Felsenstein, 2003). In practice, the branch length units do not affect the pPCA ordination because the scale does not affect the orientation of the eigenvectors and the scores are not rescaled to have the same variance as the eigenvalues.

The variance of scores in pPCA space do not equal the eigenvalues of the pPCA vectors

Note that neither the eigenvalues ($\mathbf{\Sigma_P}$) nor the evolutionary covariance matrix ($\mathbf{P_P}$) are used to produce the scores (Equation 5); only the eigenvectors of $\mathbf{P_P}$ have an actual effect on the ordination. The eigenvectors describe the direction of the pPCA axes, and thus the orientation of the data, but not their variance or scale. pPCA thus affects only the orientation of the tips in the shape space, it does not change their overall variance or the distances among them, which is why Revell (2009) cautioned that subsequent statistical analyses still require phylogenetic correction.

The issue of variance and scale in pPCA space is complicated and worth exploring. Tab. 2 summarizes the variance in the traits, evolutionary covariance matrix, eigenvalues, and scores using the same example data from Fig. 2. The total variance in the two traits is the same as reported before, but the sum of the diagonals of the evolutionary covariance matrix $\mathbf{P_P}$ is different because the trait variances have been adjusted to remove the phylogenetic covariances. The eigenvalues are calculated from $\mathbf{P_P}$ so their sum is equal to the trace of $\mathbf{P_P}$ and represents the total amount of variance among the taxa after adjusting for their phylogenetic covariances. In other words, the sum of the eigenvalues is the variance expected if the taxa had been drawn from a star phylogeny.

Importantly, the sum of the variance of the pPCA scores is identical to the sum of the variance of the traits, and to the sum of the variances of ordinary PCA scores (Tab. 2). The variance is conserved in pPCA scores because the tips are rigidly rotated into the pPCA space without consideration of the phylogenetically corrected variance. In other words, the orientation of pPCA axes has been adjusted for the effects of phylogeny, but the spacing of the taxa has not. The conservation of variance in the scores has several effects. First, the distances between tips are identical in the original trait space, PCA space, and pPCA space. We already reported that the average distance between tips in the first two of these spaces was 0.730, and so is their average distance in pPCA space. pPCA thus preserves the distance between objects and their proportionality in exactly the same way that PCA does, meaning that many kinds of analysis, such as morphological disparity, will be unaffected by the choice of ordination technique (see examples below).

Equally importantly, the sum of the variance of the pPCA scores is not the same as the sum of the eigenvalues (Tab. 2). The direction of the ordination of the tips has been adjusted for phylogenetic covariance, but their variance and distances have not. Thus, the distances between tips in the pPCA ordination is just as much affected by phylogeny as in PCA. For this reason, Revell (2009) cautioned that phylogenetic statistics should be used to analyze pPCA scores in cases where one wants to remove the effects of phylogenetic covariance.

pPCA scores are correlated

Even though pPCA axes are orthogonal with respect to each other, pPCA scores are, in fact, highly correlated between axes. In our example (Fig. 2), the scores on PC 1 and PC 2 are uncorrelated (R = 0.0), but the scores on pPC 1 and pPC 2 have a substantial correlation (R = 0.39). The eigenvectors extracted from $\mathbf{P_P}$ are orthogonal, but the scores of the tips are not, which makes them very different from normal PCA shape variables. This fact is an important consideration for further analyses where statistical independence between variables may be required.

Humerus morphology in mammalian carnivores: a worked example

Here, we provide a worked example of a phylogenetic PCA using a humeral dataset from musteloid carnivorans (red pandas, skunks, raccoons, weasels, badgers, otters, and allies) assembled by one of us (A.-C. F.). This sample of small to medium-sized carnivorans is ideal for demonstrating the utility of pPCA, as it includes terrestrial, fossorial, arboreal, and aquatic species, and thus represents considerable ecological and morphological diversity. The humerus is well known to reflect locomotory ecology, and provides a simple example of the relationship between shape, ecology, and phylogeny. However, morphometric analyses of single elements are not without complications. In particular, landmark-based geometric morphometric approaches may be insufficient for analysing structures without points of clear homology, such as articular surfaces of long bones. In order to more accurately represent the morphology of the humerus in the musteloid sample, we gathered landmarks across the humerus, as well as semi-landmarks from the distal articular surface, as described below. This study thus provides an example of pPCA with a highly multivariate dataset that is becoming increasingly common with the growing availability of advanced biological imaging tools.

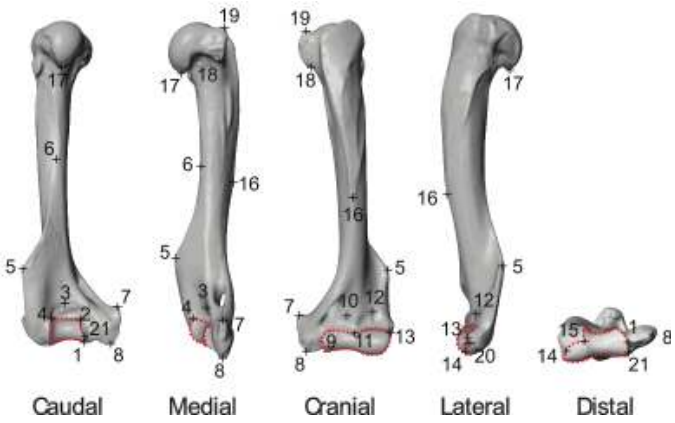


Figure 4 – Three-dimensional landmarks and semilandmarks of the humerus used in the example study. Landmarks are numbered and semilandmarks were placed on the surface outlined in red.

Specimens

Humeri for 29 species of musteloids, spanning their extant diversity, were obtained from the following two collections: the Muséum national d'Histoire naturelle, Paris (*Ailurus fulgens*, *Bassaricyon gabii*, *Bassariscus astutus*, *Conepatus chinga*, *Eira barbara*, *Enhydra lutris*, *Galictis vittata*, *Gulo gulo*, *Lontra felina*, *Lutra lutra*, *Lyncodon patagonicus*, *Martes foina*, *Martes martes*, *Meles meles*, *Mellivora capensis*, *Mustela eversmanii*, *Mustela lutreola*, *Mustela putorius*, *Nasua narica*, *Nasua nasua*, *Neovison vison*, *Poecilogale albinucha*, *Potos flavus*, *Procyon cancrivorus*, *Procyon lotor* and *Pteronura brasiliensis*) and the Naturhistorisches Museum, Basel (*Mydaus javanensis*, *Taxidea taxus* and *Vormela peregusna*). All specimens were adults and predominantly wild caught. Because gender information is often missing from museum specimens, specimens include both male and females. For the purposes of this worked example, only one specimen per species was used in the analyses. In order to perform pPCA on this sample, we used the relevant subset of the supertree of Nyakatura and Bininda-Edmonds (2012). While phylogenetic relationships based on supertree methods are not always congruent with the relationships that would be obtained directly from a combined phylogenetic dataset (Kluge, 1989), this supertree is broadly congruent with carnivoran relationships as they are currently understood.

Shape Coordinates

3-D surface scans of humeri were acquired with a white light fringe Breuckmann scanner (StereoSCAN) using its dedicated scanning software Optocat 2009 (http://www.breuckmann.com). Twenty-one true landmarks and 285 semi-landmarks were selected to represent humeral morphology (Fig. 4, Appendix). The landmarks were gathered using the software package Idav Landmark (Wiley et al., 2005). To generate semi-landmarks a template was created as a reference following the method of Souter et al. (2010). The 3-D sliding landmark procedure (Bookstein, 1997b; Gunz et al., 2005) was used in order to generate landmarks within the boundaries of the template by transforming sliding semi-landmarks into landmarks using Edgewarp3D 3.31 (Bookstein and Green, 2002). The semi-landmarks of the template are warped onto each new specimen within the predefined curves of the template (denoted by the red dotted line in Fig. 4) followed by spline relaxation. Both sliding and relaxation are repeated iteratively until the bending energy is minimized. After this operation has been performed, the landmarks and semi-landmarks are treated identically as variables in the equations described above.

Results

We performed several analyses on both PCA and pPCA scores to demonstrate when the choice of method makes a difference. We looked at the ordinations themselves, plus from the scores we constructed distance-based and trait-based trees, two measures of morphological disparity, and a multivariate regression to find the relationship between shape and a continuous variable, and from the eigenvalues we calculated a simple measure of modularity.

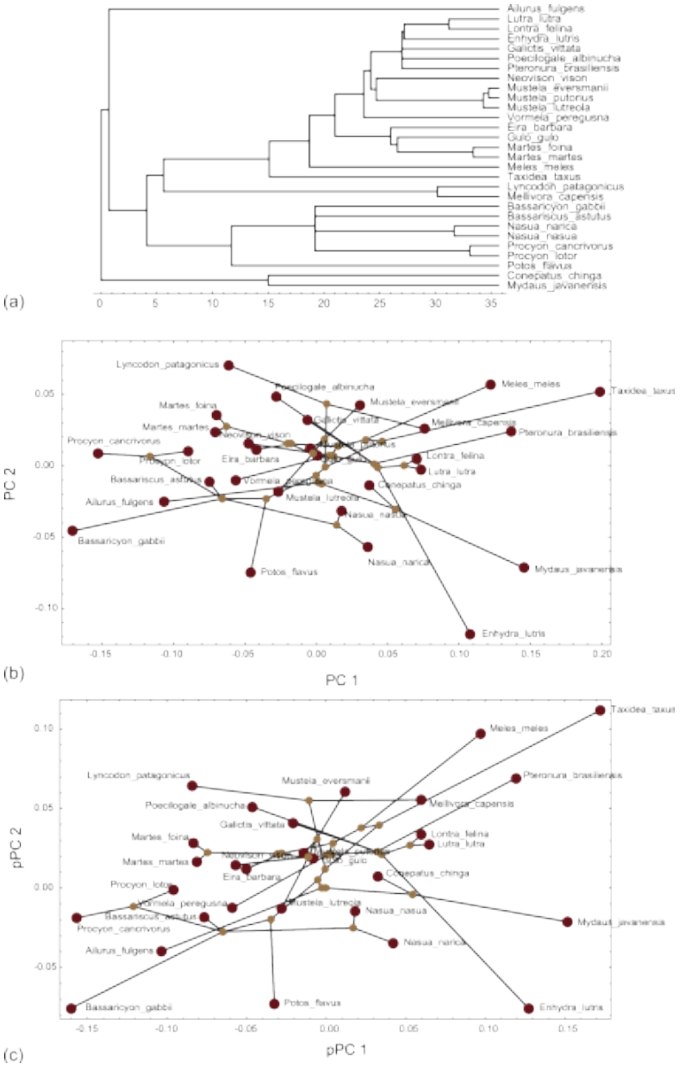


Figure 5 – (a) Phylogenetic supertree of the 29 carnivoran species used in the example data set, taken from Nyakatura and Bininda-Edmonds (2012). (b) PCA ordination of these species based on 21 landmarks and 285 semilandmarks. The phylogenetic supertree for these species has been projected into the space. (c) Phylogenetic PCA of the same data. The first two axes of these ordinations are different, but due to the considerable homoplasy in humeral shape among these taxa the two ordinations are very similar. Note that there are 26 additional dimensions to the shape space other than those shown here.

The PCA and pPCA ordinations were similar but not identical, even though the phylogenetic tree used to produce the pPCA axes has a lot of structure (Fig. 5). The fact that the two ordinations are rigid rotations (plus translation) of each other can be seen in the plots, even though only the first two dimensions of the shape spaces are visible. Inspection of the plots reveals that humeral shape has considerable homoplasy, with lineages from different clades independently colonizing some regions of shape space. The homoplasy in this data set is probably responsible for the strong similarity in the two ordinations, having masked most of the phylogenetic covariance in shape.

Distance-based trees based on the two sets of scores were identical (Fig. 6a-b). These are UPGMA trees constructed from Euclidean distance matrices of the scores of the PCA and pPCA ordinations. Because inter-object distances are preserved in both kinds of ordination, the distance matrices are identical; hence the trees are, too.

Trait-based trees were marginally different for PCA and pPCA (Fig. 6c-d). The trees shown here are maximum-likelihood trees using the continuous traits algorithm implemented in PHYLIP (Felsenstein, 1973, 2009). This algorithm treats each trait, which consists of scores on the 28 principal component axes, as independent in estimating the

tree. The PCA and pPCA trees are different because the axes differ between the analyses, and because the pPCA scores are not actually independent (uncorrelated) meaning that some of them contain redundant information that causes some aspects of shape variation to be overweighted in the estimation of the ML tree.

Table 3 – Comparison of results of follow-up analyses performed using PCA and pPCA scores as shape variables.

Analysis	PCA scores	pPCA scores
UPGMA tree	Identical	
ML tree	Different	
Disparity (Mean pairwise Procrustes distance)	0.174	0.174
Disparity (Mean Procrustes distance to mean)	0.121	0.121
R^2 of shape on centroid size	0.03	0.03
Eigenvalue variance	0.964	0.965
Eigenvalue standard deviation	0.00029	0.00003

Measures of disparity were identical between PCA and pPCA (Tab. 3). Since disparity is a measure of morphospace occupation, it is based on any one of several measures of between-specimen distances (Foote and Miller, 2007). We calculated two disparity metrics, the mean multivariate inter-species distance (which in this case is the Procrustes distance between species) and the mean multivariate distance (Procrustes distance) between species and their group mean. When all dimensions of the shape space are used, these disparity metrics are identical between PCA and pPCA because distances between objects are preserved in both these analyses. If only the first two axes were used to calculate disparity, the results would be different (but the results would not be based on actual differences in shape, only on an arbitrarily defined subcomponents of shape difference).

The proportion of variance explained by multivariate regression of shape onto log centroid size (R^2) was also identical between PCA and pPCA (Tab. 3). The regression coefficients of each individual PC differed slightly (not shown), because the PCs each represent a different component of variance in the two analyses, but the overall relationship between shape and size is conserved because the shape variables collectively describe the same shape variation in PCA and pPCA, thus the proportion of shape variance explained is identical.

Measures of eigenvalue dispersion, a straightforward and commonly used estimate of morphological integration (Pavlicev at al., 2009; Goswami and Polly, 2010) were also compared between PCA and pPCA. While eigenvalue variance was nearly identical in the two analyses (Tab. 3), eigenvalue standard deviation, which may be a more reliable indicator of integration, was tenfold higher in PCA than in pPCA. In part the difference is due to the arbitrary scaling of the covariances introduced by the branch length units in Equation 4 of pPCA, but the difference is also due to non-proportional changes in the estimated covariance structure from having removed the phylogenetic covariances. Thus eigenvalue dispersion measured from pPCA eigenvalues suggests lower integration in the musteloid humerus than does the dispersion of eigenvalues from PCA. Some of the integration in the ordinary PCA results arises from the phylogenetic covariance of the taxa.

Discussion

Phylogenetic PCA is a compromise between non-phylogenetic variance and total shape variance

The axes in pPCA describe the non-phylogenetic component of shape covariance and, as a result, they are independent from phylogeny and orthogonal to one another (Revell, 2009). On average, the eigensystem derived from the phylogenetically corrected covariance matrix is more closely related to the generating covariances than the uncorrected matrix (Revell, 2009). In some cases, such as our simulated example, the uncorrected matrix and the first PC are also highly correlated with the generating covariances, but this will depend on tree balance and mode of evolution. The scores of objects projected onto pPCA axes are neither independent of phylogeny nor orthogonal with respect to each other, however (as clearly stated by Revell 2009). The scores are obtained by rigid rotation of the original data to the pPCA eigenvectors, which preserves the inter-object shape distances, but also preserves the phylogenetic structure in the data. The only difference between a pPCA ordination and regular PCA is that the axes are oriented differently and centered on a different point. The trajectory of shape variation along the pPC 1 is by definition a trajectory parallel to the major non-phylogenetic axis of variation, but the spacing of objects along pPC 1 includes phylogenetic similarity and difference. The visual information conveyed by the spacing of objects in a pPCA plot thus contains a significant phylogenetic component.

Non-independence of scores on pPCA axes may affect some kinds of analysis but not others

Scores of objects on pPCA axes are correlated, whereas PCA scores are uncorrelated across axes. Statistical independence is a desirable property in shape coordinates that are used for tree building, statistical analysis, or shape modeling (Rohlf, 1993), but it is not a requirement if the subsequent analysis does not assume independence of its input variables. All of the analyses we applied to the example data set were based on algorithms that take into account non-independence except for maximum-likelihood tree building for continuous traits. Consequently, results from PCA and pPCA were identical except for the ML tree. pPCA scores can, therefore, be safely used as shape variables for most applications, just as partial warps scores (Bookstein, 1997a) can be used as shape variables even though they are similarly co-dependent (Zelditch et al., 2004).

Dimensionality of pPCA space is appropriate for shape analysis

Removal of the phylogenetic component of shape variance in the calculation of pPCA space does not change its dimensionality. The number of pPC axes with non-zero variance is the same as the number of ordinary PC axes, and that number is $2k-4$ for two-dimensional landmarks and $3k-7$ for three-dimensional landmarks (Rohlf and Slice, 1990; Rohlf, 1999). pPCA scores thus have appropriate dimensionality for further statistical analysis.

Adaptation versus direct environmental effects: Phylogenetic correction can be the wrong approach

Phylogenetic comparative methods are often employed for studying the relationship between phenotypes and environment, but not all environmental effects on morphology arise from non-phylogenetic sources. When the phylogenetic changes are adaptive (*sensu* Gould and Vrba 1982), then environmental components of variation are identical to phylogenetic components. Greene (1986) and Coddington (1988) provided a rigorous phylogeny-based method for studying adaptation, arguing that for a phenotypic change to be considered adaptive, it must have arisen on a phylogenetic tree at the same point as its associated change in function. In other words, functional (or environmental) change must be perfectly correlated with phylogeny. Removing phylogenetic correlations from data will remove precisely the component of shape variation that is relevant to adaptation.

Too often phylogenetic comparative methods are used to remove the effects of phylogeny when the processes being studied are, in fact, phylogenetic rather than non-phylogenetic (Westoby et al., 1995). Conflation of non-phylogenetic and environmental components of variation in the literature probably arises by false analogy with the quantitative genetics concepts of "genetic" and "environmental" components of phenotypic variance (Falconer and Mackay, 1996; Lynch and Walsh, 1998). Phylogenetic comparative methods similarly partition phenotypic variance among taxa into phylogenetic and non-phylogenetic components (Martins and Hansen, 1997; Felsenstein, 2003), but the analogy that the non-phylogenetic component is "environmental" and the phylogenetic component is not does not hold. Rather, the non-phylogenetic component of comparative data is the component associated with universal factors that affect organisms regardless of their

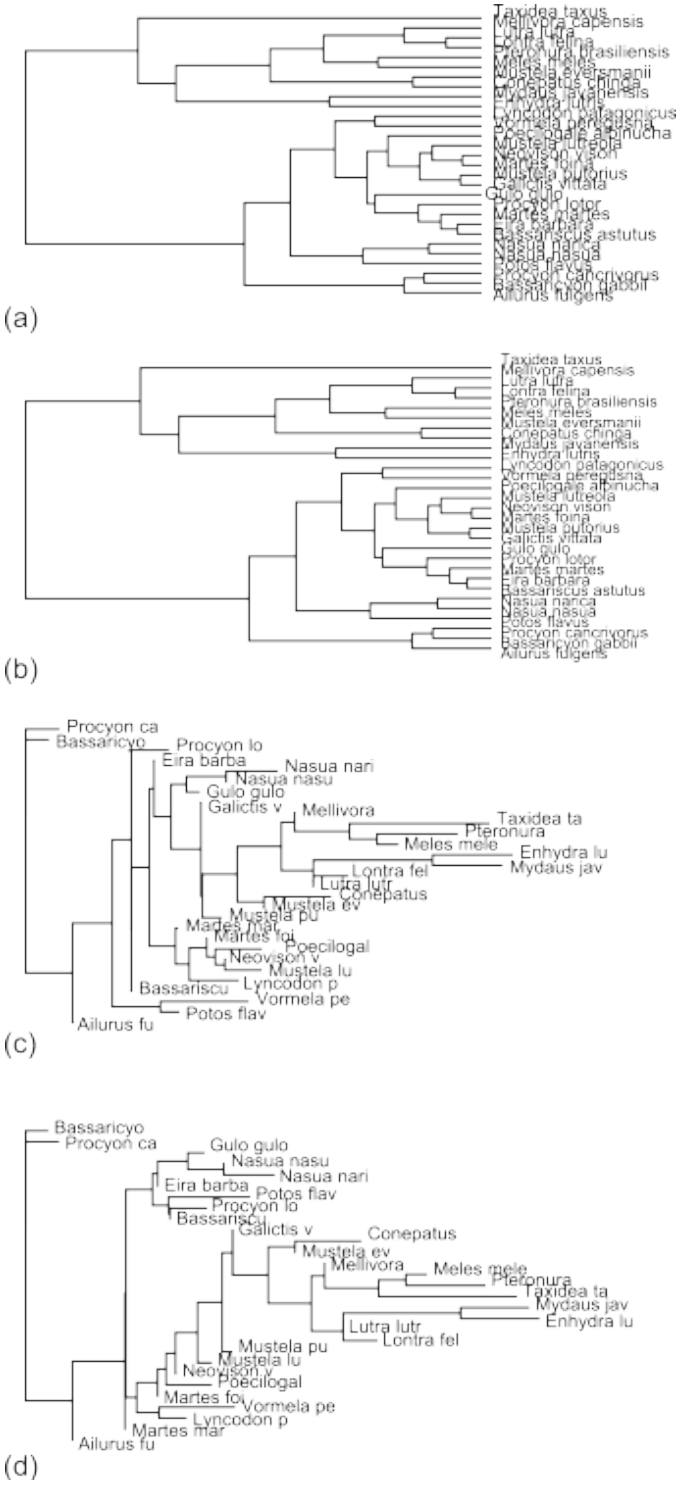


Figure 6 – Trees constructed from PCA and pPCA scores. UPGMA trees constructed from PCA scores (a) and pPCA scores (b). These trees are based on inter-object distances, which are conserved by these two methods, and so they are identical. Continuous-trait maximum-likelihood trees constructed from PCA scores (c) and pPCA scores (d). The ML method treats each PC as if it were an independent trait. Because PCA and pPCA find different axes and because pPCA scores are intercorrelated, the two ML trees are different (though only minorly so for this particular data set).

phylogenetic relationship, whereas the phylogenetic component of variation is the component shared through ancestry. pPCA may therefore not be appropriate if adaptation of morphology is being studied, or if homoplasy is being assessed.

Similarly, the removal of phylogenetic covariances in analyses of modularity or morphological integration should be carefully considered. In the musteloid humerus example, measures of eigenvalue dispersion were very similar for PCA and pPCA, but eigenvalues from pPCA did show lower standard deviation, and thus lower integration, than those from PCA. A strong relationship between generating covariances, such as those driven by genetic and development effects, and phylogenetic covariances may be expected, as evolution along any branch will be strongly influenced by the generating covariances. Moreover the genetic and developmental drivers of morphological integration and modularity inevitably correlate strongly with phylogeny. Removing the phylogenetic covariances may thus also obscure the real modularity or integration of a structure. For some analyses of modularity, it may be appropriate to remove phylogenetic effects prior to analysis of trait covariances, but, as with environmental effects, using phylogenetic comparative has the capacity to obscure, as well as to reveal, evolutionary relationships among traits.

Conclusions

Phylogenetic PCA belongs to the class of phylogenetic comparative methods that remove the expected covariance among objects prior to statistical analysis. The general purpose of such methods is to correct the statistical non-independence of data points that arises from shared phylogenetic history for tests whose p-values depend on the assumption that data points are independent (e.g., the probability that the slope of a regression line significantly differs from zero). These methods down-weight the contribution of closely related taxa to test the relationship between phenotype and associated function or environment because each of close relatives is putatively sampling the same evolutionary adaptation in their common ancestor. Phylogenetic PCA is an unusual example of phylogenetic comparative methods because it is not an analysis *per se* and it serves as a replacement for ordinary PCA, even though the latter does not depend in any way on objects being statistically independent. pPCA is a potentially confusing mixture of major axes that describe non-phylogenetic variation and scores that contain phylogenetic components of variation. The individual axes of pPCA therefore are not usually aligned with clade differences; however, phylogenetic groupings are still equally as evident in pPCA as they would be in ordinary PCA. Potentially undesirable properties of pPCA are that the scores are correlated between axes, the variance of scores on pPC axes does not necessarily decrease with sequentially higher axes, and pPCA scores are not phylogenetically corrected for purposes of subsequent analysis. Nevertheless, pPCA space can be used for geometric morphometric shape modeling and its scores have other properties that are desirable for shape variables, such as having the correct dimensionality and being complete descriptors of shape variation. Phylogenetic principal components analysis can, therefore, be included among the tools available for geometric morphometric analysis as long as it is used knowledgably.

References

Ackerly D.D., Donoghue M.J., 1998. Leaf size, sapling allometery, and Corner's rules: phylogeny and correlated evolution in maples (*Acer*). Am. Nat. 152: 767–791.

Adams D.C., Collyer M.L., 2009. A general framework for the analysis of phenotypic trajectories in evolutionary studies. Evolution 63: 1143–1154.

Bergmann P.J., Berk C.P., 2012. The evolution of positive allometry of weaponry in horned lizards (*Phrynosoma*). Evol. Biol. 39: 311–323.

Bookstein F.L., 1997a. Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge University Press: Cambridge, UK.

Bookstein F.L., 1997b. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Med. Image Anal. 3: 225–243.

Bookstein F.L., Green W.D.K., 2002. Users Manual, EWSH3.19. Available at http://brainmap.stat.washington.edu/edgewarp/

Coddington J.A., 1988. Cladistic tests of adaptational hypotheses. Cladistics 4: 3–22.

Dryden I.L., Mardia K.V., 1998. Statistical Shape Analysis. Wiley: Chichester, UK.

Falconer D.S., Mackay T.F.C., 1996. Quantitative Genetics. Benjamin Cummings: San Francisco, California.

Felsenstein J., 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Human Genet. 25: 471–492.

Felsenstein J., 2003. Inferring Phylogenies. Sinauer Associates: Sunderland, Massachusetts.

Felsenstein J., 2009. PHYLIP (Phylogeny Inference Package) version 3.6.9. Distributed by the author. Department of Genome Sciences, University of Washington: Seattle.

Foote M., Miller A.I., 2007. Principles of Paleontology. W.H. Freeman: New York.

Goswami A., Polly P.D., 2010. Methods for studying morphological integration, modularity and covariance evolution. In: Alroy J., Hunt G. (Eds.). Quantitative Methods in Paleobiology. Paleontological Society Short Course. The Paleontological Society Papers 16: 213–243.

Gould S.J., Vrba E.S., 1982. Expatation – a missing term in the science of form. Paleobiology 8: 4–15.

Greene H.W., 1986. Diet and arboreality in the Emerald Monitor, *Varanus prasinus*, with comments on the study of adaptation. Fieldiana Zool. 31: 1–12.
Gunz P., Mitteroecker P., Bookstein F.L., 2005. Semi-landmarks in three dimensions. In: Slice D. (Ed.). Modern Morphometrics in Physical Anthropology. Kluwer Academic/Plenum Publishers: New York.
Jollife I.T., 2002. Principal Components Analysis, 2nd Edition. Springer: New York.
Klingenberg C.P., Gidaszewski N.A., 2010. Testing and quantifying phylogenetic signals and homoplasy in morphometric data. Syste. Biol. 59: 245–261.
Kluge A.G., 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38: 7–25.
Kohlsdorf T., Navas C., 2012. Evolution of form and function: morphophysiological relationships and locomotor performance in tropidurine lizards. J. Zool. 288: 41–49.
Lynch M., Walsh B., 1998. Genetics and the analysis of quantitative traits. Sinauer Associates: Sunderland, Massachusetts.
Maddison W.P., 1991. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. Syst. Biol. 40: 304–314.
Martins E.P., Hansen T.F., 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149: 646–667.
Mitteroecker P., Gunz P., Bernhard M., Schaefer K., Bookstein F.L., 2004. Comparison of cranial ontogenetic trajectories among great apes and humans. J. Human Evol. 46: 679–698.
Nyakatura K., Bininda-Edmonds O.R.P., 2012. Updating the evolutionary history of Carnivora (Mammalia): a new species-level supertree complete with divergence time estimates. BMC Biology 10: 12.
Pavlicev M., Cheverud J.M., Wagner G.P., 2009. Measuring morphological integration using eigenvalue variance. Evol. Biol. 36: 157–170.
Polly P.D., 2004. On the simulation of the evolution of morphological shape: multivariate shape under selection and drift. Palaeontol. Electronica 7.2.7A: 1–28.
Polly P.D., 2008. Adaptive Zones and the Pinniped Ankle: A 3D Quantitative Analysis of Carnivoran Tarsal Evolution. In: Sargis E., Dagosto M. (Eds.). Mammalian Evolutionary Morphology: A Tribute to Frederick S. Szalay. Springer, Dordrecht: The Netherlands. 165–194.
Polly P.D., 2012a. Geometric morphometrics for Mathematica. Version 9.0. Department of Geological Sciences, Indiana University, Bloomington, Indiana. Available from: http://mypage.iu.edu/~pdpolly/Software.html.
Polly P.D., 2012b. Phylogenetics for Mathematica. Version 2.1. Department of Geological Sciences, Indiana University, Bloomington, Indiana. Available from: http://mypage.iu.edu/~pdpolly/Software.html.
Revell L.J., 2009. Size-correction and principal components for interspecific comparative studies. Evolution 63: 3258–3268.
Rohlf F.J., 1993. Relative warp analysis and an example of its application to mosquito wings. In: Marcus L.F, Bello E., Garcia Valdecasas A. (Eds.). Contributions to Morphometrics. Museo National de Ciencias Naturales, Madrid. 131–159.
Rohlf F.J., 1998. On applications of geometric morphometrics to studies of ontogeny and phylogeny. Syst. Biol. 47: 147–158.
Rohlf F.J., 1999. Shape statistics: Procrustes superimpositions and tangent spaces. J. Classif. 16: 197–223.
Rohlf F.J., 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution 55: 2143–2160.
Rohlf F.J., 2002. Geometric morphometrics and phylogeny. In: MacLeod N., Forey P.L. (Eds.). Morphology, Shape, and Phylogeny. Taylor and Francis: London. 175–193.
Rohlf F.J., 2003. Bias and error in estimates of mean shape in geometric morphometrics. J. Human Evol. 44: 665–683.
Rohlf F.J., Slice D., 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39: 40–59.
Schluter D., Price T., Mooers A.Ø., Ludwig D., 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51: 1699–1711.
Souter T., Cornette R., Pedraza J., Hutchinson J.R, Baylac M., 2010. Two applications of 3D semi-landmark morphometrics implying different template designs: the theropod pelvis and the shrew skull. C. R. Palevol 9: 411–422.
Westoby M., Leishman M.R., Lord J.M., 1995. On misinterpreting the "phylogenetic correction". J. Ecol. 83: 531–534.
Wiley D.F., Amenta N., Alcantara D.A., Ghosh D., Kil Y.J., Delson E., Harcourt-Smith W., Rohlf F.J., St John K., Hamann B., 2005. Evolutionary morphing. In: Proceedings of IEEE Visualization 2005 (VIS'05), 23-28 October 2005, Minneapolis, MN, USA.
Zelditch M.L., Swiderski D.L., Sheets H.D., Fink W.L., 2004. Geometric Morphometrics for Biologists: A Primer. Elsevier Academic Press: San Diego, California.

Associate Editor: A. Cardini

Appendix

Definitions of the 21 landmarks used in the example study of carnivoran humerus shape.

Landmark	Definition
1	Most medio-distal point of the caudal part of the capitulum
2	Most medio-proximal point of the caudal side of the capitulum
3	Point of maximum of curvature of the olecranon fossa
4	Most latero-proximal point of the caudal side of the capitulum
5	Point of maximum of convexity of the lateral epicondylar ridge
6	Point of insertion of the lateral epicondylar ridge on the diaphysis
7	Most proximal tip of the entepicondylar area
8	Most distal tip of the entepicondylar area
9	Most medio-proximal point of the cranial side of the capitulum
10	Point of maximum of curvature of the coronoid fossa
11	Most proximal point of contact between the trochlea and the capitulum
12	Point of maximum of curvature of the radial fossa
13	Most latero-proximal point of the cranial side of the capitulum
14	Most disto-lateral point of the capitulum
15	Most distal point of contact between the trochlea and the capitulum
16	Most distal point of the deltopectoral crest
17	Tip of the lesser tuberosity
18	Most proximo-medial point of the greater tuberosity
19	Most disto-medial point of the greater tuberosity
20	Most latero-distal point of the cranial side of the capitulum
21	Point of maximum of concavity of the caudo-medio-distal part of the capitulum

1993 – Il Ciocco – NATO ASI "Advances in Morphometrics". Among others: Leslie Marcus, Chris Klingenberg, Marco Corti, Jim Rohlf, Gavin Naylor, Antonio Garcia-Valdecasas, Santiago Reig, Michel Baylac, Dennis Slice, Fred Bookstein, Richard Reyment, Jean Christophe Auffray (picture by M.Corti).

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 43–58, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/6367/pdf
doi:10.4404/hystrix-24.1-6367

Research Article

Cranial integration and modularity: insights into evolution and development from morphometric data

Christian Peter Klingenberg[a,*]

[a] *Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK*

Keywords:
allometry
comparative methods
geometric morphometrics
modularity
morphological integration

Article history:
Received: 9 July 2012
Accepted: 17 January 2013

Acknowledgements
I thank Andrea Cardini for inviting me to write this review for the special issue of Hystrix, and him, David Polly and Anna Loy for helpful comments on a previous version of the manuscript.

Abstract

Morphological integration and modularity have become central concepts in evolutionary biology and geometric morphometrics. This review summarizes the most frequently used methods for characterizing and quantifying integration and modularity in morphometric data: principal component analysis and related issues such as the variance of eigenvalues, partial least squares, comparison of covariation among alternative partitions of landmarks, matrix correlation and ordinations of covariance matrices. Allometry is often acting as an integrating factor. Integration and modularity can be studied at different levels: developmental integration is accessible through analyses of covariation of fluctuating asymmetry, genetic integration can be investigated in different experimental protocols that either focus on effects of individual genes or consider the aggregate effect of the whole genome, and several phylogenetic comparative methods are available for studying evolutionary integration. Morphological integration and modularity have been investigated in many species of mammals. The review gives a survey of geometric morphometric studies in some of the groups for which many studies have been published: mice and other rodents, carnivorans, shrews, humans and other primates. This review demonstrates that geometric morphometrics offers an established methodology for studying a wide range of questions concerning integration and modularity, but also points out opportunities for further innovation.

Introduction

The parts of skulls are integrated with each other because they develop, function and evolve jointly. Integration of cranial parts is inevitable because parts share developmental precursors, are packed together tightly in the head, because different functions place demands on different pats of the skull, and because all parts of the head share an evolutionary history (Moore, 1981; Depew et al., 2002; Lieberman, 2011). Yet this integration is not total, but is structured as modules that are relatively independent within the overall integration of the head as a whole (Klingenberg, 2008, 2010). The concepts of integration and modularity are therefore closely linked to each other and need to be discussed jointly.

The methods of geometric morphometrics are suitable for investigating morphological integration and modularity for several reasons. Geometric morphometrics offers a range of flexible and powerful statistical tools for addressing a range of specific biological questions concerning modularity and integration. The combination of geometry with multivariate statistics, which is at the core of geometric morphometrics, ensures that the shape of a structure is characterized completely and without redundancy. These analyses automatically take into account all the spatial relations among morphological landmarks or other geometric features included in the analysis, without the need for the investigator to define, a priori, a set of "traits" to include in the analysis. Which specific aspects of shape are important in answering a particular question is therefore a part of the results of a morphometric analysis. Finally, the various possibilities for visualization of results (Klingenberg, this issue) enable researchers to interpret findings directly in the context of cranial anatomy.

Analyses of morphological integration and modularity have been conducted with geometric morphometric approaches for somewhat more than a decade now (e.g. Klingenberg and Zaklan 2000; Klingenberg et al. 2001a; Bookstein et al. 2003; Klingenberg et al. 2003; Badyaev and Foresman 2004; Bastir and Rosas 2005; Monteiro et al. 2005; Goswami 2006b; Young 2006; Young and Badyaev 2006; Cardini and Elton 2008a; Zelditch et al. 2008; Klingenberg 2009; Ivanović and Kalezić 2010; Jamniczky and Hallgrímsson 2011; Jojić et al. 2012; Martínez-Abadías et al. 2012a). Several different approaches have been used, and new methods continue to be developed by various investigators. It is therefore timely to take stock of the existing work and survey the challenges that remain or new opportunities for innovative research that have emerged recently. This paper briefly summarizes the concepts of morphological integration and modularity, surveys the morphometric methods for studying them, and provides an overview of empirical results from such studies. Because this special issue of Hystrix is devoted to geometric morphometrics, this review will pay special attention to studies of integration and modularity that use geometric morphometric methods. There is a large literature on morphological integration and modularity that is based on traditional morphometric methods. This literature will only be covered as far as it provides unique information and insight on the topic. Likewise, because Hystrix is a journal of mammalogy, this review will primarily concentrate on studies in mammals (including humans) and will only mention a few of the many studies that have been conducted with other organisms.

Morphological integration

Different parts of organisms are coordinated in their sizes and shapes to make up a functional whole. This idea of integration goes back to the beginnings of the study of morphology, where it played a central role in the thinking of pioneers such as Georges Cuvier (Mayr, 1982). The current concepts of integration were conceived by researchers in the 20th century, such as the seminal book by Olson and Miller (1958), and further transformed by linking them width ideas from other fields, such as evolutionary quantitative genetics and developmental biology (Cheverud, 1982a, 1996a; Wagner and Altenberg, 1996; Hallgrímsson et al., 2002; Klingenberg, 2008). Through its developmental and quantitative genetic aspects, morphological integration also closely relates to

*Corresponding author
Email address: cpk@manchester.ac.uk (Christian Peter Klingenberg)

developmental and evolutionary constraints (Cheverud, 1984, 1996a; Raff, 1996; Wagner, 1996; Wagner and Altenberg, 1996; Arthur, 2001; Klingenberg, 2008; Futuyma, 2010).

Morphological integration can be observed at a variety of different levels, depending on the origin of variation under consideration and the observations under study (Cheverud, 1982a; Klingenberg, 2008). For instance, integration within a species at a particular ontogenetic stage can be investigated by analyzing a sample of individuals of that species, ontogenetic integration can be studied from data derived from different growth stages of one species, or evolutionary integration can be investigated by examining how evolutionary changes in multiple parts are coordinated across a set of related species. Other levels focus more on processes, for example functional integration and developmental integration (Breuker et al., 2006a). The levels of integration mentioned here are not meant as a full enumeration of all possible types, or even of the levels of integration that exist in the literature. At every level, integration is concerned in some way with covariation among the parts or traits, but the origin of variation as well as the conceptual context and focus of investigation differ from level to level.

At each level, morphological integration arises from some sort of interaction among parts. The kind of interaction differs according to the type of integration – it can be developmental interactions for developmental integration, functional interactions for functional integration, shared inheritance by pleiotropy or linkage for genetic integration, concerted evolution among parts for evolutionary integration, and so on.

The interactions that constitute the mechanisms responsible for morphological integration at the different levels are mostly not directly observable. Morphological integration is manifest, however, in the covariation among morphological traits. With appropriate study designs, underlying mechanisms can be inferred from the covariation of morphometric measurements and hypotheses about their effects can be tested. Depending on the level of integration that is of interest, study designs differ, so that covariation of different origins can be analysed and the respective patterns of integration can be compared. For instance, genetic integration requires quantitative genetic designs, such as breeding experiments or measurements from individual for which pedigrees are available (Klingenberg and Leamy, 2001; Myers et al., 2006; Klingenberg et al., 2010; Adams, 2011; Martínez-Abadías et al., 2012a). Developmental integration can be inferred from covariation of fluctuating asymmetry (Klingenberg, 2003), which has been widely used for this purpose (Klingenberg and Zaklan, 2000; Klingenberg et al., 2001a, 2003; Laffont et al., 2009; Ivanović and Kalezić, 2010; Jojić et al., 2011, 2012). For studies of evolutionary integration, data about interspecific variation are used and phylogenetic comparative analyses are used (Monteiro et al., 2005; Drake and Klingenberg, 2010; Monteiro and Nogueira, 2010). Multiple levels of integration can be included in single studies and compared to each other (Klingenberg and McIntyre, 1998; Debat et al., 2000; Klingenberg et al., 2003; Monteiro et al., 2005; Willmore et al., 2005; Young and Badyaev, 2006; Drake and Klingenberg, 2010; Ivanović and Kalezić, 2010; Monteiro and Nogueira, 2010; Gonzalez et al., 2011c; Jamniczky and Hallgrímsson, 2011; Jojić et al., 2012; Klingenberg et al., 2012).

Modularity

Morphological integration is not uniform throughout entire organisms, but tends to be concentrated in certain complexes of parts that are tightly integrated internally, but are relatively independent of other such complexes. Such complexes are called modules. Modularity, therefore, is a concept that is closely connected to integration (e.g. Klingenberg 2008).

In particular, modularity has been prominent in the area of evolutionary developmental biology, where several authors have argued that it provides the flexibility for different traits to evolve more easily under differing or even conflicting functional demands (Raff, 1996; Wagner and Altenberg, 1996; Gerhart and Kirschner, 1997; Kirschner and Gerhart, 1998). Modularity, because of the weak integration among modules, can therefore mitigate the effects of constraints that would apply if variation were fully integrated. Some authors have hypothesized that genetic and developmental modularity should evolve to match functional modularity (Cheverud, 1984, 1996a; Wagner, 1996; Wagner and Altenberg, 1996). Such tests will be especially fruitful where developmental and functional considerations result in different hypotheses of modularity (Breuker et al., 2006a). So far, there is only limited evidence available, and testing this hypothesis with morphometric data has only begun recently (e.g., Klingenberg et al. 2010).

Modularity has become a very popular research topic in recent years and the concept applies to a wide range of fields in biology and beyond (Schlosser and Wagner, 2004; Callebaut and Rasskin-Gutman, 2005). In many of those applications, the starting information is about the actual interactions of interest, for instance, in biochemical or gene regulatory networks, or networks of social or ecological interactions and so on. The analyses therefore dissect those known networks into modules based on the patterns of connectivity among the parts. The context of morphometric studies of modularity differs from these network analyses in that the network of interactions is not directly observable, but that modularity needs to be inferred from patterns of covariation among parts. Depending on the biological question that is to be addressed, the study design can be chosen so that variation and covariation from which modules and interactions are inferred are at the appropriate level. Most often, this is intraspecific variation among individuals of a population, but it is possible to combine this level with the analysis of fluctuating asymmetry for inferring developmental modularity, with quantitative genetic analyses, or with phylogenetically informed comparative analyses across species to study evolutionary modularity (Hallgrímsson et al., 2009; Klingenberg, 2009; Drake and Klingenberg, 2010; Jojić et al., 2011, 2012; Martínez-Abadías et al., 2012a; Sanger et al., 2012).

A wide variety of morphometric methods have been used to study modularity in mammalian skulls. This diversity partly reflects differences in the concepts of modularity that different investigators use. The details of the definitions that are used can make a substantial difference for the choice of methods and the interpretations of results.

Morphometric methods

Morphometric studies of integration and modularity use a variety of methods that address specific questions, for instance, concerning the patterns of integration within a single structure or the strength of integration between different parts. Usually, a single study includes several different analyses, so that their results can be combined to provide a comprehensive understanding of integration in the structure under study. Some of these methods are the standard tools of geometric morphometrics, such as principal component analysis, whereas others are specialized methods that were developed specifically for this purpose. This overview will focus exclusively on landmark methods (for other overviews of morphometric methods to study integration and modularity, see Goswami and Polly 2010b; Zelditch et al. 2012).

Overall integration in a structure

Integration in a morphological structure means that different parts covary with each other. At the extreme, all components are perfectly correlated, so that variation of the relative positions of landmarks in any small region is sufficient to predict perfectly the variation of the relative positions of the remaining landmarks. If relations are linear, this also means that all variation is contained in a single dimension of shape space. From this reasoning, it follows that principal component analysis, with its natural focus on dimensionality of variation, is an important tool for investigating morphological integration.

Principal component analysis is a traditional method in multivariate statistics (e.g. Jolliffe 2002). In geometric morphometrics, it has sometimes been called relative warp analysis if the analysis uses the shape descriptors from the thin-plate spline approach (Rohlf, 1993), but, in the vast majority of analyses, this is identical to a principal component analysis of landmark coordinates after a Procrustes superimposition. Principal component analysis provides two main results that are particularly relevant for the study of morphological integration: the eigenvalues, which indicate the variance for which each principal component

accounts, and the eigenvectors, which indicate the shape features associated with the principal components.

The eigenvalues are the variances of the principal component scores. Because principal components are defined to maximize, successively, the variance for which they account, the first eigenvalue is the largest variance for any linear combination of the variables included in the analysis (subject to a scaling constraint). Similarly, the last eigenvalue is the smallest variance for any linear combination. As a consequence, the eigenvalues offer a convenient way to assess the degree to which variation is concentrated in just one or a few dimensions or spread across many dimensions. In many geometric morphometric studies, a large proportion of the total variation is concentrated in just a few dimensions, so that only the first few eigenvalues are relatively large and drop fairly quickly, before tapering off toward zero (e.g. Klingenberg and McIntyre 1998; Young 2004; Polly 2007; Cooper et al. 2010; Gómez-Robles et al. 2011a; Klingenberg et al. 2012). There are exceptions where variation is spread across very many dimensions and no principal component accounts for a very large share of the total shape variation (e.g. Martínez-Abadías et al. 2012a).

Shape changes associated with principal components, which can be obtained from the eigenvectors, can also be informative about integration and modularity. They indicate whether the dominant aspects of shape variation affect the entire structure or are mostly contained in particular parts or regions. This type of interpretation provides an intuitive idea of the patterns of variation, integration and modularity, but interpreting the shape changes associated with principal components can be quite difficult and clear inference about integration and modularity is often impossible. In particular, the eigenvectors are well defined only if the eigenvalues are clearly distinct. If two or more eigenvalues are identical, the corresponding eigenvectors can be rotated freely in the subspace that they span, so that these principal components are associated with varying combinations of shape features. Therefore, this type of interpretation should be used with considerable caution.

Principal component analysis is also important for the understanding of morphological integration for another reason: an important class of indices of integration are derived from the variance of eigenvalues of a correlation matrix of measurements (Wagner, 1984; Pavlicev et al., 2009; Haber, 2011). If there is complete integration, all variation in the data is in a single dimension of the phenotypic space and the variance of eigenvalues of the correlation matrix is maximal, whereas at the opposite extreme of the spectrum, with not integration whatsoever, all eigenvalues are equal and their variance therefore is zero. Although these indices have primarily been applied with traditional morphometric measurements (lengths, etc.), the index was modified for the context of geometric morphometrics (Young, 2006). As usual in geometric morphometrics, the covariance matrix and not the correlation matrix is used for this purpose (this is, among other reasons, because the eigenvalues of the correlation matrix of Procrustes coordinates are not invariant under rotation of the landmark configurations relative to the coordinate system). To obtain an index of integration that does not depend on the total amount of shape variation in the sample, the eigenvalues of the covariance matrix can be standardized by the total variance in the sample (Young, 2006). This index of integration, sometimes in variants that differ in the way eigenvalues were scaled, has been used in a growing number of studies (Willmore et al., 2006a; Hallgrímsson et al., 2009; Jamniczky and Hallgrímsson, 2009; Ivanović and Kalezić, 2010; Gonzalez et al., 2011b; Jojić et al., 2011; Gómez-Robles and Polly, 2012).

Integration between structures or their parts: partial least squares

Whereas overall integration within a structure is an important topic, many morphometric studies address more specific questions of the integration between specific parts within an overall structure (e.g. the face and neurocranium in the skull) or between entirely separate structures (e.g. the mandible and cranium). Again, both the strength of association between parts and the patterns of covariation are of interest in studies of morphological integration.

The most widely used morphometric method in studies of the patterns of covariation between sets of landmarks is partial least squares (e.g. Tabachnick and Bookstein 1990; Klingenberg and Zaklan 2000; Rohlf and Corti 2000; Klingenberg et al. 2001a; Bookstein et al. 2003; Klingenberg et al. 2003; Mitteroecker and Bookstein 2007; Hautier et al. 2012). This is a method that has a number of common features with principal component analysis, but instead of decomposing the overall variation in a configuration of landmarks into components according to the amounts of variation with which they are associated, partial least squares decomposes a matrix of covariances between two landmark configurations into pairs of axes (one axis for each configuration) that are shape features showing successively maximal covariance with each other. Both the covariance and the shape features are of interest in the context of morphometric integration.

Just as for principal components, partial least squares analysis provides information about the shape change associated with each axis. Shape changes can be visualized and interpreted anatomically in the same way as principal components and other results from morphometric analyses (Klingenberg et al., 2012). The difference is that partial least squares analysis yields the shape features that account for the most covariation between parts rather than overall variation throughout the entire structure under study.

If the analysis examines covariation between two sets of landmarks within a single configuration, such as face and neurocranium, there is a choice whether the analysis should use separate Procrustes superimpositions for the parts (Bastir and Rosas, 2005; Marugán-Lobón and Buscalioni, 2006; Laffont et al., 2009; Bastir et al., 2010; Gkantidis and Halazonetis, 2011; Jamniczky and Hallgrímsson, 2011; Martínez-Abadías et al., 2011; Parsons et al., 2011; Singh et al., 2012) or the data from a single Procrustes fit for the entire structure (Klingenberg and Zaklan, 2000; Bookstein et al., 2003; Klingenberg et al., 2003; Monteiro et al., 2005; Mitteroecker and Bookstein, 2008; Makedonska et al., 2012; Klingenberg and Marugán-Lobón, 2013). Both options are valid, but they may produce markedly different results because they differ in which aspects of covariation they consider (Klingenberg, 2009; Kulemeyer et al., 2009; McCane and Kean, 2011). The approach with separate Procrustes superimpositions focuses on the covariation between the shapes of the parts, each taken separately, and does not consider the covariation that is due to coordinated variation in the relative sizes or arrangement of the parts. By contrast, the method that uses a joint Procrustes fit for both parts examines covariation between parts in the context of the structure as a whole and therefore considers all aspects of covariation, including the relative sizes and arrangement of the parts. Because this additional component of covariation can amount to a significant proportion of the total covariation between parts of a configuration, the difference between the two types of analyses can be fairly substantial (Klingenberg, 2009; Kulemeyer et al., 2009; McCane and Kean, 2011). The effect of a joint Procrustes fit also needs to be taken into account for statistical testing of covariation, for instance by including a new Procrustes fit in each iteration of a permutation test (Klingenberg et al., 2003; Klingenberg, 2009).

Most partial least squares analyses consider two blocks of variables (Rohlf and Corti, 2000). A few studies, however, have used three blocks simultaneously (Bookstein et al., 2003; Monteiro et al., 2005; Gunz and Harvati, 2007). The properties of these multi-block analyses are similar to the two-block analyses, but differ in some details. The shape features extracted for each block of variables are those that most strongly covary across all blocks simultaneously.

To quantify the strength of covariation between parts, it is possible to compute indices such as the RV coefficient (Escoufier, 1973; Klingenberg, 2009). Because the RV coefficient quantifies the total covariation as the sum of all squared covariances between blocks, it is compatible with the framework of partial least squares, where this sum of squared covariances (identical to the sum of squared singular values) also plays an important role (Rohlf and Corti, 2000). The RV coefficient is a multivariate generalization of the squared correlation coefficient but, rather than indicating the degree of association between two variables, it quantifies the strength of association between two sets of

variables. It can be used to assess how strong the integration between two sets of landmarks is, either within a single structure where both sets have undergone a joint Procrustes fit or for two configurations of landmarks with separate Procrustes fits. Although the RV coefficient can be computed in both these situations, it is important to note that there can be substantial differences between them, because a simultaneous Procrustes superimposition includes covariance due to variation in the relative sizes and arrangement of the two sub-configurations (Klingenberg, 2009). The RV coefficient is compatible with partial least squares analysis because it is based on related algebra, with the sum of all squared covariances between the two sets of variables as the measure of covariation. As a scalar measure of association between sets of variables, the RV coefficient can be used as an intuitive tool to assess the strength of integration between structures, but it is also useful as a test statistic for permutation tests of association in the context of partial least squares (e.g. as implemented in the MorphoJ software, Klingenberg 2011) or for analyses of modularity (Klingenberg, 2009). All these usages of the RV coefficient are increasingly widespread in various contexts (e.g. Laffont et al. 2009; Drake and Klingenberg 2010; Ivanović and Kalezić 2010; Gómez-Robles and Polly 2012; Jojić et al. 2012; Renaud et al. 2012; Sanger et al. 2012; Klingenberg and Marugán-Lobón 2013).

The RV coefficient is not the only index of association between sets of variables. The simplest way of quantifying covariation is to compute the sum of squared covariances between the two blocks of variables, which is the same as the sum of squared singular values that is computed as part of a partial least squares analysis (Bookstein et al., 1990; Rohlf and Corti, 2000). It is related to the RV coefficient, which is a scaled variant of this sum (scaled by the total within-block covariation). Other measures of covariation between sets of variables, such as the trace correlation (Hooper, 1959; Mardia et al., 1979), which has been used as a measure of integration in geometric morphometrics (Klingenberg et al., 2003, 2004), can show undesirable statistical behaviour. It is therefore recommended to avoid such measures and to use the RV coefficient instead.

Associations between distance matrices

Studying patterns and strength of integration with partial least squares analysis is the most widespread, but not the only approach. An alternative method is based on computing a matrix of all pairwise Procrustes distances between the study units (species average shapes, individuals, etc.) for each part and then comparing the distance matrices for different parts using matrix correlation and Mantel tests (Monteiro et al., 2005). The result of this type of analysis is an overall measure of association between distance matrices of the parts, which can be further used, for example, in cluster analyses to explore the structure of covariation among parts (Monteiro et al., 2005). This method does not provide a direct visualization of the patterns of covariation in the shapes of the parts, which is why these analyses are sometimes are supplemented by a partial least squares analysis that provides those patterns (Monteiro et al., 2005).

The analyses with this approach can be conducted at various levels, for instance among individual within species or among taxa in a phylogenetic comparison (Monteiro and Nogueira, 2010) or variation among individuals and fluctuating asymmetry (Zelditch et al., 2008). This approach has been applied in a range of studies (Monteiro et al., 2005; Zelditch et al., 2008, 2009; Monteiro and Nogueira, 2010; Muñoz-Muñoz et al., 2011; Webster and Zelditch, 2011a,b; Renaud et al., 2012).

Comparing covariance matrices

Patterns of integration can be compared by quantifying the resemblance between covariance matrices of shape variables. A widely used index is matrix correlation, the product moment correlation between corresponding elements in two covariance matrices. Matrix correlations can be tested with a matrix permutation procedure against the null hypothesis that the two covariance matrices are totally unrelated. Both the computation of matrix correlations and the matrix permutation procedure have been specially adapted for the context of geometric morphometrics (Klingenberg and McIntyre, 1998). These adaptations concern the choice whether the diagonal blocks of the matrices, containing the variances and covariances of the landmark coordinates within landmarks, should be included or not. The difference between the two versions sometimes can provide insight into the nature of the resemblance between two covariance matrices, but often both versions provide largely similar results (e.g. Breuker et al. 2006b). Moreover, the matrix permutation test needs to permute landmarks rather than individual coordinates in order to simulate the null hypothesis in a realistic manner (Klingenberg and McIntyre, 1998).

Whereas the matrix correlation is an intuitive measure of resemblance between covariance matrices, it is not the only such measure. Several measures of distances between covariance matrices have been described, which could be used in a similar way (Dryden et al., 2009; Mitteroecker and Bookstein, 2009). For statistical inference based on these measures, however, tests other than the matrix permutation tests will be needed.

For comparison among multiple covariance matrices, it is possible to obtain an ordination of covariance matrices by principal coordinate analysis based on a distance measurement derived from matrix correlation or on other distance measurements among covariance matrices (Debat et al., 2006, 2008, 2009; Jamniczky and Hallgrímsson, 2009; Mitteroecker and Bookstein, 2009; Breno et al., 2011; Gonzalez et al., 2011a). This type of analysis provides scatter plots where the relationships among covariance matrices can be interpreted graphically. It should be noted, however, that the distance measures are not well understood and several others have been described (Dryden et al., 2009). These distance measures differ in which aspects of the covariance matrices they consider. For instance, the distance measure based on matrix correlation is invariant to changes of scale and focuses exclusively on the patterns of covariation, whereas the Riemannian metric (Dryden et al., 2009; Mitteroecker and Bookstein, 2009) considers both the pattern and scale of variation. Two studies compared ordinations derived from different distance measures with the same set of covariance matrices and found some agreement, but also considerable differences (Breno et al., 2011; Gonzalez et al., 2011a). More theoretical work for understanding the nature of the different distance measures and more empirical comparisons of distances are needed.

For all these types of comparisons, a possible complication arises if the structure under study is symmetric in itself, such as the mammalian skull. This type of symmetry is known as object symmetry, and differs from matching symmetry, where there are separate structures on the left and right body sides, such as human hands (Mardia et al., 2000; Klingenberg et al., 2002). For structures with object symmetry, there are two types of landmarks: single landmarks on the midline or midplane and paired landmarks on either side of it. Also, the total shape variation can be divided into separate components of symmetric variation and asymmetry, which occupy orthogonal subspaces of the shape tangent space (Klingenberg et al., 2002). Special care is needed if covariance matrices derived from the two components need to be compared to each other, for instance to compare the patterns of symmetric variation among individual and of fluctuating asymmetry. Because the symmetry and asymmetry components occupy orthogonal subspaces, the whole covariance matrices are uncorrelated, even if there are clear relations in the patterns of shape changes (in whatever dimension there is variation in one component, there is none in the opposite component). A possible solution is to concentrate exclusively on the paired landmarks from one body side (i.e. on one landmark from each pair, after computing Procrustes tangent coordinates for the symmertry or asymmetry components from the entire landmark configurations) and to ignore the landmarks on the midline (Klingenberg et al., 2002). This approach leaves out those aspects of variation that cannot agree between the two components (e.g. for unpaired landmarks, symmetric variation is in the midline or midplane and asymmetry is perpendicular to it) and focuses on those features of shape variation that may or may not match. This limitation and the solution applies to matrix correlation as well as to the computation of the various distance measures for

principal coordinate analyses. A growing number of studies have used this approach for computing matrix correlations in animals and plants (Klingenberg et al., 2002; Drake and Klingenberg, 2010; Ivanović and Kalezić, 2010; Breno et al., 2011; Jojić et al., 2011; Klingenberg et al., 2012). It is implemented in the MorphoJ software package (Klingenberg, 2011) and is automatically used by the program in appropriate situations; in other software, users need to make the necessary adjustments manually.

Assessing modularity

Hypotheses of modularity can originate from many types of reasoning, grounded in anatomical, developmental, functional or genetic arguments (Breuker et al., 2006a; Willmore et al., 2006a; Klingenberg, 2008). In the context of geometric morphometrics, hypotheses of modularity are stated in terms of the landmarks that belong to the putative modules. Under such a hypothesis, the relative positions of landmarks belonging to the same module should be integrated relatively tightly, whereas integration between modules should be weaker. By contrast, if the landmarks are partitioned into subsets differently, so that their divisions do not coincide with the boundaries of true modules, the strong within-module covariation contributes to the covariation among subsets, and the overall covariation among subsets of landmarks is therefore expected to be stronger. This prediction can be assessed empirically by quantifying covariation among the sets of landmarks belonging to the different hypothesized modules and comparing this with the strength of covariation among subsets of landmarks that have been partitioned in different ways (Klingenberg, 2009).

To quantify the strength of covariation, the RV coefficient between the Procrustes coordinates of the sets of landmarks can be used if there are just two hypothesized modules or, for more than two modules, a multi-set RV coefficient can be computed by averaging the RV coefficients for all pairwise combinations of the subsets of landmarks (Klingenberg, 2009). The RV coefficient or multi-set RV coefficient is first computed for the partition of landmarks into subsets that coincide with the hypothesis of modularity, and then for a number of alternative partitions of the landmarks. Depending on the number of hypothesized modules and landmarks, this may be a full enumeration of all possible partitions or a large number of random partitions.

Also, depending on the biological context, investigators may choose to restrict the alternative partitions to those that are spatially contiguous. If disjoint sets of landmarks are not considered to be plausible candidates for modules, for instance for developmental modules where the interactions responsible for the integration within modules are tissue-bound processes such as the diffusion of signalling factors, it may be undesirable to include partitions of landmarks with disjoint sets as part of the comparison. A criterion for the spatial contiguity of subsets can be defined by requiring subsets to be connected by the edges of an graph that specifies which landmarks are considered to be adjacent to each other (Klingenberg, 2009). Whether disjoint subsets of landmarks are plausible as potential modules or whether each module needs to be spatially contiguous depends entirely on the biological context of each specific analysis.

A growing number of studies have used this method for testing hypotheses of modularity (Hallgrímsson et al., 2009; Klingenberg, 2009; Bruner et al., 2010; Drake and Klingenberg, 2010; Ivanović and Kalezić, 2010; Klingenberg et al., 2010; Jojić et al., 2011; Burgio et al., 2012a; Jojić et al., 2012; Kimmel et al., 2012; Lewton, 2012; Martínez-Abadías et al., 2012a; Sanger et al., 2012; Sydney et al., 2012; Klingenberg and Marugán-Lobón, 2013). Beyond biological structures, the method has even been used to evaluate modularity in archaeological artefacts (González-José and Charlin, 2012).

Several other methods also exist that assess models of modularity in shape data within a variety of different statistical frameworks (e.g. Mitteroecker and Bookstein 2007; Márquez 2008. Jojić et al. (2012) conducted a direct comparison, using the same data, between the method described above and traditional morphometric approaches (Cheverud, 1995, 1996b) and found that both approaches produced compatible results and supported the same conclusions concerning modularity. This comparison is encouraging and suggests that comparisons of results across studies are possible.

Whereas testing a-priori hypotheses of modularity is an important aspect of the study of modularity, investigators often want to find modules with exploratory analyses when no such hypotheses are available. This means that the task is to search for a partition of landmarks into subsets so that the covariation between subsets is minimal and strong covariation of relative landmark positions is mainly confined within subsets. It is tempting to use the combinatorial approach that is part of the approach outlined above, and simply to search for that partition of landmarks that results in the weakest covariation between subsets (lowest RV coefficient or multi-set RV coefficient). However, the problem with this idea is that there always is at least one partition that provides the weakest covariation, even if there is no modularity at all in the structure under study. Accordingly, this approach is not suitable for an exploratory search for modules, but should be restricted to the test of a-priori hypotheses (Klingenberg, 2009).

A different strategy is to use some clustering technique as an exploratory approach to find modules based on a measure of covariation among landmarks or regions of the structure under study (e.g. Monteiro et al. 2005; Goswami 2006a,b; Willmore et al. 2006a; Goswami 2007; Zelditch et al. 2008, 2009; Makedonska et al. 2012). Goswami (2006a,b, 2007) used a scalar measure of covariance among landmarks, which combined information about both the degree of association and the angles between the relative shifts of landmarks, in a hierarchical cluster analysis to identify modules in mammalian skulls. Other studies used cluster analysis to examine the larger-scale structure of variation among parts of a structure (Monteiro et al., 2005; Zelditch et al., 2008, 2009). Because clustering algorithms always produce clusters, even when the data do not contain such a hierarchical structure (Sneath and Sokal, 1973), some considerable caution is required when interpreting the results from such exploratory analyses. It is tempting to use the approach outlined above for testing the resulting modules, but there is a danger of circular reasoning if the clustering itself was based on information about covariation among landmarks. Much further work is required to develop methods for testing for modular structure in situations where no a-priori hypotheses are available.

Allometry as an integrating factor

Allometry, the association between size and shape, is a factor that contributes to morphological integration (Klingenberg et al., 2001a; Mitteroecker and Bookstein, 2007; Klingenberg, 2009). Because the relationship between size and shape is often nearly linear, allometry generates shape variation that is primarily in one direction in shape tangent space. If there is a relatively large amount of size variation and allometry is sufficiently strong, this size-related component of shape variation can make up a substantial proportion of shape variation. As a result, a considerable proportion of the total shape variation may be concentrated in the direction of the allometric effects. If allometric variation is sufficiently abundant relative to variation from other sources, it manifests itself as increased integration of shape variation. Also, allometry often affects all parts of a structure or organism jointly, leading to integrated change of the entire configuration of landmarks under study. Integration from allometry can vary in its effects, and it may or may not interfere with the analysis of modularity (Klingenberg, 2009; Jojić et al., 2012).

In geometric morphometrics, the most straightforward and most frequently used approach for analysing allometry is multivariate regression of shape on size (e.g. Loy et al. 1998; Monteiro 1999; Cardini 2003; Drake and Klingenberg 2008; Rodríguez-Mendoza et al. 2011; Sidlauskas et al. 2011; Weisensee and Jantz 2011; Klingenberg et al. 2012; Martínez-Abadías et al. 2012b; Ponssa and Candioti 2012). The most widely used shape measures are centroid size and log-transformed centroid size (log-transformed centroid size is particularly useful when the range of sizes is very large and much of the size-related shape change occurs among smaller specimens). The direction of allometry can be characterized by the vector of regression coefficients, which is the shape change that is expected for an increase of size by one unit

(usually centroid size or log-transformed centroid size is used; note that the magnitude of this vector depends on the units of measurement or the basis used for the log transformation). To correct for the effects of allometry in a dataset, investigators can compute the residuals from the regression of shape on size and use those residuals in further analyses of integration or modularity (e.g. Klingenberg 2009; Jojić et al. 2012; Klingenberg and Marugán-Lobón 2013).

Depending on the context of a study, different levels of allometry may be relevant: ontogenetic allometry from size variation due to growth, evolutionary allometry due to evolutionary size changes among species, and static allometry due to individual size variation at a given ontogenetic stage within a single population (e.g. Cock 1966; Gould 1966; Cheverud 1982b; Klingenberg and Zimmermann 1992; Klingenberg 1996, 1998). When data for multiple levels are available within a single study, comparing them explicitly and relating them to patterns of integration may provide insight into the origins of observed patterns of variation. Examples of such studies include comparisons of static and ontogenetic allometry (Drake and Klingenberg, 2008; Weisensee and Jantz, 2011), ontogenetic and evolutionary allometry (Gonzalez et al., 2011c) and of static allometry within populations and evolutionary allometry among taxa (Klingenberg et al., 2012). All of these types of allometry can contribute to integration at the respective levels of variation.

Martínez-Abadías et al. (2012b) used multivariate regression to relate skull shape not only to the centroid size of the skull, but also to chondrocranial length and to estimated brain volume in mice and humans. Like allometry, these relations to additional quantities, which are linked with important developmental factors and processes, contribute to variation across the skull and can contribute to overall integration. The interplay of the different processes is likely to be complex and difficult to untangle with morphometric methods (e.g. Hallgrímsson et al. 2009).

Sample sizes

Integration and modularity concern the patterns of variation and covariation about the mean of shape or other morphological traits. Because morphometric data usually have a high dimensionality, estimating the patterns of covariation is inherently a quite ambitious statistical undertaking and requires a substantial sample size. Theoretical considerations and simulations indicate that estimates of mean shape using Procrustes superimposition are well behaved, provided shape variation is sufficiently concentrated around the mean shape (Dryden and Mardia, 1998; Rohlf, 2003), which applies for most biological data even on large taxonomic scales (Marcus et al., 2000). This is much less clear, however, for measures of variation around the mean, which are of central importance for studies of integration and modularity.

Some studies have used rarefaction methods, subsampling from the original data, to examine the statistical behaviour of estimates of integration and modularity (Polly, 2005; Goswami, 2006a; Cardini and Elton, 2007; Goswami and Polly, 2010b). There is a possibility, however, that rarefaction procedures provide an overoptimistic view of the required sample sizes because they ignore the fact that even the total dataset is a sample drawn from some population. An alternative to rarefaction is the bootstrap, which is a resampling method that explicitly considers that the total dataset itself has been obtained by sampling (Efron and Tibshirani, 1993). This method has also been used to assess the reliability of summary statistics such as centroid size and the total variance of shape and suggested that relatively modest sample sizes can give reasonable estimates of those statistics (Cardini and Elton, 2007).

There are also some simple theoretical considerations that can guide the choice of sample sizes. Because most analyses of integration use covariance matrices in one way or another, sample sizes need to provide sufficiently reliable estimates of covariance matrices. The usual recommendation for multivariate analyses is that sample size should be substantially larger than the dimensionality of the data, which is often very large in geometric morphometrics because of the high dimensionality of the shape tangent space (nearly two or three times the number of landmarks, depending on whether the data are in 2D or 3D – but only about half if the data have object symmetry and only the symmetric or asymmetric component of variation are used; Klingenberg et al. 2002). If integration is strong, so that the vast majority of variation is contained within a lower-dimensional subspace of the shape tangent space, a considerably lower sample size may be sufficient. This reasoning suggests an informal procedure for determining whether sample size is sufficient: a preliminary principal component analysis is used to check the dimensionality of the data, and if the sample size is much larger than the number of principal components that account for most of the total shape variance (e.g. 95%), the sample size should be sufficient for analyses of integration and modularity.

Levels of integration and study designs

Morphological integration applies at a range at different levels, which can provide information about different biological processes concerning variation in a structure (Klingenberg, 2008). This is similar to allometry, where levels such as static, ontogenetic and evolutionary allometry have long been known and discussed extensively (Cock, 1966; Gould, 1966; Cheverud, 1982b; Klingenberg and Zimmermann, 1992; Klingenberg, 1996, 1998). For allometry, the levels are defined by the process that produces the size variation involved in the allometric effects (individual variation within populations at a given ontogenetic stage for static allometry, growth for ontogenetic allometry, evolutionary change of size for evolutionary allometry). Similarly, the levels of morphological integration are defined by the processes responsible for the morphological variation.

The different levels apply to the same structures, but examine integration and modularity in different contexts (Klingenberg, 2008). It is therefore possible to study integration at multiple levels jointly with the same set of landmarks or other morphometric data. Depending on which level is of interest, the data have to be collected according to different study designs—data from multiple related species are needed for studies of evolutionary integration, genetic information is needed to infer genetic integration, an understanding of biomechanical interactions provides information about functional integration, and so on. It is therefore important to tailor the study design to the specific questions being asked. Often, however, a single study design can provide information on integration at multiple levels. This possibility is easily overlooked because the vast majority of studies of integration have focussed entirely on the level of intraspecific or intra-population variation.

Some authors have hypothesized that genetic and developmental modularity evolve adaptively to match functional modularity—traits that are involved jointly in particular functions should evolve to share common developmental pathways and common genetic control of morphological variation (Cheverud, 1984, 1996a; Wagner, 1996; Wagner and Altenberg, 1996). This "matching hypothesis" can be tested by comparing modularity across levels: functional versus developmental and genetic modularity (Breuker et al., 2006a). Such tests are expected to be particularly fruitful if there are reasons to expect differences between levels in how traits are grouped together as modules, so that it will become apparent if modularity at one level "wins" over another level (Breuker et al., 2006a). Only few analyses have conducted such tests of the matching hypothesis so far (Breuker et al., 2006a; Klingenberg et al., 2010), but there are many opportunities for future studies. Further topics for comparison across levels comes from the idea that the dominant direction of genetic variation in shape space, the first principal component of the genetic covariance matrix, within populations is a "line of least resistance" for evolutionary change (Schluter, 1996; Renaud and Auffray, this issue) or, if multiple principal components account for substantial variation, that there may a plane or subspace of least resistance (Martínez-Abadías et al., 2012a). Overall, therefore, studies comparing patterns of integration and modularity across different levels are a promising area for future studies.

Integration within populations and species is important because it constitutes a basis for comparison and figures prominently in evolutionary theory (e.g., Steppan et al. 2002). Accordingly, it is justified that the vast majority of studies of integration and modularity have fo-

cussed on this level of variation (often, such studies do not specify the level of variation, but authors simply write about "modularity" or "morphological integration" without any additional qualifier). Even studies of modularity at large taxonomic scale mostly have concentrated on comparisons of intraspecific modularity (e.g. Steppan 1997a,b; Goswami 2006a,b, 2007; Sanger et al. 2012). Whereas this interest in intraspecific integration and modularity is clearly justified, other levels of variation are providing interesting information as well, and this information is often available from the same samples.

The following sections will focus on three of the other levels of integration that have been relatively well studied: developmental integration (accessible via analyses of integration in fluctuating asymmetry), genetic integration and evolutionary integration.

Fluctuating asymmetry and developmental integration

The developmental basis of morphological integration is critically important for understanding evolutionary processes, but it cannot be observed directly from morphological data and needs to be inferred. Morphological integration can be the result of developmental interactions between precursors of the traits, or it can result from environmental variation or genetic differences that affect multiple traits simultaneously, even if those traits are separate and do not interact in their development. A tool that can be used for inferring direct interactions between the developmental pathways that produce different morphological traits is fluctuating asymmetry (Klingenberg, 2003, 2005). Because fluctuating asymmetry originates from random perturbations in developmental processes, it can only produce covariation of asymmetry between two traits if the effects of the perturbations are transmitted between precursors of the traits by developmental interactions. If there are no such interactions, asymmetries of the traits are uncorrelated. The integration of fluctuating asymmetry can therefore be used as a tool to assess the developmental origin of morphological integration.

Methods for the study of fluctuating asymmetry have been firmly established in geometric morphometrics, both for the situation where there are two separate structures on the left and right body sides (matching symmetry) and for object symmetry, where the entire structure is symmetric in itself because the axis or plane of symmetry passes through it (Klingenberg and McIntyre, 1998; Mardia et al., 2000; Klingenberg et al., 2002). For matching symmetry, there are separate landmark configurations for the left and right sides that are subjected jointly to a Procrustes superimposition after reflection of all configurations from one side (e.g. all configurations from the left side). Asymmetry can be computed from the differences between the shapes and sizes of the left and right sides. For object symmetry, a copy of each landmark configuration is reflected to its mirror image and paired landmarks are relabelled so that the reflected copy can be fit together with the untransformed copy, and all copies are then used together in a Procrustes superimposition. Asymmetry can be computed from the differences between the original and reflected and relabelled configurations after the Procrustes fit (for more details, see Klingenberg et al. 2002). This type of analysis has also been generalized for complex types of symmetry (Savriama and Klingenberg, 2011; Savriama et al., 2012).

Fluctuating asymmetry is used in an increasing number of analyses to investigate the developmental basis of integration (Klingenberg and Zaklan, 2000; Klingenberg et al., 2001a, 2003; Hallgrímsson et al., 2004; Breuker et al., 2006b; Young and Badyaev, 2006; Zelditch et al., 2008; Klingenberg, 2009; Laffont et al., 2009; Zelditch et al., 2009; Drake and Klingenberg, 2010; Ivanović and Kalezić, 2010; Klingenberg et al., 2010; Jamniczky and Hallgrímsson, 2011; Jojić et al., 2011; Webster and Zelditch, 2011a; Jojić et al., 2012; Klingenberg et al., 2012). Other studies have collected the same type of data for different purposes, such as comparing the patterns of developmental instability and canalization (Debat et al., 2000; Santos et al., 2005; Willmore et al., 2005, 2006a,b; Debat et al., 2008, 2009; Breno et al., 2011). Together, these studies constitute a growing database that can be used to assess the importance of direct developmental interactions for integration in morphological structures. Unfortunately, the results of these studies are quite heterogeneous at the moment and do not allow any generalizations. This heterogeneity is at least partly due to true biological differences, but it is possible that heterogeneity of experimental designs, rearing procedures and morphometric analyses also contributes to the variation. At this time, it seems advisable to investigate the developmental origin of integration on a case-by-case basis—it is possible, however, that general trends will emerge if more case studies are done.

Genetic integration

Because patterns of genetic variation and covariation are critical determinants of evolutionary change (Lande, 1979; Roff, 1997), patterns of genetic integration have long received considerable attention in evolutionary biology. In particular, many studies have focused on model systems such as the mouse mandible (Atchley et al., 1985; Atchley and Hall, 1991; Atchley, 1993; Cheverud, 1996a; Mezey et al., 2000; Klingenberg et al., 2001b, 2004; Burgio et al., 2012a) and the *Drosophila* wing (Fernández Iriarte et al., 2003; Houle et al., 2003; Mezey and Houle, 2005; Santos et al., 2005). More and more studies are specifically addressing questions concerning genetic integration of complex structures in these models, but also in a growing number of non-model organisms including humans (Martínez-Abadías et al., 2009, 2012a). Whereas most genetic studies have used traditional morphometric approaches, the methods of geometric morphometrics are becoming more widespread in genetics.

Several distinct approaches exist for genetic studies that relate to morphological integration. Perhaps the most straightforward approach is to examine the effects of specific mutations on the strength and patterns of integration. Studies of this kind have been conducted in mice (Hallgrímsson et al., 2006; Willmore et al., 2006b; Hallgrímsson et al., 2009; Martínez-Abadías et al., 2011; Parsons et al., 2011) and *Drosophila* wings (Debat et al., 2006, 2009, 2011). These experiments have the potential to reveal possible mechanisms that contribute to integration, but most studies so far have yielded complex results, so that it is difficult or impossible to draw clear-cut conclusions.

A second approach is to use gene mapping techniques to find quantitative trait loci (QTLs) that affect the shape of a structure and to examine integration and modularity of their effects. QTL studies require information on genetic markers and shape data for a suitably structured study population (e.g. the F2 or a later generation from a cross between two inbred lines). Such studies have been done in mice with traditional morphometric methods (Cheverud et al., 1997; Leamy et al., 1999; Mezey et al., 2000; Ehrich et al., 2003) and with geometric morphometric approaches (Klingenberg et al., 2001b; Workman et al., 2002; Klingenberg et al., 2004). Although the results of these studies depend considerably on the specific methods and data used, they have tended to show that the genetic architecture of shape is quite complex, with many QTLs affecting shape, and that nonadditive effects are important. Other studies used different genetic designs where phenotypic changes can be related to substitutions of chromosomal regions between strains, have broadly supported these results (Burgio et al., 2009; Boell et al., 2011; Burgio et al., 2012a,b).

Yet another strategy is to investigate genetic variation in a population without separating the effects of individual loci at all, but to focus on the aggregate effect of the whole genome on a set of morphological traits. The patterns and amount of genetic variation can be obtained from genetic covariance matrices, which can be estimated from shape data in a breeding experiment or in a study population of individuals for which mutual genealogical relationships are known. These requirements can be met relatively easily for many types of organisms and, accordingly, quantitative genetic studies of shape have been conducted in a wide range of plants and animals, including humans (Klingenberg and Leamy, 2001; Mezey and Houle, 2005; Willmore et al., 2005; Myers et al., 2006; Gómez et al., 2009; Martínez-Abadías et al., 2009; Klingenberg et al., 2010; Adams, 2011; Martínez-Abadías et al., 2012a). Several studies have specifically considered questions about genetic integration or modularity. Because the genetic covariance matrix can be used to predict the response to selection, it is possible to simulate localized selection focused on a particular part and examine

whether the response is also local or global throughout the entire structure under study (Klingenberg and Leamy, 2001; Martínez-Abadías et al., 2009; Klingenberg et al., 2010; Martínez-Abadías et al., 2012a). Also, it is possible to conduct tests of modularity using genetic covariance matrices (Klingenberg et al., 2010; Martínez-Abadías et al., 2012a).

If integration of the total genetic variation in a population is strong, so that the first principal component of the genetic covariance matrix accounts for a disproportionate share of the total genetic variation, then evolution by drift or selection can occur more easily in the direction of that first principal component than in other directions. In other words, this first principal component acts as a "genetic line of least resistance" (Schluter, 1996). If there is not a single principal component that accounts for an unusually large share of the genetic variation, but two or a few principal components together account for much of the total genetic variation, they can instead form a plane or subspace of least resistance (Martínez-Abadías et al., 2012a). In this way, genetic integration may have a substantial potential influence on long-term evolutionary outcomes.

Even though genetic studies of integration and modularity have become easier with advances in statistical methods and genotyping technology, investigating the genetic basis of morphological integration and modularity remains challenging. Most studies are limited to some extent by statistical power to detect genetic effects or by the uncertainty about estimates of genetic parameters due to limited sample sizes. Further work, using large experimental designs or extensive pedigree data, will be necessary to characterize the exact nature of genetic integration and modularity in complex structures such as the mammalian skull.

Evolutionary integration: comparative methods

Integration and modularity not only apply to the patterns of covariation within populations and species, but also to the patterns of covariation among evolutionary changes of shape in a clade of related taxa. At this macroevolutionary scale, studies of integration and modularity need to use comparative methods to take into account the phylogenetic structure of variation. With this proviso, all the morphometric methods for investigating patterns of covariation and modularity can also be used at this level (Klingenberg and Marugán-Lobón, 2013).

To assess whether there is phylogenetic signal in the morphometric data, a permutation test has been described that simulates the null hypothesis of a complete lack of phylogenetic structure by randomly swapping the shape data across the taxa in the phylogeny (Klingenberg and Gidaszewski, 2010). This test is now widely used in comparative studies of shape (e.g. Gidaszewski et al. 2009; Figueirido et al. 2010; Álvarez et al. 2011a; Fortuny et al. 2011; Meloro et al. 2011; Perez et al. 2011; Álvarez and Perez 2013; Brusatte et al. 2012; Klingenberg et al. 2012; Meloro and Jones 2012; Sanger et al. 2012; Klingenberg and Marugán-Lobón 2013). In the majority of studies, there is a significant phylogenetic signal, so that phylogenetic comparative methods should normally be used for analyses of evolutionary integration and modularity (Klingenberg and Gidaszewski, 2010; Klingenberg and Marugán-Lobón, 2013).

Phylogenetic comparative methods are now well known and widely available, particularly independent contrasts and phylogenetic generalized least squares (Felsenstein, 1985; Rohlf, 2001; Felsenstein, 2004). Both these methods have been shown to provide equivalent results (Rohlf, 2001; Blomberg et al., 2012), both can easily accommodate multivariate data, and both can therefore be used in the context of geometric morphometrics.

The analysis of evolutionary integration and modularity can use covariance matrices computed from independent contrasts or the estimated matrix of evolutionary covariances from phylogenetic generalized linear models with the same morphometric tools as they are used for other levels of variation. Overall patterns and the amounts and dimensionality of evolutionary variation can be studied with principal component analysis of the covariance matrix of independent contrasts (Klingenberg et al., 2012; Klingenberg and Marugán-Lobón, 2013). Estimating evolutionary allometry and correcting for allometric effects of evolutionary changes in shape can be achieved by multivariate regression of independent contrasts of shape onto independent contrasts of size (usually centroid size or log-transformed centroid size; Figueirido et al. 2010; Swiderski and Zelditch 2010; Perez et al. 2011; Klingenberg et al. 2012; Klingenberg and Marugán-Lobón 2013). Evolutionary integration between different structures or parts can be studied with partial least squares analysis of independent contrasts (Bastir et al., 2010; Klingenberg and Marugán-Lobón, 2013) or with partial least squares analysis of covariance matrices from phylogenetic generalized least squares (Dornburg et al., 2011; Meloro et al., 2011). Modularity tests also can be conducted with covariance matrices computed from independent contrasts (Drake and Klingenberg, 2010; Klingenberg and Marugán-Lobón, 2013). These analyses use the same computations for evolutionary integration and modularity that are used in analyses at different levels and the results are directly comparable across levels. Phylogenetic comparative methods are only beginning to be used in geometric morphometrics, but it is likely that they will become much more widespread in the near future.

These methods are related to the method of mapping morphometric data onto phylogenies using squared change parsimony, which has been used increasingly in recent years (Klingenberg and Ekau, 1996; Rohlf, 2002; Nicola et al., 2003; Linde et al., 2004; Macholán, 2006; Sidlauskas, 2008; Astúa, 2009; Gidaszewski et al., 2009; Figueirido et al., 2010; Klingenberg and Gidaszewski, 2010; De Esteban-Trivigno, 2011a,b; Dornburg et al., 2011; Fortuny et al., 2011; Monteiro and Nogueira, 2011; Brusatte et al., 2012; Klingenberg et al., 2012; Meloro and Jones, 2012; Klingenberg and Marugán-Lobón, 2013). This approach provides a direct visualization of evolutionary changes in shape space, and therefore is a useful tool for understanding the evolutionary history of morphological structures. Whereas analyses using comparative methods such as independent contrasts provide summary information about patterns of shape changes on all branches of the phylogeny, this graphical approach is complementary because it provides information about changes on particular branches and the occupation of the shape space by different subclades within the group under study. The two approaches are therefore complementary and can be used in combination with each other (Klingenberg and Marugán-Lobón, 2013).

As an alternative to phylogenetic comparative methods, other studies have used an approach based on taxonomic hierarchy, using distances between average shapes for different taxonomic levels (within species, among species within genera, among genera, etc.) to construct distance matrices for different parts of a structure and then to characterize integration using the relations among distance matrices (Monteiro et al., 2005; Monteiro and Nogueira, 2010).

Integration and modularity of cranial shape in mammals

So far, this review has focused on the methodology for analysing morphological integration and modularity. For the remaining part of this paper, however, I will organize the discussion around the primary mammalian systems that have been used in morphometric studies and what their results imply for our understanding of mammalian evolution. The summary that follows is an overview of work on morphological integration that has been done in the main mammalian study systems (both model and "non-model" systems), mostly with geometric morphometric methods. Because the literature relating to integration and modularity in mammalian skulls has become so large, it is impossible to present a complete survey. I apologise to the authors whose work I had to omit.

High-level comparisons

Morphometric analyses at large phylogenetic scales can be challenging because of difficulties in identifying homologous landmarks and the sheer scale of variation, which can pose challenges to morphometric procedures such as the tangent space approximation to Kendall's shape space. Nevertheless, empirical analyses have demonstrated that analyses of skull shape can be feasible, even for analyses across all mammalian orders (Marcus et al., 2000).

Some studies compared modularity and integration across mammals with different methods and found conserved features (Goswami, 2006a; Porto et al., 2009) and identified possible consequences for macroevolutionary processes (Marroig et al., 2009; Goswami and Polly, 2010a). Other studies focused on comparing between specific, phylogenetically remote groups, such as the comparison between didelphid marsupials and New-World monkeys (Shirai and Marroig, 2010) or carnivorans and primates (Goswami and Polly, 2010a). Some studies have compared aspects of integration and allometry between mice and humans or other primates to identify commonalities in the developmental origin of variation (Hallgrímsson et al., 2004; Martínez-Abadías et al., 2012b). Other studies compared major clades in terms of evolutionary integration to examine whether variation in diet brings about similar responses in different clades (Wroe and Milne, 2007; De Esteban-Trivigno, 2011a; Goswami et al., 2011). A general interpretation of the results is difficult, because these studies used very different methods and sampling strategies.

These studies demonstrate that analyses of integration at very large phylogenetic scale are feasible. Much more work is required to establish patterns of evolutionary integration across the mammalian phylogeny and within-taxon patterns of integration.

Mouse and other rodents

The mouse mandible has long been a model for development and evolution of complex morphological structures (Atchley and Hall, 1991; Klingenberg and Navarro, 2012). It is composed of several units with distinct developmental origins, functional roles and, to some extent, separate inheritance and evolutionary history. Yet, the mandible is still relatively simple, by comparison to structures such as the cranium, so that it can be studied relatively easily. Accordingly, many studies have investigated the mandible in the context of development, genetics and evolution.

A particular focus of the research on morphological integration and modularity in the mouse mandible was the hypothesis that the alveolar region and ascending ramus are two distinct modules. Sometimes, these modules were further subdivided into smaller elements (Atchley and Hall, 1991; Atchley, 1993). Evidence for modularity first came from studies using traditional morphometrics in combination with different quantitative genetic approaches in laboratory mice (Atchley et al., 1985; Leamy, 1993; Cheverud, 1996a; Cheverud et al., 1997; Mezey et al., 2000; Ehrich et al., 2003) and later also from studies using geometric morphometrics (Klingenberg and Leamy, 2001; Klingenberg et al., 2001b, 2003, 2004; Klingenberg, 2009; Burgio et al., 2012a; Renaud et al., 2012). Studies in hybrids between different subspecies of house mice indicate that hybridization is affecting different parts of the mandible differently, so that the overall effect on the shape of the entire mandible is complex (Renaud et al., 2012). Experimental studies suggest that bone remodelling under mechanical loading is important for integration and modularity in the mandible (Lightfoot and German, 1998; Tagliaro et al., 2009; Renaud et al., 2010; Vecchione et al., 2010). Histological studies further support the subdivision of the mouse mandible into two modules, as the difference between alveolar region and ascending ramus is also apparent in the processes of bone growth and remodelling in postnatal development (Martinez-Maza et al., 2012).

For wild populations of the house mouse, both plasticity and integration have been shown to be important factors for the evolution of mandible shape (Renaud and Auffray, 2009; Boell and Tautz, 2011; Siahsarvie et al., 2012). Allometry has been reported to be an integrating factor in mandibular shape variation in a Robertsonian chromosome polymorphism in the house mouse, but differential response of the alveolar region and ascending ramus was also found (Sans-Fuentes et el., 2009; Muñoz-Muñoz et al., 2011). Similar modularity was found in the mandible of yellow-necked field mice (Jojić et al., 2007, 2012). Renaud et al. (2007) reported allometry in the evolution of mandible outline shape in murids, indicating that size is an integrating factor, but also found a strong effect of dietary specialization. A detailed study of integration and modularity in a species of deer mouse found complex patterns for individual variation and fluctuating asymmetry (Zelditch et al., 2008).

Studies of integration in the mandible have also been conducted in a wide range of other rodents. In marmots and squirrels (Sciuridae), integration in the mandible has been investigated with a range of different morphometric approaches and both a modular structure of covariation and allometry have been reported (Velhagen and Roth, 1997; Cardini, 2003; Cardini and Tongiorgi, 2003; Cardini and Thorington, 2006; Zelditch et al., 2009; Swiderski and Zelditch, 2010). A series of papers examined morphological variation and integration in the mandible of spiny rats (Echimyidae) using a range of geometric morphometric methods and in relation to phylogeny, geography and ecological variables (Monteiro et al., 2003a, 2005; Monteiro and dos Reis, 2005; Perez et al., 2009). Similar analyses were also conducted for the whole group of caviomorph rodents (Álvarez et al., 2011b,a; Álvarez and Perez, 2013). Hautier et al. (2012) studied variation in the mandible and cranium of hystricognathous rodents and the integration between mandible and cranium.

The whole skull also has been used in many studies of morphological integration, particularly in laboratory mouse, where various experimental approaches have been used to investigate the mechanisms involved in cranial integration. Because the mouse is one of the classical "model organisms", its craniofacial development is known in great detail and many genetic and other experimental resources are available for it (Depew et al., 2002; Chai and Maxson, 2006). In particular, for laboratory mice, it is possible to investigate development directly by conducting morphometric studies including a range of ontogenetic stages (Willmore et al., 2006a; Zelditch et al., 2006). This is now possible even for embryos (Young et al., 2007; Boughner et al., 2008; Parsons et al., 2008, 2011), although careful attention to experimental procedures is required because of the potential for artefacts (Schmidt et al., 2010). These direct analyses of development complement genetic approaches, where mutations are used that disrupt specific developmental processes (Hallgrímsson et al., 2006; Willmore et al., 2006b; Hallgrímsson et al., 2009; Martínez-Abadías et al., 2011; Parsons et al., 2011) or where QTLs for different skull regions are sought for understanding the genetic origin of modularity (Leamy et al., 1999). In addition, experimental approaches can be used for investigating specific processes, for example to demonstrate the importance of bone remodelling for cranial morphology (Lightfoot and German, 1998; Vecchione et al., 2010). Note that integration does not only concern the skull itself, but that there is also extensive integration between the skull and surrounding soft tissue (Jamniczky and Hallgrímsson, 2011).

Overall, these studies support the idea that the complexity of cranial integration reflects the complexity of cranial development (Hallgrímsson et al., 2009). In addition, patterns and strengths of integration in the rodent skull change over ontogeny (Zelditch, 2005; Willmore et al., 2006a; Zelditch et al., 2006; Gonzalez et al., 2011a).

Integration in the skull has been studied extensively with a range of methods. Several studies quantified the degree of integration in the rodent skull with a measure derived from the variance of eigenvalues of the covariance matrix of shape variables (Willmore et al., 2006a; Hallgrímsson et al., 2009; Jamniczky and Hallgrímsson, 2009; Gonzalez et al., 2011b; Jojić et al., 2011). Some studies used partial least squares analysis to investigate patterns of integration between different skull regions such as the dorsal and ventral sides (Rohlf and Corti, 2000; Corti et al., 2001; Monteiro et al., 2003a; Macholán et al., 2008; Burgio et al., 2009). Tests of modularity have yielded mixed results, with some supporting and others inconsistent with a-priori hypotheses (Hallgrímsson et al., 2009; Jamniczky and Hallgrímsson, 2011; Jojić et al., 2011). The patterns of integration for individual variation and fluctuating asymmetry appear to be related in some cases and totally dissimilar in others (Debat et al., 2000; Hallgrímsson et al., 2004; Breno et al., 2011; Jojić et al., 2011), indicating that the role of developmental interactions in determining patterns of integration in the skull is unclear.

Allometry appears to be an important contributing factor to integration in rodent skulls and can produce strong integration for large-scale

phylogenetic comparison, but not necessarily at smaller scales (Roth, 1996; Monteiro et al., 1999; Cardini and O'Higgins, 2004; Cardini and Thorington, 2006; Hautier et al., 2012; Martínez-Abadías et al., 2012b). Note, however, that cranial allometries themselves can evolve and therefore differ among taxa (Cardini and Thorington, 2006; Wilson and Sánchez-Villagra, 2010).

Integration has not only been studied in the mandible and cranium of rodents, but also in the teeth, both for the entire molar row and for individual molar teeth. Laffont et al. (2009) found significant covariation among mandibular molars of voles for individual variation but not for fluctuating asymmetry, and thus suggested that factors other than direct developmental interactions were responsible for integration. A search for QTLs affecting the shape of the mandibular molar row in laboratory mice found 18 putative QTLs, which had effects involving complex shape changes involving all three molars in combination, rather than changes limited mostly to a single tooth (relative to the remaining landmarks), and thus suggested that the genetic architecture of shape is integrated throughout the molar row (Workman et al., 2002). Patterns of within-population integration of molar tooth shape have been shown to coincide with directions of evolutionary diversification, suggesting that these patterns of integration can function as lines of least resistance (Renaud et al., 2005, 2011; Renaud and Auffray, this issue).

Although a multitude of studies have addressed questions on craniofacial integration and modularity in rodents, many gaps remain in our knowledge. Because so much is known about mouse development and genetics, furthering our understanding of mechanisms and consequences of integration and modularity in rodents must remain a priority for further research.

Carnivorans

Studies of cranial integration and modularity and of the evolution of skull shape in carnivorans are quite abundant. Patterns of cranial integration in carnivorans vary, partly in accordance with phylogenetic relatedness and also with diet (Goswami, 2006b). A considerable degree of evolutionary integration was found in studies that combined samples of carnivorans and carnivorous marsupials, and also appeared to be associated with differences in diet (Wroe and Milne, 2007; Goswami et al., 2011). The patterns of integration uncovered by these studies feature general contrasts between relatively short, broad and elongate, slender skulls that have also been reported in studies of carnivorans alone (Drake and Klingenberg, 2010). That diet can be a crucial factor for evolution of skull shape has been shown by studies examining transitions to herbivory, which are accompanied by marked shifts in morphology and functional aspects such as bite force (Christiansen and Wroe, 2007; Figueirido et al., 2010, 2011, 2012). In the extreme, such as for the evolution of sabre-toothed cats, such morphological and functional changes can bring about major shifts in the patterns of integration by comparison to other, related taxa (Christiansen, 2008, 2012).

For evolutionary change across the Carnivora, a pattern of modularity between the face and braincase has been observed, which also holds within species in the gray wolf and in domestic dogs (Drake and Klingenberg, 2010). This modular pattern of integration is also found for fluctuating asymmetry, indicating that it has arises from direct developmental interactions within the modules of the face and braincase, but that there are few such interactions between face and braincase (Drake and Klingenberg, 2010). The same modular structure of covariation in the skull also appears to have facilitated the explosive evolution of skull shapes in domestic dogs under domestication and selection by breeders (Drake and Klingenberg, 2008, 2010; Drake, 2011). Integrated evolution also has been shown for the ramus and corpus of the mandible (Meloro et al., 2011).

Shrews

Patterns of integration for variation among individuals, fluctuating asymmetry and variation among species in the mandible of shrews were found to correspond clearly, and intraspecific patterns of integration were similar in different species (Badyaev and Foresman, 2000, 2004; Young and Badyaev, 2006). These results indicate that patterns of integration are evolutionarily stable and that developmental interactions are important in determining patterns of integration within and among species. Functional aspects, assessed by the locations of muscle insertions on the mandible, appear to be important determinants of the patterns of integration (Badyaev and Foresman, 2000, 2004; Young and Badyaev, 2006) and bone remodelling in the mandible under muscle loading is related to morphological variation (Young and Badyaev, 2010).

A study combining morphometric analyses of shrew molar teeth in several populations and a numerical model of development of the tooth showed clear similarities of phenotypic variation in tooth shape among populations and correspondence to the patterns of variation produced by the developmental model (Polly, 2005).

Humans and other primates

The literature on morphological integration in the skull of humans and other primates is massive and dates back several decades. Accordingly, it is also very heterogeneous in terms of the morphometric methods and biological concepts that were used in different studies. Many studies use traditional morphometric methods for investigating morphological integration in primates, as approaches for investigating integration in the context of geometric morphometrics have only been developed relatively late. Primates were used as the study system in several key papers that revived the interest in morphological integration and related subjects such as allometry (e.g. Gould 1975; Cheverud 1982a,b; Cheverud et al. 1989; Cheverud 1995; Ackermann and Cheverud 2000; Marroig and Cheverud 2001; Ackermann 2005; Marroig and Cheverud 2005; Shirai and Marroig 2010). After geometric morphometric techniques were established, applications concerning cranial integration and allometry in primates, including humans, have become increasingly widespread (e.g. O'Higgins and Jones 1998; Collard and O'Higgins 2001; Lieberman et al. 2002; Penin et al. 2002; Singleton 2002; Bookstein et al. 2003; Frost et al. 2003; Bastir et al. 2004; Hallgrímsson et al. 2004; Mitteroecker et al. 2004; Bastir and Rosas 2005; Willmore et al. 2005; Mitteroecker and Bookstein 2008; Martínez-Abadías et al. 2009; Makedonska et al. 2012; Martínez-Abadías et al. 2012a; Singh et al. 2012). In accordance with the focus of this special issue of Hystrix, I will concentrate primarily on papers about cranial integration and modularity that use geometric morphometric methods.

Many studies of morphological integration in the primate skull are concerned with the covariation of specific anatomical or developmental components, which has been a been the subject of extensive discussion in physical anthropology and primatology (e.g.Lieberman 2011). A wide variety of different divisions of the skull into parts and study designs have been used, but many studies use partial least squares to extract the patterns of covariation between parts from morphometric data. Examples include parts such as the cranium and mandible (Bastir et al., 2005), face and neurocranium (Mitteroecker and Bookstein, 2008), the face, cranial base and cranial vault (Bookstein et al., 2003; Bastir and Rosas, 2006; Gkantidis and Halazonetis, 2011; McCane and Kean, 2011; Makedonska et al., 2012; Singh et al., 2012), cranial base, cranial fossa and mandibular ramus (Bastir and Rosas, 2005), the parietal, occipital and temporal bones in the posterior braincase (Gunz and Harvati, 2007), oral and zygomatic components of the face (Makedonska et al., 2012), or between parts of the mandible (Harvati et al., 2011). Baab et al. (2010) used partial least squares to study the association of cranial shape robustness, a general feature of cranial morphology that might itself be a form of integration. In general, these analyses tend to show that integration between different cranial parts is fairly strong, tests against the null hypothesis of independence usually yield statistically significant results if sufficiently large sample sizes are available, and the patterns of covariation revealed by the shape changes associated with partial least squares axes often suggest biological explanations for the origin or adaptive value of integration.

Partial least squares analysis has also been used for studying integration between skull and soft tissues, for example between the brain and skull (Bastir et al., 2010). An association between the soft tissues of the

face and the underlying skull was also found in this way (McCane and Kean, 2011), as well as with a different method (Halazonetis, 2007).

Tests of hypotheses on modularity, using comparisons of the strength of association between subsets of landmarks (Klingenberg, 2009), have been conducted for two examples in humans. In a study of the shape of the human brain in the midsagittal plane cortical and subcortical regions behaved as separate modules (Bruner et al., 2010). By contrast, the face, cranial base and cranial vault in human skulls did not behave as separate modules, neither for genetic nor phenotypic covariation, but variation appeared to be integrated throughout the entire skull (Martínez-Abadías et al., 2012a). Hypotheses of modularity can also be formulated for postcranial traits—one study used similar methods for studying modularity in the pelvis across a spectrum of primates (Lewton, 2012).

Because allometry has long been a central theme in primate ontogeny and evolution (e.g. Gould 1975), many morphometric studies have examined allometry through growth and static allometry within populations (e.g. O'Higgins and Jones 1998; Collard and O'Higgins 2001; Ponce de León and Zollikofer 2001; Lieberman et al. 2002; Penin et al. 2002; Strand Viðarsdóttir et al. 2002; Bastir and Rosas 2004; Berge and Penin 2004; Cobb and O'Higgins 2004; Mitteroecker et al. 2004; Zollikofer and Ponce de León 2004; Mitteroecker et al. 2005; Bastir et al. 2006; Bulygina et al. 2006; Bastir et al. 2007; Marroig 2007; Sardi et al. 2007; Morimoto et al. 2008; Baab and McNulty 2009; Gonzalez et al. 2010, 2011c; Weisensee and Jantz 2011; Martínez-Abadías et al. 2012b; Sardi and Ramírez Rozzi 2012). Other studies have considered only adult specimens, but from several species, and therefore concern evolutionary allometry, with a possible contribution from static allometry within species (e.g. Singleton 2002; Frost et al. 2003; Rosas and Bastir 2004; Cardini and Elton 2008b; Bastir et al. 2010; Elton et al. 2010; Bastir et al. 2011; Gilbert 2011; Ito et al. 2011). It is also possible to use these allometric approaches to investigate growth of the human face (Hennessy and Moss, 2001; Velemínská et al., 2012) or allometry of the brain (Bruner et al., 2010) and, related to it, the endocranial cavity (Neubauer et al., 2009; Gunz et al., 2010; Neubauer et al., 2010; Gunz et al., 2012). In most of these studies, pronounced allometry has been found, and allometry may therefore be an important integrating factor in the primate head. Some studies have therefore applied size correction, by using residuals from the multivariate regression of shape on size, before further analyses of integration (e.g. Bastir et al. 2011; Martínez-Abadías et al. 2012a).

Several studies have compared ontogenetic trajectories among populations and species (e.g. Collard and O'Higgins 2001; Ponce de León and Zollikofer 2001; Penin et al. 2002; Strand Viðarsdóttir et al. 2002; Berge and Penin 2004; Cobb and O'Higgins 2004; Mitteroecker et al. 2004, 2005; Mitteroecker and Bookstein 2008; Gunz et al. 2010, 2012; Sardi and Ramírez Rozzi 2012). If data from multiple populations or species are available and ages of specimens are known, it is possible to compare ontogenetic and evolutionary allometry to assess whether evolution occurred by ontogenetic scaling (Gonzalez et al., 2011c). This kind of study is important because it directly relates to explanations of evolutionary change by ontogenetic scaling, possibly through processes such as heterochrony (Klingenberg, 1998; Lieberman, 2011). In turn, these evolutionary processes can be crucial determinants for macroevolutionary patterns, such as evolutionary allometry. Evolutionary phenomena, in turn, depend on the availability of genetic variation on which natural selection and drift can act. So far, only few analyses combine geometric morphometric approaches with quantitative genetic analyses, but two of these studies have been conducted in primates: macaques (Willmore et al., 2005) and humans (Martínez-Abadías et al., 2009, 2012a). Willmore et al. (2005) studied the developmental basis of genetic and phenotypic integration by comparing covariance patterns of genetic and environmental variation as well as fluctuating asymmetry. Correspondences between patterns of variation were not very strong, but statistically significant and therefore suggest that direct developmental interactions are involved in shaping genetic integration, but perhaps make only a relatively small contribution. Martínez-Abadías et al. (2009, 2012a) used hypothetical selection for localized shape changes in the human skull and consistently found that the predicted responses affected the entire skull, indicating that genetic variation for human skull shape is highly integrated. Consistent with this finding, a test rejected the hypothesis that the face, cranial base and cranial vault are distinct modules concerning genetic variation (Martínez-Abadías et al., 2012a). Given the far-reaching implications of genetic integration for the evolution of the human head, further analyses of the genetic basis of cranial shape variation in primates are urgently needed.

The strong genetic integration for cranial shape that is apparent in the results on humans (Martínez-Abadías et al., 2009, 2012a) has implications for the understanding of primate and particularly human evolution. Because simulations of selection for different localized shape features gave similar responses involving global shape changes throughout the skull (Martínez-Abadías et al., 2012a), it is clear that the shape change in response to selection is not a reliable guide for inferring the specific features under selection and, above all, that there is no direct correspondence between the shape change that was selected for and the resulting evolutionary response. This adds to the complexities for inferring past selection, for instance, in human evolution (Lieberman, 2008, 2011). The second implication is that it is doubtful whether the skull can be divided into anatomical or functional subunits that are independent of each other in their evolution and therefore can be used as distinct cladistic characters for inferring phylogeny (Skelton and McHenry, 1992; Strait et al., 1997; Cardini and Elton, 2008a; González-José et al., 2008). If the genetic variation of skull shape observed in this one human population is representative of past populations (for which it is the best evidence that is currently available), the evolution in all parts of the skull is highly interdependent. Different putative modules in the skull are therefore unlikely to provide independent information on phylogeny. Moreover, because cranial integration is associated with strong evolutionary constraints (Martínez-Abadías et al., 2012a), it also makes it plausible that independent evolutionary changes in different evolutionary lineages produce similar shape changes, which would help to account for the homoplasy that has made it difficult to infer phylogenies from craniodental characters (e.g. Skelton and McHenry 1992; Strait et al. 1997; Wood and Lonergan 2008).

Some studies have found that different parts of the human skull, such as the face, temporal bone or cranial base, reflect factors such as population history and adaptation to climates to different degrees (Harvati and Weaver, 2006; Perez and Monteiro, 2009; von Cramon-Taubadel, 2011b; von Cramon-Taubadel and Smith, 2012). This implies some degree of modularity at the evolutionary level. To reconcile this observation with the strong genetic integration in the skull, one can hypothesize that this evolutionary modularity results from differential effects of selection on different parts of the skull, each in its own developmental and functional contexts. Any such hypothesis must necessarily remain rather speculative because of the various difficulties inherent in inferring past selection regimes (or tests of selection versus the null hypothesis of random drift). For instance, tests of such hypotheses almost inevitably make unrealistic assumptions, such as constancy of genetic covariance matrices over time or proportionality of genetic and phenotypic covariance matrices (e.g. Perez and Monteiro 2009; von Cramon-Taubadel 2009; Smith 2011).

There have been numerous morphometric studies of teeth in humans and other primates (Robinson et al., 2001; Martinón-Torres et al., 2006; Kieser et al., 2007; White, 2009; Gómez-Robles et al., 2011b; Grieco et al., 2013). Integration in the dentition has clear functional significance because it relates directly to occlusion during biting and chewing. Therefore, it is not surprising that integration has been shown among the shapes of premolar teeth and throughout the premolars and molars (Gómez-Robles et al., 2011a; Gómez-Robles and Polly, 2012). It appears that evolutionary allometry in the shape of teeth can be quite strong and evolution by allometric scaling has also been demonstrated (Martinón-Torres et al., 2006; Singleton et al., 2011).

Studies in humans offer some special opportunities for investigating factors that contribute to craniofacial shape variation. Some studies have investigated plasticity in the skull by following the consequences

of transitions in the mode of subsistence in human populations, for instance from hunter-gatherer to agricultural modes, and have found shape changes in the skull and mandible (Paschetta et al., 2010; von Cramon-Taubadel, 2011a). Because these shape changes are likely to relate to differences in the consistency of food and requirements for mastication, they can be informative about functional effects of mastication and bone remodelling under mechanical loads, and are therefore relevant for this source of integration. A different approach is to take advantage of opportunities like the quasi-experimental modifications of the skull through intentional deformation (Cheverud et al., 1992; Kohn et al., 1993; Perez, 2007; Martínez-Abadías et al., 2009). And finally, the data concerning dysmorphologies from diseases with known aetiologies and localized origins are a very rich, so far greatly underexploited, resource for the study of integration in the skull (e.g. Tobin et al. 2008; Richtsmeier and DeLeon 2009; Heuzé et al. 2012).

Conclusions

This article has reviewed some of the vast and still rapidly growing literature on morphological integration and modularity. Two primary conclusions arise from this survey: first, there is now an established and diverse set of tools for investigating morphological integration and modularity within the framework of geometric morphometrics and, second, a large and increasing number of studies have applied these methods to mammalian systems so that a general picture starts to emerge, but many opportunities remain for filling in big gaps in current knowledge. Also, there are several major challenges for innovations in methodology and development of new experimental protocols to tie the study of integration and modularity to functional, genetic and phylogenetic aspects of craniofacial evolution.

Although numerous studies on integration and modularity have been published, many of them using mammals, it is surprisingly difficult to use this information for making comparisons or developing generalizations across larger taxa. Authors differ in how many "routine" statistics they provide in their papers: statistics such as eigenvalues or matrix correlations between covariance matrices for individual variation and fluctuating asymmetry, which are all relevant for studies of morphological integration, are reported in some papers but not in others. It is fairly simple for authors to include such statistics in their papers; the incentive for them is that the paper may be cited a few times more often as a result (which is also an argument that might convince editors and reviewers anxious to improve the impact factor of their journal). Also, there is a need for carrying out analyses of integration in additional species, even if a similar study already exists in a related species. This will facilitate comparative studies on the evolution of integration and modularity in the mammalian skull, which will benefit from the rapid growth of knowledge in this area and will in turn contribute to it.

Challenges that require new developments remain in several areas. In many of the cases I will mention here, it will also be useful to gain a better understanding and to provide more accessible explanations of how the existing methods work and what assumptions they make. For comparisons of integration across taxa, it will be important to develop methods for comparing many covariance matrices simultaneously without losing too much information. Improving the methodology for exploratory searches for modules in morphometric data also remains a challenge, and it clearly is a daunting one, both from biological and statistical points of view. There are many possibilities for incorporating phylogenetic comparative approaches and quantitative genetics into analyses of integration and modularity to encompass the spectrum from micro- to macroevolutionary perspectives. There are promising new possibilities in the emerging synthesis of morphometric and biomechanical approaches (e.g. O'Higgins et al. 2011), which may lead to new ways of understanding the functional aspects of integration and modularity. Finally, there are new opportunities to relate the study of integration and modularity to insights and experimental protocols from developmental biology (Young et al., 2010; Parsons et al., 2011; Kimmel et al., 2012).

I am optimistic that morphometric studies on integration and modularity in the mammalian skull will contribute substantially to a comprehensive and unified understanding of the developmental, functional and historical aspects of the evolution of complex morphological structures (Breuker et al., 2006a; Klingenberg, 2010). This is a major contribution of geometric morphometrics to evolutionary biology, which will in turn consolidate geometric morphometrics as an important discipline in 21st-century biology.

References

Ackermann R.R., 2005. Ontogenetic integration of the hominoid face. J. Hum. Evol. 48: 175–197.

Ackermann R.R., Cheverud J.M., 2000. Phenotypic covariance structure in tamarins (genus *Saguinus*): a comparison of variation patterns using matrix correlation and common principal component analysis. Am. J. Phys. Anthropol. 111: 489–501.

Adams D.C., 2011. Quantitative genetics and evolution of head shape in *Plethodon* salamanders. Evol. Biol. 38: 278–286.

Álvarez A., Perez S.I., 2013. Two- versus three-dimensional morphometric approaches in macroevolution: insight from the mandible of caviomorph rodents. Evol. Biol. 40(1): 150–157. doi:10.1007/s11692-012-9194-3

Álvarez A., Perez S.I., Verzi D.H., 2011a. Early evolutionary differentiation of morphological variation in the mandible of South American caviomorph rodents (Rodentia, Caviomorpha). J. Evol. Biol. 24: 2687–2695.

Álvarez A., Perez S.I., Verzi D.H., 2011b. Ecological and phylogenetic influence on mandible shape variation of South American caviomorph rodents (Rodentia: Hystricomorpha). Biol. J. Linn. Soc. 102: 828–837.

Arthur W., 2001. Developmental drive: an important determinant of the direction of phenotypic evolution. Evol. Dev. 3: 271–278.

Astúa D. 2009. Evolution of scapular size and shape in didelphid marsupials (Didelphimorphia: Didelphidae). Evolution 63: 2438–2456.

Atchley W.R., 1993. Genetic and developmental aspects of variability in the mammalian mandible. In: Hanken J., Hall B.K. (Eds.) The skull. University of Chicago Press, Chicago. 207–247.

Atchley W.R., Hall B.K., 1991. A model for development and evolution of complex morphological structures. Biol. Rev. 66: 101–157.

Atchley W.R., Plummer A.A., Riska B., 1985. Genetics of mandible form in the mouse. Genetics 111: 555–577.

Baab K.L., Freidline S.E., Wang S.L., Hanson T., 2010. Relationship of cranial robusticity to cranial form, geography and climate in *Homo sapiens*. Am. J. Phys. Anthropol. 141: 97–115.

Baab K.L., McNulty K.P., 2009. Size, shape, and asymmetry in fossil hominins: the status of the LB1 cranium based on 3D morphometric analyses. J. Hum. Evol. 57: 608–622.

Badyaev A.V., Foresman K.R., 2000. Extreme environmental change and evolution: stress-induced morphological variation is strongly concordant with patterns of evolutionary divergence in shrew mandibles. Proc. R. Soc. B Biol. Sci. 267: 371–377.

Badyaev A.V., Foresman K.R., 2004. Evolution of morphological integration. I. Functional units channel stress-induced variation in shrew mandibles. Am. Nat. 163: 868–879.

Bastir M., O'Higgins P., Rosas A., 2007. Facial ontogeny in Neanderthals and modern humans. Proc. R. Soc. Lond. B Biol. Sci. 274: 1125–1132.

Bastir M., Rosas A., 2004. Facial heights: evolutionary relevancve of postnatal ontogeny for facial orientation and skull morphology in humans and chimpanzees. J. Hum. Evol. 47: 359–381.

Bastir M., Rosas A., 2005. Hierarchical nature of morphological integration and modularity in the human posterior face. Am. J. Phys. Anthropol. 128: 26–34.

Bastir M., Rosas A., 2006. Correlated variation between the lateral basicranium and the face: A geometric morphometric study in different human groups. Arch. Oral Biol. 51: 814–824.

Bastir M., Rosas A., Gunz P., Peña-Melian A., Manzi G., Harvati K., Kruszynski R., Stringer C.B., Hublin J.-J., 2011. Evolution of the base of the brain in highly encephalized human species. Nat. Commun. 2: 588.

Bastir M., Rosas A., Kuroe K., 2004. Petrosal orientation and mandibular ramus breadth: evidence for an integrated petroso-mandibular developmental unit. Am. J. Phys. Anthropol. 123: 340–350.

Bastir M., Rosas A., O'Higgins P., 2006. Craniofacial levels and the morphological maturation of the human skull. J. Anat. 209: 637–654.

Bastir M., Rosas A., Sheets H.D., 2005. The morphological integration of the hominoid skull: A partial least squares and PC analysis with implications for European middle pleistocene mandibular variation. In: Slice D.E. (Ed.) Modern Morphometrics in Physical Anthropology. Kluwer Academic, New York. 265–284.

Bastir M., Rosas A., Stringer C.B., Cuétara J.M., Kruszynski R., Weber G.W., Ross C.F., Ravosa M.J., 2010. Effects of brain and facial size on basicranial form in human and primate evolution. J. Hum. Evol. 58: 424–431.

Berge C., Penin X., 2004. Ontogenetic allometry, heterochrony, and interspecific differences in the skull of African apes, using tridimensional Procrustes analysis. Am. J. Phys. Anthropol. 124: 124–138.

Blomberg S.P., Lefevre J.G., Wells J.A., Waterhouse M., 2012. Independent contrasts and PGLS regression estimators are equivalent. Syst. Biol. 61: 382–391.

Boell L., Gregorová S., Forejt J., Tautz D., 2011. A comparative assessment of mandible shape in a consomic strain panel of the house mouse (*Mus musculus*) – implications for epistasis and evolvability of quantitative traits. BMC Evol. Biol. 11: 309.

Boell L., Tautz D., 2011. Micro-evolutionary divergence patterns of mandible shapes in wild house mouse (*Mus musculus*) populations. BMC Evol. Biol. 11: 306.

Bookstein F.L., Gunz P., Mitteroecker P., Prossinger H., Schaefer K., Seidler H., 2003. Cranial integration in *Homo*: singular warps analysis of the midsagittal plane in ontogeny and evolution. J. Hum. Evol. 44: 167–187.

Bookstein F.L., Sampson P.D., Streissguth A.P., Barr H.M., 1990. Measuring "dose" and "response" with multivariate data using partial least squares techniques. Comm. Statist. Theor. Meth. 19: 765–804.

Boughner J.C., Wat S., Diewert V.M., Young N.M., Browder L.W., Hallgrímsson B., 2008. Short-faced mice and developmental interactions between the brain and the face. J. Anat. 213: 646–662.

Breno M., Leirs H., Van Dongen S., 2011. No relationship between canalization and developmental stability of the skull in a natural population of *Mastomys natalensis* (Rodentia: Muridae). Biol. J. Linn. Soc. 104: 207–216.

Breuker C.J., Debat V., Klingenberg C.P., 2006a. Functional evo-devo. Trends Ecol. Evol. 21: 488–492.

Breuker C.J., Patterson J.S., Klingenberg C.P., 2006b. A single basis for developmental buffering of *Drosophila* wing shape. PLoS ONE 1(1): e7. doi:10.1371/journal.pone.0000007

Bruner E., Martin-Loeches M., Colom R., 2010. Human midsagittal brain shape variation: patterns, allometry and integration. J. Anat. 216: 589–599.

Brusatte S.L., Sakamoto M., Montanari S., Harcourt Smith W.E.H., 2012. The evolution of cranial form and function in theropod dinosaurs: insights from geometric morphometrics. J. Evol. Biol. 25: 365–377.

Bulygina E., Mitteroecker P., Aiello L. 2006. Ontogeny of facial dimorphism and patterns of individual development within one human population. Am. J. Phys. Anthropol. 131: 432–443.

Burgio G., Baylac M., Heyer E., Montagutelli X., 2009. Genetic analysis of skull shape variation and morphological integration in the mouse using interspecific recombinant congenic strains between C57BL/6 and mice of the *Mus spretus* species. Evolution 63: 2668–2686.

Burgio G., Baylac M., Heyer E., Montagutelli X., 2012a. Exploration of the genetic organization of morphological modularity on the mouse mandible using a set of interspecific recombinant congenic strains between C57BL/6 and mice of the *Mus spretus* species. G3 (Bethesda) 2: 1257–1268.

Burgio G., Baylac M., Heyer E., Montagutelli X., 2012b. Nasal bone shape is under complex epistatic genetic control in mouse interspecific recombinatnt congenic strains. PLoS ONE 7(5): e37721. doi:10.1371/journal.pone.0037721

Callebaut W., Rasskin-Gutman D. (Eds.), 2005. Modularity: Understanding the Development and Evolution of Natural Complex Systems. MIT Press, Cambridge, MA.

Cardini A., 2003. The geometry of the marmot (Rodentia: Sciuridae) mandible: phylogeny and patterns of morphological evolution. Syst. Biol. 52: 186–205.

Cardini A., Elton S., 2007. Sample size and sampling error in geometric morphometric studies of size and shape. Zoomorphology (Berl.) 126: 121–134.

Cardini A., Elton S., 2008a. Does the skull carry a phylogenetic signal? Evolution and modularity in the guenons. Biol. J. Linn. Soc. 93: 813–834.

Cardini A., Elton S., 2008b. Variation in guenon skulls (I): species divergence, ecological and genetic differences. J. Hum. Evol. 54: 615–637.

Cardini A., O'Higgins P., 2004. Patterns of morphological evolution in *Marmota* (Rodentia, Sciuridae): geometric morphometrics of the cranium in the context of marmot phylogeny, ecology and conservation. Biol. J. Linn. Soc. 82: 385–407.

Cardini A., Thorington R.W., Jr., 2006. Postnatal ontogeny of marmot (Rodentia, Sciuridae) crania: allometric trajectories and species divergence. J. Mammal. 87: 201–215.

Cardini A., Tongiorgi P., 2003. Yellow-bellied marmots (*Marmota flaviventris*) "in the shape space" (Rodentia, Sciuridae): sexual dimorphism, growth and allometry of the mandible. Zoomorphology (Berl.) 122: 11–23.

Chai Y., Maxson R.E., Jr., 2006. Recent advances in craniofacial morphogenesis. Dev. Dyn. 235: 2353–2375.

Cheverud J.M., 1982a. Phenotypic, genetic, and environmental morphological integration in the cranium. Evolution 36: 499–516.

Cheverud J.M., 1982b. Relationships among ontogenetic, static, and evolutionary allometry. Am. J. Phys. Anthropol. 59: 139–149.

Cheverud J.M., 1984. Quantitative genetics and developmental constraints on evolution by selection. J. Theor. Biol. 110: 155–171.

Cheverud J.M., 1995. Morphological integration in the saddle-back tamarin (*Saguinus fuscicollis*) cranium. Am. Nat. 145: 63–89.

Cheverud J.M., 1996a. Developmental integration and the evolution of pleiotropy. Amer. Zool. 36: 44–50.

Cheverud J.M., 1996b. Quantitative genetic analysis of cranial morphology in the cottontop (*Saguinus oedipus*) and saddle-back (*S. fuscicollis*) tamarins. J. Evol. Biol. 9: 5–42.

Cheverud J.M., Kohn L.A.P., Konigsberg L.W., Leigh S.R., 1992. Effects of fronto-occipital artificial cranial vault modification on the cranial base and face. Am. J. Phys. Anthropol. 88: 323–345.

Cheverud J.M., Routman E.J., Irschick D.J., 1997. Pleiotropic effects of individual gene loci on mandibular morphology. Evolution 51: 2006–2016.

Cheverud J.M., Wagner G.P., Dow M.M., 1989. Methods for the comparative analysis of variation patterns. Syst. Zool. 38: 201–213.

Christiansen P., 2008. Evolution of skull and mandible shape in cats (Carnivora: Felidae). PLoS ONE 3(7): e2807. doi:10.1371/journal.pone.0002807

Christiansen P. 2012. The making of a monster: postnatal ontogenetic changes in the craniomandibular shape in the great sabercat *Smilodon*. PLoS ONE 7(1): e29699. doi:10.1371/journal.pone.0029699

Christiansen P., Wroe S., 2007. Bite forces and evolutionary adaptations to feeding ecology in carnivores. Ecology 88: 347–358.

Cobb S.N., O'Higgins P., 2004. Hominins do not share a common postnatal facial ontogenetic shape trajectory. J. Exp. Zool. B Mol. Dev. Evol. 302: 302–321.

Cock A.G., 1966. Genetical aspects of metrical growth and form in animals. Q. Rev. Biol. 41: 131–190.

Collard M., O'Higgins P., 2001. Ontogeny and homoplasy in the papionin monkey face. Evol. Dev. 3: 322–331.

Cooper W.J., Parsons K.J., McIntyre A., Kern B., McGee-Moore A., Albertson R.C., 2010. Bentho-pelagic divergence of cichlid feeding architecture was prodigious and consistent during multiple adaptive radiations within African rift-lakes. PLoS ONE 5(3): e9551. doi:10.1371/journal.pone.0009551

Corti M., Aguilera M., Capanna E., 2001. Size and shape changes in the skull accompanying speciation of South American spiny rats (Rodentia: *Proechimys* spp.). Journal of Zoology 253: 537–547.

De Esteban-Trivigno S., 2011a. Buscando patrones ecomorfológicos comunes entre ungulados actuales y xenartros extintos. Ameghiniana 48(2): 189–209. [In Spanish]

De Esteban-Trivigno S., 2011b. Ecomorfología de xenartros extintos: análisis de la mandíbula con métodos de morfometría geométrica. Ameghiniana 48(3): 381–398. [In Spanish]

Debat V., Alibert P., David P., Paradis E., Auffray J.-C., 2000. Independence between developmental stability and canalization in the skull of the house mouse. Proc. R. Soc. Lond. B Biol. Sci. 267: 423–430.

Debat V., Bloyer S., Faradji F., Gidaszewski N.A., Navarro N., Orozco-terWengel P., Ribeiro V., Schlötterer C., Deutsch J.S., Peronnet F., 2011. Developmental stability: a major role for *Cyclin G* in *Drosophila melanogaster*. PLoS Genet. 7(10): e1002314. doi:10.1371/journal.pgen.1002314

Debat V., Cornette R., Korol A.B., Nevo E., Soulet D., David J.R., 2008. Multidimensional analysis of *Drosophila* wing variation in Evolution Canyon. J. Genet. 87: 407–419.

Debat V., Debelle A., Dworkin I., 2009. Plasticity, canalization, and developmental stability of the *Drosophila* wing: joint effects of mutations and developmental temperature. Evolution 63: 2864–2876.

Debat V., Milton C.C., Rutherford S., Klingenberg C.P., Hoffmann A.A., 2006. Hsp90 and the quantitative variation of wing shape in *Drosophila melanogaster*. Evolution 60: 2529–2538.

Depew M.J., Tucker A.S., Sharpe P.T., 2002. Craniofacial development. In: Rossant J., Tam P.P.L. (Eds.) Mouse development: patterning, morphogenesis, and organogenesis. Academic Press, San Diego. 421–498.

Dornburg A., Sidlauskas B., Santini F., Sorenson L., Near T.J., Alfaro M.E., 2011. The influence of an innovative strategy on the phenotypic diversification of triggerfish (family: Balistidae). Evolution 65: 1912–1926.

Drake A.G., 2011. Dispelling dog dogma: an investigation of heterochrony in dogs using 3D geometric morphometric analysis of skull shape. Evol. Dev. 13: 204–213.

Drake A.G., Klingenberg C.P., 2008. The pace of morphological change: historical transformation of skull shape in St. Bernard dogs. Proc. R. Soc. Lond. B Biol. Sci. 275: 71–76.

Drake A.G., Klingenberg C.P., 2010. Large-scale diversification of skull shape in domestic dogs: disparity and modularity. Am. Nat. 175: 289–301.

Dryden I.L., Koloydenko A., Zhou D., 2009. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. Ann. Appl. Stat. 3: 1102–1123.

Dryden I.L., Mardia K.V., 1998. Statistical shape analysis. Wiley, Chichester.

Efron B., Tibshirani R.J., 1993. An introduction to the bootstrap. Chapman & Hall, New York.

Ehrich T.H., Vaughn T.T., Koreishi S.F., Linsey R.B., Pletscher L.S., Cheverud J.M., 2003. Pleiotropic effects on mandibular morphology I. Developmental morphological integration and differential dominance. J. Exp. Zool. B Mol. Dev. Evol. 296: 58–79.

Elton S., Dunn J., Cardini A., 2010. Size variation facilitates population divergence but does not explain it all: an example study from a widespread African monkey. Biol. J. Linn. Soc. 101: 823–843.

Escoufier Y., 1973. Le traitement des variables vectorielles. Biometrics 29: 751–760.

Felsenstein J., 1985. Phylogenies and the comparative method. Am. Nat. 125: 1–15.

Felsenstein J., 2004. Inferring phylogenies. Sinauer Associates, Sunderland, MA.

Fernández Iriarte P., Céspedes W., Santos M., 2003. Quantitative-genetic analysis of wing form and bilateral asymmetry in isochromosomal lines of *Drosophila subobscura* using Procrustes methods. J. Genet. 82: 95–113.

Figueirido B., Palmqvist P., Pérez-Claros J.A., Dong W., 2011. Cranial shape transformation in the evolution of the giant panda *(Ailuropoda melanoleuca*). Naturwissenschaften 98: 107–116.

Figueirido B., Serrano-Alarcón F.J., Palmqvist P., 2012. Geometric morphometrics shows differences and similarities in skull shape between the red and giant pandas. J. Zool. (Lond.) 296: 293–302.

Figueirido B., Serrano-Alarcón F.J., Slater G.J., Palmqvist P., 2010. Shape at the crossroads: homoplasy and history in the evolution of the carnivoran skull towards herbivory. J. Evol. Biol. 23: 2579–2594.

Fortuny J., Marcé Nogué J., De Esteban-Trivigno S., Gil L., Galobart À., 2011. Temnospondyli bite club: ecomorphological patterns of the most diverse group of early tetrapods. J. Evol. Biol. 24: 2040–2054.

Frost S.R., Marcus L.F., Bookstein F.L., Reddy D.P., Delson E., 2003. Cranial allometry, phylogeography, and systematics of large-bodied papionins (Primates: Cercopithecinae) inferred from geometric morphometric analysis of landmark data. Anat. Rec. 275A: 1048–1072.

Futuyma D.J., 2010. Evolutionary constraint and ecological consequences. Evolution 64: 1865–1884.

Gerhart J., Kirschner M., 1997. Cells, embryos, and evolution: toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptability. Blackwell Science, Malden, MA.

Gidaszewski N.A., Baylac M., Klingenberg C.P., 2009. Evolution of sexual dimorphism of wing shape in the *Drosophila melanogaster* subgroup. BMC Evol. Biol. 9: 110.

Gilbert C.C., 2011. Phylogenetic analysis of the African papionin basicranium usinf 3-D geometric morphometrics: the need for improved methods to account for allometric effects. Am. J. Phys. Anthropol. 144: 60–71.

Gkantidis N., Halazonetis D.J., 2011. Morphological integration between the cranial base and the fasce in children and adults. J. Anat. 218: 426–438.

Gómez J.M., Abdelaziz M., Muñoz-Pajares J., Perfectti F., 2009. Heritability and genetic correlation of corolla shape and size in *Erysimum mediohispanicum*. Evolution 63: 1820–1831.

Gómez-Robles A., Martinón-Torres M., Bermúdez de Castro J.M., Prado-Simón L., Arsuaga J.L., 2011a. A geometric morphometric analysis of hominin upper premolars. Shape variation and morphological integration. J. Hum. Evol. 61: 688–702.

Gómez-Robles A., Olejniczak A.J., Martinón-Torres M., Prado-Simón L., Bermúdez de Castro J.M., 2011b. Evolutionary novelties and losses in geometric morphometrics: a practical approach through hominin molar morphology. Evolution 65: 1772–1790.

Gómez-Robles A., Polly P.D., 2012. Morphological integration in the hominin dentition: evolutionary, developmental, and functional factors. Evolution 66: 1024–1043.

Gonzalez P.N., Hallgrímsson B., Oyhenart E.E., 2011a. Developmental plasticity in covariance structure of the skull: effects of prenatal stress. J. Anat. 218: 243–257.

Gonzalez P.N., Oyhenart E.E., Hallgrímsson B., 2011b. Effects of environmental perturbations during postnatal development on the phenotypic integration of the skull. J. Exp. Zool. B Mol. Dev. Evol. 316: 547–561.

Gonzalez P.N., Perez S.I., Bernal V., 2010. Ontogeny of robusticity of craniofacial traits in modern humans: a study of South American populations. Am. J. Phys. Anthropol. 142: 367–379.

Gonzalez P.N., Perez S.I., Bernal V., 2011c. Ontogenetic allometry and cranial shape diversification among human populations from South America. Anat. Rec. 294: 1864–1874.

González-José R., Charlin J., 2012. Relative importance of modularity and other morphological attributes of different types of lithic point weapons: assessing functional variations. PLoS ONE 7(10): e48009. doi:10.1371/journal.pone.0048009

González-José R., Escapa I., Neves W.A., Cúneo R., Pucciarelli H.M., 2008. Cladistic analysis of continuous modularized traits provides phylogenetic signals in *Homo* evolution. Nature 453: 775–778.
Goswami A., 2006a. Cranial modularity shifts during mammalian evolution. Am. Nat. 168: 270–280.
Goswami A., 2006b. Morphological integration in the carnivoran skull. Evolution 60: 169–183.
Goswami A., 2007. Phylogeny, diet and cranial integration in australodelphian marsupials. PLoS ONE 2(10): e995. doi:10.1371/journal.pone.0000995
Goswami A., Milne N., Wroe S., 2011. Biting through constraints: cranial morphology, disparity and convergence across living and fossil carnivorous mammals. Proc. R. Soc. Lond. B Biol. Sci. 278: 1831–1839.
Goswami A., Polly P.D., 2010a. The influence of modularity on cranial morphological disparity in Carnivora and Primates (Mammalia). PLoS ONE 5(3): e9517. doi:10.1371/journal.pone.0009517
Goswami A., Polly P.D., 2010b. Methods for studying morphological integration and modularity. In: Alroy J., Hunt G. (Eds.) Quantitative methods in paleobiology. Paleontological Society, Ithaca, NY. 213–243.
Gould S.J. 1966. Allometry and size in ontogeny and phylogeny. Biol. Rev. 41: 587–640.
Gould S.J., 1975. Allometry in primates, with emphasis on scaling and the evolution of the brain. Contrib. Primatol. 5: 244–292.
Grieco T.M., Rizk O.T., Hlusko L.J., 2013. A modular framework characterizes micro- and macroevolution of Old World monkey dentitions. Evolution 67(1): 241–259. doi: 10.1111/j.1558-5646.2012.01757.x
Gunz P., Harvati K., 2007. The Neanderthal "chignon": variation, integration, and homology. J. Hum. Evol. 52: 262–274.
Gunz P., Neubauer S., Golovanova L., Doronichev V., Maureille B., Hublin J.-J., 2012. A uniquely modern human pattern of endocranial development. Insights from a new cranial reconstruction of the Neandertal newborn from Mezmaiskaya. J. Hum. Evol. 62: 300–313.
Gunz P., Neubauer S., Maureille B., Hublin J.-J., 2010. Brain development differs between Neanderthals and modern humans. Curr. Biol. 20: R921–R922.
Haber A., 2011. A comparative analysis of integration indices. Evol. Biol. 38: 476–488.
Halazonetis D.J., 2007. Morphometric correlation between facial soft-tissue profile shape and skeletal pattern in children and adolescents. Am. J. Orthod. Dentofacial Orthop. 132: 450–457.
Hallgrímsson B., Brown J.J.Y., Ford-Hutchinson A.F., Sheets H.D., Zelditch M.L., Jirik F.R., 2006. The brachymorph mouse and the developmental-genetic basis for canalization and morphological integration. Evol. Dev. 8: 61–73.
Hallgrímsson B., Jamniczky H.A., Young N.M., Rolian C., Parsons T.E., Boughner J.C., Marcucio R.S., 2009. Deciphering the palimpsest: studying the relationship between morphological integration and phenotypic covariation. Evol. Biol. 36: 355–376.
Hallgrímsson B., Willmore K., Dorval C., Cooper D.M.L., 2004. Craniofacial variability and modularity in macaques and mice. J. Exp. Zool. B Mol. Dev. Evol. 302: 207–225.
Hallgrímsson B., Willmore K., Hall B.K., 2002. Canalization, developmental stability, and morphological integration in primate limbs. Yearb. Phys. Anthropol. 45: 131–158.
Harvati K., Singh N., Nicholson López E., 2011. A three-dimensional look at the Neanderthal mandible. In: Condemi S., Weniger G.-C. (Eds.) Continuity and discontinuity in the peopling of Europe: one hundred fifty years of Neanderthal study. Springer, Berlin. 179–192.
Harvati K., Weaver T.D., 2006. Human cranial anatomy and the differntial preservation of population history and climate signatures. Anat. Rec. 288A: 1225–1233.
Hautier L., Lebrun R., Cox P.G., 2012. Patterns of covariation in the masticatory apparatus of hystricognathous rodents: implications for evolution and diversification. J. Morphol. 273: 1319–1337.
Hennessy R.J., Moss J.P., 2001. Facial growth: separating shape from size. Eur. J. Orthod. 23: 275–285.
Heuzé Y., Martínez-Abadías N., Stella J.M., Senders C.W., Boyadjiev S.A., Lo L.-J., Richtsmeier J.T., 2012. Unilateral and bilateral expression of a quantitative trait: asymmetry and symmetry in coronal craniosynostosis. J. Exp. Zool. B Mol. Dev. Evol. 318: 109–122.
Hooper J.W., 1959. Simultaneous equations and canonical correlation theory. Econometrica 27: 245–256.
Houle D., Mezey J.G., Galpern P., Carter A., 2003. Automated measurement of *Drosophila* wings. BMC Evol. Biol. 3: 25.
Ito T., Nishimura T., Takai M., 2011. Allometry and interspecific differences in the facial cranium of two closely related macaque species. Anat. Res. Int. 2011: 849751.
Ivanović A., Kalezić M.L., 2010. Testing the hypothesis of morphological integration on a skull of a vertebrate with a biphasic life cycle: a case study of the alpine newt. J. Exp. Zool. B Mol. Dev. Evol. 314: 527–538.
Jamniczky H.A., Hallgrímsson B., 2009. A comparison of covariance structure in wild and laboratory muroid crania. Evolution 63: 1540–1556.
Jamniczky H.A., Hallgrímsson B., 2011. Modularity in the skull and cranial vasculature of laboratory mice: implications for the evolution of complex phenotypes. Evol. Dev. 13: 28–37.
Jojić V., Blagojević J., Ivanović A., Bugarski-Stanojević V., Vujošević M., 2007. Morphological integration of the mandible in yellow-necked field mice: the effects of B chromosmes. J. Mammal. 88: 689–695.
Jojić V., Blagojević J., Vujošević M., 2011. B chromosomes and cranial variability in yellow-necked field mice (*Apodemus flavicollis*). J. Mammal. 92: 396–406.
Jojić V., Blagojević J., Vujošević M., 2012. Two-module organization of the mandible in the yellow-necked mouse: a comparison between two different morphometric approaches. J. Evol. Biol. 25: 2489–2500.
Jolliffe I.T., 2002. Principal component analysis. Springer-Verlag, New York.
Kieser J.A., Bernal V., Waddell J.N., Raju S., 2007. The uniqueness of the human anterior dentition: a geometric morphometric analysis. J. Forensic Sci. 52: 671–677.
Kimmel C.B., Hohenlohe P.A., Ullmann B., Currey M., Cresko W.A., 2012. Developmental dissociation in morphological evolution of the stickleback opercle. Evol. Dev. 14: 326–337.
Kirschner M., Gerhart J., 1998. Evolvability. Proc. Natl. Acad. Sci. USA 95: 8420–8427.
Klingenberg C.P., 1996. Multivariate allometry. In: Marcus L.F., Corti M., Loy A., Naylor G.J.P., Slice D.E. (Eds.) Advances in morphometrics. Plenum Press, New York. 23–49.
Klingenberg C.P., 1998. Heterochrony and allometry: the analysis of evolutionary change in ontogeny. Biol. Rev. 73(1): 79–123.
Klingenberg C.P., 2003. Developmental instability as a research tool: using patterns of fluctuating asymmetry to infer the developmental origins of morphological integration. In: Polak M. (Ed.) Developmental instability: causes and consequences. Oxford University Press, New York. 427–442.
Klingenberg C.P., 2005. Developmental constraints, modules and evolvability. In: Hallgrímsson B., Hall B.K. (Eds.) Variation: a central concept in biology. Elsevier, Burlington, MA. 219–247.
Klingenberg C.P., 2008. Morphological integration and developmental modularity. Annu. Rev. Ecol. Evol. Syst. 39: 115–132.
Klingenberg C.P., 2009. Morphometric integration and modularity in configurations of landmarks: tools for evaluating a-priori hypotheses. Evol. Dev. 11: 405–421.
Klingenberg C.P., 2010. Evolution and development of shape: integrating quantitative approaches. Nat. Rev. Genet. 11: 623–635.
Klingenberg C.P., 2011. MorphoJ: an integrated software package for geometric morphometrics. Mol. Ecol. Resour. 11: 353–357.
Klingenberg C.P., 2013. Visualizations in geometric morphometrics: how to read and how to make graphs showing shape changes. Hystrix 24(1) (Online First) doi:10.4404/hystrix-24.1-7691
Klingenberg C.P., Badyaev A.V., Sowry S.M., Beckwith N.J., 2001a. Inferring developmental modularity from morphological integration: analysis of individual variation and asymmetry in bumblebee wings. Am. Nat. 157: 11–23.
Klingenberg C.P., Barluenga M., Meyer A., 2002. Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. Evolution 56: 1909–1920.
Klingenberg C.P., Debat V., Roff D.A., 2010. Quantitative genetics of shape in cricket wings: developmental integration in a functional structure. Evolution 64: 2935–2951.
Klingenberg C.P., Duttke S., Whelan S., Kim M., 2012. Developmental plasticity, morphological variation and evolvability: a multilevel analysis of morphometric integration in the shape of compound leaves. J. Evol. Biol. 25: 115–129.
Klingenberg C.P., Ekau W., 1996. A combined morphometric and phylogenetic analysis of an ecomorphological trend: pelagization in Antarctic fishes (Perciformes: Nototheniidae). Biol. J. Linn. Soc. 59(2): 143–177.
Klingenberg C.P., Gidaszewski N.A., 2010. Testing and quantifying phylogenetic signals and homoplasy in morphometric data. Syst. Biol. 59: 245–261.
Klingenberg C.P., Leamy L.J., 2001. Quantitative genetics of geometric shape in the mouse mandible. Evolution 55: 2342–2352.
Klingenberg C.P., Leamy L.J., Cheverud J.M., 2004. Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible. Genetics 166: 1909–1921.
Klingenberg C.P., Leamy L.J., Routman E.J., Cheverud J.M., 2001b. Genetic architecture of mandible shape in mice: effects of quantitative trait loci analyzed by geometric morphometrics. Genetics 157: 785–802.
Klingenberg C.P., Marugán-Lobón J., 2013. Evolutionary covariation in geometric morphometric data: analyzing integration, modularity and allometry in a phylogenetic context. Syst. Biol.(Advance Online) doi:10.1093/sysbio/syt025
Klingenberg C.P., McIntyre G.S., 1998. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. Evolution 52: 1363–1375.
Klingenberg C.P., Mebus K., Auffray J.-C., 2003. Developmental integration in a complex morphological structure: how distinct are the modules in the mouse mandible? Evol. Dev. 5: 522–531.
Klingenberg C.P., Navarro N., 2012. Development and genetics of the mouse mandible: a model system for complex morphological structures. In: Macholán M., Baird S.J.E., Munclinger P., Piálek J. (Eds.) Evolution of the house mouse. Cambridge University Press, Cambridge. 135–149.
Klingenberg C.P., Zaklan S.D., 2000. Morphological integration between developmental compartments in the *Drosophila* wing. Evolution 54: 1273–1285.
Klingenberg C.P., Zimmermann M., 1992. Static, ontogenetic, and evolutionary allometry: a multivariate comparison in nine species of water striders. Am. Nat. 140(4): 601–620.
Kohn L.A.P., Leigh S.R., Jacobs S.C., Cheverud J.M., 1993. Effects of annular cranial vault modification on the cranial base and face. Am. J. Phys. Anthropol. 90: 147–168.
Kulemeyer C., Asbahr K., Gunz P., Frahnert S., Bairlein F., 2009. Functional morpology and integration of corvid skulls – a 3D geometric morphometric approach. Front. Zool. 6: 2.
Laffont R., Renvoisé E., Navarro N., Alibert P., Montuire S., 2009. Morphological modularity and assessment of developmental processes within the vole dental row (*Microtus arvalis*, Arvicolinae, Rodentia). Evol. Dev. 11: 302–311.
Lande R., 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33: 402–416.
Leamy L., 1993. Morphological integration of fluctuating asymmetry in the mouse mandible. Genetica 89: 139–153.
Leamy L.J., Routman E.J., Cheverud J.M., 1999. Quantitative trait loci for early- and late-developing skull characters in mice: a test of the genetic independence model of morphological integration. Am. Nat. 153: 201–214.
Lewton K.L., 2012. Evolvability of the primate pelvic girdle. Evol. Biol. 39: 126–139.
Lieberman D.E., 2008. Speculations about the selective basis for modern human craniofacial form. Evol. Anthropol. 17: 55–68.
Lieberman D.E., 2011. The evolution of the human head. Harvard University Press, Cambridge, MA.
Lieberman D.E., McBratney B.M., Krovitz G., 2002. The evolution and development of cranial form in *Homo sapiens*. Proc. Natl. Acad. Sci. USA 99: 1134–1139.
Lightfoot P.S., German R.Z., 1998. The effects of muscular dystrophy on craniofacial growth in mice: a study of heterochrony and ontogenetic allometry. J. Morphol. 235: 1–16.
Linde M., Palmer M., Gómez-Zurita J., 2004. Differential correlates of diet and phylogeny on the shape of the premaxilla and anterior tooth in sparid fishes (Perciformes: Sparidae). J. Evol. Biol. 17: 941–952.
Loy A., Mariani L., Bertelletti M., Tunesi L., 1998. Visualizing allometry: geometric morphometrics in the study of shape changes in the early stages of the two-banded sea bream, *Diplodus vulgaris* (Perciformes, Sparidae). J. Morphol. 237: 137–146.
Macholán M., 2006. A geometric morphometric analysis of the shape of the first upper molar in mice of the genus *Mus* (Muridae, Rodentia). J. Zool. (Lond.) 270: 672–681.
Macholán M., Mikula O., Vohralík V., 2008. Geographic phenetic variation of two eastern-Mediterranean non-commensal mouse species, *Mus macedonicus* and *M. cypriacus* (Ro-

dentia: Muridae) based on traditional and geometric approaches to morphometrics. Zool. Anz. 247: 67–80.
Makedonska J., Wright B.W., Strait D.S., 2012. The effect of dietary adaptation on cranial morphological integration in capuchins (order Primates, genus *Cebus*). PLoS ONE 7(10): e40398. doi:10.1371/journal.pone.0040398
Marcus L.F., Hingst-Zaher E., Zaher H., 2000. Application of landmark morphometrics to skulls representing the orders of living mammals. Hystrix 11(1): 27–47. doi:10.4404/hystrix-11.1-4135
Mardia K.V., Bookstein F.L., Moreton I.J., 2000. Statistical assessment of bilateral symmetry of shapes. Biometrika 87: 285–300.
Mardia K.V., Kent J.T., Bibby J.M., 1979. Multivariate analysis. Academic Press, London.
Márquez E.J., 2008. A statistical framework for testing modularity in multidimensional data. Evolution 62: 2688–2708.
Marroig G., 2007. When size makes a difference: allometry, life-history and morphological evolution of capuchins (*Cebus*) and squirrels (*Saimiri*) monkeys (Cebinae, Platyrrhini). BMC Evol. Biol. 7: 20.
Marroig G., Cheverud J.M., 2001. A comparison of phenotypic variation and covariation patterns and the role of phylogeny, ecology, and ontogeny during cranial evolution of New World monkeys. Evolution 55: 2676–2600.
Marroig G., Cheverud J.M., 2005. Size as a line of least evolutionary resistance: diet and adaptive morphological radiation in New World monkeys. Evolution 59: 1128–1142.
Marroig G., Shirai L.T., Porto A., de Oliveira F.B., De Conto V., 2009. The evolution of modularity in the mammalian skull II: evolutionary consequences. Evol. Biol. 36: 136–148.
Martínez-Abadías N., Esparza M., Sjøvold T., González-José R., Santos M., Hernández M., Klingenberg C.P., 2012a. Pervasive genetic integration directs the evolution of human skull shape. Evolution 66: 1010–1023.
Martínez-Abadías N., Heuzé Y., Wang Y., Jabs E.W., Aldridge K., Richtsmeier J.T., 2011. FGF/FGFR signaling coordinates skull development by modulating magnitude of morphological integration: evidence from Apert syndrome mouse models. PLoS ONE 6(10): e26425. doi:10.1371/journal.pone.0026425
Martínez-Abadías N., Mitteroecker P., Parsons T.E., Esparza M., Sjøvold T., Rolian C., Richtsmeier J.T., Hallgrímsson B., 2012b. The developmental basis of quantitative craniofacial variation in humans and mice. Evol. Biol. 39: 554–567.
Martínez-Abadías N., Paschetta C., de Azevedo S., Esparza M., González-José R., 2009. Developmental and genetic constraints on neurocranial globularity: insights from analyses of deformed skulls and quantitative genetics. Evol. Biol. 36: 37–56.
Martinez-Maza C., Montes L., Lamrous H., Ventura J., Cubo J., 2012. Postnatal histomorphogenesis of the mandible in the house mouse. J. Anat. 220: 472–483.
Martinón-Torres M., Bastir M., Bermúdez de Castro J.M., Gómez A., Sarmiento S., Muela A., Arsuaga J.L., 2006. Hominin lower second premolar morphology: evolutionary inferences through geometric morphometric analysis. J. Hum. Evol. 50: 523–533.
Marugán-Lobón J., Buscalioni Á.D., 2006. Avian skull morphological evolution: exploring exo- and endocranial covariation with two-block partial least squares. Zoology (Jena) 109: 217–230.
Mayr E., 1982. The growth of biological thought: diversity, evolution, and inheritance. Harvard University Press, Cambridge, MA.
McCane B., Kean M.R., 2011. Integration of parts in the facial skeleton and cervical vertebrae. Am. J. Orthod. Dentofacial Orthop. 139: e13–e30.
Meloro C., Jones A.G., 2012. Tooth and cranial disparity in the fossil relatives of *Sphenodon* (Rhynchocephalia) dispute the persistent "living fossil" label. J. Evol. Biol. 25: 2194–2209.
Meloro C., Raia P., Carotenuto F., Cobb S.N., 2011. Phylogenetic signal, function and integration in the subunits of the carnivoran mandible. Evol. Biol. 38: 465–475.
Mezey J.G., Cheverud J.M., Wagner G.P., 2000. Is the genotype-phenotype map modular? A statistical approach using mouse quantitative trait loci data. Genetics 156: 305–311.
Mezey J.G., Houle D., 2005. The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. Evolution 59: 1027–1038.
Mitteroecker P., Bookstein F.L., 2007. The conceptual and statistical relationship between modularity and morphological integration. Syst. Biol. 56: 818–836.
Mitteroecker P., Bookstein F.L., 2008. The evolutionary role of modularity and integration in the hominoid cranium. Evolution 62: 943–958.
Mitteroecker P., Bookstein F.L., 2009. The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans. Evolution 63: 727–737.
Mitteroecker P., Gunz P., Bernhard M., Schaefer K., Bookstein F.L., 2004. Comparison of cranial ontogenetic trajectories among great apes and humans. J. Hum. Evol. 46: 679–698.
Mitteroecker P., Gunz P., Bookstein F.L. 2005. Heterochrony and geometric morphometrics: a comparison of cranial growth in *Pan paniscus* versus *Pan troglodytes*. Evol. Dev. 7: 244–258.
Monteiro L.R., 1999. Multivariate regression models and geometric morphometrics: the search for causal factors in the analysis of shape. Syst. Biol. 48: 192–199.
Monteiro L.R., Bonato V., dos Reis S.F., 2005. Evolutionary integration and morphological diversification in complex morphological structures: Mandible shape divergence in spiny rats (Rodentia, Echimyidae). Evol. Dev. 7: 429–439.
Monteiro L.R., dos Reis S.F., 2005. Morphological evolution in the mandible of spiny rats, genus *Trinomys* (Rodentia: Echimyidae). J. Zool. Syst. Evol. Res. 43: 332–338.
Monteiro L.R., Duarte L.C., dos Reis S.F., 2003a. Environmental correlates of geographical variation in skull and mandible shape of the punaré rat *Thrichomys apereoides* (Rodentia: Echimyidae). J. Zool. (Lond.) 261: 47–57.
Monteiro L.R., Lessa L.G., Abe A.S., 1999. Ontogenetic variation in skull shape of *Thrichomys apereoides* (Rodentia: Echimyidae). J. Mammal. 80: 102–111.
Monteiro L.R., Nogueira M.R., 2010. Adaptive radiations, ecological specialization, and the evolutionary integration of complex morphological structures. Evolution 64: 724–744.
Monteiro L.R., Nogueira M.R., 2011. Evolutionary patterns and processes in the radiation of phyllostomid bats. BMC Evol. Biol. 11: 137.
Moore W.J., 1981. The mammalian skull. Cambridge University Press, Cambridge.
Morimoto N., Ogihara N., Katayama K., Shiota K., 2008. Three-dimensional ontogenetic shape changes in the human cranium during the fetal period. J. Anat. 212: 627–635.
Muñoz-Muñoz F., Sans-Fuentes M.A., López-Fuster M.J., Ventura J., 2011. Evolutionary modularity of the mouse mandible: dissecting the effect of chromosomal reorganizations and isolation by distance in a Robertsonian system of *Mus musculus domesticus*. J. Evol. Biol. 24: 1763–1776.
Myers E.M., Janzen F.J., Adams D.C., Tucker J.K., 2006. Quantitative genetics of plastron shape in slider turtles (*Trachemys scripta*). Evolution 60: 563–572.
Neubauer S., Gunz P., Hublin J.-J., 2009. The pattern of endocranial ontogenetic shape changes in humans. J. Anat. 215: 240–255.
Neubauer S., Gunz P., Hublin J.-J., 2010. Endocranial shape changes during growth in chimpanzees and humans: a morphometric analysis of unique and shared aspects. J. Hum. Evol. 59: 555–566.
Nicola P.A., Monteiro L.R., Pessôa L.M., Von Zuben F.J., Rohlf F.J., Dos Reis S.F., 2003. Congruence of hierachical, localized variation in cranial shape and molecular phylogenetic structure in spiny rats, genus *Trinomys* (Rodentia: Echimyidae). Biol. J. Linn. Soc. 80: 385–396.
O'Higgins P., Cobb S.N., Fitton L.C., Gröning F., Phillips R., Liu J., Fagan M.J., 2011. Combining geometric morphometrics and functional simulation: an emerging toolkit for virtual functional analyses. J. Anat. 218: 3–15.
O'Higgins P., Jones N., 1998. Facial growth in *Cercocebus torquatus*: an application of three-dimensional geometric morphometric techniques to the study of morphological variation. J. Anat. 193: 251–272.
Olson E.C., Miller R.L., 1958. Morphological integration. University of Chicago Press, Chicago.
Parsons T.E., Kristensen E., Hornung L., Diewert V.M., Boyd S.K., German R.Z., Hallgrímsson B., 2008. Phenotypic variability and craniofacial dysmorphology: increased shape variance in a mouse model for cleft lip. J. Anat. 212: 135–143.
Parsons T.E., Schmidt E.J., Boughner J.C., Jamniczky H.A., Marcucio R.S., Hallgrímsson B., 2011. Epigenetic integration of the developing brain and face. Dev. Dyn. 240: 2233–2244.
Paschetta C., De Azevedo S., Castillo L., Martínez-Abadías N., Hernández M., Lieberman D.E., González-José R., 2010. The influence of masticatory loading on craniofacial morphology: a test case across technological transitions in the Ohio valley. Am. J. Phys. Anthropol. 141: 297–314.
Pavlicev M., Cheverud J.M., Wagner G.P., 2009. Measuring morphological integration using eigenvalue variance. Evol. Biol. 36: 157–170.
Penin X., Berge C., Baylac M., 2002. Ontogenetic study of the skull in modern humans and the common chimpanzees: neotenic hypothesis reconsidered with a tridimensional Procrustes analysis. Am. J. Phys. Anthropol. 118: 50–62.
Perez S.I., 2007. Artificial cranial deformation in South America: a geometric morphometrics approximation. J. Archaeol. Sci. 34: 1649–1658.
Perez S.I., Diniz-Filho J.A.F., Rohlf F.J., dos Reis S.F., 2009. Ecological and evolutionary factors in the morphological diversification of South American spiny rats. Biol. J. Linn. Soc. 98: 646–660.
Perez S.I., Klaczko J., Rocatti G., dos Reis S.F., 2011. Patterns of cranial shape diversification during the phylogenetic branching process of New World monkeys (Primates: Platyrrhini). J. Evol. Biol. 24: 1826–1835.
Perez S.I., Monteiro L.R., 2009. Nonrandom factors in modern human morphological diversification: a study of craniofacial variation in southern South American populations. Evolution 63: 978–993.
Polly P.D., 2005. Development and phenotypic correlations: the evolution of tooth shape in *Sorex araneus*. Evol. Dev. 7: 29–41.
Polly P.D., 2007. Phylogeographic differentiation in *Sorex araneus*: morphology in relation to geography and karyotype. Russ. J. Theriol. 6(1): 73–84.
Ponce de León M.S., Zollikofer C.P.E., 2001. Neanderthal cranial ontogeny and its implications for late homonid diversity. Nature 412: 534–538.
Ponssa M.L., Candioti M.F.V., 2012. Patterns of skull development in anurans: size and shape relationship during postmetamorphic cranial ontogeny in five species of the *Leptodactylus fuscus* group (Anura: Leptodactylidae). Zoomorphology (Berl.) 131: 349–362.
Porto A., De Oliveira F.B., Shirai L.T., De Conto V., Marroig G., 2009. The evolution of modularity in the mammalian skull I: Morphological integration patterns and magnitudes. Evol. Biol. 36: 118–135.
Raff R.A., 1996. The shape of life: genes, development and the evolution of animal form. University of Chicago Press, Chicago.
Renaud S., Alibert P., Auffray J.-C., 2012. Modularity as a source of new morphological variation in the mandible of hybrid mice. BMC Evol. Biol. 12: 141.
Renaud S., Auffray J.-C., 2009. Adaptation and plasticity in insular evolution of the house mouse mandible. J. Zool. Syst. Evol. Res. 48: 138–150.
Renaud S., Auffray J.-C., 2013. The direction of main phenotypic variance as a channel to evolution: cases in murine rodents. Hystrix 24(1) (Online Fisrt) doi:10.4404/hystrix-24.1-6296
Renaud S., Auffray J.-C., de la Porte S., 2010. Epigenetic effects on the mouse mandible: common features and discrepancies in remodeling due to muscular dystrophy and response to food consistency. BMC Evol. Biol. 10: 28.
Renaud S., Auffray J.-C., Michaux J., 2006. Conserved phenotypic variation patterns, evolution along lines of least resistance, and departure due to selection in fossil rodents. Evolution 60: 1701–1717.
Renaud S., Chevret P., Michaud J., 2007. Morphological vs. molecular evolution: ecology and phylogeny both shape the mandible of rodents. Zool. Scr. 36: 525–535.
Renaud S., Pantalacci S., Auffray J.-C., 2011. Differential evolvability along lines of least resistance of upper and lower molars in island house mice. PLoS ONE 6(5): e18951. doi:10.1371/journal.pone.0018951
Richtsmeier J.T., DeLeon V.B., 2009. Morphological integration of the skull in craniofacial anomalies. Orthod. Craniofac. Res. 12: 149–158.
Robinson D.L., Blackwell P.G., Stillman E.C., Brook A.H., 2001. Planar Procrustes analysis of tooth shape. Arch. Oral Biol. 46: 191–199.
Rodríguez-Mendoza R., Muñoz M., Saborido-Rey F., 2011. Ontogenetic allometry of the bluemouth, *Helicolenus dactylopterus dactylopterus* (Teleostei: Scorpaenidae), in the Northeast Atlantic and Mediterranean based on geometric morphometrics. Hydrobiologia 670: 5–22.
Roff D.A., 1997. Evolutionary quantitative genetics. Chapman & Hall, New York.
Rohlf F.J. 1993. Relative warp analysis and an example of its application to mosquito wings. In: Marcus L.F., Bello E., García-Valdecasas A. (Eds.) Contributions to morphometrics. Museo Nacional de Ciencias Naturales, Madrid. 131–159.
Rohlf F.J., 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution 55: 2143–2160.

Rohlf F.J., 2002. Geometric morphometrics and phylogeny. In: MacLeod N., Forey P.L. (Eds.) Morphology, shape, and phylogeny. Taylor & Francis, London. 175–193.
Rohlf F.J., 2003. Bias and error in estimates of mean shape in geometric morphometrics. J. Hum. Evol. 44: 665–683.
Rohlf F.J., Corti M., 2000. The use of two-block partial least-squares to study covariation in shape. Syst. Biol. 49: 740–753.
Rosas A., Bastir M., 2004. Geometric morphometric analysis of allometric variation in the mandibular morphology of the hominids of Atapuerca, Sima de los Huesos site. Anat. Rec. 278A: 551–560.
Roth V.L., 1996. Cranial integration in the Sciuridae. Amer. Zool. 36: 14–23.
Sanger T.J., Mahler D.L., Abzhanov A., Losos J.B., 2012. Roles for modularity and constraint in the evolution of cranial diversity among *Anolis* lizards. Evolution 66: 1525–1542
Sans-Fuentes M.A., Ventura J., López-Fuster M.J., Corti M., 2009. Morphological variation in house mice from the Robertsonian polymorphism area of Barcelona. Biol. J. Linn. Soc. 97: 555–570.
Santos M., Fernández Iriarte P., Céspedes W., 2005. Genetics and geometry of canalization and developmental stability in *Drosophila subobscura*. BMC Evol. Biol. 5: 7.
Sardi M.L., Ramírez Rozzi F.V., 2012. Different cranial ontogeny in Europeans and Southern Africans. PLoS ONE 7(4): e35917. doi:10.1371/journal.pone.0035917
Sardi M.L., Ventrice F., Ramírez Rozzi F., 2007. Allometries throughout the late prenatal and early postnatal human craniofacial ontogeny. Anat. Rec. 290: 1112–1120.
Savriama Y., Gómez J.M., Perfectti F., Klingenberg C.P., 2012. Geometric morphometrics of corolla shape: dissecting components of symmetric and asymmetric variation in *Erysimum mediohispanicum* (Brassicaceae). New Phytol. 196: 945–954.
Savriama Y., Klingenberg C.P., 2011. Beyond bilateral symmetry: geometric morphometric methods for any type of symmetry. BMC Evol. Biol. 11: 280.
Schlosser G., Wagner G.P. (Eds.), 2004. Modularity in development and evolution. University of Chicago Press, Chicago.
Schluter D., 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50: 1766–1774.
Schmidt E.J., Parsons T.E., Jamniczky H.A., Gitelman J., Trpkov C., Boughner J.C., Logan C.C., Sensen C.W., Hallgrímsson B., 2010. Micro-computed tomography-based phenotypic approaches in embryology: procedural artifacts on assessments of embryonic craniofacial growth and development. BMC Dev. Biol. 10: 18.
Shirai L.T., Marroig G., 2010. Skull modularity in Neotropical marsupials and monkeys: size variation and evolutionary constraint and flexibility. J. Exp. Zool. B Mol. Dev. Evol. 314: 663–683.
Siahsarvie R., Auffray J.-C., Darvish J., Rajabi-Maham H., Yu H.-T., Agret S., Bonhomme F., Claude J., 2012. Patterns of morphological evolution in the mandible of the house mouse *Mus musculus* (Rodentia: Muridae). Biol. J. Linn. Soc. 105: 635–647.
Sidlauskas B., 2008. Continuous and arrested morphological diversification in sister clades of characiform fishes: a phylomorphospace approach. Evolution 62: 3135–3156.
Sidlauskas B.L., Mol J.H., Vari R.P., 2011. Dealing with allometry in linear and geometric morphometrics: a taxonomic case study in the *Leporinus cylindriformis* group (Characiformes: Anostomidae) with description of a new species from Suriname. Zool. J. Linn. Soc. 162: 103–130.
Singh N., Harvati K., Hublin J.-J., Klingenberg C.P., 2012. Morphological evolution through integration: a quantitative study of cranial integration in *Homo*, *Pan*, *Gorilla* and *Pongo*. J. Hum. Evol. 62: 155–164.
Singleton M., 2002. Patterns of cranial shape variation in the Papionini (Primates: Cercopithecinae). J. Hum. Evol. 42: 547–578.
Singleton M., Rosenberger A.L., Robinson C., O'Neill R., 2011. Allometric and metameric shape variation in *Pan* mandibular molars: a digital morphometric analysis. Anat. Rec. 294: 322–334.
Skelton R.R., McHenry H.M., 1992. Evolutionary relationships among early hominids. J. Hum. Evol. 23: 309–349.
Smith H.F., 2011. The role of genetic drift in shaping modern human cranial evolution: a test using microevolutionary modeling. Int. J. Evol. Biol. 2011: 145262.
Sneath P.H.A., Sokal R.R., 1973. Numerical taxonomy: the principles and practice of numerical classification. W. H. Freeman, San Francisco.
Steppan S.J., 1997a. Phylogenetic analysis of phenotypic covariance structure. I. Contrasting results from matrix correlation and common principal component analyses. Evolution 51: 571–586.
Steppan S.J., 1997b. Phylogenetic analysis of phenotypic covariance structure. II. Reconstructing matrix evolution. Evolution 51: 587–594.
Steppan S.J., Phillips P.C., Houle D., 2002. Comparative quantitative genetics: evolution of the G matrix. Trends Ecol. Evol. 17: 320–327.
Strait D.S., Grine F.E., Moniz M.A., 1997. A reappraisal of early hominid phylogeny. J. Hum. Evol. 32: 17–82.
Strand Viðarsdóttir U., O'Higgins P., Stringer C.B., 2002. A geometric morphometric study of regional differences in the ontogeny of the modern human facial skeleton. J. Anat. 201: 211–229.
Swiderski D.L., Zelditch M.L., 2010. Morphological diversity despite isometric scaling of lever arms. Evol. Biol. 37: 1–18.
Sydney N.V., Machado F.A., Hingst-Zaher E., 2012. Timing of ontogenetic changes of two cranial regions in *Sotalia guianensis* (Delphinidae). Mamm. Biol. 77: 397–403.
Tabachnick R.E., Bookstein F.L., 1990. The structure of individual variation in Miocene *Globorotalia*. Evolution 44: 416–434.
Tagliaro M.L., Mattos de Oliveira R., Pereira Padilha D.M., Callegiari-Jacques S.M., Jeckel-Neto E.A., 2009. Morphological changes in the mandible of male mice associated with aging and biomechanical stimulus. Anat. Rec. 292: 431–438.
Tobin J.L., Di Franco M., Eichers E., May-Simera H., Garcia M., Yan J., Quinlan R., Justice M.J., Hennekam R.C.M., Briscoe J., Tada M., Mayor R., Burns A.J., Lupski J.R., Hammond P., Beales P.L., 2008. Inhibition of neural crest migration underlies craniofascial dysmorphology and Hirschsprung's disease in Bardet-Biedl syndrome. Proc. Natl. Acad. Sci. USA 105: 6714–6719.
Vecchione L., Miller J., Byron C., Cooper G.M., Barbano T., Cray J., Losee J.E., Hamrick M.W., Sciote J.J., Mooney M.P., 2010. Age-related changes in craniofacial morphology in GDF-8 (myostatin)-deficient mice. Anat. Rec. 293: 32–41.
Velemínská J., Bigoni L., Krajíček V., Borský J., Šmahelová D., Cagáňová V., Peterka M., 2012. Surface facial modelling and allometry in relation to sexual dimorphism. Homo 63: 81–93.
Velhagen W.A., Roth V.L., 1997. Scaling of the mandible in squirrels. J. Morphol. 232: 107–132.
von Cramon-Taubadel N., 2009. Congruence of individual cranial bone morphology and neutral molecular affinity patterns in modern humans. Am. J. Phys. Anthropol. 140: 205–215.
von Cramon-Taubadel N., 2011a. Global human mandibular variation reflects differences in agricultural and hunter-gatherer subsistence strategies. Proc. Natl. Acad. Sci. USA 108: 19546–19551.
von Cramon-Taubadel N., 2011b. The relative efficacy of functional and developmental cranial modules for reconstructing global human population history. Am. J. Phys. Anthropol. 146: 83–93.
von Cramon-Taubadel N., Smith H.F., 2012. The relative congruence of cranial and genetic estimates of hominoid taxon relationships: implications for the reconstruction of hominin phylogeny. J. Hum. Evol. 62: 640–653.
Wagner G.P., 1984. On the eigenvalue distribution of genetic and phenotypic dispersion matrices: evidence for a nonrandom organization of quantitative character variation. J. Math. Biol. 21: 77–95.
Wagner G.P., 1996. Homologues, natural kinds and the evolution of modularity. Am. Zool. 36: 36–43.
Wagner G.P., Altenberg L., 1996. Complex adaptations and the evolution of evolvability. Evolution 50: 967–976.
Webster M., Zelditch M.L., 2011a. Evolutionary lability of integration in Cambrian ptychoparioid trilobites. Evol. Biol. 38: 144–162.
Webster M., Zelditch M.L., 2011b. Modularity of a Cambrian ptychoparioid trilobite cranidium. Evol. Dev. 13: 96–109.
Weisensee K.E., Jantz R.L., 2011. Secular change in craniofacial morphology of the Portuguese using geometric morphometrics. Am. J. Phys. Anthropol. 145: 548–559.
White J., 2009. Geometric morphometric investigation of molar shape diversity in modern lemurs and lorises. Anat. Rec. 292: 701–719.
Willmore K.E., Klingenberg C.P., Hallgrímsson B., 2005. The relationship between fluctuating asymmetry and environmental variance in rhesus macaque skulls. Evolution 59: 898–909.
Willmore K.E., Leamy L., Hallgrímsson B., 2006a. Effects of developmental and functional interactions on mouse cranial variability through late ontogeny. Evol. Dev. 8: 550–567.
Willmore K.E., Zelditch M.L., Young N., Ah-Seng A., Lozanoff S., Hallgrímsson B., 2006b. Canalization and developmental stability in the brachyrrhine mouse. J. Anat. 208: 361–372.
Wilson L.A.B., Sánchez-Villagra M.R., 2010. Diversity trends and their ontogenetic basis: an exploration of allometric disparity in rodents. Proc. R. Soc. Lond. B Biol. Sci. 277: 1227–1234.
Wood B., Lonergan N., 2008. The hominin fossil record: taxa, grades and clades. J. Anat. 212: 354–376.
Workman M.S., Leamy L.J., Routman E.J., Cheverud J.M., 2002. Analysis of quantitative trait locus effects on the size and shape of mandibular molars in mice. Genetics 160: 1573–1586.
Wroe S., Milne N. 2007. Convergence and remarkably consistent constraint in the evolution of carnivore skull shape. Evolution 61: 1251–1260.
Young N., 2004. Modularity and integration in the hominoid scapula. J. Exp. Zool. B Mol. Dev. Evol. 302: 226–240.
Young N.M., 2006. Function, ontogeny and canalization of shape variance in the primate scapula. J. Anat. 209: 623–636.
Young N.M., Chong H.J., Hu D., Hallgrímsson B., Marcucio R.S., 2010. Quantitative analyses link modulation of sonic hedgehog signaling to continuous variation in facial growth and shape. Development 137: 3404–3409.
Young N.M., Wat S., Diewert V.M., Browder L.W., Hallgrímsson B., 2007. Comparative morphometrics of embryonic facial morphogenesis: Implications for cleft-lip etiology. Anat. Rec. 290: 123–139.
Young R.L., Badyaev A.V., 2006. Evolutionary persistence of phenotypic integration: influence of developmental and functional relationships on complex trait evolution. Evolution 60(6): 1291–1299.
Young R.L., Badyaev A.V., 2010. Developmental plasticity links local adaptation and evolutionary diversification in foraging morphology. J. Exp. Zool. B Mol. Dev. Evol. 314: 434–444.
Zelditch M.L., 2005. Developmental regulation of variability. In: Hallgrímsson B., Hall B.K. (Eds.) Variation: a central concept in biology. Elsevier, Burlington, MA. 249–276.
Zelditch M.L., Mezey J.G., Sheets H.D., Lundrigan B.L., Garland T., Jr., 2006. Developmental regulation of skull morphology II: ontogenetic dynamics of covariance. Evol. Dev. 8: 46–60.
Zelditch M.L., Swiderski D.L., Sheets H.D., 2012. Geometric morphometrics for biologists: a primer. Elsevier, Amsterdam.
Zelditch M.L., Wood A.R., Bonett R.M., Swiderski D.L., 2008. Modularity of the rodent mandible: integrating bones, muscles, and teeth. Evol. Dev. 10: 756–768.
Zelditch M.L., Wood A.R., Swiderski D.L., 2009. Building developmental integration into functional systems: function-indiced integration of mandibular shape. Evol. Biol. 36: 71–87.
Zollikofer C.P.E., Ponce de León M.S., 2004. Kinematics of cranial ontogeny: heterotopy, heterochrony, and geometric morphometric analysis of growth models. J. Exp. Zool. B Mol. Dev. Evol. 302: 322–340.

Associate Editor: A. Cardini

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 59–66, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/6369/pdf
doi:10.4404/hystrix-24.1-6369

Research Article

A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology

Philipp Mitteroecker[a,*], Philipp Gunz[b], Sonja Windhager[c,d], Katrin Schaefer[c]

[a] *Department of Theoretical Biology, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria*
[b] *Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04107 Leipzig, Germany*
[c] *Department of Anthropology, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria*
[d] *Konrad Lorenz Institute for Evolution and Cognition Research, Adolf Lorenz Gasse 2, A-3422, Altenberg, Austria*

Keywords:
allometry
centroid size
face shape
form space
Procrustes
shape space

Article history:
Received: 9 July 2012
Accepted: 16 April 2013

Acknowledgements
We thank Anna Loy and Andrea Cardini for inviting us to contribute to this special issue. We are further grateful to Chris Klingenberg, Andrea Cardini, Barbara Fischer, and Jürgen Herler for helpful comments on the manuscript. Philipp Mitteroecker was supported by the Focus of Excellence "Biometrics of EvoDevo" of the Faculty of Life Sciences, University of Vienna.

Abstract

We briefly and informally review the concepts of size, shape, and form and how they are estimated in geometric morphometrics using Procrustes analysis. We demonstrate how deformation grids and reconstructed shapes or forms can be used as powerful tools to visualize shape and form differences. Complex patterns of individual or group differences can be studied effectively by ordinations of shape space or form space. Allometry, the statistical relationship between size and shape, is estimated in geometric morphometrics by regression of the Procrustes shape coordinates on centroid size. We illustrate these methods by an application to human face shape. We reveal shape cues to body size in the adult male face, partly resembling ontogenetic allometry. Facial allometry might thus be an important confounding factor in studies of face perception and human mate choice.

During the last two decades, geometric morphometrics became the state-of-the-art method for statistical shape analysis in biology (Rohlf and Marcus, 1993; Bookstein, 1996; Adams et al., 2004; Mitteroecker and Gunz, 2009; Klingenberg, 2010). In this paper we briefly and informally review the concepts of size, shape, and form, and how they are estimated in geometric morphometrics. We further discuss how the classic concept of allometry is analyzed in contemporary geometric morphometrics and illustrate this by an application to human facial form. In the Appendix we provide some algebraic details on the visualization of shape and form differences.

Size and shape

It is a common practice in morphometrics to distinguish between the size and the shape of a biological structure. The *shape* of an object are the geometric properties that are invariant to translation, rotation, and scaling. In other words, the shape of an object is unaffected by changes in the position, the orientation, and the size of the object. Two objects have the same shape if they can be translated, rescaled, and rotated to each other so that they match exactly. For instance, the shape of a rectangle, which can be described by the ratio of the two different side lengths, does neither depend on the size of the rectangle, nor on its position and orientation. By contrast, the term *form* refers to the geometric properties invariant only to translation and rotation. Hence, form can be considered as "size-and-shape" (e.g., Dryden and Mardia 1998). The form of a rectangle (measured, for instance, by the two side lengths or by one side length together with the ratio of side lengths) does not depend on the position and orientation of the rectangle, but both on its size and shape.

The size or scale of different objects is easy to quantify when the objects all have the same shape. For a set of rectangles, all of the same shape, any side length or diagonal length would be an equally suitable size measure. But whenever shape varies, size is an ambiguous concept and no unique quantification exists. For a set of rectangles varying both in size and shape, a single side length as size measure would be affected both by changes in shape (ratio of side lengths) and by changes in scale. Composite size measures, such as functions of the sum or of the product of the side lengths, may be less affected by shape changes than single measurements. Ultimately, it is up to the researcher to decide, based on biological considerations, whether a single measurement or a composite variable is the most useful size measure.

In contrast to classical morphometric approaches based on linear distances and angles, geometric morphometric methods are based on the Cartesian coordinates of measurement points, so-called *landmarks*. Landmarks have a position (two or three coordinates) and a name, expressing some sort of homology across all measured specimens (Bookstein, 1991). Since the geometry of the measured landmark configuration is preserved by the set of landmark coordinates, geometric morphometrics allows for effective visual representations of statistical results as actual shapes/forms or shape/form deformations (e.g., Bookstein 1991, 1996; Rohlf and Marcus 1993; Zollikofer and Ponce de Leon 2002; Adams et al. 2004; Zelditch et al. 2004; Slice 2007; Mitteroecker and Gunz 2009).

The most common measure of size used in geometric morphometrics is *centroid size* (CS): the square root of the summed squared distances between all landmarks and their centroid (Fig. 1). The centroid of a landmark configuration is the average (arithmetic mean) of all land-

*Corresponding author
Email address: philipp.mitteroecker@univie.ac.at (Philipp Mitteroecker)

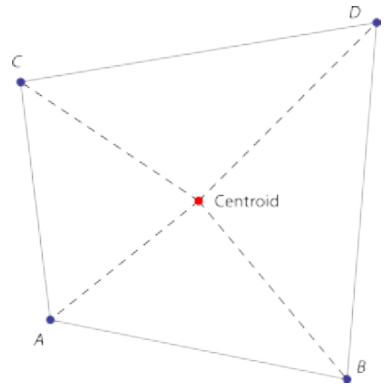


Figure 1 – A configuration of four landmarks (A-D) with their centroid, which is equal to the average landmark position. Centroid size, the size measure used in geometric morphometrics, is equal to the square root of the summed squared distances between the landmarks and their centroid (square root of the summed squared lengths of the dashed lines).

marks. Centroid size is a composite size measure based on all landmarks and is proportional to the square root of the summed squared interlandmark distances. It has been shown to be uncorrelated with shape for small isotropic variation at each landmark (Bookstein, 1991; Dryden and Mardia, 1998). Isotropic variation (independent, identically distributed circular variation at each landmark) is the usual null-model corresponding to pure noise. It is an important property of a statistical method that pure noise does not induce an apparent "signal" (in this case a correlation between size and shape). Real biological data, however, may deviate considerably from isotropy and hence CS usually is not unrelated to shape.

The raw landmark coordinates do not only comprise information on size and shape of the landmark configurations, but also on their position and orientation. Landmark coordinates hence are not directly suitable for statistical analysis. The most common approach for separating shape from size and the "nuisance parameters" position and orientation is Generalized Procrustes Analysis (Gower, 1975; Rohlf and Slice, 1990). This method comprises three steps: translating all landmark configurations to the same centroid, scaling all configurations to the same centroid size, and iteratively rotating all configurations until the summed squared distances between the landmarks and their corresponding sample average is a minimum (Fig. 2). The coordinates of the resulting superimposed landmark configurations are called *Procrustes shape coordinates* as they only contain information about the shape of the configurations. (Note that scaling the configurations to unit CS resembles the usual approaches to size correction but is not the actual least-squares solution; it has thus been referred to as partial Procrustes fitting. Also, CS is slightly modified in the course of tangent space projection. For more details see Rohlf and Slice 1990; Dryden and Mardia 1998; Rohlf 1999).

The standardization for position, scale, and orientation is based on all landmarks. Alternatively, Procrustes analysis may be based on just a subset of landmarks, for example when some landmarks are known to vary in the sample or are subject to some treatment, whereas the other landmarks are relatively stable (e.g., Bookstein et al. 1999). When the registration is based on two landmarks only, the resulting shape coordinates are called two-point shape coordinates or Bookstein shape coordinates (Bookstein, 1991). Alternatives to the commonly used least-squares oriented Procrustes analysis are maximum likelihood Procrustes analysis (Theobald and Wuttke, 2006) and robust Procrustes analysis based on medians instead of means ("Resistant Fit"; Slice 1996).

Shape space and form space

Mathematical spaces are widely used in the sciences to represent relationships between complex objects (e.g., Stadler et al. 2001; Mitteroecker and Huttegger 2009). In a *shape space* the shapes of different objects are represented by single points, related by some notion of distance or proximity (as a measure of shape difference). The mathematical properties of the shape space for landmark configurations, usually referred to as Kendall's shape space, have been studied intensively (e.g., Kendall 1984; Bookstein 1991, 1996; Goodall 1991; Small 1996; Dryden and Mardia 1998; Rohlf 1999; Slice 2001). For p landmarks in k dimensions, it is a nonlinear Riemannian manifold of dimension $pk - k - k(k-1)/2 - 1$, which can be approximated locally by a Euclidean space of the same dimension, a so-called tangent space. Marcus et al. (2000) showed that the Euclidean approximation of shape space is appropriate for most biological data sets. The metric on Kendall's shape space is *Procrustes distance*, which is approximated by the Euclidean distance between two sets of Procrustes shape coordinates (square root of the summed squared distances between the corresponding landmarks after Procrustes superimposition). Procrustes distance is a measure of shape difference between two landmark configurations. It is zero only if the configurations have the same shape, and larger than zero otherwise. Standard multivariate methods, such as principal component analysis, can thus be applied to Procrustes shape coordinates in order to yield a low-dimensional representation (a so-called ordination) of shape space.

Figure 3a shows the first two principal components for the Procrustes shape coordinates of the landmark configurations in Fig. 2. Each point in this plot corresponds to the shape of one landmark configuration; the plot thus is a low-dimensional representation of shape space. The principal component loadings (eigenvectors of the covariance matrix of Procrustes shape coordinates) can be visualized by deformation grids ("relative warps"; Bookstein 1991).

A space relating the forms of different objects is called *form space* or *size-and-shape space* (or simply size-shape space). For landmark configurations, form space can be constructed in two principal ways: first, by standardizing the raw landmark configurations for position and orientation as in the usual Procrustes analysis, but not standardizing for size (e.g., Dryden and Mardia 1998, chapter 8), and second, by augmenting the Procrustes shape coordinates with the natural logarithm of centroid size ($\ln$ CS) as an additional variable (Mitteroecker et al., 2004a). The log transformation guarantees that for isotropic landmark variation the distribution in size-and-shape space is isotropic as well. Either set of variables can be used for principal component analysis (PCA) and other statistical analyses, and the Euclidean distance in these spaces can be interpreted as a measure of form difference. Both approaches yield similar results for small variation in size and shape, but they may differ if shape variation is large. We recommend constructing form space by augmenting the Procrustes coordinates by $\ln$ CS, following Mitteroecker et al. (2004a), because size is explicitly represented by centroid size in this approach and the loadings or coefficients

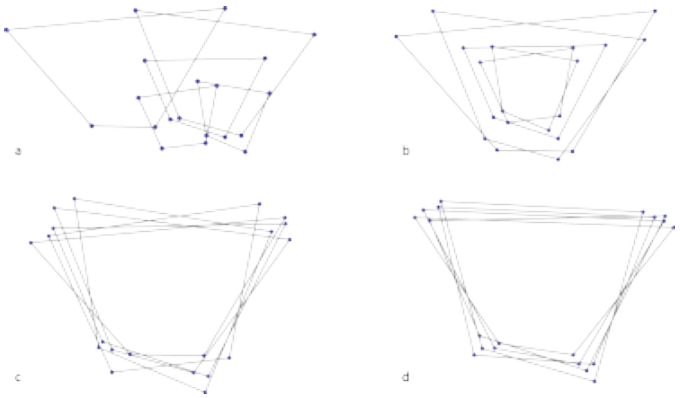


Figure 2 – Procrustes superimposition consists of three steps: translation, scaling, and rotation. As an example, take five configurations of four landmarks each. The raw landmark configurations in (a) are translated so that they all have the same centroid (b). The centered configurations then are scaled to the same centroid size (c) and iteratively rotated until the summed squared distances between the landmarks and their corresponding sample average position is a minimum (d). The resulting landmark coordinates are called Procrustes shape coordinates because variation in position, size, and orientation is removed.

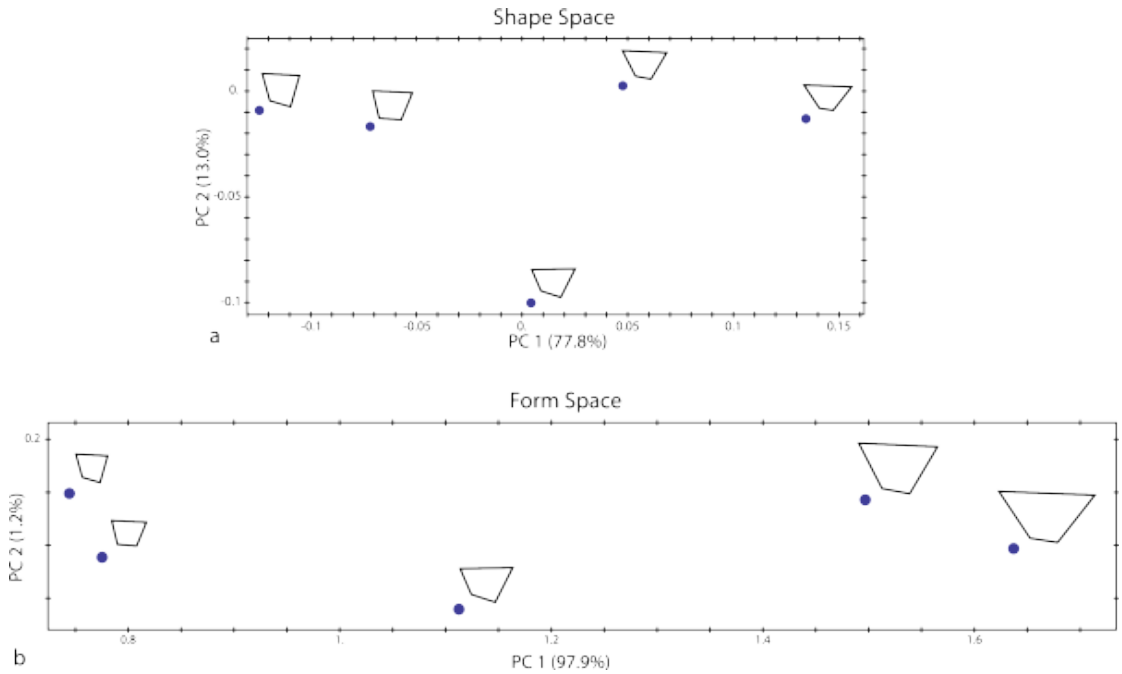


Figure 3 – (a) The first two principal components (PCs) of the eight Procrustes shape coordinates for the landmark configurations shown in Fig. 2. This plot is a low-dimensional representation of shape space, in which every point represents one shape. The actual shapes are drawn next to the corresponding points. (b) The first two PCs of the Procrustes shape coordinates and the natural logarithm of centroid size (nine variables in total). This plot is a low-dimensional representation of form space: every form is represented by a single point. The forms (scaled shapes) are drawn next to the corresponding points. Note that differences in size and allometric shape are closely aligned with PC 1 of form space (ln CS has a loading of 0.97 on the first eigenvector). The first PC thus accounts for a larger fraction of variance in form space than in shape space.

for size and shape, resulting from statistical methods such as PCA or multivariate regression, can directly be compared and interpreted.

In Fig. 3b, the first two PCs of form space (Procrustes shape coordinates plus ln CS) are shown for the landmark configurations in Fig. 2. This plot is a low-dimensional representation of form space. The forms are drawn next to the corresponding points. Note that these configurations are the same as those in Fig. 3a, but scaled to their original size. See the Appendix for more details on the visualization of principal components in form space.

Allometry

Allometry is the statistical association between size and shape (Mosimann, 1970), or as formulated by (Gould, 1966, p. 587) with a stronger emphasis on causality, "the study of size and its consequences". Allometry has been an influential concept in biology and morphometrics since Huxley's seminal treatise "Problems of Relative Growth" in 1932 (for reviews see, e.g., Gould 1966, 1977; Klingenberg 1998). Huxley's original version as well as its multivariate generalization by Jolicoeur (1963) were based on multiplicative growth models for size measurements such as bone lengths and other linear distances, areas, organ weights or volumes. Allometry was thus expressed as a power function between traits, or equivalently, as a linear relationship between the log-transformed traits. Logarithmic transformation of variables is also common as a way of overall size correction and to account for different units (e.g., Gould 1966; Bookstein et al. 1985; Marcus 1990).

Geometric morphometric studies of allometry require a methodological approach that differs in several aspects from the classic allometry studies. Procrustes shape coordinates are shape variables, not size variables, and they have no natural zero point (they are on interval scales, not ratio scales) and hence cannot be log transformed. Allometry is thus expressed as a – usually linear – function of the Procrustes shape coordinates, estimated by multivariate regression of the shape coordinates on centroid size or the logarithm of centroid size (Fig. 4; see also Bookstein 1991; Dryden and Mardia 1998; Klingenberg 1998; Monteiro 1999; Mitteroecker et al. 2004a). Note that in the classic approach to allometry a constant linear slope between two size variables indicates a constant size ratio, i.e., isometric size increase, whereas a constant linear slope between a shape variable and a size variable indicates allometric size increase. The vector of regression coefficients resulting from the multivariate regression of the shape coordinates on CS describes how shape changes in response to one unit size increase. This coefficient vector, or an arbitrary multiple of this vector, can be visualized as a shape deformation.

Statistical significance tests for allometry, i.e., for the dependence of shape on size, usually are multivariate tests based on all shape coordinates (or a subset of shape coordinates). Since allometry is computed as a multivariate regression of the shape coordinates on CS, significance levels can be computed either by the usual multivariate parametric methods (e.g., MANOVA) or by resampling tests, which do not require normally distributed variables (e.g., Good 2000; Mitteroecker and Gunz 2009; Bookstein in press).

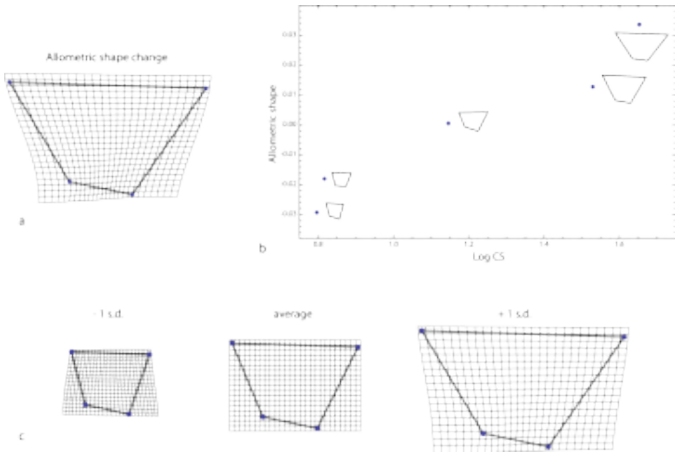


Figure 4 – (a) The multivariate regression (shape regression) of the Procrustes shape coordinates on ln centroid size is visualized by a TPS deformation grid. It represents the allometric shape change that corresponds to a size increase of 2 standard deviations. (b) Scatter plot of allometric shape (scores along the vector of regression coefficients) versus ln CS. (c) Visualization of PC 1 in form space (Fig. 3b). The left grid corresponds to a deformation from the average form to -1 standard deviation (s.d.) along PC 1 and the right grid to a deformation from the average form to +1 s.d. along PC 1. Note that the configurations change both in shape and size along this direction in form space. The shape changes depicted in (a) and (c) are very similar.

In many morphometric data sets, allometry is the most dominant factor of shape and form variation within one group. Thus allometry often is closely aligned with the first (within-group) principal component in shape space and particularly in form space (Mitteroecker et al. 2004a; see also Fig. 4c). However, since this is not guaranteed to be the case for all data sets, the actual multivariate regression of shape on CS is the more direct and reliable method to determine allometry and should be preferred over the first PC in shape space or form space.

Similarly, it has become common practice in geometric morphometrics to plot principal component scores of shape versus CS in order to assess or test for allometry. But this approach is likewise not ideal because single principal components are not necessarily good representations of allometric shape, especially if the PCA is over multiple groups. Instead of plotting PC scores, the score along the allometry vector (vector of regression coefficients) can be plotted as an "allometric shape score" against CS or other variables (Fig. 4; see also the Appendix). Statistical tests for allometry, the dependence of shape on size, should not be based on selected PCs, but should be multivariate tests based on all shape coordinates or all principal components. Of course, statistical tests for the dependence of *form* on size make no sense, since form comprises both shape and size.

Ontogenetic allometry, the association between size and shape across different age stages, often is used as an estimate of a population's ontogenetic trajectory (average growth pattern), especially when the individual calendar ages are not known. *Static allometry*, the association between size and shape within a single age stage (usually in adults), has often been used to explain the coevolution of size and shape (i.e., *evolutionary allometry* – the association between size and shape across multiple species) and as a model for functional and behavioral adaptations (e.g., Gould 1966, 1977; Lande 1979; Cheverud 1982; Klingenberg 1998; Marroig and Cheverud 2005; Schaefer et al. 2004; Gunz 2012).

In the classic concept of allometry, a trait is considered as negatively allometric if it increases less in size than other traits or overall size do. The trait's relative size thus decreases with increasing overall size. A trait with positive allometry, by contrast, increases more in size than other traits do. For instance, head size, compared to body size, is negatively allometric during human growth, whereas limb length is positively allometric. The variables used in geometric morphometrics are the shape coordinates of the landmarks. Positive or negative allometry cannot be inferred from single shape coordinates (they are shape variables, not size variables, and depend on the actual superimposition). All shape coordinates must be visualized together in order to draw inferences about the relative size increase or decrease of specific parts described by the landmarks (see the example below).

Comparison of allometric relationships

When comparing two or more groups of individuals, the question might arise how allometry differs across the groups. In recent years, a large body of literature on the morphometric comparison of growth patterns and ontogenetic allometries arose, advocating different morphometric and statistical approaches (e.g., O'Higgins et al. 2001; Ponce de Leon and Zollikofer 2001; Zelditch et al. 2003; Mitteroecker et al. 2004a,b, 2005; Gerber et al. 2007; Gerber 2011; Adams and Collyer 2009; Piras et al. 2011; Gunz 2012; Collyer and Adams this issue; Sheets and Zelditch this issue).

A simple and effective way to compare ontogenetic or static allometry across groups is the visual comparison of deformation grids or of series of reconstructed shapes representing group-specific allometry (see Fig. 4 and the example below). Since these deformations can easily be described in qualitative morphological terms, the comparison of deformations often leads to useful biological inferences.

Growth patterns and allometric relationships can be compared in a more abstract way by assessing the geometry of the corresponding vectors in shape space or form space. This can be particularly effective when comparing more than two trajectories (see the example below), but the biological meaning of such quantifications, especially of single parameters such as the angle between two trajectories, is not always clear (Mitteroecker et al., 2004b, 2005; Huttegger and Mitteroecker, 2011). Ordinations such as principal component analysis or between-group principal component analysis (Mitteroecker and Bookstein, 2011) can be useful to explore ontogenetic and allometric trends in different groups. Growth trajectories and allometric vectors can be plotted within a principal component plot, either as linear vectors or as nonlinear estimates such as local linear regressions (e.g., Bulygina et al. 2006; Coquerelle et al. 2011). Alternatively, multiple allometric vectors can directly be compared by principal component analysis (for examples see the analysis below and Schaefer et al. 2004).

Parallel ontogenetic trajectories or allometry vectors in two or more groups indicate that the groups have the same linear pattern of (relative) growth during the observed age periods, even if they differ in the initial or in the adult morphology. If the average morphology of a species differs from its ancestral average morphology mainly by the extension, truncation, or the developmental timing of an otherwise conserved ontogenetic trajectory, the underlying evolutionary process is referred to as *allometric scaling* or *heterochrony* (Gould, 1977; Alberch et al., 1979; Klingenberg, 1998; Mitteroecker et al., 2004a, 2005; Gerber et al., 2007).

Ontogenetic trajectories or allometry vectors differing in direction indicate different growth patterns or ontogenetic allometries. Many studies advocated the use of the angle between two trajectories in shape space as a measure of difference in allometry or growth. While this can be useful, one should be careful in interpreting angles without the starting positions of the trajectories. For example, the angle alone is not sufficient to distinguish between diverging, converging, or intersecting trajectories. Furthermore, if two or more pairs of trajectories deviate in different directions in shape space or form space, the angles may not be comparable in a biologically meaningful way because they also depend on the number and spatial distribution of landmarks (Huttegger and Mitteroecker, 2011).

Example: Ontogenetic and static allometry of male faces

In this empirical example we apply the methods described above to assess allometry in human faces. A large number of studies investigated how face shape leads to masculinity attributions and other social inference, such as dominance, attractiveness, or trustworthiness (e.g., Zebrowitz and Montepare 2008; Schaefer et al. 2009; Little et al. 2011). Head size and body height are positively correlated (e.g., Geraedts et al. 2011), and body height has been found to play an important role in studies of social inference and mate choice (Pawlowski et al., 2000; Courtiol et al., 2010). Therefore, human facial allometry – aspects of face shape reflecting body size – seems to be of central relevance to these questions. However, it has received surprisingly little attention in the literature.

Our sample comprises frontal photographs of 19 boys (age 6–11 years) and 25 men (age 17–33 years). A camera with a 200 mm lens was positioned at eye height, 3.5 m away from the face. The heads were adjusted according to the Frankfort Horizontal Plane, and a ruler was placed next to one ear. This careful procedure is a prerequisite for the reliable assessment of undistorted face shape and size (e.g., Schneider et al. 2012).

On each image we digitized 35 anatomical landmarks and 34 semilandmarks to describe overall facial form (see Windhager et al. 2011 for details). Semilandmarks are points on smooth curves, for which the exact location on the curve cannot be identified and hence is statistically estimated. We used the sliding landmark algorithm for this purpose, which minimizes the bending energy, a measure of local shape difference, between each individual and the sample average (Bookstein, 1997; Gunz et al., 2005; Gunz and Mitteroecker, this issue). This approach allows for the joint analysis of biologically homologous points (anatomical landmarks) and curves (represented by geometrically corresponding semilandmarks).

After sliding the semilandmarks, all 44 landmark configurations were superimposed by a Generalized Procrustes Analysis and symmetrized by averaging each landmark configuration with its relabeled

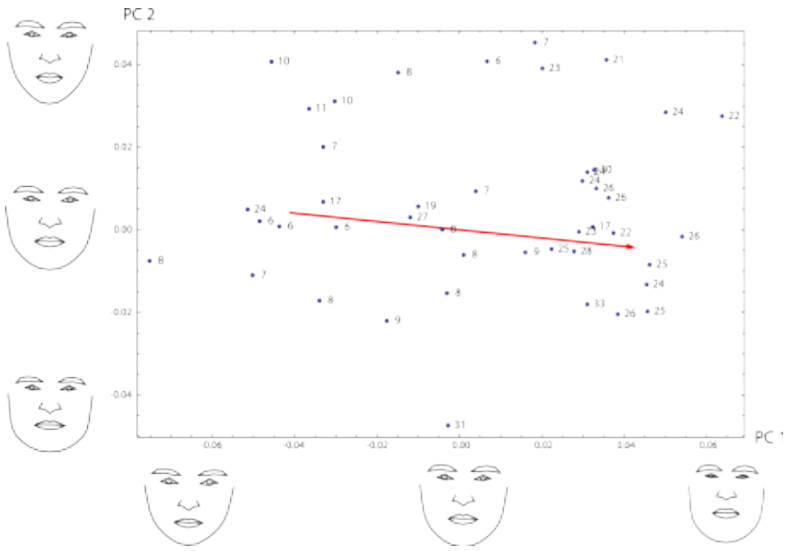


Figure 5 – Scatterplot of the first two principal components (PC) of face shape. The individuals are labelled by their age. The red arrow represents ontogenetic allometry – the regression of shape on ln centroid size. The facial reconstructions correspond to PC 1 scores of -0.7, 0, and 0.7 and to PC 2 scores of -0.5, 0, and 0.5, respectively. They visualize the shape differences associated with the first two PCs.

reflection (Mardia et al., 2000; Mitteroecker and Gunz, 2009). Figure 5 shows a scatterplot of the first two principal components (PC 1 and PC 2) of the resulting shape coordinates, accounting for 42% and 14% of total shape variation. The difference between juvenile and adult face shape is the most dominant factor of shape variation and hence is closely aligned with the first principal component of shape – a common finding in ontogenetic samples. The red arrow represents the coefficient vector for the regression of shape on ln centroid size, i.e., actual ontogenetic allometry, which is likewise aligned along PC 1. The three facial configurations visualizing shape differences along PC 1 show an overall elongation of the midface and the lower face, a relative decrease in eye size, as well as a thickening and lowering of the eyebrows. Shape changes associated with PC 2 reflect the differences between more gracile and more robust faces (Fig. 5) – a pattern known to correlate with prenatal testosterone exposure (Fink et al., 2005; Meindl et al., 2012).

Figure 6 shows the first two principal components of facial form (Procrustes shape coordinates augmented by ln CS). They account for 79% and 6% of total form variation, respectively. The shape differences corresponding to PC 1 are similar to the first PC in shape space (Fig. 5), but PC 1 in form space additionally comprises differences in size. As in most other morphometric datasets, PC 1 in form space accounts for a larger fraction of total variance and is even more closely aligned with allometry (the red arrow) than PC 1 in shape space.

Because facial growth comprises changes in both shape and size, facial form more closely reflects an individual's age than facial shape. The first PC in form space accounts for 71% of variance in age, whereas PC 1 in shape space accounts for 42% only.

Figure 7a contrasts ontogenetic allometry (the regression of shape on ln CS in the full ontogenetic sample) with static allometry (the regression of shape on ln CS in adults). Ontogenetic allometry, which is close to PC 1, reflects the relative size increase (positive allometry) of the lower face and the resulting decrease of relative eye size (negative allometry). A permutation test shows that ontogenetic allometry is statistically significant (Tab. 1). The pattern of static allometry is less pronounced and the regression is not significant, indicating a weaker allometric relationship within male adults and less variation in face size.

Behavioral biological theory primarily refers to sexual dimorphism separately in body size and in face shape, but face size is itself negatively allometric with respect to body size during human ontogeny (for our data, ln face CS and ln body height have a correlation of 0.47 and a slope of 0.14 in adults). We thus also regressed face shape on body height, both in the ontogenetic sample and in adults only. Furthermore, we regressed face shape on age in the two samples (Fig. 7b,c). All three ontogenetic regressions are very similar, reflecting the high correlation between face size, body height, and age during ontogeny (the pairwise correlation coefficients range from 0.84 to 0.92). The static regression of shape on body height to some degree resembles the ontogenetic pattern, but the static regression on age seems to reflect another process—aging as opposed to growth. Static and ontogenetic regressions on body height as well as on age are all significant (Tab. 1).

Figure 8 shows a more abstract comparison of the three ontogenetic and the three static patterns of allometry. All six coefficient vectors, standardized to unit length, are subjected to a principal component analysis. The three ontogenetic regressions cluster together closely, reflecting the similar deformation grids in Fig. 7, whereas the static regressions are heterogeneous and differ from the ontogenetic ones.

Discussion

Size, shape, and form

In this paper we briefly reviewed the concepts of size, shape, and form, and how they are estimated in geometric morphometrics. The shape of a landmark configuration is what remains after position, orientation, and size have been filtered out. In geometric morphometrics, the shape parameters of a set of landmark configurations usually are estimated by Generalized Procrustes Analysis, and the scale of the configurations is quantified separately by centroid size.

Size and shape are classic geometric concepts, and it is a long-standing tradition in morphometrics to analyze and interpret variation in shape separately from variation in size. Yet every organism, and every part of an organism, has a certain form, i.e., a certain size *and* shape. Modern geometric morphometrics hence offers two modes of analysis: an analysis of shape separately from size, and the joint analysis of size and shape in a single form space. But when should one study the shape of objects, separately from size, and when their form? Form is the more comprehensive description of an object than shape alone; it should be used, for instance, in classification studies whenever groups of organisms differ both in size and shape. Discrimination and

Table 1 – Explained variance and statistical significance for the ontogenetic and static regressions of shape on the natural logarithm of centroid size (ln CS), body height, and age. The *p*-values were estimated by permutation tests using 1000 random permutations and the explained variance as the test statistic.

	ontogenetic regressions		static regressions	
	***p*-value**	**expl. var.**	***p*-value**	**expl. var.**
ln CS	$p<0.001$	22.4%	$p=0.76$	2.8%
body height	$p<0.001$	23.5%	$p<0.001$	7.8%
age	$p<0.001$	20.1%	$p<0.001$	7.0%

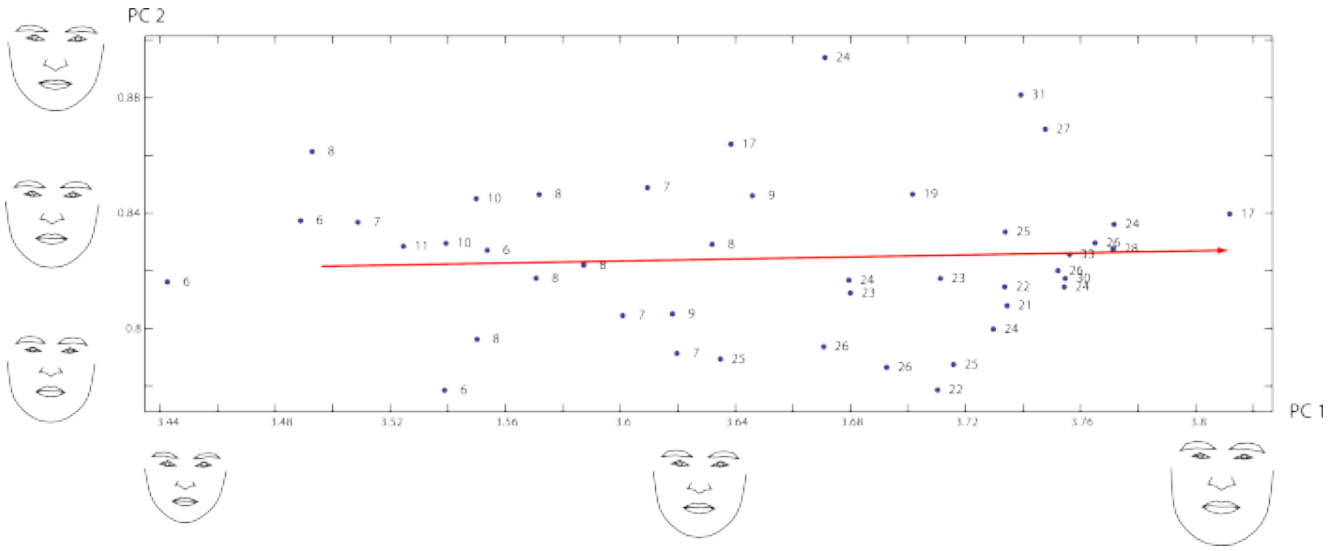


Figure 6 – Scatterplot of the first two principal components (PCs) of facial form (Procrustes shape coordinates augmented by ln CS). The individuals are labelled by their age. The redarrow represents ontogenetic allometry (allometric shape and size) – the regression of form on ln CS. The facial reconstructions visualize the form differences associated with the first two PCs and correspond to PC 1 scores of 3.44, 3.63, and 3.82 and to PC 2 scores of 0.78, 0.84, and 0.90, respectively.

classification based on form usually is more successful than based on shape alone (e.g., Mitteroecker and Bookstein 2011). Likewise, when predicting a variable, such as age in our example on facial morphology, form is the better predictor than shape whenever size is related to the variable as well. In most studies of growth and development, changes both in size and shape may be of scientific relevance and can be analyzed jointly.

When variation in shape is explained differently from variation in size, or when size differences are used as an explanation of shape differences (such as in studies of allometry), shape and size should be quantified and analyzed separately. In some studies, particularly on organisms with indeterminate growth, both geometric size and allometric shape might intentionally be neglected and removed from the data. When the size of the measured objects is not well preserved (e.g., because of unstandardized photographs), shape should of course be analyzed separately from size.

In general, it is a useful exploratory approach to contrast ordinations in shape space to that in form space. For example, ontogenetic trajectories that overlap or are parallel in shape space but differ in direction in form space indicate a dissociation of size and shape during ontogeny (e.g., Mitteroecker et al. 2004a, 2005; Gerber et al. 2007). An *a priori* limitation of a morphometric analysis to either shape or size should be justified explicitly.

Geometric morphometrics offers powerful techniques for the visualization of both shape differences and form differences. Deformation grids and series of reconstructed shapes or forms can easily be interpreted within a biological context (e.g., Bookstein 1991). Ordinations and other multivariate statistical analyses of high-dimensional shape spaces or form spaces can be useful when comparing relationships between multiple groups, but meaningful biological interpretations of such analyses are more difficult (e.g., Mitteroecker et al. 2004b, 2005; Huttegger and Mitteroecker 2011; Bookstein in press). A careful ordination analysis, consilient with a biological explanation, can be more convincing than lists of geometric parameters and significance tests (McCloskey and Ziliak, 2009; Bookstein, in press). Furthermore, in contrast to deformation grids, most multivariate analyses and geometric parameters in shape or form space do not account for the spatial relationship among landmarks (they ignore the mean shape), even though the information is present in the shape coordinates. For example, all shape coordinates contribute equally to multivariate statistics such as covariance matrices or the angle between two trajectories, regardless of whether the landmarks are far apart or whether they almost have the same position (e.g., Mitteroecker 2009; Huttegger and

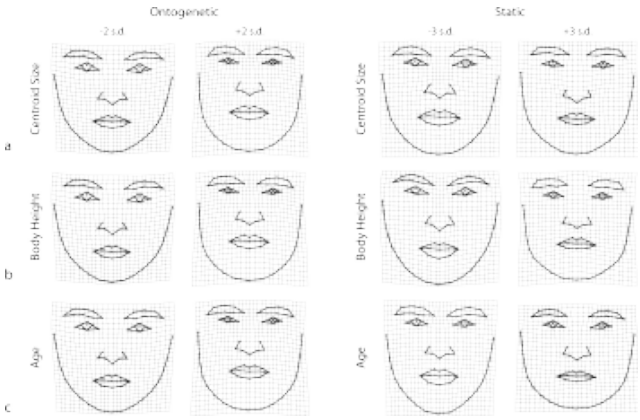


Figure 7 – Visualization of the ontogenetic and static shape regressions. (a) Facial allometry estimated via regressions of face shape on ln CS in the full ontogenetic sample (ontogenetic allometry) and in the subsample of adults (static allometry). The displayed grids are deformations from the mean shape to shapes corresponding to -2 standard deviations (s.d.) and +2 s.d. of centroid size. (b) Regression of face shape on body height in the full sample and in adults only. (c) Regression of face shape on age in the full sample and in adults only.

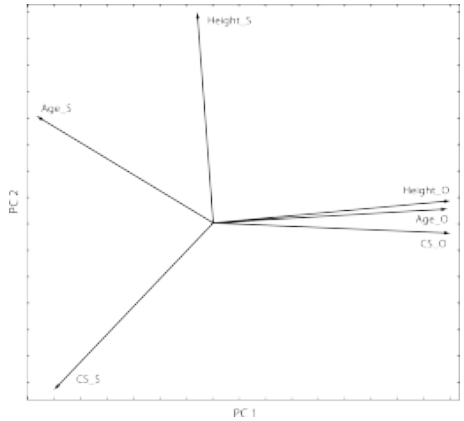


Figure 8 – Principal component analysis of the six shape regression vectors (standardized to the same length) shown in Fig. 7. The ontogenetic regressions are very similar and hence point into a similar direction in this ordination, whereas the three static regressions differ considerably, with the vector of the static regression on body heigh located closest to the ontogenetic ones.

Mitteroecker 2011). Further theoretical and statistical developments may lead to more powerful morphometric analyses.

Facial allometry

We have applied geometric morphometric methods to study ontogenetic and static allometry in human faces. Despite its obvious importance for studies in face perception and evolutionary psychology, facial allometry has received little attention in the corresponding literature. The ontogenetic relationship between body size, face size, and face shape is no big surprise since they all change considerably with age. But our results further reveal shape cues to body size in the adult male face, partly resembling ontogenetic allometry. Face size, by contrast, is less clearly and not significantly related to face shape in our adult sample.

This has far reaching consequences for several lines of face research. Shorter men on average have more childlike facial features than taller men, which in turn affect trait attributions and social stereotyping (Zebrowitz and Montepare, 2008). Facial allometry might thus be an important confounding factor in studies of facial masculinity, dominance, and the like.

We further found age-related shape changes in the adult male face, resembling the qualitative descriptions of Albert et al. (2007): the face becomes wide relative to its length and the lips become thinner, even after correcting for body weight. These are features typically associated with and perceived as mature and male, even in inanimate objects (Windhager et al., 2008).

Effects of aging on face shape differ from the subadult growth pattern, even in our sample with an adult age range from 17 to only 33 years. This may confound estimates of ontogenetic allometry and linear growth trajectories when pooling subadult individuals with adults of a large age range.

Appendix

Here we give some details about the visualization of regression and principal component analysis in shape space and in form space. To our knowledge, the visualization in form space has not been published before.

Let $\mathbf{s}_i$ be a vector of the pk shape coordinates of the ith individual, where p is the number of landmarks in k dimensions, and $i = 1, \ldots, n$. Let further $\mathbf{a}$ be a vector of pk regression coefficients (resulting, for example, from a regression of the shape coordinates on CS in order to estimate allometry) or a vector of principal component loadings in shape space. The coefficient vector $\mathbf{a}$ can be visualized as a deformation grid between a reference shape, usually the mean shape $\overline{\mathbf{s}_i}$, and $\overline{\mathbf{s}_i} + f\mathbf{a}$, the reference shape plus a convenient multiple f of the coefficient vector $\mathbf{a}$. It is effective to contrast two deformations, one from the reference to $\overline{\mathbf{s}_i} - f\mathbf{a}$ and one from the reference to $\overline{\mathbf{s}_i} + f\mathbf{a}$ (such as in Fig. 7), or to present a series of reconstructed shapes, such as $\overline{\mathbf{s}_i} - f\mathbf{a}$, $\overline{\mathbf{s}_i}$, and $\overline{\mathbf{s}_i} + f\mathbf{a}$ (Fig. 5). Scores along the vector $\mathbf{a}$, that is, a variable reflecting allometric shape (Fig. 4), can be computed as the linear combination $\mathbf{a}'\mathbf{s}_i$. When $\mathbf{a}$ is scaled to unit length, the linear combination is equal to an orthogonal projection on the coefficient vector.

Vectors in form space, resulting from a regression of form on some variable or from PCA in form space, can be visualized as form deformations or series of reconstructed forms, comprising differences in both size and shape. When form space is constructed based on the pk coordinates of the centered and rotated – but not scaled – landmark configurations, the visualization can proceed as described above for shape, but now $\mathbf{a}$ might induce a change in size.

When form space is based instead on the shape coordinates (centered, scaled, and rotated configurations) together with the natural logarithm of centroid size ($pk + 1$ variables in total), the visualization is slightly more complex. A vector $\mathbf{b}$ of regression coefficients or of principal component loadings in form space consists of pk coefficients or loadings for the Procrustes shape coordinates and one for ln CS. Let $\mathbf{b}_{1\ldots pk}$ denote the elements of $\mathbf{b}$ corresponding to the shape coordinates and $\mathbf{b}_{pk+1}$ the element corresponding to ln CS. The visualization of $\mathbf{b}$ must be based on a reference form, which usually is the product $\overline{\mathbf{s}_i} e^{\overline{\ln \rho_i}}$, that is, the mean shape $\overline{\mathbf{s}_i}$ scaled by the mean size $\overline{\ln \rho_i}$, where ρ_i is the centroid size of the ith individual. Note that when size is measured by ln CS, the average shape must be scaled by the exponential function of the average ln CS, which is equal to the geometric mean of CS. The target of the deformation is $(\overline{\mathbf{s}_i} + f\mathbf{b}_{1\ldots pk}) e^{\overline{\ln \rho_i} + f\mathbf{b}_{pk+1}}$, i.e., the shape $\overline{\mathbf{s}_i} + f\mathbf{b}_{1\ldots pk}$ scaled by $\overline{\ln \rho_i} + f\mathbf{b}_{pk+1}$. For the visualization of form differences the absolute size usually does not matter, just the relative size difference associated with some shape difference is presented. The reference and the target forms can thus be rescaled by the same size factor and the terms above can be simplified to $\overline{\mathbf{s}_i}$ for the reference form and to $(\overline{\mathbf{s}_i} + f\mathbf{b}_{1\ldots pk}) e^{f\mathbf{b}_{pk+1}}$ for the target form.

In order to visualize allometry as a shape change together with the corresponding size change, one can add the vector of slopes resulting from the regression of the shape coordinates on ln CS to a reference form, and increase the scale of this reference form (measured as ln CS) by one unit. This is equivalent to regressing form (shape coordinates plus ln CS) on ln CS and to follow the procedure outlined above (the slope for ln CS would be 1). The same regression of form on ln CS (i.e., the regression slopes of shape on ln CS plus one element containing 1) can be used to project the vector of allometry (vector of regression slopes) into a PCA plot of form space (e.g., such as in Fig. 6).

References

Adams D.C., Collyer M.L., 2009. A general framework for the analysis of phenotypic trajectories in evolutionary studies. Evolution 63: 1143–1154.

Adams D.C., Rohlf F.J., Slice D.E., 2004. Geometric morphometrics: Ten years of progress following the "revolution". Italian Journal of Zoology 71: 5–16.

Alberch P., Gould S.J., Oster G.F., Wake D.B., 1979. Size and shape in ontogeny and phylogeny. Paleobiology 5: 296–317.

Albert A.M., Ricanek K. Jr., Patterson E., 2007. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. Forensic Sci. Int. 172(1): 1–9.

Bookstein F., 1991. Morphometric tools for landmark data: geometry and biology. Cambridge University Press, Cambridge (UK); New York.

Bookstein F., 1996. Biometrics, biomathematics and the morphometric synthesis. Bull. Math. Biol. 58(2): 313–365.

Bookstein F., 1997. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Med. Image Anal. 1(3): 225–243.

Bookstein F.L., (in press). Reasoning and Measuring: Numerical Inference in the Sciences. Cambridge University Press, Cambridge (UK).

Bookstein F., Chernoff B., Elder R.L., Humphries J.M., Smith G.R., Strauss R.E., 1985. Morphometrics in Evolutionary Biology: The geometry of size and shape change, with examples from fishes. Academy of Sciences of Philadelphia Special Publication 15.

Bookstein F., Schäfer K., Prossinger H., Seidler H., Fieder M., Stringer C., Weber G.W., Arsuaga J.L., Slice D.E., Rohlf F.J., Recheis W., Mariam A.J., Marcus L.F., 1999. Comparing frontal cranial profiles in archaic and modern homo by morphometric analysis. Anat Rec 257(6): 217–224.

Bulygina E., Mitteroecker P., Aiello L.C., 2006. Ontogeny of facial dimorphism and patterns of individual development within one human population. Am. J. Phys. Anthropol. 131(3): 432–443.

Cheverud J.M., 1982. Relationships among ontogenetic, static, and evolutionary allometry. Am. J. Phys. Anthropol. 59: 139–149.

Collyer M.L., Adams D.C., 2013. Phenotypic trajectory analysis: comparison of shape change patterns in evolution and ecology. Hystrix 24(1) (online first) doi:l0.4404/hystrix-24.l-6298

Coquerelle M., Bookstein F.L., Braga J., Halazonetis D.J., Weber G.W., Mitteroecker P., 2011. Sexual dimorphism of the human mandible and its association with dental development. Am. J. Phys. Anthropol. 145: 192–202

Courtiol A., Raymond M., Godelle B., Ferdy J.-B., 2010. Mate choice and human stature: Homogamy as a unified framework for understanding mating preferences. Evolution 64(8): 2189–2203.

Dryden I.L., Mardia K.V., 1998. Statistical Shape Analysis. John Wiley and Sons, New York.

Fink B., Grammer K., Mitteroecker P., Gunz P., Schaefer K., Bookstein F.L., Manning J.T., 2005. Second to fourth digit ratio and face shape. Proc. R. Soc. Lond. B 272(1576): 1995–2001.

Gerber S., 2011. Comparing the differential filling of morphospace and allometric space through time: the morphological and developmental dynamics of Early Jurassic ammonoids. Paleobiology 37: 369–382.

Gerber S., Neige P., Eble G.J., 2007. Combining ontogenetic and evolutionary scales of morphological disparity: a study of early Jurassic ammonites Evolution & Development 9(5): 472–482.

Geraedts E.J., van Dommelen P., Caliebe J., Visser R., Ranke M.B., van Buuren S., Wit J.M., Oostdijk W., 2011. Association between head circumference and body size. Horm. Res. Paediatr. 75(3): 213–219.

Good P., 2000. Permutation tests: a practical guide to resampling methods for testing hypotheses (2nd ed.) Springer Series in Statistics. Springer, New York.

Goodall C., 1991. Procrustes methods in statistical analysis of shape (with discussion and rejoinder). J. Roy. Stat. Soc. 53: 285–339.

Gould S.J., 1966. Allometry and size in ontogeny and evolution. Biological Reviews 41: 587–640.

Gould S.J., 1977. Ontogeny and Phylogeny. Harvard University Press, Cambridge.

Gower J., 1975. Generalized procrustes analysis. Pyschometrika 40: 33–51.

Gunz P., 2012. Evolutionary relationships among robust and gracile australopiths: an "evo-devo" perspective. Evolutionary Biology 39: 472–487.

Gunz P., Mitteroecker P., 2013. Semilandmarks: a method for quantifying curves and surfaces. Hystrix 24(1) (online first) doi:l0.4404/hystrix-24.l-6292

Gunz P., Mitteroecker P., Bookstein F.L., 2005. Semilandmarks in three dimensions. In: Slice D.E. (Ed.) Modern Morphometrics in Physical Anthropology. Kluwer Press, New York. 73–98.

Huttegger S., Mitteroecker P., 2011. Invariance and Meaningfulness in Phenotype Spaces. Evolutionary Biology 38: 335–352.

Jolicoeur P., 1963. The multivariate generalization of the allometry equation. Biometrics 19: 497–499.

Kendall D., 1984. Shape manifolds, Procrustean metrics and complex projective spaces. Bull. London Math. Soc. 16(2): 81–121.

Klingenberg C.P., 1998. Heterochrony and allometry: the analysis of evolutionary change in ontogeny. Biological Reviews 73: 70–123.

Klingenberg C.P., 2010. Evolution and development of shape: integrating quantitative approaches. Nature Reviews Genetics 11: 623–635.

Lande R., 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33: 402–416.

Little A.C., Jones B.C., DeBruine L.M., 2011. Facial attractiveness: evolutionary based research. Philos. Trans. R. Soc. B-Biol. Sci. 366(1571): 1638–1659.

Marcus L.F., 1990. Traditional morphometrics. In: Rohlf F.J., Bookstein F.L. (Eds.). Proc. Michigan Morphometrics Workshop. Univ. Michigan Museums, Ann Arbor, Michigan. 77–122.

Marcus L.F., Hingst-Zaher E., Zaher H., 2000. Application of landmark morphometrics to skulls representing the orders of living mammals. Hystrix 11(1): 27–47. doi:10.4404/hystrix-11.1-4135

Mardia K.V., Bookstein F., Moreton I., 2000. Statistical assessment of bilateral symmetry of shapes. Biometrika 87: 285–300.

Marroig G., Cheverud J.M., 2005. Size as a line of least evolutionary resistance: diet and adaptive morphological radiation in New World monkeys. Evolution 59(5): 1128–1142.

McCloskey D.M., Ziliak S.Z., 2009. The Unreasonable Ineffectiveness of Fisherian "Tests" in Biology, and Especially in Medicine. Biological Theory 4(1): 44–53.

Meindl K., Windhager S., Wallner B., Schaefer K., 2012. Second-to-fourth digit ratio and facial shape in boys: the lower the digit ratio, the more robust the face. Proc. R. Soc. B-Biol. Sci. 279: 2457–2463.

Mitteroecker P., 2009. The Developmental Basis of Variational Modularity: Insights from Quantitative Genetics, Morphometrics, and Developmental Biology. Evolutionary Biology 36: 377–385.

Mitteroecker P., Bookstein F.L., 2011. Classification, linear discrimination, and the visualization of selection gradients in modern morphometrics. Evolutionary Biology. 38: 100–114.

Mitteroecker P., Gunz P., 2009. Advances in geometric morphometrics. Evolutionary Biology 36: 235–247.

Mitteroecker P., Gunz P., Bernhard M., Schaefer K., Bookstein F., 2004. Comparison of cranial ontogenetic trajectories among great apes and humans. Journal of Human Evolution 46: 679–697.

Mitteroecker P., Gunz P., Weber G.W., Bookstein F.L., 2004b. Regional dissociated heterochrony in multivariate analysis. Annals of Anatomy 186(5–6): 463–470.

Mitteroecker P., Gunz P., Bookstein F.L., 2005. Heterochrony and geometric morphometrics: A comparison of cranial growth in *Pan paniscus* versus *Pan troglodytes*. Evolution & Development 7(3): 244–258.

Mitteroecker P., Huttegger S., 2009. The Concept of Morphospaces in Evolutionary and Developmental Biology: Mathematics and Metaphors. Biological Theory 4(1): 54–67.

Monteiro L.R., 1999. Multivariate Regression Models and Geometric Morphometrics: The Search for Causal Factors in the Analysis of Shape. Systematic Biology 48: 192–199.

Mosimann J.E., 1970. Size allometry: Size and shape variables with characterizations of the lognormal and generalized gamma distributions. Journal of the American Statistical Association 65: 930–945.

O'Higgins P., Chadfield P., Jones N., 2001. Facial growth and the ontogeny of morphological variation within and between the primates *Cebus apella* and *Cercocebus torquatus*. J. Zool., Lond. 254: 337–357.

Pawlowski B., Dunbar R.I.M., Lipowicz A., 2000. Evolutionary fitness: Tall men have more reproductive success. Nature 403(6766): 156.

Piras P., Salvi D., Ferrar S., Maiorino L., Delfino M., Pedde L., Kotsakis T., 2011. The role of post-natal ontogeny in the evolution of phenotypic diversity in *Podarcis* lizards. Journal of Evolutionary Biology 24(12) :2705–2720.

Ponce de Leon M.S., Zollikofer C.P., 2001. Neanderthal cranial ontogeny and its implications for late hominid diversity. Nature 412(6846): 534–538.

Rohlf F.J., 1999. Shape Statistics: Procrustes Superimpositions and Tangent Spaces. J. Classification 16: 197–223.

Rohlf F. J., Marcus L. F. 1993. A Revolution in Morphometrics. Trends Ecol. Evol. 8(4): 129–132.

Rohlf F.J., Slice D.E., 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39: 40–59.

Schaefer K., Mitteroecker P., Gunz P., Bernhard M., Bookstein F.L., 2004. Craniofacial Dimorphism Patterns and Allometry among Hominoids. Annals of Anatomy 186(5–6): 471–478.

Schaefer K., Mitteroecker P., Fink B., Bookstein F.L., 2009. Psychomorphospace – from biology to perception, and back: Towards an integrated quantification of facial form variation. Biological Theory 4(1): 98–106.

Schneider T.M., Hecht H., Carbon C.-C., 2012. Judging body weight from faces: The height – weight illusion. Perception 41(1): 121–124.

Sheets H.D., Zelditch M.L., 2013. Studying ontogenetic trajectories using resampling methods and landmark data. Hystrix 24(1) (online first) doi:10.4404/hystrix-24.1-6332

Slice D.E., 2007. Geometric Morphometrics. Ann. Rev. Anthropol. 36: 261–281.

Slice D.E., 2001. Landmark coordinates aligned by Procrustes analysis do not lie in Kendall's shape space. Systematic Biology 50(1): 141–149.

Slice D.E., 1996. Three-dimensional, generalized resistant fitting and the comparison of least-squares and resistant-fit residuals. In: Marcus L.F., Corti M., Loy A., Nayler G.J.P., Slice D.E. (Eds.). Advances in Morphometrics. Plenum Press, New York.

Small C., 1996. The statistical theory of shape. Springer, New York.

Stadler B.M.R., Stadler P.F., Wagner G., Fontana W., 2001. The topology of the possible: Formal spaces underlying patterns of evolutionary change. Journal of Theoretical Biology 213: 241–274.

Theobald D.L., Wuttke D.S., 2006. Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian Procrustes problem. Proc. Natl. Acad. Sci. U.S.A. 103: 18521–18527.

Windhager S., Schaefer K., Fink B., 2011. Geometric morphometrics of male facial shape in relation to physical strength and perceived attractiveness, dominance, and masculinity. Am. J. Hum. Biol. 23(6): 805–814.

Windhager S., Slice D.E., Schaefer K., Oberzaucher E., Thorstensen T., Grammer K., 2008. Face to face – The perception of automotive designs. Hum. Nat. – Interdiscip. Biosoc. Perspect. 19(4): 331–346.

Zebrowitz L.A., Montepare J.M., 2008. Social psychological face perception: why appearance matters. Social and Personality Psychology Compass 2(3): 1497–1517.

Zelditch M.L., Sheets H.D., Fink W.L., 2003. The ontogenetic dynamics of shape disparity. Paleobiology 29: 139–156.

Zelditch M.L., Swiderski D.L., Sheets D.S., Fink W.L., 2004. Geometric Morphometrics for Biologists. Elsevier Academic Press, San Diego.

Zollikofer C.P., Ponce de Leon M.S., 2002. Visualizing patterns of craniofacial shape variation in *Homo sapiens*. Proc. Biol. Sci. 269(1493): 801–807.

Associate Editor: A. Loy

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 67–73, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/6332/pdf
doi:10.4404/hystrix-24.1-6332

Research Article

Studying ontogenetic trajectories using resampling methods and landmark data

H. David Sheets[a,*], Miriam L. Zelditch[b]

[a] *Dept. of Physics, Canisius College, 2001 Main St, Buffalo, NY 14208, USA*
Dept. of Geology, SUNY at Buffalo, 411 Cooke Hall, Buffalo, NY 14260, USA
[b] *Museum of Paleontology, University of Michigan, Ann Arbor, Michigan 48109, USA*

Keywords:
ontogeny
shape
permutation
bootstrapping
resampling
MANCOVA

Article history:
Received: 20 June 2012
Accepted: 19 September 2012

Acknowledgements
We would like to thank Anna Loy, Philipp Mitteroecker and Andrea Cardini for helpful reviews of this manuscript.

Abstract

Comparative studies of ontogenies play a crucial role in the understanding of the processes of morphological diversification. These studies have benefited from the appearance of new mathematical and statistical tools, including geometric morphometrics, resampling statistics and general linear models. This paper presents an overview of how resampling methods may be applied to linear models of ontogenetic trajectories of landmark-based geometric morphometric data, to extract information about ontogeny. That information can be used to test hypotheses about the changes (or differences) in rate, direction, duration and starting point of ontogenetic trajectories that led to the observed patterns of morphological diversification.

Introduction

A central goal of evolutionary morphology is to explain the origin of morphological diversity. That diversity is now often termed "disparity" to distinguish it from the proliferation of species, a distinction that is important because the proliferation of species may not explain the proliferation of novel morphologies (e.g., Foote 1993; Adams et al. 2009). The proximate cause of disparity is evolutionary change in ontogeny; consequently, to understand the processes generating disparity we need to understand how ontogenies evolve (Zelditch et al., 2003; Adams and Nistri, 2010; Drake, 2011; Frederich and Vandewalle, 2011; Gerber, 2011; Ivanovic et al., 2011; Piras et al., 2011). Comparative studies of ontogenies play a critical role in such studies of disparity because they can discern which modifications of ontogeny are responsible for disparity, including the modifications that increase disparity, those that decrease it, and those that maintain it. Disparity itself often has an ontogeny because species may closely resemble each other during early states of morphogenesis, diverging thereafter or they may differ substantially early in development then come to resemble each other. Both these patterns have been detected empirically. For example, the disparity of both body shape and diet increase over ontogeny of some damselfishes (Frederich and Vandewalle, 2011) but body shape and diet show contrasting patterns in piranhas, with body shape disparity decreasing over ontogeny as the disparity of diet increases (Zelditch et al., 2003). Such decreases in disparity over ontogeny have been found in other groups as well, including body shape of crested newts (Ivanovic et al., 2011), and feet and interdigital webbing of European cave salamanders' feet (Adams and Nistri, 2010).

By combining comparative studies of ontogeny with analyses of disparity at two or more developmental stages, it is possible to test hypotheses about the developmental origins of disparity. When there are several modifications of ontogeny, we can ask how much each one, taken separately, would contribute to disparity of juveniles and adults and then how two or more interact with each other. The aim of this paper is to present methods for characterizing the modification of ontogeny, as revealed by morphometric data, and then methods for modeling the evolution of ontogeny to discern the impact of those modifications on disparity at two or more developmental stages.

Shape change: Ontogenetic change as a phenotypic trajectory

This paper focuses on ontogeny, but the analysis of ontogenetic change is a special case of a phenotypic trajectory. A phenotypic trajectory refers to the extent, direction and rate of any shape change in response to some factor, which could be geographic (e.g., latitude) or ecological (e.g., predation pressure). The concepts and methods discussed in context of phenotypic trajectories are thus readily adaptable to any context (see Adams and Collyer 2009, and in this volume). A variety of statistical tools can reveal how phenotypic trajectories differ; our objective in this paper is to discuss regression models and resampling techniques as applied to ontogenetic trajectories. We begin by applying resampling methods to a simple bivariate regression model, extending that to a multivariate regression for a single group and then to multiple groups by multivariate analysis of covariance (MANCOVA). Once MANCOVA establishes the statistical significance of differences in ontogenetic trajectories, a series of more specialized tests are used to characterize those differences.

Ontogenetic trajectories are complex to model because they comprise two distinct categories of variables: shape variables, and size and age variables. Whereas size and age are typically univariate measurements (scalar values), shape is multivariate, comprising multiple variables such as the coordinates of many landmarks. In this paper, we will typically discuss shape's dependence on size, but the same models extend to age-based analyses. The samples of shape and size may be taken continuously throughout ontogeny or at two or more times. In the first case, shape's dependence on size is modeled by a single line; in the second case its dependence may be modeled by several line seg-

*Corresponding author
Email address: sheets@canisius.edu (H. David Sheets)

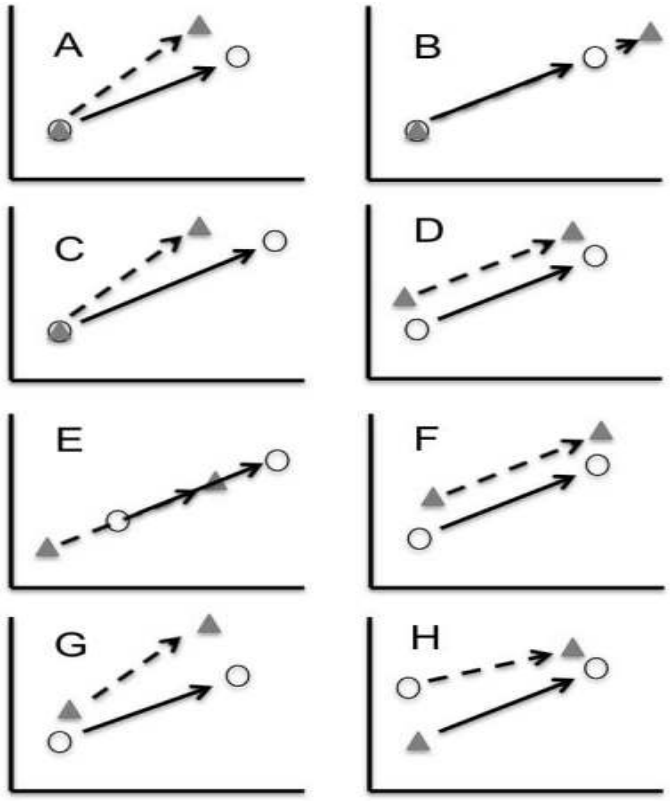


Figure 1 – Simple representation of a pairwise comparison of growth and development of two groups in a two dimensional morphospace. The circles and triangles represent the shapes of the two groups at the outset and end of the developmental stage, the arrows represent the ontogenetic trajectory. (A) Difference in the direction of ontogenetic trajectory, but equal magnitude of net change and a common shape at outset. (B) Overlapping trajectories with a common starting shape and direction, but unequal magnitude of net ontogenetic change. (C) Trajectories with differing directions and magnitudes. (D) Parallel trajectories with equal magnitudes of net change, with an elevation change in the outset shapes. (E) Overlapping parallel trajectories with a shift in outset shape along the direction of the trajectory and unequal magnitudes of change. (F) Parallel trajectories with both elevation changes and a shift along the trajectory of the shape at outset. (G) Equal length trajectories, showing divergence in shape over ontogeny. (H) Equal length trajectories showing convergence over ontogeny.

ments. This paper focuses on the analysis of a single linear segment because multiple segments may be analyzed by repeated application of the single segment models. The regression models used to analyze shape data may be readily adapted to discrete samples using dummy coding techniques. A linear model over a continuous covariate (age or size) suffices for both cases.

Fig. 1 shows a range of possible differences in growth and development of two organisms between two stages of development. The *ontogenetic trajectory* is the description of the change in shape from one stage to the next (Fig. 1). The trajectory itself may thus differ in direction, in length (magnitude) or in both. In addition to alterations in the trajectory, the *shape at the outset* of the trajectories may differ, due to alterations in development prior to the earliest stage. Differences in shape at this outset stage may be divided into *elevation* changes (Fig. 1D) and *shifts* in shape along an ontogenetic trajectory (Fig. 1E). This division of shape changes into two distinct categories at the outset becomes important when the trajectories share a common direction. Elevation changes (i.e., perpendicular to the common direction of the trajectory) produce parallel trajectories (Fig. 1D), while shifts along (i.e., parallel to) the trajectory produce overlapping trajectories (Fig. 1E), such that one group's juvenile resembles an older or younger age-class of the other group. In that case, their trajectories overlap but the age-classes differ in shape because of a shift in starting point along (parallel to) the trajectory. That can arise from differences in rate and/or duration of development prior to the youngest observed phase. A series of resampling tests based on linear models, and flowcharts illustrating a procedure for their use will be presented to address each of the possible differences in both the ontogenetic trajectory itself (direction and magnitude), differences in outset shape (elevation and shifts along a common trajectory) and duration of growth along the trajectory.

Resampling methods for testing a bivariate regression model

Classical parametric statistical methods use mathematical models of statistical distributions to calculate the distributions of test statistics. When the observed value is extreme relative to the distribution implied by the null hypothesis, it is possible to reject that null hypothesis at some calculated probability level. One alternative is to use Monte Carlo methods in which statistical models are fit to the data and then used as the bases for numerical simulations, which then can be used to determine the probability that the observed data was produced by a given null model (see Manly 1997 for examples). Resampling methods offer another approach. Resampling methods refer to a group of related methods (permutations, bootstrapping and jackknifing) that, not surprisingly, use resampling of the data to generate the distribution of a test statistic under the null hypothesis; the idea dates to Neyman (1923); Fisher (1935), and Pitman (1937) but practical application had to wait until inexpensive (and fast) computers became widely available (Efron, 1979; Efron and Tibshirani, 1998; Good, 2000, 2005).

The general process of constructing a resampling test consists of three distinct parts: (1) stating the null hypothesis, (2) determining what test statistic to use, and (3) deciding how to carry out the resampling. As a starting point, consider an ordinary bivariate linear regression model of the size of one trait (Y) relative to the size of another (X):

$$Y = MX + B + \varepsilon \quad (1)$$

M is the slope, B is the intercept and ε is a random error term (the residuals). Analytic approaches typically assume ε is an independently distributed random normal term with a mean of zero. Based on these assumptions, analytic statistics test the null hypothesis that there is no dependence of Y on X using test statistics such as the correlation coefficient R, or the estimate of the slope (M); the null is rejected if R is statistically significantly different from zero, or the confidence interval of M excludes zero. The value of R^2 also expresses the fraction of the variation in Y explained by X (which could, in principle, be used to test the null hypothesis). These analytic tests require algebraic models of the underlying distributions.

Resampling uses numerical randomization procedures to conduct the statistical tests. As noted, above, the first step is to state the hypothesis we want to test. It might seem that we have two, the first being that the correlation between Y and X is zero and the other that the slope is zero, but these are equivalent in that they assert that the model:

$$Y = B + \varepsilon \quad (2)$$

is equally effective at predicting Y as the original model which included X. The second model is called the "reduced model" because it omits a term (X) present in the full model. Whether we choose R, R^2 or M as our test statistic, the null model states that the *reduced* model will often produce a value of that statistic as extreme (i.e., as far from zero) as the *full* model (Eqn. 1). Because the omitted term is X, under the null hypothesis, the relationship between X and Y is not important. An important concept in the theory of resampling is that this makes X *exchangeable* under the null hypothesis (see discussions in Good 2000; Anderson 2001). That means that we could exchange the X value of any given specimen with that of any other specimen because, under the null hypothesis, this relationship is not important.

If X is exchangeable, the null model predicts that a randomly created version of the original data, in which the association between the X and Y values is randomized, will yield a similar distribution of values for any of our test statistics when the (full) regression model (Eqn. 1) is fit to the randomized data. Thus, if we create many such resampled versions of the data, we could generate a distribution of values of the test statistic under the null model. We can then use this distribution to determine the percentage of trials in which the observed value for that test statistic is as far from zero as the observed one. This is what we use

as our estimate of the p value. For example, if we are using R as our test statistic, and our observed value is 0.85 and a value this high never appeared in 999 trials, we would get a $p = \frac{1}{999+1} = 0.001$ or 0.1%. Note that we treat the original data as one possible resampling of the data; we resample 999 times and the 1000th value is the observed one – it therefore counts as a large value in the calculation of the p-value.

There are several different ways to resample the data that differ in two major respects. One is in how the data are selected and handled in the randomization, the other is in what exactly is permuted, the raw data X, or residuals (from either the full or reduced model). For hypothesis testing, permutation (or resampling without replacement) is thought to be the most effective approach, although permutation and bootstrap methods are thought to be asymptotically equivalent (Romano, 1989; Manly, 1997; Good, 2000). When permuting the data, the order of the X values is randomized and re-assigned to the Y values (see Manly 1997; Good 2000). When bootstrapping the data, which is resampling with replacement (Efron, 1979; Efron and Tibshirani, 1998), the individuals' values are randomly drawn, and each random draw is independent of all the others so a given individual's value may be used more than once or not at all. When jackknifing the data, a relatively small percentage of the specimens (from one specimen up to as many as 50% of the total) at a time are removed and the calculation is repeated. In addition to the distinction between the resampling procedure, methods differ according to what they permute (or bootstrap or jackknife). An alternative to permuting or bootstrapping the observed values (X) is to permute the residuals (ε) of the reduced model (Eqn. 2, which in this case implies that there is no slope, only a mean value and random variation around the mean). In this approach, the reduced model is fitted to the data and the residuals are computed and used in the permutation, as the null model implies that there is no ordering or relationship of the residuals relative to X or Y. Resampling residuals is thought to be more effective than simply permuting the observed (raw) variables (Anderson and ter Braak, 2003).

Example: Bivariate regression using different resampling methods

The following example is meant to show a range of different resampling methods applied to a simple bivariate regression model. If we start out with a set of measured values of the dependent and independent variables:

$$\begin{aligned} X =&[2.10\ 2.71\ 3.15\ 3.44\ 4.06\ 4.34\ 5.18\ 5.27 \\ &5.79\ 6.34\ 6.41\ 7.79] \end{aligned} \tag{3}$$

$$\begin{aligned} Y =&[6.54\ 7.69\ 9.10\ 9.15\ 10.86\ 11.47\ 13.46 \\ &13.67\ 14.60\ 16.00\ 16.14\ 19.40] \end{aligned} \tag{4}$$

fitting the full linear regression model, $Y = MX + B + \varepsilon$ we get estimates for the slope and intercept, and an R value

$$M = 2.2615 \qquad B = 1.6772 \qquad R = 0.9992 \tag{5}$$

We can then use the model to find the predicted values of y under the full model, and the residuals $\varepsilon = Y - Y_{predicted}$

$$\begin{aligned} Y_{predicted,full} =&[6.43\ 7.81\ 8.80\ 9.46\ 10.86\ 11.49\ 13.39 \\ &13.60\ 14.77\ 16.01\ 16.17\ 19.29] \end{aligned} \tag{6}$$

$$\begin{aligned} \varepsilon_{full} =&[0.11\ 0.12\ 0.30\ -0.31\ 0.00\ -0.02\ 0.07 \\ &0.08\ -0.17\ -0.02\ -0.03\ 0.11] \end{aligned} \tag{7}$$

For the reduced model (with no slope) $Y_{predicted,reduced} = B + \varepsilon$

$$Y_{predicted,reduced} = 12.34 \tag{8}$$

$$\begin{aligned} \varepsilon =&[-5.80\ -4.65\ -3.24\ -3.19\ -1.48\ -0.87 \\ &1.12\ 1.33\ 2.26\ 3.66\ 3.80\ 7.06] \end{aligned} \tag{9}$$

Based on these, we can see how the various types of resampled sets are formed. If we wanted to permute the variable itself, we would randomize the ordering of Y and regress this permutation set on X

$$\begin{aligned} Y_{perm,variables} =&[13.46\ 11.47\ 19.40\ 9.10\ 16.00\ 13.67 \\ &16.14\ 10.86\ 9.15\ 6.54\ 7.69\ 14.60] \end{aligned} \tag{10}$$

To permute residuals under the reduced model, we would permute them and add the permuted values to the predicted Y value ($Y_{predicted} = B$) under the reduced model

$$\begin{aligned} Y_{perm,residuals} =& [1.33\ 2.26\ 7.06\ -4.65\ 3.66\ -3.24 \\ &1.12\ -3.19\ -5.80\ 3.80\ -1.48\ -0.87] \\ &+ 12.34 \\ &= [13.67\ 14.60\ 19.40\ 7.69\ 16.00\ 9.10 \\ &13.46\ 9.15\ 6.54\ 16.14\ 10.86\ 11.47] \end{aligned} \tag{11}$$

A bootstrapping of the variables themselves might result in

$$\begin{aligned} Y_{bootstrap,variables} =&[16.00\ 9.10\ 19.40\ 10.86\ 9.10\ 9.15 \\ &13.67\ 11.47\ 10.86\ 16.00\ 13.67\ 13.46] \end{aligned} \tag{12}$$

Notice that in $Y_{bootstrap,variables}$, several values (16.00, 9.10, 10.86, 13.67) appear several times, while other values are omitted altogether. In contrast, in $Y_{permutation}$, residuals, each value in the original data set appears once and only once. If we bootstrap the residuals of the reduced model and add them to $Y_{predicted}$, we get

$$\begin{aligned} Y_{bootstrap,residuals} =& [7.06\ -3.19\ 3.66\ 3.66\ -1.48\ 1.12 \\ &-5.80\ -5.80\ 1.12\ 3.66\ 7.06\ -4.65] \\ &+ 12.34 \\ &= [19.40\ 9.15\ 16.00\ 16.00\ 10.86 \\ &13.46\ 6.54\ 6.54\ 13.46\ 16.00\ 19.40\ 7.69] \end{aligned} \tag{13}$$

and we can again see that some residual values (e.g., l7.06, 3.66, 1.12, -5.80) appear two or three times in the bootstrap set.

The original correlation was high, R = 0.992, and perhaps not surprisingly, in 9999 trials of each permutation and bootstrap, no resampled sets ever produced an R greater than or equal to 0.992. Since one of the 10000 (i.e., the original data) *did* have an R that large, the estimated p value is $\frac{1}{10000}$, or p = 0.0001. There are 12! (just over 479 million) possible permutations of the 12 values or residuals for this example, so 10000 trials does not come close to exhausting all possible combinations. Some authors urge using 1000 to 2000 trials in permutation tests (e.g., Manly 1997), but an alternative is to start with a relatively small number of trials and work upwards. If your p-value is 0.40 for 100 trials, running more trials to determine that p = 0.3874 is not productive. However, when the p-value is low relative to the desired α value, it is advisable to run several repetitions at 1000 to 2000 trials to see if that estimate is stable and reliable, increasing the number of trials until a stable estimate of p is obtained. Because resampling methods involve random processes, variance in the exact p-value obtained is expected. It may be necessary to run a relatively large number of trials to ensure that the variability in estimates of p are well below the desired α level. While the variation in p-value may seem worrisome and less precise than those from analytic tests, it is important to remember that analytic estimates of p-values are influenced by violations of the assumptions in the underlying analytic models, as well as by non-random sampling. Their apparent high precision may sometimes be illusionary.

Note that bootstrap and permutation methods assume that residuals are independently distributed, and that the data comprise a representative sample of the underlying population. These methods are not assumption free because they share some basic assumptions common to most statistical approaches.

Multivariate regression model

A wide range of methods have been used to capture information about the shapes of organisms. We focus here on the formalism of landmark-based geometric morphometrics (Bookstein, 1991; Dryden and Mardia, 1998; Adams et al., 2004; Zelditch et al., 2012), in which specimens are represented by a set of k landmarks measured in m dimensions; for 2D data, $m = 2$, and for 3D data, $m = 3$. One major advantage of landmark-based geometric morphometrics is the availability of a well characterized and robust distance metric, a univariate measure of the differences in shapes called a Procrustes distance.

We will assume that the data are superimposed by a Generalized Procrustes Analysis (GPA) so four degrees of freedom are used up when superimposing 2D landmarks and seven are used up when superimposing 3D landmarks. It is possible to include semilandmarks, i.e., points spaced along a curve or outline; these differ from landmarks in that they have only one degree of freedom per semilandmark for 2D data when semilandmark alignment procedures ("sliding") are used with semilandmark data (Sampson et al., 1996; Bookstein, 1997; Zelditch et al., 2012). This distinction becomes important when we consider estimating variance-covariance matrices later in this section.

In the case of shape data, the dependent variable Y is a vector quantity, denoted by $\mathbf{Y}$, each specimen is a row vector of k measurements per specimen. Our independent variable X is still a scalar, but the slope is now also a vector $\mathbf{M}$, as is the intercept $\mathbf{B}$. The error (residual) term $\mathbf{E}$ now also consists of a row vector of k values for each of the N specimens in the data set. This gives us a full model

$$\mathbf{Y} = \mathbf{M}X + \mathbf{B} + \mathbf{E} \tag{14}$$

and a reduced form

$$\mathbf{Y} = \mathbf{B} + \mathbf{E} \tag{15}$$

where $\mathbf{M}$ is the multivariate equivalent of the slope.

Like a univariate regression model, the significance of the full model may be estimated by a permutation test in which the residuals of the reduced model are permuted. The test statistic will be a version of an F-ratio because the F-ratio is the traditional statistic used in univariate analysis of variance (ANOVA and/or ANCOVA). There are many forms of F-ratios used in different experimental designs but all are ratios of sums of squares terms weighted by the degrees of freedom so they are akin to ratios of variances. In analytic models of multivariate data, the usual approach is to replace the sums of squares terms by a sum of squares and cross-products (SSCP) matrix. This is a major change relative to univariate data, not only because it requires a very large sample size to estimate the matrix reliably but because it requires a matrix inversion. For that inversion to be possible the variables must be linearly independent of one another and the degrees of freedom in the data must match the degrees of freedom in the measurements. But superimposition removes either four or seven degrees of freedom (and even more when semilandmark alignment is used) so the matrix is not of full rank. That is why partial warp scores (see Bookstein 1989) were typically used in multivariate statistical procedures that require the variance-covariance matrix to be invertible. Unfortunately, the matrix of partial warp scores is not of full rank when the data include semilandmarks. One approach is to reduce the dimensionality of the data using principal components, and to use the principal component scores in the analysis. That, however, does not solve the problem of estimating large variance-covariance matrices when the sample sizes are relatively small, and by "relative" here we mean relative to the number of landmarks plus semilandmarks.

Fortunately, there is another approach. We can work with *pseudo F-ratios* (Verdonschot and ter Braak, 1994; Legendre and Anderson, 1999), which are based on summed square distances of specimens about the mean rather than SSCP matrices. A distance metric, which for shape data is the Procrustes distance, is used to compute the sums of squares terms (SS), which are now scalars (simple distances) rather than SSCP matrices. For the simple regression model above, the F-ratio would be calculated as:

$$F = \frac{SS_{Model}/df_{Model}}{SS_{Residuals}/df_{Residuals}} \tag{16}$$

$$SS_{Model} = SS_{Total} - SS_{Residuals} \tag{17}$$

All sums of squares terms may be computed from a matrix of the pairwise distances between specimens (the outer product matrices, McArdle and Anderson 2001) because the sums of squares about the mean is proportional to the sums of squares between specimens. These methods were developed for a range of different types of statistical questions by Anderson and colleagues (Legendre and Anderson, 1999; McArdle and Anderson, 2001; Anderson and ter Braak, 2003) using a variety of different distance metrics, paralleling Goodall's (1991) derivation of the approach based specifically on Procrustes distance, adapted and generalized by Rohlf (2009) for permutation tests of regression models and MANCOVA designs. The availability of these pseudo F-tests (or Generalized Goodall's F-ratios) greatly speeds calculations in permutation tests, as well as having a number of other advantages (Anderson and ter Braak, 2003), plus the approach is readily adapted to a variety of experimental designs whether the data are univariate or multivariate. There are several different forms of Procrustes distances (see Dryden and Mardia 1998, or Zelditch et al. 2012), the form employed mostly commonly is properly called a *partial Procrustes distance*, in which the centroid size of all specimens is scaled to 1, and inference is carried out in the linear tangent space of the underlying curved space. The term "Procrustes distance" used hereafter refers to this partial Procrustes distance.

This set of ideas allows us to test the statistical significance of a linear regression model fitted to shape data; the pseudo F-ratio is determined for the full model and compared to the pseudo F-ratio derived from a large number of permutations of the residuals of the reduced model. From that we arrive at an estimated *p*-value for the pseudo-F statistic. This is precisely what we did when estimating the confidence interval for a test statistic in the univariate case.

Comparing ontogenies: Establishing evidence of differences in trajectories

To compare trajectories among two or more groups, the first step is to verify that there are statistically significant differences (of some kind) in the ontogenetic trajectories. Once that is established, we can go on to attempt to determine the nature of those differences. Simply computing the ontogenetic trajectories for each group and immediately doing pairwise comparisons of all the features of the trajectories rapidly leads to Bonferroni problems in the overall significance of the results (i.e., that there are statistically significant differences in the trajectories) because of the very large number of possible pairwise comparisons, which leads to an increased rate of false positive results if each test is done at the typical 5% alpha level. One can use a Bonferroni correction, carrying out each test at a lower alpha level to obtain overall results at the desired 5% alpha level, but another approach is a single overall test to establish overall significance at the desired alpha level. The first step is therefore a MANCOVA. To explain this, we introduce a factor A, the membership of each individual in a group, such as "species", which is a level in the factor A. In our full model $\mathbf{M}$ is now a function of A, as are the intercept terms:

$$\mathbf{Y} = \mathbf{M}(A)\mathbf{X} + \mathbf{B}(A) + \mathbf{E} \tag{18}$$

with the reduced model being:

$$\mathbf{Y} = \mathbf{M}\mathbf{X} + \mathbf{B}(A) + \mathbf{E} \tag{19}$$

This is sometimes called the common slopes model because it says that all groups share a common trajectory. The common slopes model itself has a reduced model with no slope at all:

$$\mathbf{Y} = \mathbf{B}(A) + \mathbf{E} \tag{20}$$

To test for statistically significant differences in the slopes $\mathbf{M}(A)$, we would form a pseudo F-ratio of the sums of squares (SS) explained by the model divided by the residual SS and then form permutation tests based on permutation of the residuals of the reduced model. Note that this permutation version of the F-ratio test does not assume equal variance at each landmark, nor does it assume that variances at each landmark are independent of one another. If this pseudo F-ratio is statistically significant at some desired level of confidence (based on an estimated distribution of pseudo F-ratio values obtained by a permutation or bootstrap test), it is then reasonable to proceed to a series of pairwise tests to understand the nature of the differences. Repeated use of the above procedure can determine which the levels of the group factor actually differed from one another. The tests discussed below can be used to determine the nature of those differences (rate of change, direction

of change or both). It is important to note that the common trajectory model is the null hypothesis in this procedure. The failure to reject the null may depend on sample size; the test is subject to some unknown rate of type II error, so accepting the null of common trajectory ($\mathbf{M}$) is not the same as a statistical proof that the trajectories are in fact common. Effect size in these systems (the magnitude of differences in slope, or in the variance explained) may also provide some insight, and should not be neglecting when examining the results of an F-test. Carefully structured tests based on geometric morphometric methods are thought to have high statistical power, and may detect statistically significant results when the differences in shape are too small to be of biological significance, particularly with relatively large sample sizes. Examination of the effect size involved in a comparison may thus be informative, both when the null is rejected, and when it is not.

Tests for differences in direction

Once a difference in the two vectors is known to exist, we need to determine if this is a difference in direction, in magnitude, or in both. Since we have two vectors, we can compute an angle between them,

$$\cos(\theta) = \frac{\mathbf{M_1} \cdot \mathbf{M_2}}{|\mathbf{M_1}||\mathbf{M_2}|} \tag{21}$$

where the numerator is the dot product of the two vectors and $|\mathbf{M}|$ is the magnitude of a vector $\mathbf{M}$. The angle between the two describes a difference in direction independent of the length (because the vectors are normalized to unit length), so the angle is a reasonable test statistic for differences in direction.

One null hypothesis addressing the need to determine if two ontogenetic trajectories are in the same direction may be stated as: *The observed angle between the trajectories is no larger than might be observed by randomly selecting two samples from within one of the groups*. That is, if we estimate the vector for each of two random samples drawn from a single population, and calculate the angle between them, the observed angle might not be large relative to the distribution of the angles under the null. To perform that test, we would calculate the vectors for each randomly drawn sample from a single group and compute the angle between them. A bootstrapping procedure can be used to estimate the range (or confidence interval) of angles that might appear within each sample (bootstrap methods allow for estimating confidence intervals from the range of variance within a sample via resampling). Bootstrapping is used here rather than permutation because there is no assumption of exchangeability here. Instead we are using the bootstrap to estimate the magnitude of variability in a derived measurement (the angle).

To compute the range of angles possible within a sample, the full regression model (eqn. 14) is fit to the data, and the residuals are calculated. Bootstrap samples are then produced by resampling the residuals with replacement to create two bootstrap sets, of the same sizes as the original data sets. To be conservative, both bootstrap samples of the smaller data set are limited to its sample size. The vectors are then calculated for each of the bootstrap sets and the angle between the bootstrap sets is determined. This is repeated for some large number of bootstrap sets for both groups to determine the confidence intervals of angles generated within each. The observed angle between the two groups may be judged statistically significant at the desired α if it lies beyond the $1 - \alpha$ confidence interval of both bootstrap sets. Failure to reject the null does not mean that the angle between the two trajectories actually *is* zero because the null result depends on the sample size and the unknown rate of type II error.

Tests based on distances between groups

Once the differences in ontogenetic trajectories have been identified, there are still some questions remaining about changes in ontogenies. One is whether the trajectories start at a common shape (Fig. 1B compared to 1E for example). Another is whether the adults are more, less or equally as different the juveniles are (Fig. 1G and 1H). Still another is whether one group undergoes a longer or shorter interval of net shape change than another (Fig. 1A compared to 1C, as one possibility). These questions can all be answered using differences in Procrustes distances between the means of groups, as shown in Fig. 2. Obtaining adequate estimates of the populations at these juvenile and adult stages is the difficult part of the process. The ideal situation is to have good collections of specimens at each stage, the second option is to attempt to estimate the mean shape and variation in the population at these stages based on the regression model and an estimate of the value of the independent variable X at each stage (see Frederich and Sheets 2009; Zelditch et al. 2012).

To test for differences in the mean shapes of two groups (at either juvenile or adult stage), the pseudo F-test based on the regression model discussed above may be used. In this pairwise test, the group membership of specimens is dummy coded as the independent variable X. Members of the first group are assigned an X value of $1/N_1$, members of the second group are assigned a value of -$1/N_2$. Other approaches to dummy coding are possible, this approach simply yields a mean value of zero for the dummy codes. Shape is then regressed on these dummy codes (X), and the permutation F-test is used to determine if the regression is statistically significant, indicating a difference in the mean juvenile shape. This procedure is equivalent to a permutation test based on Goodall's F, which has also been used to study ontogenies. The same test could be used to test for differences in adult shapes.

Confidence intervals for the Procrustes distance between means of two groups may be constructed by bootstrapping the residuals around the mean shape of each group. Both groups are resampled with replacement, and the distances between means recalculated for each bootstrap set. This allows for comparisons of shape differences from juvenile to adult shapes (i.e., the length of the ontogenetic vector), or from adult to adult ($\mathbf{A_{1-2}}$), or from juvenile to juvenile ($\mathbf{J_{1-2}}$), as shown in Fig. 1G, 1H and 2.

The net shape change during growth for two groups could be compared by comparing the distances from mean juvenile to mean adult shape for the two groups ($\delta = |\mathbf{A_{1-2}}| - |\mathbf{J_{1-2}}|$). Convergence on an adult form would imply the adult-to-adult distance for the two groups is smaller than the juvenile-to-juvenile distance (Fig. 1H). Conversely, divergence implies that the juveniles are more similar than the adults (Fig. 1G). It is possible to construct a bootstrap test of the observed differences in distances. For example, if we want to test the hypothesis that the distance from the mean of group A to the mean of group B (D_{AB}) is greater than the corresponding distance from groups C to E (D_{CE}), we could use the difference in distances, $\delta = D_{AB} - D_{CE}$ as our test statistic, measuring all distances in Procrustes units. We would then form a series of bootstrap sets of each of the sets A, B, C and E, bootstrapping within each, and then compute δ for each bootstrap set,

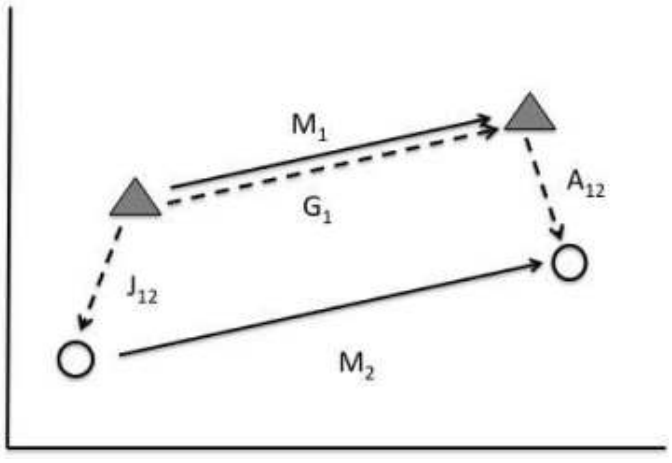


Figure 2 – Diagram of two ontogenetic trajectories with parallel ontogenetic trajectories lying along directions indicated by the multivariate regression slopes $\mathbf{M_1}$ and $\mathbf{M_2}$. The vectors $\mathbf{J_{12}}$ and $\mathbf{A_{12}}$ refer to the vector differences between the juveniles and adults of the two species.

generating a confidence interval on δ. If this interval excludes zero, then we can claim that δ is statistically significantly larger than zero.

Test of differences in elevation and shifts in starting position along the trajectory

A difference in elevation of two trajectories refers to a difference in juvenile shapes that is not along the ontogenetic trajectories, but rather perpendicular to it (Fig. 1D, 1F). If we have a common direction of the ontogenetic trajectory along the multivariate slope $\mathbf{M}$ and a difference vector between juvenile forms $\mathbf{J_{1-2}}$, then the elevation term would be the component of $\mathbf{J_{1-2}}$ perpendicular to $\mathbf{M}$ and the shift of the juvenile form along the trajectory would be the component of $\mathbf{J_{1-2}}$ parallel to $\mathbf{M}$ (Fig. 2). The parallel component is

$$\mathbf{J}_{parallel} = \frac{\mathbf{J_{1-2}} \cdot \mathbf{M}}{|\mathbf{M}|} \tag{22}$$

and

$$\mathbf{J}_{elevation} = \mathbf{J_{1-2}} - \mathbf{J}_{parallel} \tag{23}$$

If the $\mathbf{J}_{parallel}$ term is non-zero, then its dot product with $\mathbf{M}$ should always have the same sign, and should exclude zero, a hypothesis that can be tested via a bootstrap procedure. The magnitude of $\mathbf{J}_{elevation}$ could be tested computing the dot product of $\mathbf{J}_{elevation,bootstrap}$ derived from the bootstrapping procedure with the observed value of $\mathbf{J}_{elevation}$ to see if this dot product is also positive and excludes zero. Simple examination of the magnitudes (lengths) of these vectors would not necessarily be adequate, as distances are always positive. Random variation might generate small but non-zero values of these vectors, requiring the use of the dot product to detect random reversals in direction, which would not be consistent with a meaningful direction and magnitude of these components.

Overlapping trajectories

Overlapping trajectories would be a special case of parallel trajectories, but one in which the juvenile and/or adult shapes varied due to differences in rate along the trajectory or duration of shape change along the trajectory and/or shifts of the starting point along the trajectory (Fig. 1B, E). Overlapping trajectories would have a zero angle between them, and a non-significant difference in elevation, but might differ in magnitude of the ontogenetic trajectory and/or net shape change from juvenile to adult and/or exhibit shifts in juvenile shape along the trajectory.

Mitteroecker et al. (2005) looked for similar evidence of overlapping trajectories by fitting independent regression models to the specimens of each groups using the standard sum of squared residuals approach, but then used as a test criteria only the component of the residuals perpendicular to the predicted trajectory. This perpendicular component was then tested against a permutation of specimens among groups, testing the null hypothesis that group membership did not matter in predicting the perpendicular portion of the residuals, only the portion of the residuals along the trajectory as, specimens moving at different rates along the trajectory would be displaced parallel to the trajectory, not increasing the perpendicular error. If the null hypothesis is true, then the observed summed squared perpendicular errors would be consistent with the observed range of summed squared perpendicular errors generated by the permutation process.

Statistics derived from the parameters in regression models

In some situations it may be desirable to estimate derived statistics based on the results of a regression analysis. For example, if we want to compute the distance from the mean juvenile shape of one group to the mean juvenile form of a second group ($|\mathbf{J_{1-2}}|$), and use a bootstrap procedure to estimate a confidence interval on this distance, the ideal situation would be to have large sample of both groups of juveniles. In many situations, however, the researcher has a series of specimens sampled over wide range of sizes. As noted above, it is possible to estimate the predicted shape at a given juvenile size, and to use the residuals from the regression model to estimate variation around the mean juvenile shape. The distance between the means of two groups can then be estimated based on these bootstrap samples. In carrying out such an analysis, the within sample bootstrapping should be done on the residuals from the regression model, and that model should be refit to the predicted values to re-estimate the mean juvenile shape. That should be done at each iteration of the bootstrap to take the uncertainty of the regression into account.

Disparity

This approach can be used to estimate the uncertainty in derived statistic such as the disparity of a clade, which may be computed at both adult and juvenile states as was carried out in Zelditch et al. (2003). Disparity at any ontogenetic stage may be measured as:

$$Disparity = \frac{\sum_{i=1}^{m} d_i^2}{(m-1)} \tag{24}$$

where the sum is taken over all groups i out of m, and the distance d is the Procrustes distance from the mean of the i-th group to the mean of the group. Zelditch et al. (2003) examined disparity for juvenile and adult piranhas over several groups, using continuous ontogenetic series. To compare the juveniles and adults, the expected values for given sizes and the residuals were obtained from the regression model. Disparity was then calculated based on these predicted shapes, and confidence intervals for disparity were obtained by bootstrapping the residuals of the regression models, recalculating the regressions and re-estimating the disparity for the bootstrap sets. The bootstrapping procedure here incorporated the uncertainty in the regression model. The modeling procedure also allowed for creating hypothetical ontogenies, in which species were simulated to share common juvenile shapes, directions and/or rates, to see the impact of each parameter, singly and in combination, one the diversification of morphology. The simulations used bootstrapped residuals from the regression models employed to estimate the uncertainty in the simulated trajectories and resulting disparity values, in addition to the observed trajectories and disparities.

An approach to combining the tests

The individual tests presented here may be combined to determine the types of differences in ontogenetic trajectories present in a number of distinct groups of specimens. The sequence of tests shown in Fig. 3

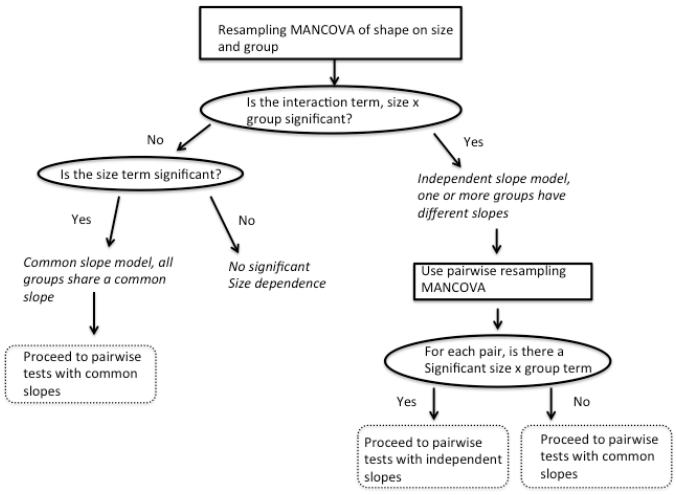


Figure 3 – Flowchart illustrating an approach to combining resampling tests to determine the types of differences in ontogenetic processes in two or more groups of specimens. Solid square boxes indicate a specific test, ellipse indicate decisions made based on the outcomes of the tests and the dashed boxes indicate what types of pairwise tests may be carried out to complete the analysis (Fig. 4 and 5).

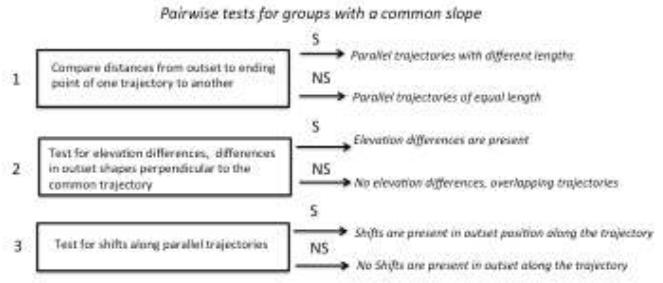


Figure 4 – The set of tests applicable to a pair of species with a shared multivariate slope, indicating parallel trajectories. The three tests may be carried out in any sequence, and need not all be used if they do not meet the goals of a study.

starts off with a MANCOVA of shape based on size and group, followed by a series of pairwise tests of the differences between groups (Fig. 4, 5). It is important to note that not all tests shown in these figures will be necessary for all analyses, the flowcharts are intended to be exhaustive in covering all possibilities. Some authors would chose to omit the pairwise MANCOVA step in favor of proceeding directly to the pairwise angle and trajectory length tests, on the grounds that the angle test adequately addresses the issue of pairwise differences in direction.

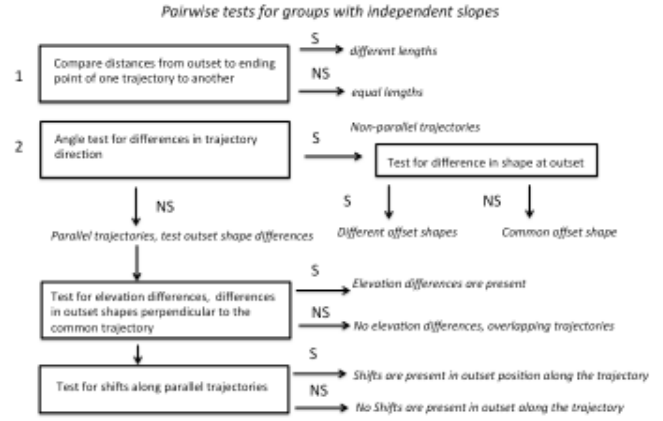


Figure 5 – The set of tests applicable to a pair of species with unequal multivariate slopes. The first and second test are independent of one another, but the result of the second test (differences in direction) does have a bearing on which of the remaining tests are applicable.

Conclusions

The combination of resampling methods, hypotheses based on general linear models and the well-established Procrustes distance as a measure of shape differences allows for a systematic and flexible approach to describing and testing how two or more ontogenetic trajectories differ. Most of the tests discussed here are available in specialized software such as the `tps` series (Rohlf, 2009) and `IMP` series (Sheets, 2001-2012; Zelditch et al., 2012). Customized bootstrapping and permutations methods are readily developed in R (see Good 2005), the `adonis` function in the `vegan` package (Oksanen et al., 2013) in R handles permutation MANCOVA, as does the `DistLM` program (Anderson, 2005). Several R scripts are either generally available or can be obtained by request from the authors that allow for testing many of these hypotheses about the evolution of ontogenies (Adams and Collyer, 2009; Gerber and Hopkins, 2011; Piras et al., 2011), making the approach outlined in this paper available to R users. The ongoing development of shape distance metrics, distance-based statistical tests, resampling methods and related software provide a steadily increasing set of analytic tools to examine morphological change in organisms, providing powerful methods for research in these areas.

References

Adams D.C., Collyer M.L., 2009. A general framework for the analysis of phenotypic trajectories in evolutionary studies. Evolution 63: 1143–1154.

Adams D.C., Nistri A., 2010. Ontogenetic convergence and evolution of foot morphology in European cave salamanders (Family: Plethodontidae). Bmc Evolutionary Biology 10: 216.

Adams D.C., Berns C.M., Kozak K.H., Wiens J.J., 2009. Are rates of species diversification correlated with rates of morphological evolution? Proceedings of the Royal Society B 276: 2729–2738.

Adams D.C., Rohlf F.J., Slice D.E., 2004. Geometric morphometrics: Ten years of progress following the "revolution". Italian Journal of Zoology 71: 5–16.

Anderson M.J., 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecology 26: 32–46.

Anderson M.J., 2005. DistLM (software). Available from http://www.stat.auckland.ac.nz/~mja/Programs.htm#Mine [5 September 2012]

Anderson M.J., ter Braak C.J.F., 2003. Permutation tests for multi-factorial analysis of variance. Journal of Statistical Computation and Simulation 73: 85–113.

Bookstein F.L., 1989. Principal warps: thing plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence. 11: 567–585.

Bookstein F.L., 1991. Morphometric tools for landmark data: Geometry and biology. Cambridge University Press, Cambridge.

Bookstein F.L., 1997. Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape. Medical Image Analysis 1: 97–118.

Drake A.G., 2011. Dispelling dog dogma: an investigation of heterochrony in dogs using 3D geometric morphometric analysis of skull shape. Evolution & Development 13: 204–213.

Dryden I.L., Mardia K.V., 1998. Statistical shape analysis. Wiley, Chichester.

Efron B., 1979. Computers and the theory of statistics: thinking the unthinkable. SIAM Review 21: 460–480.

Efron B., Tibshirani R.J., 1998. An Introduction to the Bootstrap. Chapman and Hall, Boca Raton.

Fisher R.A., 1935. The design of experiments. Oliver and Boyd, Edinburgh.

Foote M., 1993. Discordance and concordance between morphological and taxonomic diversity. Paleobiology 19: 185–204.

Frederich B., Sheets H.D., 2009. Evolution of ontogenetic allometry shaping giant species: a case study from the damselfish genus Dascyllus (Pomacentridae). Biological Journal of the Linnean Society 99: 99–117.

Frederich B., Vandewalle P., 2011. Bipartite life cycle of coral reef fishes promotes increasing shape disparity of the head skeleton during ontogeny: an example from damselfishes (Pomacentridae). Bmc Evolutionary Biology 11: 82.

Gerber S., 2011. Comparing the differential filling of morphospace and allometric space through time: the morphological and developmental dynamics of Early Jurassic ammonoids. Paleobiology 37: 369–382.

Gerber S., Hopkins M.J., 2011. Mosaic heterochrony and evolutionary modularity: The trilobite genus Zacanthopsis as a case study. Evolution 65: 3241–3252.

Good P.I., 2000. Permutation tests. Springer, New York.

Good P.I., 2005. Introduction to statistics through resampling methods and R/S-plus. Wiley, Hoboken.

Goodall C., 1991. Procrustes methods in the statistical analysis of shape. Journal of the Royal Statistical Society, Series B: Methodological 53: 285–339.

Ivanovic A., Cvijanovic M., Kalezic M.L., 2011. Ontogeny of body form and metamorphosis: insights from the crested newts. Journal of Zoology 283: 153–161.

Legendre P., Anderson M.J., 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecological Monographs 69: 1–-24.

McArdle B.H., Anderson M.J., 2001. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. Ecology 82: 290–297.

Manly B.F.J., 1997. Randomization, bootstrap and monte carlo methods in biology. Chapman and Hall, London.

Mitteroecker P., Gunz P., Bookstein F.L., 2005. Heterochrony and geometric morphometrics: a comparison of cranial growth in Pan paniscus versus Pan troglodytes. Evolution and Development 7: 244–258.

Neyman J., 1923. On the Application of Probability Theory to Agricultural Experiments. Statistical Science 5: 465–472

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Wagner H., (2013). `vegan`: Community Ecology Package. R package version 2.0-6. http://CRAN.R-project.org/package=vegan

Piras P., Salvi D., Ferrar S., Maiorino L., Delfino M., Pedde L., Kotsakis T., 2011. The role of post-natal ontogeny in the evolution of phenotypic diversity in Podarcis lizards. Journal of Evolutionary Biology 24: 2705–2720.

Pitman E.J.G,. 1937. Significance tests which may be applied to samples from any populations. Journal of the Royal Statistical Society B 4: 119–130.

Rohlf F.J., 2009. tpsRegress 1.37. (software). Ecology and evolution. State University of New York at Stony Brook. Available from http://life.bio.sunysb.edu/morph/ [5 September 2012]

Romano J.P., 1989. Bootstrap and randomization tests of some nonparametric hypotheses. The Annals of Statistics 17: 141–159.

Sampson P.D., Bookstein F.L., Sheehan H., Bolson E.L., 1996. Eigenshape analysis of left ventricular outlines from contrast ventriculograms. In Marcus L.F., Corti M., Loy A., Naylor G.J.P., Slice D.E. (Eds.) Advances in Morphometrics. NATO ASI Series A: Life Science, New York. pp. 131–152

Sheets H.D., 2001-2012. IMP software series. Available from http://canisius.edu/~sheets. [5 September 2012]

Verdonschot P.E.M., ter Braak C.J.F., 1994. An experimental manipulation of oligochaete communities in mesocosms treated with chlorpyrifos or nutrient additions: multivariate analyses with Monte Carlo permutation tests. Hydrobiologia 278: 251–266.

Zelditch M.L., Sheets H.D., Fink W.L., 2003. The ontogenetic dynamics of shape disparity. Paleobiology 29: 139–156.

Zelditch M.L., Swiderski D.L., Sheets H.D., 2012. Geometric morphometrics for biologists: a primer. Academic Press, London.

Associate Editor: A. Cardini

1994 – Jim Rohlf, Leslie Marcus, Marco Corti, Fred Bookstein, Gavin Naylor and Anna Loy editing the NATO ASI "Advances in Morphometrics" proceedings at the American Museum of Natural History (New York) (pictures by Anna Loy).

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 75–83, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/6298/pdf
doi:10.4404/hystrix-24.1-6298

Research Article

Phenotypic trajectory analysis: comparison of shape change patterns in evolution and ecology

Michael L. Collyer[a,*], Dean C. Adams[b]

[a] Department of Biology, Western Kentucky University, Bowling Green, KY 42101, USA
[b] Department of Ecology, Evolution, and Organismal Biology and Department of Statistics, Iowa State University, Ames, IA 50011, USA

Keywords:
phenotypic trajectory analysis
multi-state phenotypic change
Procrustes

Article history:
Received: 1 June 2012
Accepted: 18 January 2013

Acknowledgements
We thank A. Cardini and F.J. Rohlf for constructive comments on this manuscript. This work was supported in part by a Western Kentucky University Research and Creative Activities Program grant 12-8032 (to MLC) and an NSF grant DEB-1118884 (to DCA).

Abstract

Research using shape data from geometric morphometric (GM) methods in ecology and evolutionary biology is typically comparative, analyzing shapes and shape change over different points along ecological or evolutionary gradients. Whereas standard multivariate statistics procedures are fine for "static" variation – testing for location differences of groups in multivariate data spaces – they are limited for "dynamic" variation – testing specific differences in the ways groups change locations associated with changes in state along ecological, developmental or evolutionary gradients. In this paper, we show that continuous phenotypic change can be described by trajectories in multivariate data spaces. We describe the geometric attributes of phenotypic change trajectories (size, direction, and shape), specifically for GM data. We illustrate, with examples, how differences in such attributes can function as test statistics for comparative analyses in order to understand the mechanisms that produce dynamic differences in shape change. We demonstrate that analysis of such attributes – called phenotypic trajectory analysis (PTA) – is a general analysis that can be applied to various types of research questions concerned with measuring dynamic variation. Finally, we posit some challenges for the future for this novel analytical method.

Introduction

Most users of geometric morphometric (GM) methods are aware of the famous quote by D'Arcy Wentworth Thompson (1917), "In a very large part of morphology, our essential task lies in the comparison of related forms rather than in the precise definition of each". This preamble to Thompson's "Theory of Transformations" defines the essential task in comparative studies to be the description of morphological change between two forms rather than the mere descriptions of two separate forms. The fields of ecology and evolutionary biology often share the same objective; that is, description of a current ecological or evolutionary state is not as fascinating or important as knowing the mode and tempo (Simpson, 1944) of the change of the state. As reviewed by Adams et al. (2004, this issue), many advances in the field of geometric morphometrics in the last decade use the data from GM to identify patterns of phenotypic change in ecology and evolutionary biology. One recent advance – the result of an interesting convergence of shared principles in the fields of ecology, evolutionary biology, developmental biology, and geometric morphometrics – is phenotypic trajectory analysis, which is the description and comparison of geometric attributes of phenotypic change (Adams and Collyer, 2009). In this paper, we discuss a specific phenotype that is frequently of interest in ecology and evolutionary biology research – organismal shape – and describe how the principles on which GM is based can also be applied to the comparison of patterns of shape change in shape spaces.

On one hand, the analysis of shape change is not new. Multivariate tests for comparison of means, such as Hotelling's 1931 T^2 or multivariate analysis of variance (MANOVA) can be used to estimate the probability of shape change (difference) between two means, or the joint-change between multiple means, respectively, under a null hypothesis of no difference in means (Marcus, 1993). Multivariate tests of linear association (e.g., multivariate regression) estimate the probability associated with the amount of shape change per unit of change in a continuous independent variable, under a null hypothesis of no association (Rencher, 2002). Inherently, any statistical test for comparisons of shapes among groups or the linear associations of shape and other continuous variables is concerned with evaluating shape change relative to per unit changes in one or more independent variables. Standard multivariate test statistics (e.g., Wilks' Λ, Pillai's trace) evaluate the probability of effects – due to a single coefficient or a combination of coefficients – under a null hypothesis of no effect. For some analyses, this level of hypothesis testing is sufficient.

On the other hand, standard multivariate analyses alone are not sufficient for understanding how shape changes, or how these changes may be similar or different in distinct groups. That is, standard multivariate tests are not sufficient for understanding more precise reasons for rejecting a null hypothesis of consistency in shape change (Collyer and Adams, 2007). For example, one might analyse shape variation for four groups (e.g., species) of organisms exposed to four different treatments (e.g., different temperatures) in an experimental setting using a factorial MANOVA, including one factor for species, another factor for treatment-levels (temperature-exposure), plus an interaction of the two factors. If a significant species-treatment interaction is observed, what does this mean? For univariate response data, one might be able to make sense of the significant interaction by estimating and evaluating a response surface (Box and Draper, 1987). For multivariate response data, a significant interaction means that the coefficients for the interaction are obviously important to the factorial model, but a response surface is not possible nor will analysis of each coefficient, one by one, easily reveal how the four species differ in terms of their patterns of shape change across a temperature gradient. Nonetheless, the coefficients allow one to estimate the expected shapes that each group has at each temperature.

Another way to think about the problem is to visualize the shape patterns as trajectories in shape space (Fig. 1). Using the example described above, the four species are each represented by four points in shape space (and the corresponding tangent space; see Adams et al. this issue). The differences in positions of these four points de-

*Corresponding author
Email address: michael.collyer@wku.edu (Michael L. Collyer)

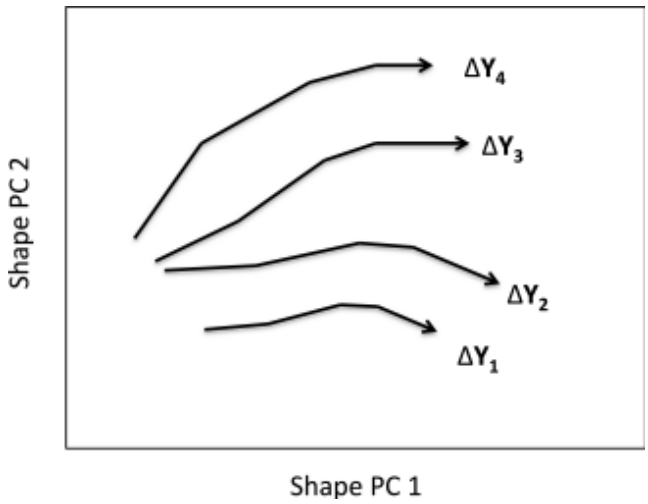


Figure 1 – Example of shape trajectories, projected onto two principal components of shape. These trajectories indicate the shape responses of four species to four different (experimental) temperature exposures. Each trajectory is denoted as $\mathbf{\Delta Y}$, comprised each of four $\mathbf{\Delta y}$ vectors (change in shape from one temperature to the next) from start point to end point (indicated by arrow direction). These trajectories illustrate differences in four geometric attributes. First, each trajectory is in a slightly different location (related to species differences in shape). Second, species 1 and 2 have trajectories of similar shape and direction, but differ in size. Third, species 2 and 3 have trajectories of similar size and shape, but differ in direction. Fourth, species 3 and 4 have trajectories of similar size and orientation, but differ in shape.

scribe the shape change of a group across a temperature gradient (four sequential, distinct temperature exposures) by the differences in their locations (in tangent space) along a trajectory. From this illustration it is easy to see why typical analyses do not describe all aspects of biological shape change. With standard multivariate analyses, one can test whether the species differ in shape (i.e., differences in their locations in tangent space), test whether shapes differ across temperature treatments, or whether there are differences in species-specific responses to particular temperatures. Yet none of these statistical components directly captures the phenotypic trajectories embodied by the shape data and shown in Fig. 1. Clearly, additional methods are required for this.

Viewing shape patterns as trajectories allows one to consider the phenotypic attributes that describe these trajectories. Specifically, trajectories in multivariate data spaces, much like landmark configurations in 2- or 3-dimensional coordinate spaces, can differ in four attributes: location, size, orientation, and shape. Differences in location between trajectories are analogous to general shape differences between groups. For the example above, a single-factor (one-way) MANOVA, followed by pairwise comparisons of species means, can be used to test for differences among species locations in shape spaces (Marcus, 1993). The other three attributes, however, describe species-specific shape change over discernable observational levels that are not encapsulated by standard multivariate analyses (Adams and Collyer, 2009). A factorial MANOVA might indicate significant variation among species-temperature means but does not implicitly test if species differ in the size, orientation, or shape of their shape trajectories. For the critical research mind, comparison of these trajectory attributes is an important and essential step in understanding why a factorial interaction is significant, and what it implies biologically.

In the following sections, we describe how such attributes of shape change can function much like test statistics to evaluate shape variation in comparative studies. We start with the simple comparison of shape change associated with an independent variable (e.g., size allometry). We then expand on this concept to introduce the comparison of shape change vectors, and finally the comparison of shape trajectories. In each section, we provide an illustrative example from empirical research. Our examples are necessarily brief and meant to highlight the conceptual advantage of examining differences in phenotypic (shape) trajectory attributes to test hypotheses. Further biological implications and specific analytical details can be found in the original sources that provide the data. Also, general analytical and statistical details can be found in Adams and Collyer (2007); Collyer and Adams (2007), and Adams and Collyer (2009).

Linear shape change associated with a continuous variable

It is common practice to compare the covariation of shape and other continuous variables collected on the same subjects (Rohlf and Corti, 2000), especially between two or more groups (e.g., Klingenberg 1996, 1998; Drake and Klingenberg 2008; Adams and Nistri 2010; Piras et al. 2010; Viscosi et al. 2010). For example, shape allometry comparisons between two or more groups – rather than comparisons of average shapes – seek to determine if groups differ in the way shape changes per unit change in size (e.g. the log of centroid size) (Klingenberg, 1996, 1998). The typical approach is to perform a "homogeneity of slopes" test. This is accomplished by comparing the log likelihoods of two different models: one model containing a group factor and a common (global) slope; another model containing a group factor and coefficients for independent slopes (i.e., including a group × size interaction; Rencher and Schaalje 2008). If there is a large difference in

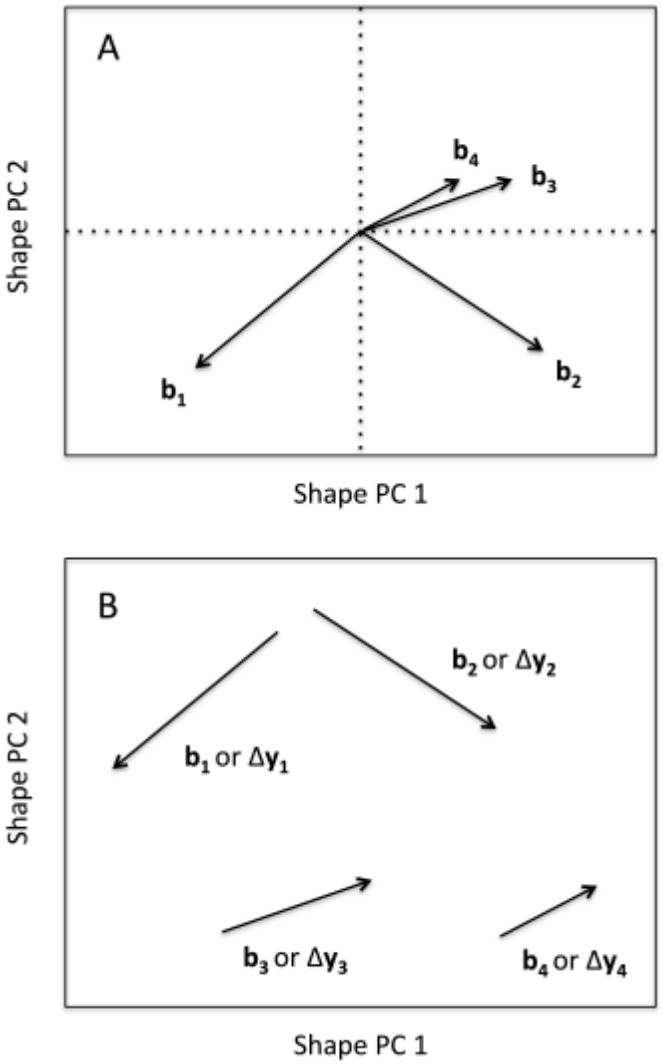


Figure 2 – Two ways to visualize shape allometry vectors. A) Allometry vectors ($\mathbf{b_i}$) as vectors of shape change from a predicted value (origin) per unit change in size. The length of the vector indicates the amount of shape change; the direction indicates the covariation of shape variables (projected onto principal components). In this case, groups 1 and 2 exhibit similar amount of shape change but the direction of shape change differs. Groups 3 and 4 change shape in similar ways with growth, but group 3 exhibits greater shape change. B) Allometry vectors ($\mathbf{b_i}$) as vectors of shape change between "small" and "large" sized estimates. Vectors are similarly scaled and orientated as vectors in A, but are located in different parts of the tangent space (based on group distinctions in shape). Such vectors could also describe shape change between discrete ecological or evolutionary states, rather than two points on a continuum ($\mathbf{\Delta y}$).

log-likelihoods between the two models (i.e., error is substantially reduced by having independent slopes), a null hypothesis of equal slopes among groups is rejected, meaning that shape allometries differ in some way. Determining significant heterogeneity in slopes is only part of the battle, as one should be compelled to understand how shape allometries differ.

For any shape defined by p shape variables, a $1 \times p$ vector of coefficients, $\mathbf{b}$, defines the linear change of each shape variable per unit change of an independent variable, X. As a vector, it has two geometric attributes: a length and a direction. Vector length describes how much shape change occurs per unit change of size; vector direction describes the relative covariations of shape variables per unit change of size (Fig. 2A). To compare these attributes between two or more groups, one can calculate the absolute difference in vector lengths (distances), $\Delta d = |d_1 - d_2|$, and the angle between vectors, θ (analytical details provided in Adams and Collyer 2007; Collyer and Adams 2007). These values can function as test statistics with expected values of Δd = 0 and θ = 0 radians or degrees, under a null hypothesis of equal allometries. The percentile of observed values of these geometric attributes from a distribution of the same values computed from a resampling experiment (i.e., generated from a null model), can be used as p-values for evaluation of the null hypothesis (see Krabbenhoft et al. 2009; Piras et al. 2010). An example based on the data from Krabbenhoft et al. (2009) is presented below. We feel it is important to point out that these attributes are calculated using all dimensions of the data space (such as the tangent space to shape space; see Adams et al. this issue) and that projection of vectors – as in Fig. 1 and 2 – onto principal components is for visual interpretation, only. Projection can distort angles and vector lengths by reducing either or both (see Collyer and Adams 2007). This phenomenon can be appreciated by envisioning an x, y, z Cartesian space containing two intersecting vectors of equal length that only differ in direction in the z dimension. An x-y projection would reveal two parallel, overlapping vectors of different length. Therefore, calculation of vector lengths and angles after projection into a space of fewer dimensions could produce erroneous results in terms of estimation, statistical evaluation, and interpretation (see also Mitteroecker et al. 2005).

It should be noted that an alternative method for describing attribute differences, which only rescales Δd and has no effect on θ, is to estimate shape at the same "small" and "large" sizes for each group (Fig. 2B). If $\hat{\mathbf{y}}$ is an $1 \times p$ vector of shape values (e.g., principal component scores) estimated from a linear regression model that describes the linear association between shape and size, then $\mathbf{b} = \hat{\mathbf{y}}_{\mathbf{large}} - \hat{\mathbf{y}}_{\mathbf{small}}$ is a vector that describes shape change between arbitrary large and small measures of size (which are consistent among groups). Differences in vector length and angles between vectors are calculated the same way. One advantage to using this approach is that changes in shape associated with growth can be shown at the locations of such points in principal component plots (Fig. 2B).

As an example, we use data originally reported in Krabbenhoft et al. (2009). These data contained landmark configurations for 868 fish, representing sister-species pairs for three genera. The three genera, *Etheostoma*, *Menidia*, and *Fundulus*, are broadly distributed in North America, but each genus contains one species endemic to the Pleistocene-originated Lake Waccamaw (North Carolina, USA) (Hubbs and Raney, 1946). These endemic species have substantially more slender body shapes in Lake Waccamaw compared to stream species occurring near Lake Waccamaw. The putative explanation for body-slendering is that higher predation in the shallow, clear waters of Lake Waccamaw is selection for streamlining, which is more energy efficient for swimming fishes (Hubbs and Raney, 1946). In order to estimate differences in linear allometric patterns, shape variables were estimated using GPA (Rohlf and Slice, 1990) performed on configurations of 12 landmarks per fish (Fig. 3A), followed by orthogonal projection and a principal component analysis. This procedure produced 20 shape variables (see Krabbenhoft et al. 2009 for further details).

One question addressed in this study was whether slender body shapes resulted from different shape allometries in Lake Waccamaw compared to streams. A homogeneity of slopes test indicated that shape allometries differed in some way (results not shown). Allometry vectors were calculated for all six species and Δd and θ were calculated between vectors for each sister-species (intra-genus) pair. Attribute differences were evaluated from sampling distributions generated from 10000 random permutations (based on a null model that lacked coefficients for independent allometries; see Krabbenhoft et al. 2009), and the null hypothesis Δd = 0 or θ = 0 was rejected if the p-value of the observed attribute difference was less than an acceptable type I error rate of α = 0.05.

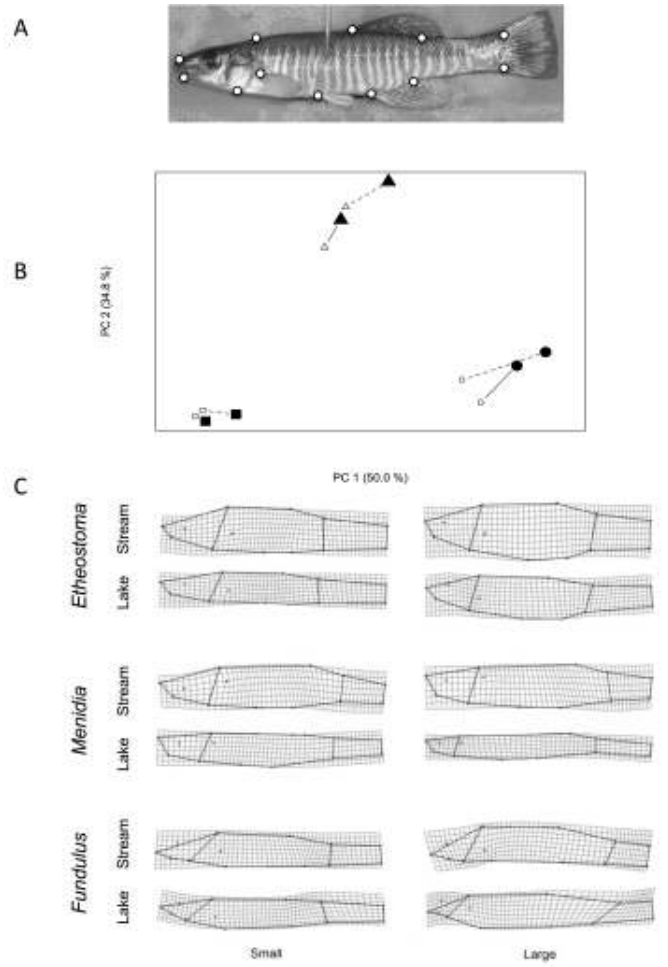


Figure 3 – Visualization of shape allometry for three species-pairs occurring in streams and Lake Waccamaw, North Carolina, USA (from Krabbenhoft et al. 2009). A) Twelve anatomical landmarks used to estimate shape (shown on a *Fundulus* specimen). B) A principal component plot (based on the covariance matrix of group means). Axes indicate the amount of inter-group variation explained by the principal component. Lines connect shape estimates at small (small, open symbols) and large (large, filled symbols) centroid sizes. Dashed lines are for Lake Waccamaw species; solid lines are for stream species. *Etheostoma* is shown as circles; *Fundulus* is shown as squares; and *Menidia* is shown as triangles. C) Transformation grids for corresponding points in the PC plot in B. Individual shapes not shown for ease of interpretation (but see Krabbenhoft et al. 2009).

It was found that genus-specific allometric differences were not consistent across genera (Fig. 3B). *Etheostoma* had a significantly higher rate of allometric shape change in the lake environment (Δd = 0.026, p = 0.0033) but allometry vectors did not significantly differ from parallel (θ = 43.65°, p = 0.5278). *Menidia* species differed significantly in vector direction (θ = 75.08°, p = 0.0198), but not length (Δd = 0.003, p = 0.9128). *Fundulus* species differed neither in length (Δd = 0.004, p = 0.3336) nor direction (θ = 30.96°, p = 0.0884). Although intergenus allometry comparisons were not statistically evaluated (as they occupied clearly different locations in tangent space), a principal component plot (Fig. 3B) illustrated that the three genera had different allometric patterns. Comparatively, *Fundulus* exhibited smaller rates of shape allometry, and *Etheostoma* and *Menidia* differed in general allometry directions (as defined by the first two PCs). Transformation grids confirmed that *Fundulus* exhibited comparatively little shape al-

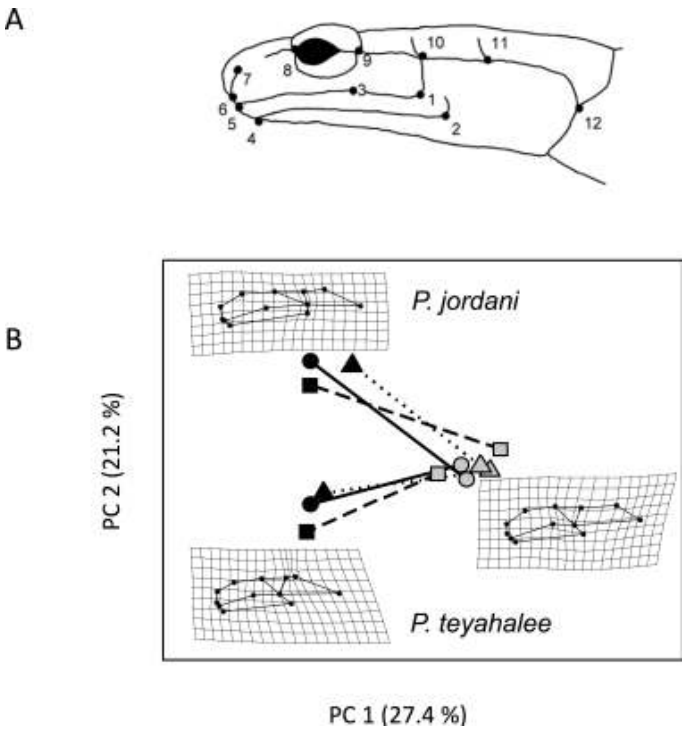


Figure 4 – Visualization of shape change vectors for three transects of repeated interspecific competition (from Adams 2010). A) The twelve anatomical landmarks used to estimate shape shown for a specimen of *Plethodon.* B) A principal component plot of shape variation (based on the covariance matrix of all individual shapes) for the two competing species of plethodontid salamanders: *Plethodon jordani* and *P. teyahalee* (percent variation explained by principal components is noted). The shape change vectors are lines in the plot; line and symbol differences correspond to different transects. Sympatric localities are represented as black symbols, allopatric localities as grey symbols. Transformation grids are shown to help facilitate an understanding of shape differences (discussed in further detail in Adams 2010). Individual shapes not shown for ease of interpretation (but see Adams 2010).

lometry, *Menidia* body shapes elongated, and *Etheostoma* body shapes deepened with growth. Within genera, body shapes were slenderer in the lake environment (Fig. 3C).

Shape change across two levels of an ordered ecological or evolutionary variable (gradient)

As indicated above, allometric vectors can either describe shape change per unit of size change, or they can describe the difference in shape between "small" and "large" sizes. The latter is simply a rescaling of the former (and a description of shape change at the location of average group shapes in a principal component plot). However, in some cases, two states are not simply points on a continuum, but represent rather a distinct change in category (i.e., a different categorical state of a qualitative independent variable). This is quite common in ecological or evolutionary biology studies. Examples include, but are not limited to, sex (e.g., Collyer and Adams 2007), experimental treatment (e.g., Hollander et al. 2006), environment type (e.g., predator/non-predator, as in Langerhans et al. 2004) and community type (e.g., allopatry/sympatry, as in Adams 2004).

For the purpose of clarity, we define a shape change vector as a vector of difference in shape between two states, a and b, $\mathbf{\Delta y} = \hat{\mathbf{y}}_\mathbf{b} - \hat{\mathbf{y}}_\mathbf{a}$, where $\hat{\mathbf{y}}$ is an estimated $1 \times p$ vector of shape, based on a categorical independent variable that describes an ecological or evolutionary gradient. (The shape change vector is the same as the phenotypic change vector, PCV, described by Adams and Collyer 2009) Conceptually, there is no difference between this vector and $\mathbf{b}$, an allometry vector, when $\mathbf{b}$ is defined for two distinct states of size (Fig. 2B). However, as a convention, we use different nomenclature to indicate that $\mathbf{\Delta y}$ represents shape change between two categorical states and $\mathbf{b}$ represents shape change associated with a continuous variable, even if defined for two fixed points on a continuum. Similar also to allometric vectors, heterogeneity in $\mathbf{\Delta y}$ is signified by a significant factor interaction, which can be determined by a likelihood ratio test that compares models that include and lack a factor interaction (Collyer and Adams, 2007). As a vector in tangent space $\mathbf{\Delta y}$, has both a length and a direction, corresponding to the amount of shape change andthe covariation of shape variables associated with a change in state for the independent variable, respectively (Adams and Collyer, 2007; Collyer and Adams, 2007). For comparison of $\mathbf{\Delta y}$ between two or more groups, Δd and θ are calculated as before, for all pairwise comparisons, and used as test statistics, which are evaluated with sampling distributions generated from null models.

As an example, we use data originally reported in Adams (2010). This study examined whether repeated interspecific competition generated parallel evolutionary divergence in phenotypes across different geographic populations of salamanders in the genus, *Plethodon*, in the Great Smoky Mountain National Park, USA. Two species, *P. jordani* and *P. teyahalee* occur in sympatry at mid-elevations in this region, where *P. jordani* is found in allopatry at higher elevations and *P. teyahalee* in allopatry at lower elevations. Various studies have documented character displacement (specifically in head shape) in this genus in sympatric populations (Adams, 2004, 2010; Adams and Rohlf, 2000; Adams and Collyer, 2007), and there is a strong genetic component to head shape (Adams, 2011), indicating that selection is capable of generating heritable, microevolutionary changes in this complex multi-dimensional trait. In the Adams (2010) study, an examination of the patterns of shape change associated with a change from allopatry to sympatry was conducted for the natural experiment of three replicated occurrences of allopatry-sympatry-allopatry gradients for these two species on different mountains.

A total of 336 salamanders comprising two species in three transects (i.e., six groups) were used in this study. Each of the six species-transect groups was found in one of two localities (sympatry or allopatry), and locality was the independent variable that described shape change within groups. Twelve anatomical landmarks were used on salamander heads and jaws (Fig. 4A) to quantify shape. The separate subset method (Adams, 1999) was used to separately perform GPA on head and jaw configurations. Shape variables from both subsets were found using the thin-plate spline and standard uniform components (Bookstein, 1991), which were combined to form a total set of 18 shape variables (see Adams 2010 for more details). (Principal components of these shape variables – referred to as relative warps [see Adams et al. 2004] – are the same as principal components of Procrustes residuals orthogonally projected into tangent space [see Adams et al. this issue]). Shape change vectors were estimated for all six sympatry-allopatry cases, and compared within species to test for differences in vector length and direction. Attribute differences were evaluated from sampling distributions generated from 10000 random permutations (based on a null model that lacked coefficients for a species-transect interaction), and the null hypothesis $\Delta d = 0$ or $\theta = 0$ was rejected if the p-value of the observed attribute difference was less than an acceptable type I error rate of $\alpha = 0.05$.

The attribute differences, Δd, and θ, for all pertinent pairwise comparisons (among transects, within species), were not significantly different from one another; though vectors between species were significantly different in direction. Thus, one can conclude that all allometry-sympatry vectors were similar in length and direction, within species (Fig. 4B; see Adams 2010 for details). Furthermore, based on locations of species-transect means in the tangent space, it was clear that allopatric localities were similar in head shape, regardless of species, but species diverged in head shape in sympatry (Fig. 4B). Together, these results demonstrate that not only did character displacement occur between these two species, but also the evolutionary mode of displacement within species was consistent among different transects (i.e., parallel evolution).

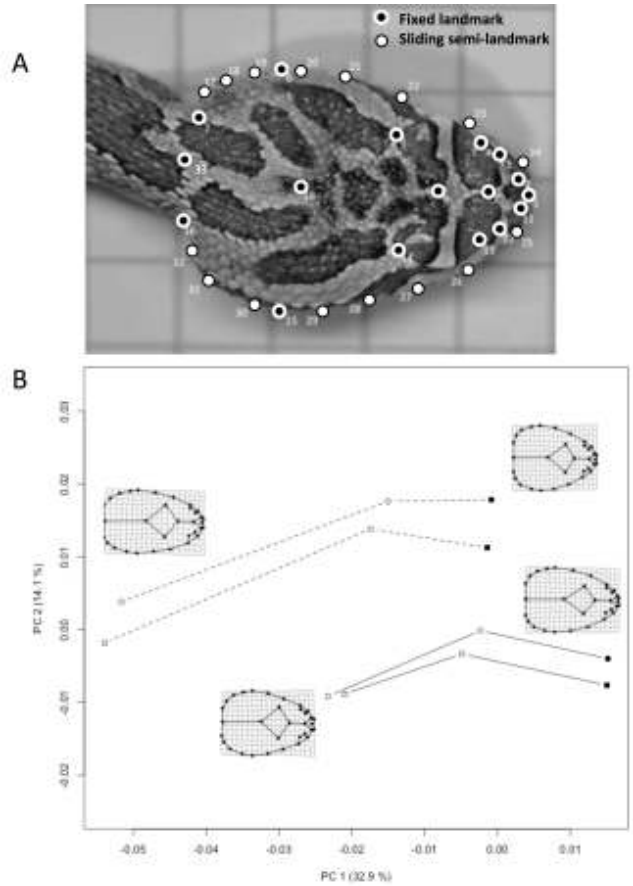


Figure 5 – Visualization of four shape trajectories for males and females from two different species of rattlesnakes (data from Davis 2012). A) The 33 landmarks used to estimate shape (fixed and sliding semi-landmarks are noted). B) Shape trajectories projected onto the first two principal components of between-group shape variation (i.e., based on covariance matrix of group means). Trajectories are shown as lines; *Crotalus viridis* trajectories are shown as dashed lines; *C. oreganus* trajectories are shown as solid lines. Filled symbols represent neonate stages (juvenile and adult can be inferred). Circles are female values; squares are male values. Transformation grids are added to facilitate an understanding of shape differences (corresponding to females end points on trajectories).

Shape change across multiple levels of an ordered ecological or evolutionary variable (gradient)

For studies that compare two or more groups that experience three or more levels of shape change associated with an ecological or evolutionary gradient, the set of geometric attributes of phenotypic trajectories increases, as well as the possible methods for describing their geometric attributes. However, the method of comparison among patterns of shape change is essentially the same (Adams and Collyer, 2009). For two ecological or evolutionary levels, shape change is described by a vector. When three or more levels are of interest, shape change forms a trajectory. These "phenotypic trajectories", in addition to location, size, and direction attributes, also have a shape (Adams and Collyer 2009; see Fig.1). (We use the term "phenotypic" rather than "shape" as a disambiguation between shape as phenotypic attribute of an organism and shape as a geometric attribute of trajectories. This disambiguation should become clear with the description of trajectory attributes below). The interpretations of differences in trajectory size and directional differences are also consistent with interpretations using shape change vectors. Trajectory size expresses the amount of shape change exhibited by a group associated with a change in ecological or evolutionary states; trajectory direction expresses the general covariation of shape variables associated with a change in ecological or evolutionary states (Adams and Collyer, 2009). The absolute difference in trajectory size and the angle between vectors of trajectory direction are values that can be evaluated from sampling distributions of random statistics (generated from null models).

Various measures of trajectory size could be used (see also the Discussion) and the choice of measure should consider the intent of the analysis. Adams and Collyer (2009) described trajectory path length – the summed lengths of vectors between sequential points in the trajectory – as a measure of trajectory size. Previous studies have also used path distance between sequential points (e.g., Dennis et al. 2011; Monnet et al. 2011; Turner et al. 2010; Frédérich et al. 2013). However, if one were interested in measuring the coverage of shape tangent space spanned by a trajectory, centroid size (Gower, 1971) or convex hull volume (Cornwell et al., 2006) might present logical alternatives for comparing the amount of shape space covered between multiple groups. (To our knowledge, no such method has yet been considered in studies that compare shape trajectories).

Adams and Collyer (2009) used the principal axis of variation among trajectory points (first principal component) as a description of trajectory direction. Large angles between principal components of compared groups indicate directional differences, which imply (under conditions that trajectories are located in similar regions of the tangent space) either ecological or evolutionary convergence or divergence (Stayton 2006; but see also Revell et al. 2007; Stayton 2006; Losos 2011; Dennis et al. 2011; Frédérich et al. 2013; Piras et al. 2012). Therefore, analyses of differences between size and direction attributes for phenotypic trajectories that describe shape change allow researchers to ascertain whether groups differ in e.g., the amount of morphological evolution, the direction of morphological evolution, or both.

The third geometric attribute of phenotypic trajectories (excluding location, as it does not describe shape change) is trajectory shape. Understanding the shape of phenotypic trajectories, and what differences in trajectory shapes describe is an area requiring additional research. Differences in trajectory shape are found as the Procrustes distances (Bookstein, 1991) between pairs of phenotypic trajectories (for details see Adams and Collyer 2009). Procrustes distances can function like test statistics to test for shape differences in trajectories among different groups. (We use D_p to denote Procrustes distance, to differentiate it from d, the distance [length] of either a shape vector or phenotypic trajectory). Unlike differences in trajectory size and orientation, differences in trajectory shapes are more challenging to interpret biologically. Differences in trajectory shape imply that, across ecological or evolutionary levels, changes in shape are accelerated or decelerated in one group relative to another, or are orientated in different directions, or both, in one portion or multiple portions of the trajectories (Fig. 1). Differences in trajectory shapes imply that there is a signal that some unique stage (or time) specific shape change is occurring. A potential useful exercise after concluding that trajectories differ in their shape is to perform qualitative pairwise comparisons of shape change vectors (or even a statistical test analogous to testing the attributes of shape change vectors), for the $k - 1$ vectors between the k points that sequentially comprise the trajectories.

As an example of phenotypic trajectory comparison, we use data originally reported by Davis (2012). These data comprise 3107 specimens of rattlesnakes in the genus, *Crotalus*. A total of 33 landmarks were digitized on the dorsal side of *Crotalus* heads from museum collections (Fig. 5A; see Davis 2012 for museum information). Although nine subspecies were analyzed in the original work, we only consider differences in ontogenetic trajectories (described for neonate-juvenile-adult sequences) between two species – prairie rattlesnakes (*C. viridis*) and western rattlesnakes (*C. oreganus*) – that have large, overlapping distributions in North America (i.e., all subspecies were pooled within species). We also consider whether there was sexual dimorphism in ontogenetic shape change in either or both species. Thus, there were four species-sex groups, each described by shape trajectories comprised of three ontogenetic stages. The three attributes of shape trajectories compared included path length, direction, and shape. The attribute differences, Δd, θ, and D_p, for all pertinent pairwise comparisons (between sex within species; between species within sex), were considered significant if the p-values from 10000 random permutations were less than an acceptable type I error rate of $\alpha = 0.05$.

Table 1 – Attribute differences, standardized scores, and *p*-values from the rattlesnake example. Pearson product-moment correlations between *p*-values and either attribute differences or standardized scores are also shown.

	Size			Direction			Shape		
Comparison	**Δd**	*Z*	*p*	**Δd**	*Z*	*p*	**Δd**	*Z*	*p*
FO-FV	0.0119	4.5368	0.0068	16.4264	5.4303	0.1342	0.2100	3.8535	0.0385
FO-MO	0.0060	3.0309	0.0662	14.4641	5.7164	0.0906	0.1568	3.5187	0.0680
FO-MV	0.0124	4.0199	0.0164	13.7274	3.9769	0.5334	0.2195	3.6074	0.0557
FV-MO	0.0178	6.6777	0.0001	23.9324	7.7021	0.0035	0.2949	5.2896	0.0033
FV-MV	0.0005	0.1660	0.9191	9.7181	2.5198	0.9890	0.0150	0.2351	0.9868
MO-MV	0.0184	5.9015	0.0007	22.7301	6.3988	0.0301	0.2971	4.8996	0.0059
correlation with *p*-value	-0.7926	-0.8542		-0.8165	-0.9388		-0.8898	-0.9386	

Both species were sexually dimorphic in terms of the amount of shape change associated with ontogeny (*C. oreganus*: Δd = 0.0060, p = 0.0001; *C. viridis*: Δd = 0.0005, p = 0.0005) (Fig. 5B). Males exhibited greater shape change in *C. viridis*, although the difference was small. Females exhibited a slightly larger amount of shape change in *C. oreganus*. Despite these results, males and females were more similar in shape at any stage of development compared to inter-stage variation in head shape, within sex (Fig. 5B). The only pertinent significant difference in directional shape change was between species, within males (θ = 22.73°, p = 0.0288), but this result was only slightly significant and the angle between principal directions was small. Therefore, the direction of shape changed was largely consistent both between males and females, and between species. *C. viridis* exhibited significantly more shape change during ontogeny (Females: Δd = 0.0119, p = 0.0069; Males: Δd = 0.0184, p = 0.0005), especially between juvenile and adult stages, confirming that differences among trajectory shapes were because of different ontogenetic patterns between species, within sex (Females: D_p = 0.21; p = 0.0405; Males: D_p = 0.21, p = 0.0048). Trajectories did not differ in shape between males and females, within species (*C. oreganus*: D_p = 0.16; p = 0.0643; *C. viridis*: D_p = 0.15; p = 0.9862).

In summary, sexual dimorphisms were small and only pertained to minor amounts of shape change, but species differed substantially in the amount and shape of shape change. These attribute differences correspond to accelerated shape change for *C. viridis* between juvenile and adult stages, compared to *C. oreganus*.

It is worth commenting that the three examples presented here were increasingly more complex in terms of the shape change gradient considered, but the analyses performed were all exactly the same. This analysis, called phenotypic trajectory analysis (PTA) is the pairwise comparison of geometric attributes – size, direction, shape – of phenotypic trajectories, and it is performed the same, irrespective of the number of phenotypic states in the trajectories. As discussed by Adams and Collyer (2009), two-state shape change is a simple case of multi-state shape change. The path length of a single vector is the vector length; the first principal component of a within-group covariance matrix is the vector that describes the difference between two states; and a vector has no shape (thus there are no trajectory shape differences among groups). Therefore, performing PTA to compare taxa that have two or more estimable shapes, corresponding to important ecological or evolutionary states, works the same for any number of sequential points in shape trajectories. Furthermore, as stated above, allometry vectors can be described as two-state shape change vectors. Although we did not discuss polynomial models (Rencher and Schaalje, 2008) as descriptions of shape allometry, one could imagine generating also multi-state trajectories to describe (potentially) non-linear shape allometries. Thus, analysis of the geometric attributes of shape trajectories is a generalized method for the comparison of any shape change associated with either qualitative or quantitative independent variables that describe important ecological or evolutionary gradients.

Qualitative comparison of geometric attributes: standardized scores of attribute differences

We have presented differences between geometric attributes of shape change as measures that function as test statistics, since the sampling distributions of the geometric attributes are created by a resampling procedure, which allows *p*-values to be estimated by the percentiles of observed attribute differences in the distributions. However, geometric attribute differences are not test statistics in the sense that they do not convey any information about the magnitude of the measure in relation to the variability of the measure. Much like many descriptive statistics can be expressed as standardized scores, geometric attribute differences can be converted to standardized scores using the standard deviations from their sampling distributions. For example, standardized scores of angles can be calculated as

$$Z_\theta = \frac{\theta - E[\theta]}{\hat{\sigma}_\theta} \quad (1)$$

where $\hat{\sigma}_\theta$ is standard deviation of angles between vectors, as estimated from the empirical distribution generated from a resampling procedure. Because the expected angle $E[\theta] = 0$ under the null hypothesis of parallel vectors, the standardized score simplifies to $Z_\theta = \theta/\hat{\sigma}_\theta$. Likewise, $Z_{\Delta d} = \Delta d/\hat{\sigma}_{\Delta d}$ and $Z_{D_p} = D_p/\hat{\sigma}_{D_p}$, for differences in trajectory size and shape, respectively. It should be noted that we use the variable, Z, as a convention because these statistics are similar to standard deviates, but we do not wish to imply that the standard normal distribution is used to find the probability of observed attribute differences. One might also think of Z as a measure analogous to Cohen's (1988) description of the standardized effect size (see also Sokal and Rohlf 2012), using the standard deviation of a sampling distribution of a statistic – i.e., the standard error – rather than a population standard deviation of a variable. Using the standard deviation of a sampling distribution (from the resampling experiment that generates random attribute differences) is necessary with the geometric attribute differences described here, because all attribute differences are scalars although the attributes, themselves, are either scalars (trajectory size), vectors (trajectory direction), or matrices (trajectory shape). Thus, although not directly analogous to Cohen's (1988) standardized effect size, the standardized scores are measures that are positively correlated to effect size. Finally, it should also be noted that we describe standardized scores based on the population of random permutations used to generate sampling distributions. One could argue that the number of resampling events is a sample of (the population of) all possible random outcomes, and standardized scores are, therefore, more analogous to t-statistics than z-statistics. However, this implication would only be a concern for unreasonably small sets of resampling iterations and would not change the important point that standardized scores are positively correlated with effect size.

Standardized scores allow one to qualitatively compare attribute differences within and between studies. This is especially useful when the value of a geometric attribute difference is unintuitive in terms of the outcome of a null hypothesis test. For example, large angles (like in the example above of comparison of allometry vectors for sister fish taxa occurring in streams and Lake Waccamaw) might not be significantly different from 0°, but small angles (like in the example above of comparison of shape trajectories between males of different species of *Crotalus*) can be significantly different from 0°, if the standard deviation of random angles is small. It is also useful if one wishes to qualitatively compare effect sizes among different geometric attributes of shape change, within the same study.

As an example, we calculated standardized scores for each of the six possible pairwise comparisons of the four groups in the *Crotalus*

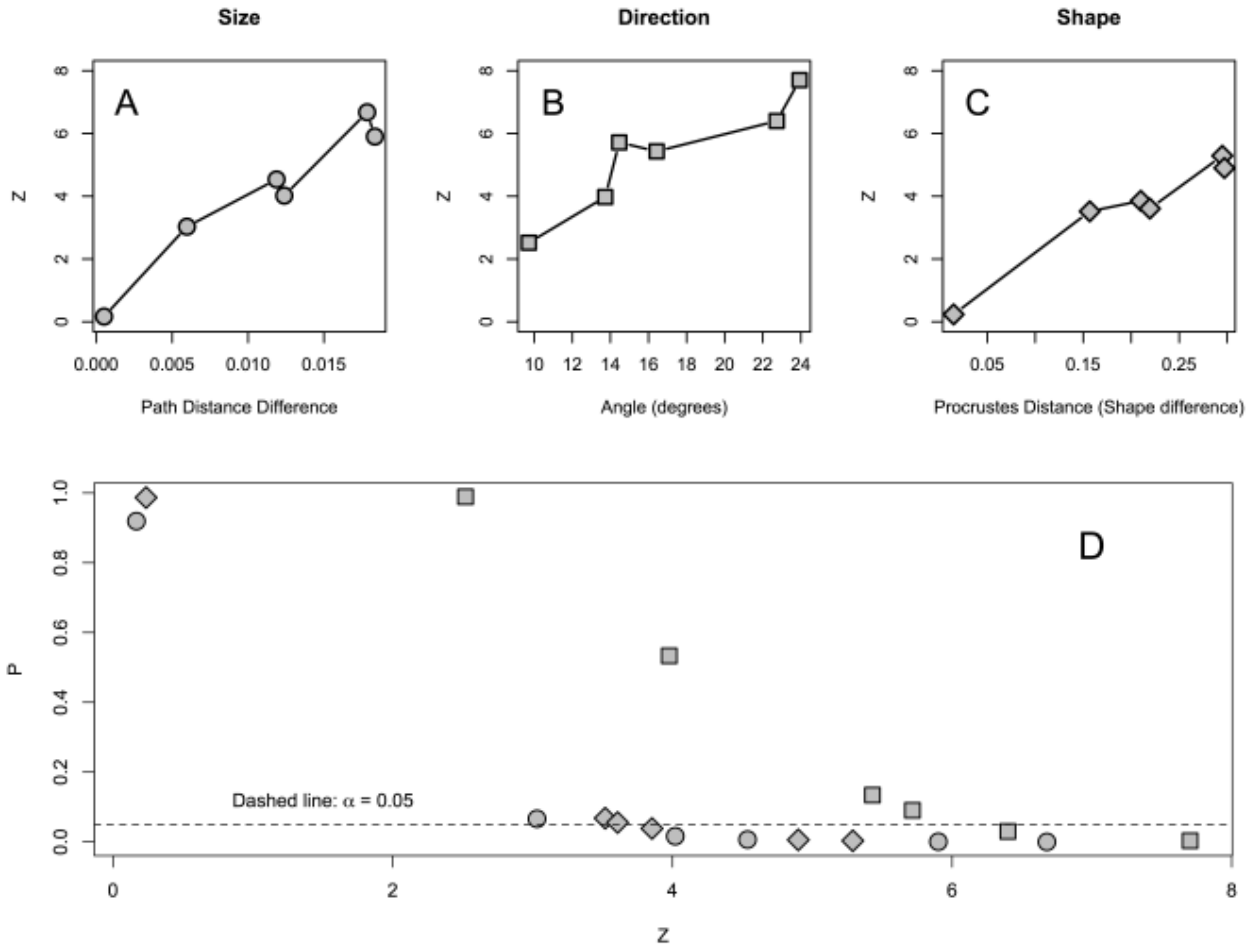


Figure 6 – Relationship between trajectory attribute differences and their standardized scores, Z, for the rattlesnake example (values given in Tab. 1). Standardized scores are plotted versus attribute differences in panels A-C. Lines connect the rank order of values for attribute differences. (A negative sloping line indicates a decrease in standardized score in spite of an increase in attribute difference). The same standardized scores are shown on the abscissa in the plot in panel D, with p-values from phenotypic trajectory analysis shown as the ordinate. Symbols in D correspond to the same symbols in A-C.

example of the previous section, for each of the geometric attributes considered. The standardized scores are shown alongside the original values and p-values in Tab. 1, and plotted in Fig. 6 to illustrate the association between standardized scores and original values, as well as their p-values. For this example, two things are clear. First, although larger attribute differences result in generally larger standardized scores, the rank orders of attribute differences and standardized scores were not exactly the same, meaning that in comparison, a larger standardized score can be found from a smaller attribute difference, or vice versa. The standardized scores presented here scale geometric attribute differences by the inverses of their standard deviations, as found from their sampling distributions. Thus, they function more as test-statistics that can be qualitatively compared within and between studies. In this example, standardized scores were more strongly correlated with p-values than attribute differences (Tab. 1), which suggests that there is some benefit to reporting them along with observed attribute differences. Second, although standardized scores might be more strongly correlated with p-values, one must use caution when comparing them between different attributes. In our example, standardized scores for directional diffrences were larger compared to size or shape differences, in general, but the null hypothesis was rejected fewer times. We suspect that this phenomenon results from using only the first principal components of randomly-generated within-trajectory covariance matrices, which might negatively bias variances of angular differences, thus inflating standardized scores. The relationship between standardized scores of geometric attribute differences and shape space dimensionality requires further research.

Discussion

As Thompson (1917) stated as an important prelude to his "Theory of transformations", in comparative studies it's the description of morphological change between two states that is intriguing rather than the precise definition of morphology in either state. Users of GM methods can appreciate that measuring shape change between two landmark configurations in terms of their Procrustes distance and the total shape change implied by it (as visualized by the thin-plate spline, Bookstein 1989), is intuitively more appealing to interpret than the meaning of the position of landmarks in one landmark configuration. Likewise, coefficients from a linear model, which describes shape variation as a function of one or more independent variables, measure the amount of expected shape change per unit change of an independent variable. As we have shown, that shape change forms a path in shape space, and has two important attributes: a size and a direction. Size is the amount of shape change associated with a per unit change in an independent variable; direction is the covariation of shape variables associated with change in an independent variable. When linear models also contain factors to describe ecological or evolutionary gradients, shape change is the cumulative sequence of vector changes in shape associated with the gradient, forming a trajectory. Therefore, a third attribute, trajectory shape, describes shape change for shape that is measured or estimated for more than two states. (Our examples all included shape data, but equally viable examples could have used other phenotypic data).

Most standard multivariate statistical analyses used in ecological and evolutionary research use linear models to estimate parameters of phenotypic change. However, standard hypothesis tests do not specifically evaluate attributes of phenotypic change, which are viable test statist-

ics themselves. Using the four species-four experimental temperatures example presented in the introduction, a factorial MANOVA evaluates the species-temperature interaction by estimating the probability of a test statistic for the comparison of two log-likelihoods for two different linear models. One model contains coefficients for just the species and temperature factors; the other contains coefficients for these factors, plus coefficients for their interaction. A large difference in log likelihoods indicates that model error is significantly reduced by including the extra coefficients (Manly and Rayner, 1987; Rencher and Schaalje, 2008). This implies that coefficients for the interaction are important, and a null hypothesis of no interaction effect is rejected (i.e., the interaction is a significant source of variation). For analyses that seek to test differences in location (e.g., species differences in shape), there is a direct connection between these test statistics and null hypotheses. We refer to such tests as tests of "static" variation – tests that evaluate variation among group locations in multivariate data spaces. By contrast, geometric attributes of phenotypic trajectories are test statistics for tests of "dynamic" variation – tests that evaluate variation among group changes in location in the multivariate data space. Standard multivariate tests might imply that dynamic variation is meaningful (e.g., a factor interaction is significant); but tests using the geometric attributes, themselves – PTA – are more direct and explain why dynamic variation is significant.

It is a safe assertion that many hypotheses in ecology and evolutionary biology describe patterns of phenotypic change, yet the standard approach to assessing these patterns is incomplete. For example, in the *Plethodon* salamander example above, a factorial MANOVA examining patterns of head shape variation revealed that a species × transect × locality interaction was significant (see Adams 2010). But how does this result confer any knowledge about the repeatability of evolutionary divergence? Alone, it does not. In fact, identifying a significant interaction term in a MANOVA is insufficient to discern among the many alternative explanations that may have generated this observed pattern. For instance, a significant interaction could be observed if there was heterogeneity among transects within species. However, by linking phenotypes across the ecological gradient, one forms trajectories of phenotypic change (in this case from allopatry to sympatry). Then, an explicit comparison of trajectory attributes using PTA allowed for direct tests of shape change and therefore of its underlying biological processes; in this case, identifying evolutionary parallelism of phenotypic change (Adams 2010, for examples of convergence see: Adams and Nistri 2010; Piras et al. 2010).

Another advantage of PTA is that it offers great flexibility and can be used for any sequence of shapes in tangent space that form a trajectory. Although trajectories tend to have logical sequences associated with ecological or evolutionary gradients, PTA works as efficiently for "configurations" of shapes in tangent space that are less logical as a sequence, but still correspond across groups. For example, one might study the shape responses of groups of organisms in an experiment or observational study due to different predator types (no predator, predator species A, predator species B, etc.). Here imposing an order or sequence to the levels is arbitrary, yet PTA can be used on the original unordered data to determine whether phenotypic responses to predators differ among groups (e.g., Hollander et al. 2006). Additionally, in many circumstances, PTA provides a complementary approach for describing patterns of change where alternative measures are more commonly used. For example, in community ecology, various dispersion metrics have been proposed to describe patterns in stable isotope data, as well as for food webs (Layman et al., 2007). Yet in these cases, trajectory analysis has proven to be a valuable complementary tool for identifying spatial and temporal patterns in isotope data spaces unexamined by standard approaches (e.g., Turner et al. 2010).

Despite the clear advantages of phenotypic trajectory analysis, PTA is still a recently developed tool. Thus, a number of issues remain to be addressed in terms of how it is implemented under particular circumstances. We feel that the following three issues will present the greatest challenges and the most interesting discoveries in the coming years. First, for certain types of data, different research designs, or different ecological or evolutionary gradients, are better alternative measures for trajectory attributes available? The initial conception of PTA was motivated by using sequential points along ecological and evolutionary gradients; thus the most appropriate size measure for such designs was path distance (Adams and Collyer, 2009). The benefit of this measure is that it produces large values from oscillatory shape changes, something centroid size or convex hull volume might fail to identify. Alternatively, if one were more interested in the amount of data space covered by a species over an ecological gradient, centroid size (Gower, 1971) or convex hull volume (Cornwell et al., 2006) might be better alternatives. With respect to directional differences, an alternative could be to analyse the angle between vectors that describe the difference between starting and end points, within groups, rather than comparing the first principal components of each. We anticipate that assessments of alternative trajectory attribute measures will be valuable in the future, especially as additional research questions prompt the need for alternative measures.

Second, for shape change across continuous variables, can PTA be used with polynomial regression or non-linear regression analyses? If so, how would such an analysis be optimized? In the discussion of allometry, we indicated that allometry vectors could also be described as two-state change vectors (between "small" and "large" size). Clearly, this formulation is a linear transformation (rescaling) of the original vector, so sampling distributions of attribute test statistics remain unchanged when the underlying trajectory is linear. However, ontogenetic trajectories can be decidedly non-linear (Mitteroecker et al., 2004), and in these circumstances, it is less obvious what quantitative representation should be utilized. For instance, one could use estimates of shape from polynomial models to form a trajectory for nonlinear allometry considerations, or describe trajectories in size-phenotype spaces (e.g., Mitteroecker et al. 2004), but how many points along the polynomial regression should the trajectory contain? Further, it remains unknown how changing the number of trajectory points would alter the sampling distributions of test statistics, and thereby affects biological interpretation. Finally, we could envision that PTA could be combined in some way with other methods for quantifying ontogenetic trajectories (e.g., the common allometric component and its residual shape variation: *sensu* Mitteroecker et al. 2004), where PTA would provide statistical tests for if, and in what manner, ontogenetic trajectories vary. Additional work is needed in this area.

Finally, what do trajectory shapes, or more specifically, differences in trajectory shapes, mean biologically? From our experience, trajectory shape differences are interpretable if trajectories have few points and are located in similar locations. Also, differences in trajectory shapes can be interpretable in the case of motion analysis (Adams and Cerney, 2007), as they describe how a motion is performed (e.g., straight-arm movements versus back-and-forth movements will generate trajectories of different shapes in shape space). However, interpreting differences in trajectory shapes when trajectories have many points or when trajectories are located in largely different regions of data spaces makes inferences more challenging (Collyer and Adams, unpublished data). Because trajectories are sequences of vectors, one has to question whether a significant trajectory shape difference implies that the entire trajectory differs in shape, or whether such patterns are isolated to particular portions of the trajectory. For instance, across trajectories representing multiple ecological or evolutionary levels, differences in trajectory shape may represent accelerated or decelerated shape change in one group relative to another, or shape changes orientated in different directions, and these may be found in one portion or multiple portions of the trajectories. Thus, for many studies, identifying significant differences in trajectory shape may represent the first step in a more in-depth and pairwise assessment of where trajectories differ, and how. As such, the shape of a trajectory might not be so much a global trajectory attribute, as it is a summary of size and directional attributes of the vectors that comprise it. More work is needed in this area.

In looking toward the future, regarding development of methods for analysis of patterns of shape change, it is perhaps instructive to look

back over the past few decades of development of GM methods. The field has certainly changed. In the past twenty years, the use of GM methods has dramatically increased, as the "Procrustes paradigm" has developed from one alternative approach of shape quantification into a rigorous discipline (Adams et al., this issue). New users of GM methods now take for granted, for example, that relative warps are typically principal components of unweighted partial warp scores (Zelditch et al. 2004, p. 423) or simple principal components of Procrustes residuals orthogonally projected into tangent space. They might not realize or appreciate the amount of attention once paid to weighting principal warps prior to generating partial warps and relative warps to develop alternative "biomathematical" strategies (e.g., Bookstein 1996). Likewise, we suspect that if another synthetic volume of progress in GM development is created ten years from now, some or all of the challenges listed in the previous paragraph will have since been rectified. Rather, the body of work on analyses of patterns of shape change – including PTA and alternative methods – will be much more comprehensive, and the analytical idiosyncrasies yet to be resolved will be fleshed out by the research questions that implore further discovery.

References

Adams D.C., 1999. Methods for shape analysis of landmark data from articulated structures. Evol. Ecol. Res 1: 959–970.

Adams D.C., 2004. Character displacement via aggressive interference in Appalachian salamanders. Ecology. 85: 2664–2670.

Adams D.C., 2010. Parallel evolution of character displacement driven by competitive selection in terrestrial salamanders. BMC Evol. Biol. 10: 1–10.

Adams D.C., 2011. Quantitative genetics and evolution of head shape in *Plethodon* salamanders. Evol. Biol. 38: 278–286.

Adams D.C., Cerney M.M., 2007. Quantifying biomechanical motion using Procrustes motion analysis. J. Biomech. 40: 437–444.

Adams D.C., Collyer M.L., 2007. The analysis of character divergence along environmental gradients and other covariates. Evolution 61: 510–515.

Adams D.C., Collyer M.L., 2009. A general framework for the analysis of phenotypic trajectories in evolutionary studies. Evolution 63: 1143–1154.

Adams D.C., Nistri A., 2010. Ontogenetic convergence and evolution of foot morphology in European cave salamanders (Family: Plethodontidae). BMC Evol. Biol. 10: 1–10.

Adams D.C., Rohlf F.J., 2000. Ecological character displacement in *Plethodon*: biomechanical differences found from a geometric morphometric study. Proceedings of the National Academy of Sciences, USA 97: 4106–4111.

Adams D.C., Rohlf F.J., Slice D.E., 2004. Geometric morphometrics: ten years of progress following the "revolution". It. J. Zool. 71: 5–16.

Adams D.C., Rohlf F.J., Slice D.E., 2013. A field comes of age: geometric morphometrics in the 21st Century. Hystrix 24(1) (Online First) doi:l0.4404/hystrix-24.l-6283

Bookstein F.L., 1989. Principal warps: thin-plate splines and the decomposition of deformations. Institute of Electrical and Electronics Engineers, Transactions on Pattern Analysis and Machine Intelligence 11: 567–585.

Bookstein F.L., 1996. Biometrics, biomathematics and the morphometric synthesis. Bulletin of Mathematical Biology 58: 313–365.

Bookstein F.L., 1991. Morphometric tools for landmark data: Geometry and Biology. Cambridge Univ. Press, New York.

Box G.E.P., Draper N.R., 1987. Least squares for response surface work. In: Box G.E.P., Draper N.R., 1987. Empirical model building and response surfaces. John Wiley & Sons, New York. 34–103

Cohen J., 1988. Statistical power analysis for the behavioural sciences, 2nd edition. Erlbaum, Hillsdale, NJ.

Collyer, M. L., and D. C. Adams. 2007. Analysis of two-state multivariate phenotypic change in ecological studies. Ecology 88:683-692.

Cornwell W.K.,Schwilk D.W., Ackerly D.A., 2006. A trait-based test for habitat filtering: convex hull volume. Ecology 87: 1465–1471.

Davis M.A., 2012. Morphometrics, molecular ecology, and multivariate environmental niche define the evolutionary history of the western rattlesnake (*Crotalus viridis*) complex. Ph. D. Dissertation, University of Illinois, Champaign, IL.

Dennis S.R., Carter M.J., Hently W.T., Beckerman A.P., 2011. Phenotypic convergence along a gradient of predation risk. Proc. Roy. Soc. B-Biol. Sci. 278: 1687–1696.

Drake A.G., Klingenberg C.P., 2008. The pace of morphological change: Historical transformation of skull shape in St Bernard dogs. P. Roy. Soc. B-Biol. Sc. 275:71–76.

Frédérich B., Sorenson L., Santini F., Slater G.J., Alfaro M.E., 2013. Iterative ecological radiation and convergence during the evolutionary history of damselfishes (Pomacentridae). Am. Nat, 181: 94–113.

Gower J.C., 1971. Statistical methods of comparing different multivariate analyses of the same data. In: Hodson F.R., Kendall D.G., Tautu P. (Eds.) Mathematics in the archaeological and historical sciences. Edinburgh Univ. Press, Edinburgh. 138–149.

Hollander J., Collyer M.L., Adams D.C., Johannesson K., 2006. Phenotypic plasticity in two marine snails: constraints superseding life-history. J. Evol. Biol. 19: 1861–1872.

Hotelling H., 1931. The generalization of the student's ratio. Ann. Math. Stat. 2: 360–378.

Hubbs C.L., Raney E.C., 1946. Endemic fish fauna of Lake Waccamaw, North Carolina. Misc. Pu. Univ. of Mich. Museum Zool. No. 65. Ann Arbor, MI.

Klingenberg K.P., 1996. Multivariate allometry. In: Marcus L.F., Corti M., Loy A., Naylor G.J.P., Slice D.E. (Eds.) Advances in morphometrics. Plenum Press, New York. 23–49.

Klingenberg K.P., 1998. Heterochrony and allometry: the analysis of evolutionary change in ontogeny. Biol. Rev. 73: 79–123.

Krabbenhoft T.J., Collyer M.L., Quattro J.M., 2009. Differing evolutionary patterns underlie convergence on elongate morphology in endemic fishes of Lake Waccamaw, North Carolina. Biol. J. Linn. Soc. 3: 636–645.

Langerhans R.B., Layman C.A., Shokrollahi A.M., DeWitt T.J., 2004. Predator-driven phenotypic diversification in *Gambusia affinis*. Evolution 58: 2305–2318.

Layman C.A., Arrington D.A., Montana C.G., Post D.M., 2007. Can stable isotope ratios provide for community wide measures of trophic structure? Ecology 88: 42–48.

Losos J.B., 2011. Convergence, adaptation, and constraint. Evolution 65: 1827–1840.

Manly B.F.T., Rayner J.C.W., 1987. The comparison of sample covariance matrices using likelihood ratio tests. Biometrika 74: 841–847.

Marcus L.F., 1993. Some aspects of multivariate statistics for morphometrics. In: Marcus L.F., Bello E., Garcia-Valdecasas A. (Eds.) Contributions to morphometrics. Monografías del Museo Nacional de Ciencias Naturales 8, Madrid. 95–130.

Mitteroecker P., Gunz P., Bernhard M., Schaefer K., Bookstein F.L., 2004. Comparison of cranial ontogenetic trajectories among great apes and humans. J. Hum. Evol. 46: 679–698.

Mitteroecker P., Gunz P., Bookstein F.L., 2005. Heterochrony and geometric morphometrics: a comparison of cranial growth in *Pan paniscus* versus *Pan troglodytes*. Evol. Develop. 7: 244–258.

Monnet C., De Baets K., Klug C., 2011. Parallel evolution controlled by adaptation and covariation in ammonoid cephalopods. BMC Evol. Biol. 11: 115.

Piras P., Colangelo P., Adams D.C., Buscalioni A., Cubo J., Kotsakis T., Meloro C., Raia P., 2010. The Gavialis-Tomistoma debate: the contribution of skull ontogenetic allometry and growth trajectories to the study of crocodylian relationships. Evol. Develop. 12: 568–579.

Piras P., Sansalone G., Teresi L., Kotsakis T., Colangelo P., Loy A., 2012. Testing convergent and parallel adaptations in talpids humeral mechanical performance by means of geometric morphometrics and finite element analysis. J. Morphol. 273: 696–711.

Rencher A.C., 2002. Methods of multivariate analysis, 2nd edition. Wiley Series in Probability and Statistics, John Wiley and Sons, Hoboken, NJ.

Rencher A.C., Schaalje C.B., 2008. Linear models in statistics, 2nd edition. John Wiley & Sons, Hoboken, NJ.

Revell L.J., Johnson M.A., Schulte J.A., Kolbe J.J., Losos J.B., 2007. A phylogenetic test for adaptive convergence in rock-dwelling lizards. Evolution 61: 2898–2912.

Rohlf F.J., Corti M., 2000. The use of partial least-squares to study covariation in shape. Syst. Biol. 49: 740–753.

Rohlf F.J., Slice D.E., 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39: 40–59.

Simpson G.E., 1944. Tempo and mode in evolution. Columbia Univ. Press, New York.

Sokal R.R., Rohlf F.J., 2012. Biometry, 4th edition. W. H. Freeman, New York.

Stayton C.T., 2006. Testing hypotheses of convergence with multivariate data: morphological and functional convergence among herbivorous lizards. Evolution 60: 824–841.

Thompson D.W., 1917. On Growth and Form. Cambridge, London.

Turner T.F., Collyer M.L., Krabbenhoft T.J., 2010. A general hypothesis-testing framework for stable isotope ratios in ecological studies. Ecology 88: 2227–2233.

Viscosi V., Loy A., Fortini P., 2010. Geometric morphometric analysis as a tool to explore covariation between shape and other quantitative leaf traits in European white oaks. In: Nimis P.L., Vignes L.R. (Eds.) Tools for identifying biodiversity: progress and problems. EUT Edizioni Università di Trieste. 257–261.

Zelditch M. L., Swiderski D.L., Sheets H.D., Fink W.L., 2004. Geometric morphometrics for biologists. Elselvier, San Diego, CA, USA.

Associate Editor: A. Loy

1996 – Budapest – V International Congress of Systematics and Evolutionary Biology. From left: Anna Loy, Fred Bookstein, Leslie Marcus, Marco Corti and Carlo Fadda (picture by Anna Loy).

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 85–93, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/6296/pdf
doi:10.4404/hystrix-24.1-6296

Research Article

The direction of main phenotypic variance as a channel to morphological evolution: case studies in murine rodents

Sabrina Renaud[a,*], Jean-Christophe Auffray[b]

[a] *Laboratoire de Biométrie et Biologie Evolutive, UMR 5558, CNRS, Université Lyon 1, Campus de la Doua, 69622 Villeurbanne, France*
[b] *Institut des Sciences de l'Evolution, UMR 5554, CNRS, Université Montpellier 2, 34095 Montpellier, France*

Keywords:
morphometrics
Murinae
evolutionary lineages
molar
mandible
Fourier outline analysis

Article history:
Received: 1 June 2012
Accepted: 14 December 2012

Acknowledgements
This study benefited from numerous discussions with colleagues. Without mentioning them all, we are particularly indebted to Sophie Pantalacci, Vincent Laudet, Julien Claude, and Paul Alibert. The comments by an anomymous reviewer, Anna Loy and Andrea Cardini significantly improved the manuscript. Laboratory material was provided by M. Texeira from the PBES of the Ecole Normale Supérieure de Lyon. We thank Jean-Pierre Quéré for access to some material considered in this study. This work was partly supported by the ANR project Bigtooth (ANR-11-BSV7-008). This is contribution ISEM 2012-262.

Abstract

A key issue in evolutionary studies is the means by which evolution can be channeled by intrinsic processes such as genetic and development. Studying the phenotypic variation in a population can shed light on these constraints, because phenotypic variation, being the product of genetic and developmental processes, is the target of both selective screening and random sampling. The main phenotypic variance in populations ("$Pmax$") could thus act as a "line of least resistance to evolution". Based on morphometric analysis of molar evolution in several fossil lineages and modern murine rodents, the role of $Pmax$ as line of least resistance to evolution is investigated: Does evolution along lineages actually occur along $Pmax$? Does this line of least resistance facilitate parallel evolution? What is the relationship of $Pmax$ to developmental processes and functional constraints? Case studies on murine rodent teeth are complemented by examples focusing on mouse mandibles. Compared to teeth, which are mineralized early during development, the mandible, as a bone, is prone to shape changes through remodeling in relation to masticatory muscles and other tissues. Mandible shape may thus vary throughout an animal's life due to allometric growth and, more generally, because of environmental influences. This may lead the mandible's $Pmax$ to align with the direction of plastic and allometric variation. However, the same kind of shape change may also be produced by genetic changes. These examples illustrate how studying patterns of phenotypic variance using geometric morphometrics can help to identify evolutionary processes, bridging several evolutionary levels from intra-group variation to inter-group evolution, and therefore can contribute to an integrated view of phenotypic evolution.

Introduction

The means by which evolution is constrained and channeled by intrinsic processes such as development is a key issue in evolutionary studies, as these processes might condition the evolvability of traits and their flexibility in response to selection as well as drift (e.g. Beldade et al. 2002; Brakefield 2006). Evolutionary patterns such as parallel evolution may have different interpretations depending on whether or not they have been channeled by intrinsic processes. A similar pattern can be the product of parallel responses to strong comparable and selective pressures, or correspond to similar outputs facilitated by common intrinsic constraints.

The variation existing within a population has the potential to provide clues for deciphering the role of these constraints. The expression of genetic variance is modulated by many genetic, epigenetic, and environmental features which interact with developmental networks to produce the phenotypic variation characteristic of a population. Recognizing the importance of development in conditioning the phenotypic outcome of a given genotype has revolutionized the simplistic view of the genotype-phenotype relationship (e.g. Jernvall 2000; Kavanagh et al. 2007; Salazar-Ciudad and Jernvall 2010; Skinner and Gunz 2010). By integrating both genetic and developmental components, the study of phenotypic variation thus appears fundamental when revisiting morphological evolution with an "evo-devo" perspective.

Furthermore, phenotypic variation is itself a key feature in evolution. Not only is it the phenotypic variation existing in a population on which natural selection operates, but even the output of random processes such as drift depends on this variation, since widespread variants will have a higher chance of being sampled. The evolution of a trait in a given direction may be facilitated when this kind of variation is already present in a population, i.e. as an important component of intra-population variance. Hence, the main direction of intra-population variance has been suggested to constitute a "line of least resistance" to evolution (Schluter, 1996). Evaluating which directions of variance are produced preferentially, their stability over time and space, and their relationship with developmental processes, may thus shed precious light on the role and strength of intrinsic constraints in directing short and long term evolution (e.g. Marroig and Cheverud 2001, 2005; Renaud et al. 2006; Hunt 2007). The aim of the present study is to exemplify how studying the main directions of phenotypic variance, as potential lines of least evolutionary resistance, might help for a better understanding of morphological evolution. Starting from a conceptual background, including methodological issues, case studies of rodent evolution will be used to illustrate the potential of this type of investigation in evolutionary studies.

Conceptual background

The idea that the main direction of variance may constitute a line of least resistance to evolution was first proposed for genetic variance (Schluter, 1996). The direction of greatest genetic variation (or $Gmax$) corresponds to the major axis of the genetic variance-covariance (VCV) matrix, or $\mathbf{G}$ matrix. This role of $Gmax$ as line of least evolutionary resistance was supported in several studies (e.g. Bégin and Roff 2004; Steppan et al. 2002; McGuigan et al. 2005).

An accurate assessment of the role of the $\mathbf{G}$ matrix and $Gmax$ in evolution requires well known genealogies (Steppan et al., 2002),

*Corresponding author
Email address: Sabrina.Renaud@univ-lyon1.fr (Sabrina Renaud)

which are difficult or almost impossible to obtain in wild populations. In contrast, evaluating the phenotypic variance-covariance matrix, or $\mathbf{P}$ matrix, requires measurements of traits in a sample of individuals from a population (Fig. 1A). This type of information may be much easier to obtain. Thus, using the $\mathbf{P}$ matrix as a surrogate to the $\mathbf{G}$ matrix potentially greatly expands the range of evo-devo applications by making it possible to study both wild populations (Cheverud, 1988; Ackermann and Cheverud, 2000; Marroig and Cheverud, 2001), and fossils, for which a direct estimate of the $\mathbf{G}$ matrix is generally impossible (Renaud et al., 2006; Hunt, 2007).

The $\mathbf{P}$ matrix is related to the $\mathbf{G}$ matrix by the equation $\mathbf{P} = \mathbf{H} \cdot \mathbf{G}$, where $\mathbf{H}$ is the heritability matrix (Polly 2004, and references therein). Data about heritability of morphometric characters are scarce, especially in the context of geometric morphometrics. Univariate estimates provide intermediate values (Cheverud, 1988). Multivariate estimates suggest that indeed, $\mathbf{P}$ is significantly correlated to $\mathbf{G}$ and that their main axes of variation ($Pmax$ and $Gmax$) also have similar directions (Siahsarvie, 2012). Such results provide support for the use of the $\mathbf{P}$ matrix as a surrogate of the $\mathbf{G}$ matrix in evolutionary studies. Besides, the $\mathbf{P}$ matrix is interesting in itself, as it contains information not only on the genetic variance but also on non-heritable, environmental and developmental components, which are a central focus of evo-devo studies.

The study of the main direction of variance ($Pmax$ or $Gmax$) provides a conceptual and methodological framework to bridge the gap between different evolutionary scales. $Pmax$, estimated at the intra-population level, can be compared to long-term evolutionary trajectories (Fig. 1A, B) to assess the role of genetic/developmental constraints. The main direction of variance and its relation with evolutionary trajectories can further be interpreted in the context of adaptive landscapes (Fig. 1C, D). This representation plots the fitness (z-axis) as a function of two traits (life-history traits, or morphological traits, which may be axes from geometric morphometrics) (Arnold et al., 2001; Polly, 2008). High adaptive regions are represented by peaks, and unfavorable areas, in terms of fitness, by valleys (Fig. 1C). The evolution of populations can be described by trajectories in the adaptive landscape, with the nearest peak attracting populations towards a local optimum of adaptation (e.g. Arnold et al. 2001; Polly 2008). However, the main direction of variance might sometimes constrain and sometimes facilitate evolution towards a one or the other among neighboring peaks (Fig. 1D). The relation between patterns of variance and observed evolutionary trajectories in the adaptive landscape is seldom investigated in real cases because of the difficulties in measuring fitness changes due to subtle multivariate morphometric variations.

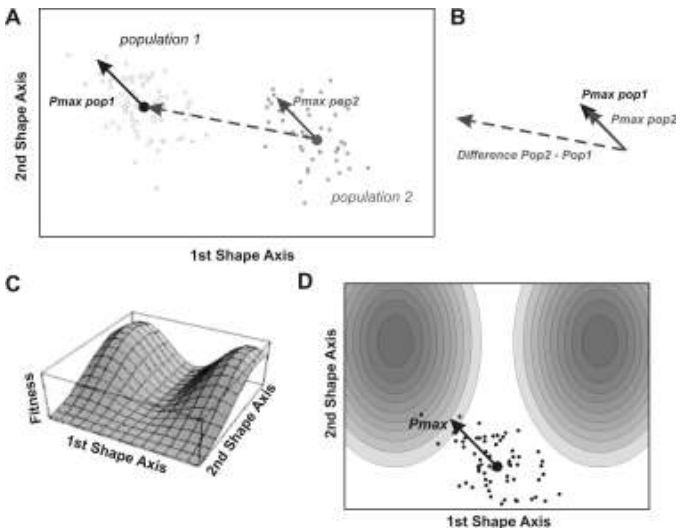


Figure 1 – . Theoretical illustration of the geometric framework for the study of $Pmax$ and its relationship with the adaptive landscape. A) $Pmax$ and direction of evolution in the morphometric space. A set of morphometric variables can be summarized using independent axes representing directions of variance of decreasing magnitude (1st, 2nd shape axes, etc.). A population can be visualized as a cloud of points in this morphospace. The direction of maximum dispersion of this population is represented by $Pmax$ (first eigenvector of the intra-group VCV matrix). B) The direction of main variance in two populations can be compared using the correlation between their $Pmax$ vectors. $Pmax$ can be further compared with evolutionary directions, which can be evaluated as the difference between group means of two populations, or by the first axis of variation among the means of a set of populations. C) Any morphotype in the morphospace is characterized by its fitness value. An adaptive landscape represents the variations in fitness (values along the vertical z-axis) as a function of traits (for instance along the 1st and 2nd axes of the morphospace) in a population. Two adaptive peaks are shown in the example. D) The variation present in a population (e.g, $Pmax$) can promote evolution towards one or the other adaptive peak.

Methodological issues

The main direction of phenotypic variance ($Pmax$) can be estimated from quantitative morphological variables by computing the major axes of their variance-covariance (VCV) matrix (the $\mathbf{P}$ matrix). This corresponds to performing a principal component analysis on the variation within the considered sample (e.g. a population, or a fossil assemblage). Successive principal axes describe statistically independent directions of variation. The first one (V1, or $Pmax$) describes the greatest intra-group variance, the second one (V2) describes the second most important direction of intra-group variance, and so forth. Several populations, or species, can be analyzed and represented in a morphospace. Their corresponding $Pmax$ can be projected and compared in this space (Fig. 1A). The direction of $Pmax$ in the different groups can also be quantitatively compared (Fig. 1B) using vector angles (the arccosine of the inner product of the two vector elements). The inner product ranges between -1 (vectors pointing in totally opposite directions) and $+1$ (vectors perfectly pointing in the same direction), similar to simple correlation. Comparing $Pmax$ to other trajectories requires estimating their direction: (1) as difference between two endpoints in evolution (for instance, the difference between an ancestor and its descendent); (2) as the main direction of inter-group variation; or (3) as a direction of morphological change set by its covariation with other factors (e.g., environmental gradients, diet variation, etc.). For instance, the morphological effect of a treatment (e.g., mice bred on standard food vs. mice fed exclusively on soft food) or of a genetic mutation (e.g., a normal strain vs. a genetically manipulated one) can be summarized by a vector connecting the mean of the "control" population to the mean of the "treated" population (e.g. Renaud et al. 2010). The significance of the angle between two vectors is finally estimated using non-parametric models. Among these, a fairly straightforward procedure is to conduct simulations to compute angles between random vectors of the same dimensionality as those being tested (Klingenberg, 1996; Renaud et al., 2006; Marroig and Cheverud, 2010). The corresponding distribution of angles simulates the null hypothesis of no relationship between vectors. If the observed angle is an outlier relative to this distribution, then it can be concluded that it is significantly smaller than expected by chance. A drawback is that random vectors may not accurately represent the distribution of real vectors in the morphospace as these are likely to share some common structure which is not taken into account by the simulation.

An alternative model for the null hypothesis could be using the correlations among a set of real morphometric vectors assumed to randomly explore all directions of the morphometric space (Boell et al., 2011). This approach would take into account the commonalities between vectors describing similar morphological features (say, a rodent mandible). A drawback could be that the distribution of the correlations determining the null hypothesis depends on the set of vectors chosen as a reference. If they are not distributed at random in the morphological space, the vectors will be themselves correlated and will not provide an adequate distribution for assessing the correlation of other vectors of interest. Both of these approaches are designed to compare vectors, such as $Pmax$ of two or more groups. However, it is important to also consider the structure of the entire $\mathbf{P}$ matrix, which can be compared using a Mantel-test. The degree of similarity between matrices can also be evaluated using common principal component analyses (CPCA). Using this method, matrices can be shown to be related in different ways: proportional (when eigenvectors are equal and eigenvalues proportional); characterized by common prin-

cipal components (with equal eigenvectors equal but different eigenvalues); or completely unrelated (with both different eigenvectors and eigenvalues) (e.g. Arnold et al. 2008).

All these methods for comparing vectors and matrices, however, do not take into account the phylogenetic relatedness between populations and species, an issue that will require the development of techniques within the broader context of phylogenetic comparative methods (Klingenberg and Gidaszewski, 2010). Also, these tests assume that vectors or matrices are reliably estimated. Sampling error, however, may severely affect estimates of means, variances, and angles (Polly, 2005; Cardini and Elton, 2007). The uncertainty in the estimates of these parameters can be assessed using bootstrap methods.

Case studies: the molar tooth and the mandible of murine rodents

Rodents are the most diverse order of mammals, with ca. 2000 species including nearly half of all mammalian species. Among them, the subfamily of murine rodents (Murinae, or Old World mice and rats) includes today ca. 120 genera and 550 species (Wilson and Reeder, 2005). Their radiation involved numerous morphological and life-history traits, among which the diversification in diet caused considerable variation in the selective pressure on morphological traits related to food processing, such as teeth and mandibles (e.g. Misonne 1969; Michaux 1971; Michaux et al. 2007) (Fig. 2). The house mouse (*Mus musculus*) belongs to the Murinae, and, as an emblematic laboratory model, a rich background on its genetics and development is available from experimental studies (e.g. Klingenberg et al. 2001, 2003; Workman et al. 2002; Shimizu et al. 2004; Kassai et al. 2005; Kavanagh et al. 2007; Boell et al. 2011).

The first model investigates the evolution of molar tooth shape in fossil and modern representatives of murine rodents. The following issues will be addressed: (1) Is the main direction of variance conserved across lineages? This is a prerequisite for a potential role as a line of least evolutionary resistance. (2) Does the main direction of variance actually parallel the evolutionary trajectory along a lineage? This provides correlational evidences for $Pmax$ being a line of least resistance to evolution. (3) If the main direction of variance is shared across lineages, and serves as a line of least evolutionary resistance, can it contribute to facilitate parallel evolution in different lineages? (4) Both selection and random processes can "surf" on these lines of least resistance. Are there means to disentangle their signature on morphological evolution? (5) How does the main direction of phenotypic variance relate to developmental processes and function? These questions will be answered using molar teeth, which mineralize early in development and are not prone to change with late growth except for wear. Whether plasticity in bones affects the main direction of variance and its role as line of least resistance to evolution will be, in contrast, investigated using the house mouse mandibles, which are subject to remodeling throughout life.

The material investigated therefore includes a set of fossil and modern first upper molars (UM1) of murine rodents, and mandibles of modern house mice (Fig. 2). Molar samples include specimens of murine rodents from Western Europe from the Miocene to present day (Tab. 1). They document the molar shape evolution along three fossil lineages which, starting with an ancestral form, *Progonomys*, lead independently to *Stephanomys* (Renaud et al., 1996, 2005, 2006), to *Paraethomys* (Renaud et al., 1999a) and to the wood mouse *Apodemus sylvaticus* (Renaud et al., 2005). The paleontological record was completed by two modern populations of wood mice, as well as two populations of the house mouse *Mus musculus domesticus*, which are used to exemplify evolution on islands (Renaud et al., 2011). Mandible data, in contrast, are from a sample of laboratory mice from an outbred strain (OF1), bred in controlled conditions at the PBES (Ecole Normale Supérieure, Lyon, France) and sacrificed at various ages from weaning (22 days) up to six months of age. This cross-sectional ontogenetic series was compared to variation in a sample from a natural population (Gardouch, France).

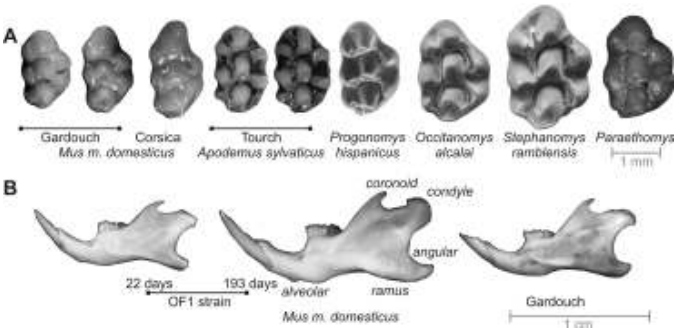


Figure 2 – Examples of first upper molars (A) and mandibles (B) of murine rodents considered in this study. A) First upper molars (lingual side to the right) in various modern and fossil murine rodents. From left to right, teeth exemplify variation in a house mouse continental population (Gardouch, France); the insular population of Corsica; variation in a wood mouse population (Tourch, France); evolution along the *Stephanomys* lineage, and *Paraethomys*. B) House mouse mandibles. The two mandibles on the left visualize ontogenetic variation in the OF1 laboratory strain (left, a specimen sacrificed at weaning; right, a six-months old specimen). The mandible to the right corresponds to a wild population (Gardouch, France).

Shape was quantified using a Fourier analysis of the 2D outline (basis of the crown for the molar and labial view of the bone for the mandible). Using this approach, each outline was described by successive trigonometric functions of decreasing wavelength, the harmonics. Each was weighted by Fourier coefficients constituting shape variables after size standardization. Consideration of the first seven harmonics appeared as a satisfactory compromise between information content and number of variables for both characters (e.g. Renaud and Michaux 2007). The 14 resulting Fourier coefficients (2 Fourier coefficients per 7 harmonics) were used as shape variables.

The main direction of phenotypic variance, $Pmax$, was calculated based on the variance-covariance (VCV) matrix of the 14 shape variables. It was evaluated at the intra-group level, with a group corresponding to a population of modern specimens, an assemblage of fossil teeth, or a taxon including several fossil deposits or modern populations. Directions of evolution were calculated for each lineage as the first axis of the inter-group VCV matrix, calculated on the group means.

$\mathbf{P}$ matrices were compared using Mantel t-tests. Similarity between vectors ($Pmax$ and evolutionary directions) was assessed by comparing their observed correlation R to the distribution of R from fifty thousand simulated random vectors. For vectors of 14 elements, this provided the following significance threshold values for the absolute value of R (a significant probability meaning that the observed R is larger than expected based on the distribution of R between random vectors): $p < 0.01$, $R = 0.651$ (*); $p < 0.001$, $R = 0.770$ (**); $p < 0.0001$, $R = 0.860$ (***). Note that the absolute value of R was considered, because the +/- direction of $Pmax$ (and of any eigenvector) is arbitrary.

Impact of sampling on Pmax estimate

As a preliminary analysis, in order to investigate the effect of sampling on the estimate of the $\mathbf{P}$ matrix and $Pmax$, the structure of morphological variance-covariance was computed in two samples of house mice (Fig. 3) and its variation assessed by bootstrapping. The samples were molars from the Fango population in Corsica ($N = 53$) and mandibles from the French Gardouch population ($N = 68$). Each sample was bootstrapped 100 times. The bootstrap procedure was repeated in random subsamples with $N = 50$ (for the Gardouch population starting from $N = 68$), $N = 25$ and $N = 10$. This demonstrated that $Pmax$ from bootstrapped samples were in a vast majority of cases significantly correlated to the observed $Pmax$ (Tab. 2). However, when N decreased, some of the estimates of $Pmax$ show large differences and become inaccurate. The percentage of variance explained by $Pmax$ varied considerably even in bootstrapped samples with the original sample size (Tab. 2). It tended to be slightly overestimated when sample size decreased. Thus, overall, bootstrap analyses confirm previous findings suggesting that reliably assessing $Pmax$ and $\mathbf{P}$ matrices does require large number of specimens per population (Prôa et al., 2013).

Table 1 – Sampling of the fossil deposits and modern localities that delivered the first upper molars (UM1) considered in this study. Age in million years is provided for the fossil localities, together with the number of first upper molars measured (UM1). Data from Renaud et al. 1996, 1999a,b, 2005, 2006, 2011.

Group/Lineage	Locality	Abbreviation	Genus	Species	Age (myr)	UM1
Stephanomys	La Roma 4B	ROM4B	*Progonomys*	*hispanicus*	9.6	4
	La Roma 4C	ROM4C	*Progonomys*	*hispanicus*	9.5	8
	Masia Del Barbo 2B	MBB	*Progonomys*	*hispanicus*	9.2	20
	Peralejos D	PERD	*Progonomys*	*hispanicus*	8.7	16
	Dionay	DIO-PH	*Progonomys*	*hispanicus*	8.6	15
	Puente Minero	PM	*Occitanomys*	*sondaari*	8.3	20
	Tortajada A	TOA	*Occitanomys*	*sondaari*	8.1	20
	Masada Del Valle 2	MDV2	*Occitanomys*	*adroveri*	7.3	20
	Concud 3	CC3	*Occitanomys*	*adroveri*	7.0	20
	Los Mansuetos	LM	*Occitanomys*	*adroveri*	6.9	20
	Valdecebro 3	VDC3	*Stephanomys*	*ramblensis*	6.3	21
	Las Casiones	KS	*Stephanomys*	*ramblensis*	6.1	20
	La Gloria 4	GLO4	*Stephanomys*	*dubari*	5.9	12
	Castelnou 3	C3	*Stephanomys*	*dubari*	5.6	15
	La Tour	LT	*Stephanomys*	*dubari*	5.6	5
	Sète	STE-SD	*Stephanomys*	*donnezani*	3.1	79
	Lo Fournas 13	LF13	*Stephanomys*	*donnezani*	3.0	30
	Balaruc 2	BAL2-SC	*Stephanomys*	*calveti*	2.7	44
	Pla De La Ville	PLV-SC	*Stephanomys*	*calveti*	2.5	101
	Seyne	SEY-ST	*Stephanomys*	*thaleri*	2.5	30
	Moreda 1B	MOR	*Stephanomys*	*minor*	2.4	60
	Balaruc 6	BAL6	*Stephanomys*	*thaleri*	2.3	30
	Lo Fournas 4	LF4	*Stephanomys*	*thaleri*	2.0	30
	Iles Medas	ILM	*Stephanomys*	*balcellsi*	1.9	44
	Casablanca 1	CAS	*Stephanomys*	*progressus*	1.8	30
Apodemus	Dionay	DIO-PL	*Parapodemus*	*lugdunensis*	8.6	18
	Sète	STE-AD	*Apodemus*	*dominans*	3.1	39
	Balaruc 2	BAL2-AD	*Apodemus*	*dominans*	2.7	43
	Pla De La Ville	PLV-AD	*Apodemus*	*dominans*	2.5	10
	Seynes	SEY-AD	*Apodemus*	*dominans*	2.5	20
	Vergranne	VER	*Apodemus*	*sylvaticus*	0.45	13
	Orgnac 3	OR3	*Apodemus*	*sylvaticus*	0.35	8
	Montpellier	MTP	*Apodemus*	*sylvaticus*	modern	14
	Tourch	TOU	*Apodemus*	*sylvaticus*	modern	88
Paraethomys	Oued Tabia	OTAB	*Progonomys*	*cathalai*	9.5	3
	Afoud8	AF8	*Paraethomys*	sp.	5.2	2
	Wanou	WAN	*Paraethomys*	*miocaenicus*	7.8	2
	Khendek El Ouaich	KEO	*Paraethomys*	*miocaenicus*	7.7	2
	Amama 2	AMA2	*Paraethomys*	*miocaenicus*	7.6	1
	Azib	AZB	*Paraethomys*	*pusillus*	5.3	4
	Amama 3	AMA3	*Paraethomys*	*anomalus*	2.9	7
	Irhoud DV	IDV	*Paraethomys*	*darelbeidae*	1.0	30
	Sidi Abdallah 1	SABH1	*Paraethomys*	*rbiae*	1.5	1
	Irhoud Neand.	IRHN	*Paraethomys*	*filfilae*	0.6	13
Mus musculus	Gardouch	GARD	*Mus*	*musc. dom.*	modern	68
	Corsica	CO	*Mus*	*musc. dom.*	modern	62

Molar shape: Stability of Pmax across time and phylogeny

The phenotypic signature of conserved genetic/developmental constraints should be indicated by a relative invariance of $Pmax$ in different lineages (Badyaev and Foresman, 2000; Marroig and Cheverud, 2001). $Pmax$ and $\mathbf{P}$ matrix of the first upper molar (UM1) were estimated and shown to be conserved in two lineages spanning over 10 million years of evolution (Renaud et al., 2006) (Tabs. 3 and 4). A second study (Renaud et al., 2009) demonstrated that the same pattern of variance is conserved also when the wood mouse (*Apodemus sylvaticus*) is compared to the house mouse (*Mus musculus domesticus*), two species that diverged some 10 million years ago (Lecompte et al., 2008). Whatever the basic tooth shape characteristic of the species, the main direction of intra-population variance corresponds to a trend from narrow to broad teeth (Figs. 3 and 4).

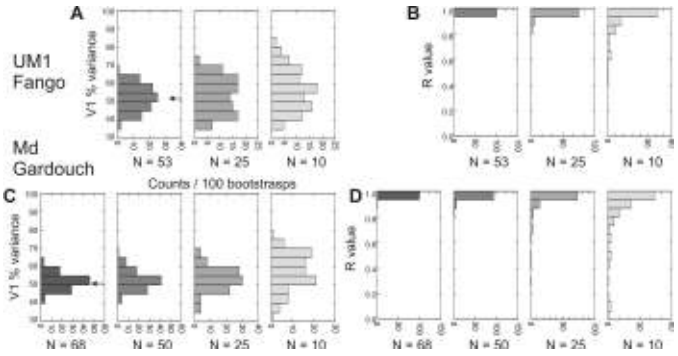


Figure 3 – Impact of sampling on the estimate of $Pmax$. Two populations with fairly large sample size were considered: Corsica for first upper molars (N = 53; A and B) and Gardouch for mandibles (N = 68; C and D). The initial samples were bootstrapped 100 times. The number of specimens in the bootstrapped samples was then progressively decreased to 50 (starting from 68 for the mandible set), 25 and 10 specimens, and bootstraps repeated. The first eigenvector was extracted in all corresponding VCV matrix, providing estimates of $Pmax$ that were compared to the initial $Pmax$. The distribution of the percentage of variance explained by $Pmax$ (A, C) and the correlation R with the original vector (B, D) are shown. Initial percentage of variance is represented by an arrow (UM1: 51.3%; mandible: 50.6%).

Pmax as a line to least resistance to molar shape evolution

Murine rodents diversified in Europe around 10 million years ago from the primitive, generalist *Progonomys* (Michaux, 1971; Renaud et al., 1999b). One lineage developed a peculiar dental specialization termed stephanodonty (Schaub, 1938), characterized by longitudinal crests connecting the transverse rows of cusps on the upper molars (Fig. 2) which slide in corresponding gutters on the occluding lower molars. Teeth also became larger and higher-crowned along the lineage. Supposed to increase masticatory efficiency, these morphological changes have been interpreted as adaptations to a more abrasive diet, probably grass. This interpretation is supported by comparative studies in extant murines with similar teeth and diets (Renaud and Michaux, 2004;

Table 2 – Effect of sampling on estimate of $Pmax$. $Pmax$ was computed in two samples with large sample size of house mice, for first upper molar shape variation (Fango, Corsica) and for mandible shape variation (Gardouch, mainland France). A bootstrap procedure (100 replications) was repeated in random subsamples decreasing from initial sample size to N = 50, N = 25, and N = 10. Mean, standard deviation (SD), maximum and minimum of the distribution in bootstrapped samples are provided for the percentage of variance explained by $Pmax$ (%$V1$) and the correlation (absolute value of R, in order to take into account the arbitrary +/- direction of the eigenvector) between bootstrapped and initial estimates of $Pmax$ (including the percentage of significant correlations [%*], with a threshold of R = 0.651 corresponding to $p < 0.01$.

		% V1				R			
		Initial		N = 53	N = 25	N = 10		N = 53	N = 25	N = 10
UM1 Fango	Mean	51.3		52.4	53.6	56.6		0.988	0.972	0.910
	SD			6.8	9.5	11.1		0.012	0.034	0.111
	Max			67.0	71.4	82.7		0.999	0.998	0.997
	Min			37.6	36.5	36.0		0.943	0.839	0.463
	% *							100	100	95
Md Gardouch			N = 68	N = 50	N = 25	N = 10	N = 68	N = 50	N = 25	N = 10
	Mean	50.6	52.0	52.5	53.3	57.1	0.981	0.974	0.923	0.829
	SD		4.2	4.9	6.4	9.6	0.017	0.024	0.113	0.220
	Max		63.3	66.6	69.4	76.0	0.997	0.998	0.996	0.989
	Min		41.6	40.6	38.6	34.9	0.909	0.863	0.274	0.057
	% *						100	100	97	86

Renaud et al., 2005). The origin of these specialized phenotypes has been related to climatic changes, which caused a shift from dominantly closed landscapes towards more open environments (Fox and Koch, 2003; deMenocal, 2004).

The possible role of $Pmax$ in constraining morphological evolution in response to these environmental changes was tested by comparing $Pmax$ with the directions of molar shape changes in the lineage (Renaud et al., 2006). Indeed, the lineage leading from *Progonomys* to *Stephanomys dubari* (from around 10 to 5 million years ago) evolved along a direction of shape change parallel to $Pmax$ in the ancestor population (Fig. 4) (correlation between inter-group $V1$ and $Pmax$ of *Progonomys hispanicus*, R = 0.896***). Yet, the same lineage also showed one case of departure from a model of evolution along lines of least resistance: the transition from the Miocene *Stephanomys* to the Pliocene representatives of the genus (between 7 and 3.5 million years) implied a drastic change of direction along an evolutionary trajectory statistically independent from the ancestral $Pmax$ (Fig. 4A). As a consequence, the direction of evolution along the whole lineage (*Stephanomys*-total on Fig. 4B) was only marginally correlated with the $Pmax$ in the ancestral population (correlation between inter-group $V1$ and $Pmax$ *Progonomys hispanicus*, R = 0.635) (Fig. 4B). An alternation of evolutionary modes, either along lines of least resistance or not, seems to suggest changes in the selection regime. Responses to weak or even intermediate selective pressures might be facilitated along the lines of least resistance, since they represent the major pattern of covariation among phenotypic traits. Yet, in the case of strong selection in favor of phenotypes expressing rare covariation among traits, constraints can be overriden (Beldade et al., 2002) and evolution might follow directions unrelated to the lines of least resistance.

Based on paleoenvironmental proxies (Zachos et al., 2001; Fox and Koch, 2003), we may infer that the environmental trend driving morphological evolution along the *Stephanomys* lineage was regular and of limited magnitude from 10 to 6 million years. The climatic trend accelerated afterwards leading to extremes of variation in the Pleistocene. An initial response, occurring along the lines of least resistance and mainly corresponding to a broadening of the teeth, might have been sufficient to deal with small environmental changes. In contrast, the evolution of the stephanodont pattern away from the direction of least resistance set by the main structure of variance and covariance in teeth morphology may have been crucial for adaptation to the much more pronounced change in the environment which occurred later in the history of the lineage. This extreme specialization was an evolutionary dead-end, and *Stephanomys* did not survive the extreme Pleistocene climatic fluctuations and went extinct about 1.2 million years ago. That specialists might be more prone to extinction in times of environmental change seems to be a general occurrence in the evolution of life on Earth (e.g. Leonard et al. 2007) and one which is likely important to understand how living species are and will be affected by global changes in our climate and environment (Clavel et al., 2011).

Lines of least resistance and parallel evolution

Related species, especially if under similar ecological and/or environmental pressures, are expected to show a variable extent of parallel evolution. Parallelism in evolution might become especially pronounced if genetic and developmental constraints, common to different lineages, channel evolution towards a similar preferential direction. Thus, we compared molar shape evolution along several loosely related lineages to estimate whether they occurred in a similar direction which is consistent with our assessment of $Pmax$. Evolving in the same time period as *Stephanomys* from another representative of the primitive genus *Progonomys*, a lineage developed, mainly in North Africa from the Miocene to the Pleistocene, which led to an independent acquisition of stephanodont crests (Jaeger et al., 1975) in *Paraethomys* (Fig. 2). This genus underwent an evolutionary acceleration between 3 and 1 million years ago, probably in relation to climate change, and went extinct soon after, in the Late Pleistocene (Renaud et al., 1999a). Beyond the acquisition of stephanodont crests, which suggests a common adaptive response to grass eating, it is still an open question whether the evolution of molar shape in *Paraethomys* occurred along lines of least resistance, which might have been the same as the one leading to the unusual morphology of *Stephanomys*.

To answer this question, we compared directions of evolution in the two lineages (Fig. 4). Although characterized by idiosyncratic molar shape and independent evolution, the *Paraethomys* lineage displayed a direction of evolution highly correlated with that of the *Stephanomys* lineage (directions of evolution estimated by inter-group $V1$; *Paraethomys* vs. total *Stephanomys* lineage: R = 0.815**; vs. early *Stephanomys* lineage: R = 0.926***). This direction of evolution is similar to that of $Pmax$ estimated in various murine species: in the ancestor population of the *Stephanomys* lineage (*Progonomys hispanicus*: R = 0.875***), the wood mouse (*Apodemus sylvaticus* Tourch, R = 0.877***), and even the house mouse (Gardouch population: R = 0.844**). Thus, overall, these results are consistent with the idea of commonalities in processes (genetic and/or developmental ones), which might facilitate parallel evolution in related linages.

Selection and random processes "surfing" along lines of least resistance

Phenotypic evolution is in essence multivariate. Focusing exclusively on $Pmax$ (= $V1$ of the $\mathbf{P}$ matrix) neglects subsequent components of the variation ($V2$, $V3$, etc.), which may also represent significant directions of evolutionary changes. Selection might tend to favor evolution mostly along one specific direction, which seems often to coincide, to a large extent, with $Pmax$. Random processes, however, should not occur along preferential trajectories and changes should simply be proportional to the variance-covariance structure in the population. Thus, one can try to disentangle the effects of random processes from those of selection by comparing the proportions of variance on successive axes between intra- and inter-group VCV matrices (Roff, 2000; Ackermann and Cheverud, 2004; Arnold et al., 2008). With this aim, we compared

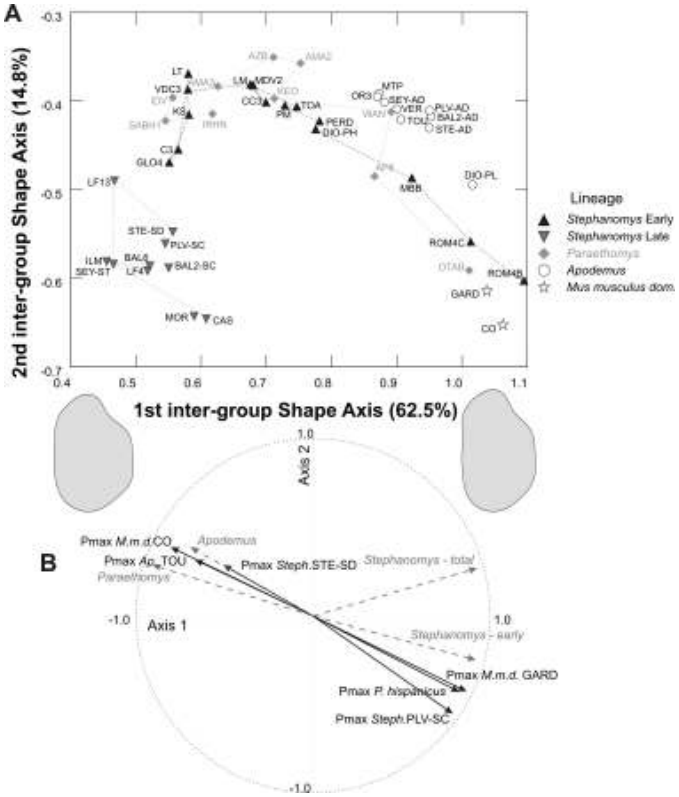


Figure 4 – $Pmax$ and evolution of the first upper molar of murine rodents. A) Temporal and phylogenetic differentiation of the first upper molar along three lineages of murine rodents: *Stephanomys*, *Paraethomys*, and *Apodemus*, and two house mouse populations. Each symbol corresponds to the mean shape of a population, plotted in a morphospace defined by the first two principal axes of the total inter-group variation. Along the first axis, molar outlines visualize the shape changes from a PC1 score of 0.0 to one of 2.0 (0.5 for PC2). B) Relationships between $Pmax$ (first vector of intra-group variance, full black arrows) and directions of evolution (first vector of inter-group variance, dotted grey arrows). All vectors are projected on the axes represented on (A). Vectors pointing in a similar directions (+ and - arbitrary) suggest that molar shape changes share common components. Note that the morphospace was constructed using 14 shape variables; the correlation of the vectors is thus expressed on a 14-dimensional space. All vectors are here scaled to unity (shown using a circle of radius 1): vectors shorter than unity point into a multivariate direction out of the plane (e.g. $Pmax$ of *Stephanomys* in STE-SD).

(Tab. 5): (1) intra-group variance in several populations/species; (2) inter-group variance in lineages, where selection is assumed to have played a major role; (3) inter-group variance in lineages, where random processes are more likely to have occurred.

The first axis describing the inter-group variance along lineages evolving under directional selection (e.g. *Stephanomys* or *Paraethomys*) is expected to represent significantly more variance than its counterpart at the intra-group level. Indeed, $Pmax$ represented between 30% and 55% of variance in all groups considered (Tab. 5), an appreciably smaller percentage than the first axis of inter-group variance in *Paraethomys* (66%) and *Stephanomys* (77% in the early part of the lineage).

In contrast, the lineage of the wood mouse does not seem to have evolved under strong directional selection. After originating from the primitive *Progonomys*, it did not undergo much morphological change and evolved into a group of taxa related to the modern wood mouse (*Apodemus*), which is still living in Europe today (Michaux et al., 1997). Presumably, while *Stephanomys* colonized niches in the new open habitats, *Apodemus*, a generalist, survived the climatic fluctuations by tracking its forest habitat in a mosaic landscape (Renaud et al., 2005). If this hypothetical reconstruction is correct, it is reasonable to assume that stabilizing selection maintained a fairly constant pattern of tooth morphology in this lineage. Consistent with this expectation, the first axis of inter-group variance in *Apodemus* explained a comparable amount of variation (47%) to those within groups (Tab. 5).

In conclusion, the approach exemplified in this study seems promising and provides clues on the selection regime which might have been the main driver of evolution in these groups (Ackermann and Cheverud, 2004; Marroig and Cheverud, 2010). It is important to bear in mind, however, that this approach requires a large number of groups to reliably estimate the matrix of inter-group variances and its structure, and an extensive sampling of specimens for estimating the intra-group variance-covariance matrix, which will be used to compute the percentages of variance explained by different components (Fig. 3).

Beyond lines of least resistance: genetics and function

Some phenotypes seem more widespread than others in a population. This might be because the corresponding genotype is more common or because developmental processes are channelling phenotypic variation in a specific direction. Indeed, using quantitative trait loci (QTL) analyses, it has been shown that the mouse mandible is characterized by some recurrent patterns of shape change which are associated with specific genetic traits (Klingenberg et al., 2001). In the murine first upper molar, the pattern of associated with $Pmax$ corresponds to a trend from slender to broad molars (Renaud et al., 2006, 2009) (Figs. 2, 4). The position of the cusps is determined early during embryogenesis by the position of signaling centers (enamel knots), and the size of their surrounding inhibitory field (Jernvall, 2000). A broadening of the molar can be triggered by a concerted increase in breadth of the developmental field, which will later become a tooth, together with an increase in the lateral spacing of the primary enamel knots and their inhibitory field. This might be mediated by changes in the regulation of genes controlling tooth development (e.g. Mustonen et al. 2003). Such effects should be global and concern all molars and all cusps of a tooth. Indeed, strong integration was found in all six molars (upper and lower

Table 3 – Correlation among $\mathbf{P}$ matrices, estimated using Mantel tests. Above the diagonal, p values; below the diagonal, R values. In bold significant probabilities.

R/p	*Prog.hisp.*	*Steph.* STE	*Steph.* PLV	*Apod.* TOU	*M.m.d.* GARD	*M.m.d.* CO
Prog.hisp.	-	**0.005**	**< 0.001**	**0.004**	**< 0.001**	0.088
Steph. STE	**0.272**	-	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
Steph. PLV	**0.512**	**0.610**	-	**< 0.001**	**< 0.001**	**< 0.001**
Apod. TOU	**0.285**	**0.400**	**0.392**	-	**0.001**	**< 0.001**
M.m.d. GARD	**0.570**	**0.547**	**0.513**	**0.339**	-	**< 0.001**
M.m.d. CO	0.142	**0.556**	**0.475**	**0.361**	**0.566**	-

Table 4 – Correlation between $Pmax$ in various groups. Below the diagonal, correlation of the vectors R = inner product of the two vector elements. Above the diagonal, significance of the correlation, obtained by comparing the observed R to the distribution of R between random vectors. In bold significant correlations (p < 0.01, R = 0.651).

R/p	*Prog.hisp.*	*Steph.* STE	*Steph.* PLV	*Apod.* TOU	*M.m.d.* GARD	*M.m.d.* CO
Prog.hisp.	-		***	***	***	*
Steph. STE	0.461	-		*		***
Steph. PLV	**0.960**	0.545	-	***	***	**
Apod. TOU	**0.860**	**0.749**	**0.891**	-	***	***
M.m.d. GARD	**0.939**	0.629	**0.963**	**0.917**	-	**
M.m.d. CO	0.721	**0.876**	**0.779**	**0.898**	**0.848**	-

Table 5 – Structure of the inter-group and intra-group variance in different cases of evolution of the first upper molar in murine rodents. Upper panel, inter-group variance (VCV matrix) estimated on group means of a set of fossil and/or modern populations documenting the evolution along the lineages of *Stephanomys* (total: from *Progonomys* to *Stephanomys progressus*; early: from *Progonomys hispanicus* to *Stephanomys dubari*), *Paraethomys*, and *Apodemus*. Lower panel, intra-group variance ($\mathbf{P}$ matrix) in a set of fossil populations from the *Stephanomys* lineage, and in modern populations of the wood mouse (*Apodemus sylvaticus*, Tourch, France) and the house mouse (*Mus musculus domesticus*, Gardouch and Corsica, France). N, number of items (specimens for intra-group and group means for inter-group) used for the calculation of the VCV matrix. $V1$ (= $Pmax$), $V2$, $V3$: % of variance explained by the first three eigenvectors.

	Model		N	V1	V2	V3
Inter-group	*Stephanomys*	total	24	59.9	23.0	7.9
	Stephanomys	early	14	77.4	15.2	2.3
	Paraethomys	total	10	66.2	14.5	6.7
	Apodemus	total	9	47.4	28.0	11.7
Intra-group	*Prog. hisp.*	total	63	54.3	15.6	8.5
	Stephanomys	STE	79	35.5	25.8	9.6
	Stephanomys	PLV	101	34.1	26.8	11.3
	Apodemus	Tourch	88	36.8	14.6	11.1
	Mus musc. dom.	Gardouch	68	38.9	19.9	12.7
	Mus musc. dom.	Corsica	62	49.2	17.0	10.5

ones) in mice Renaud et al. (2009), such that when upper teeth become broader, lower molars do the same.

A broadly similar pattern of variation was also found in insular populations of the house mouse. Compared with continental populations, the first upper molar of Corsican mice is slender. This elongation is nevertheless not related to a narrowing of the latitudinal rows of cusps on the tooth, but to a local, anterior elongation of the tooth, to the point of the appearance of an additional cusplet (Renaud et al., 2011). This anterior elongation was a recurrent pattern in several insular populations. It was speculated that it might involve a differential incorporation of a vestigial bud, anterior to the developmental field of the first molar, which usually aborts as the first molar forms (Prochazka et al., 2010; Renaud et al., 2011).

The morphological signature of the two patterns of variance (narrowing of the tooth vs. anterior elongation) suggests a discrepancy. One corresponds to a generalized effect on all molars (narrowing/broadening; Renaud et al. 2009), the other one to a localized change in the first upper molar (anterior elongation; Renaud et al. 2011). They were attributed to different candidate developmental mechanisms. Yet, they also seem to share some common features: they both affect the shape of the tooth without changing the longitudinal alignment of the cusps. Indeed, $Pmax$ in Corsican populations (related to the anterior elongation) is similar to $Pmax$ in continental populations of house mice and to that found in several other murines, that are not characterized by an anterior elongation (Tabs. 3 and 4). This might be explained by a functional constraint. Murine rodents are characterized by a longitudinal chewing movement (propalinal direction), which is achieved by arranging cusps in longitudinal rows that slide into gutters on the occluding tooth (Lazzari et al., 2008). This mechanism inevitably constrains the arrangement of the cusps because any change of the longitudinal arrangement would disrupt function and will therefore be strongly counter-selected. However, a global narrowing/broadening of the tooth, achieved by changing the spacing of the longitudinal rows in a concerted way between occluding teeth, does not perturb this arrangement. In a different way, the anterior elongation of the first upper molar is also consistent with functional requirements. In conclusion, the congruence between functional expectations, inter-group evolutionary trajectories, and $Pmax$ suggests that genetic and developmental systems coevolved in order to match the requirements of the propalinal masticatory movement (Butler, 1985; Lazzari et al., 2008).

Pmax and plasticity of the mouse mandible

In murine rodents, the molar tooth shape is determined during prenatal development and this is particularly evident for the first upper molar, which is the first to develop and the one which influences all the others in a cascade of spatial interactions along the molar row (Kavanagh et al., 2007). Once erupted, murine molar teeth remain unchanged throughout life except that they wear with use. Mandibles, in contrast, are bony structures and they are actively remodeled by their constant interactions with the muscles and other tissues during and after prenatal development (Katsaros et al., 2001; Mavropoulos et al., 2004, 2005). Mice for instance have only reached about 80% of their adult skull size at weaning (Zelditch et al., 2003), which leaves room for further growth and remodeling. The importance of shape change late during growth is illustrated using a sample of laboratory mice bred in controlled conditions (Fig. 5). All these animals had the third molars fully erupted and would have been considered as adult in a wild population. Their $Pmax$ (Fig. 5C) was largely in the direction of allometry (correlation of $Pmax$ with direction of allometric variation, estimated by regressing shape onto mandibular size, R = 0.997***) and similar to $Pmax$ of wild animals, with fully erupted teeth and hence considered as subadults and adults (Gardouch, France: R = 0.713*). Thus, for mandibles, which are highly plastic, $Pmax$ can strongly be influenced by growth and environmental factors. Shape changes among populations of a same species could occur following $Pmax$ by mere differences in their age structure, which could be enough to generate shape differences because of allometry. Size differences are likely to produce shape variation in mandibles even simply because of physics and the non-linear changes in the forces required during mastication to move mandibles of different sizes (Satoh, 1997; Cardini and Tongiorgi, 2003; Michaux et al., 2007). Size is seen as a highly labile evolutionary character (Bünger and Hill, 1999; Dupont and Holzenberger, 2003), and accordingly it might display rapid divergence among populations or related species (Nevo, 1989; Ganem et al., 1995; Dayan and Simberloff, 1998; Kingsolver and Pfennig, 2004; Cardini et al., 2007). The importance of size-related shape changes can thus makes $Pmax$ collinear with allometry within populations, so that size and allometric shape changes become a line of least evolutionary resistance (Marroig and Cheverud, 2005, 2010).

Because of interactions with muscles and surrounding tissues, mandible shape may also vary in response to environmental factors such as food consistency (e.g. Katsaros et al. 2001; Mavropoulos et al. 2004; Renaud et al. 2010). This plastic effect was demonstrated in laboratory mice and it was also shown to be collinear to allometry and $Pmax$ estimated in wild populations (Renaud and Auffray, 2010). The same areas of the mandible, especially the zone of insertion of the masticatory muscles, seem to recurrently emerge as very variable. It is thus

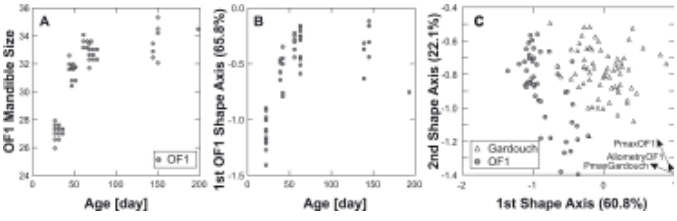


Figure 5 – Variation in mandible size and shape in the house mouse (*Mus musculus domesticus*). A) Mandible size increase with age of the animal in a laboratory strain (OFI). B) Mandible shape changes along growth of the same OFI mice. Shape is estimated by scores on $Pmax$. C) Morphospace including OFI mice and a wild-trapped population (Gardouch, France). $Pmax$ and allometry of OFI mice, and $Pmax$ of a wild population, were projected into a common morphospace (length of the vectors arbitrary).

strongly contributing to $Pmax$. However, this zone is also strongly involved in allometric shape changes, plastic response to food consistency, genetic (Klingenberg et al., 2001), and biogeographic variation (Renaud and Michaux, 2003; Siahsarvie et al., 2012). Such areas might be under strong functional requirements, making them prone to both genetic and plastic variation, and preferential targets of adaptive evolution. This illustrates that $Pmax$, even when including a component of plastic variation, might be relevant to infer complex evolutionary processes beyond patterns of morphological differences.

Concluding remarks

The geometric morphometric framework we used in this study has a great potential for investigating the role of intrinsic constraints in channeling evolution. Thus, it can help to bridge evolutionary studies at different scales, from micro- (intra- and inter-population variance) to macro-evolution (species and supra-specific differences). It can also be instrumental in identifying the mechanisms involved in evolutionary divergence (e.g. mutational effect, developmental mechanism, effect of a treatment). However, the mechanisms behind morphological change in evolution are diverse and complex, and might involve genetics, plasticity, developmental and environmental factors. Experimental studies will be needed together with descriptive approaches to fully understand the relative roles of these components and better disentangle processes from patterns.

References

Ackermann R.R., Cheverud J.M., 2000. Phenotypic covariance structure in tamarins (Genus *Saguinus*): a comparison of variation patterns using matrix correlation and common principal component analysis. American Journal of Physical Anthropology 111: 489–501.

Ackermann R.R., Cheverud J.M., 2004. Detecting genetic drift versus selection in human evolution. Proceedings of the National Academy of Sciences, USA 101: 17946–17951.

Arnold S.J., Bürger R., Hohenlohe P.A., Ajie B.C., Jones A.G., 2008. Understanding the evolution and stability of the G-matrix. Evolution 62: 2451–2461.

Arnold S.J., Pfrender M.E., Jones A.G., 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. Genetica 112–113: 9–32.

Badyaev A.V., Foresman K.R., 2000. Extreme environmental change and evolution: stress-induced morphological variation is strongly concordant with patterns of evolutionary divergence in shrew mandibles. Proceedings of the Royal Society, London B. 267: 371–377.

Bégin M., Roff D.A., 2004. From micro- to macroevolution through quantitative genetic variation: positive evidence from field crickets. Evolution 58: 2287–2304.

Beldade P., Koops K., Brakefield P.M., 2002. Developmental constraints versus flexibility in morphological evolution. Nature 416: 844–847.

Boell, L. Gregorova S., Forejt, J., Tautz D., 2011. A comparative assessment of mandible shape in a consomic strain panel of the house mouse (*Mus musculus*) – implications for epistasis and evolvability of quantitative traits. BMC Evolutionary Biology 11: 309.

Brakefield P.M., 2006. Evo-devo and constraints on selection. Trends in Ecology and Evolution 21: 362–368.

Bünger L., Hill W., 1999. Role of growth hormone in the genetic change of mice divergently selected for body weight and fatness. Genetical Research, Cambridge 74: 351–360.

Butler P.M., 1985. Homologies of molar cusps and crests, and their bearing of assessment of rodent phylogeny. In: Luckett W.P., Hartenberger J.-L. (Eds.). Evolutionary relationships among Rodents. A multidisciplinary analysis, Vol. 92. NATO ASI Series, Series A. Plenum Press, New York and London. 381–401.

Cardini A., Elton S., 2007. Sample size and sampling error in geometric morphometric studies of size and shape. Zoomorphology 126: 121–134.

Cardini A., Tongiorgi P., 2003. Yellow-bellied marmots (*Marmota flaviventris*) "in the shape space" (Rodentia, Sciuridae): sexual dimorphism, growth and allometry of the mandible. Zoomorphology 122: 11–23.

Cardini A., Jansson, A.-U., Elton S., 2007. A geometric morphometric approach to the study of ecogeographical and clinal variation in vervet monkeys. Journal of Biogeography 34: 1663–1678.

Cheverud J.M., 1988. A comparison of genetic and phenotypic correlations. Evolution 42: 958–968.

Clavel J., Julliard R., Devictor V., 2011. Worldwide decline of specialist species: toward a global functional homogenization? Frontiers in Ecology and the Environment 9: 222–228.

Dayan T., Simberloff D., 1998. Size patterns among competitors: ecological character displacement and character release in mammals, with special reference to island populations. Mammal Review 28: 99–124.

deMenocal P.B., 2004. African climate and faunal evolution during the Pliocene-Pleistocene. Earth and Planetary Science Letters 220: 3–24.

Draghi J.A., Whitlock M.C., 2012. Phenotypic plasticity facilitates mutational variance, genetic variance, and evolvability along the major axis of environmental variation. Evolution 66(9): 2891–2902.

Dupont J., Holzenberger M., 2003. Biology of Insulin-like growth factors in development. Birth Defects Research (Part C) 69: 257–271.

Fox D.L., Koch P.L., 2003. Tertiary history of C4 biomass in the Great Plains, USA. Geology 31: 809–812.

Ganem G., Granjon L., Ba K., Duplantier J.-M., 1995. Body size variability and water balance: A comparison between mainland and island populations of *Mastomys huberti* (Rodentia: Muridae) in Senegal. Experientia 51: 402–410.

Hunt G., 2007. Evolutionary divergence in directions of high phenotypic variance in the ostracode genus *Poseidonamicus*. Evolution 61: 1560–1576.

Jaeger J.-J., Michaux J., Thaler L., 1975. Présence d'un rongeur muridé nouveau, *Paraethomys miocaenicus* nov. sp., dans le Turolien supérieur du Maroc et d'Espagne. Implications paléogéographiques. Comptes Rendus de l'Académie des Sciences, série D 280: 1673–1676. [in French]

Jernvall J., 2000. Linking develoment with generation of novelty in mammalian teeth. Proceedings of the National Academy of Sciences, USA 97: 2641–2645.

Kassai Y., Munne P., Hotta Y., Penttilä E., Kavanagh K. D., Ohbayashi N., Takada S., Thesleff I., Jernvall J., Itoh N., 2005. Regulation of mammalian tooth cusp patterning by ectodin. Science 309: 2067–2070.

Katsaros C., Berg R., Kiliardis S., 2001. Influence of masticatory muscle function on transverse skull dimension in the growing rat. Journal of Orofacial Orthopedics 1: 5–13.

Kavanagh K.D., Evans A.R., Jernvall J., 2007. Predicting evolutionary patterns of mammalian teeth from development. Nature 449: 427–432.

Kingsolver J.G., Pfennig D.W., 2004. Individual-level selection as a cause of Cope's rule of phyletic size increase. Evolution 58: 1608–1612.

Klingenberg C.P., 1996. Multivariate allometry. In: Marcus L.F., Corti M., Loy A., Naylor G.J.P., Slice D.E. (Eds.). Advances in Morphometrics, Vol. 284. NATO ASI Series A: Life Sciences. Plenum Press, New York. 23–49.

Klingenberg C.P., Gidaszewski N.A., 2010. Testing and quantifying phylogenetic signals and homoplasy in morphometric data. Systematic Biology 59(3): 245–261.

Klingenberg C.P., Leamy L.J., Routman E.J., Cheverud J.M., 2001. Genetic architecture of mandible shape in mice: effects of quantitative trait loci analyzed by geometric morphometrics. Genetics 157: 785–802.

Klingenberg C.P., Mebus K., Auffray J.-C., 2003. Developmental integration in a complex morphological structure: how distinct are the modules in the mouse mandible? Evolution and Development 5: 522–531.

Lazzari V., Tafforeau P., Aguilar J.-P., Michaux J., 2008. Topographic maps applied to comparative molar morphology: the case of murine and cricetine dental plans (Rodentia, Muroidea). Paleobiology 34: 46–64.

Lecompte E., Aplin K., Denys C., Catzeflis F., Chades M., Chevret P., 2008. Phylogeny and biogeography of African Murinae based on mitochondrial and nuclear gene sequences, with a new tribal classification of the subfamily. BMC Evolutionary Biology 8: 199.

Leonard J.A., Vila C., Fox-Dobbs K., Koch P.L., Wayne R.K., Van Valkenburgh B., 2007. Megafaunal extinctions and the disappearance of a specialized wolf ecomorph. Current Biology 17: 1146–1150.

Marroig G., Cheverud J.M., 2001. A comparison of phenotypic variation and covariation patterns and the role of phylogeny, ecology, and ontogeny during cranial evolution of new world monkeys. Evolution 55: 2576–2600.

Marroig G., Cheverud J.M., 2005. Size as line of least evolutionary resistance: diet and adaptive morphological radiation in New World monkeys. Evolution 59: 1128–1142.

Marroig G., Cheverud J.M., 2010. Size as a line of least resistance II: Direct selection on size or correlated response due to constraints? Evolution 64: 1470–1488.

Mavropoulos A., Ammann P., Bresin A., Kiliardis S., 2005. Masticatory demands induce region-specific changes in mandibular bone density in growing rats. Angle Orthodontist 75: 625–630.

Mavropoulos A., Bresin A., Kiliardis S., 2004. Morphometric analysis of the mandible in growing rats with different masticatory functional demands: adaptation to an upper posterior bite block. European Journal of Oral Sciences 112: 259–266.

McGuigan K., Chenoweth S.F., Blows M.W., 2005. Phenotypic divergence along lines of genetic variance. The American Naturalist 165: 32–43.

Michaux J., 1971. Muridae (Rodentia) néogènes d'Europe sud-occidentale. Evolution et rapports avec les formes actuelles. Paléobiologie continentale, Montpellier II: 1–67. [in French]

Michaux J., Aguilar J.-P., Montuire S., Wolff A., Legendre S., 1997. Les Murinae (Rodentia, Mammalia) néogènes du Sud de la France: évolution et paléoenvironnements. Geobios, Mémoire Spécial 20: 379–385. [in French]

Michaux J., Chevret P., Renaud S., 2007. Morphological diversity of Old World rats and mice (Rodentia, Muridae) mandible in relation with phylogeny and adaptation. Journal of Zoological Systematics and Evolutionary Research 45: 263–279.

Misonne X., 1969. African and Indo-Australian Muridae. Evolutionary trends. Musée Royal de l'Afrique Centrale, Tervuren, Belgique. [in French]

Mitteroecker P., Bookstein F., 2009. The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans. Evolution 63: 727–737.

Mustonen T., Pispa J., Mikkola M.L., Pummila M., Kangas A.T., Pakkasjärvi L., Jaatinen R., Thesleff I., 2003. Stimulation of ectodermalorgan development by Ectodysplasin-A1. Developmental Biology 259: 123–136.

Nevo E., 1989. Natural selection of body size differentiation in Spiny mice, *Acomys*. Zeitschrift für Säugetierkunde 54: 81–99.

Polly P.D., 2004. On the simulation of the evolution of morphological shape: multivariate shape under selection and drift. Palaeontologia Electronica 7(7A): 1–28.

Polly P.D., 2005. Development and phenotypic correlations: the evolution of tooth shape in *Sorex araneus*. Evolution and Development 7: 29–41.

Polly P.D., 2008. Developmental dynamics and G-matrices: Can morphometric spaces be used to model phenotypic evolution? Evolutionary Biology 35: 83–96.

Pôa M., O'Higgins P., Monteiro L., 2013. Type I error rates for testing genetic drift with phenotypic covariance matrices: a simulation study. Evolution 67(1): 185–195. doi: 10.1111/j.1558-5646.2012.01746.x

Prochazka J., Pantalacci S., Churava S., Rothova M., Lambert A., Lesot H., Klein O., Peterka M., Laudet V., Peterková R., 2010. Patterning by heritage in mouse molar row development. Proceedings of the National Academy of Sciences, USA 107: 15497–15502.

Renaud S., Auffray J.-C., 2010. Adaptation and plasticity in insular evolution of the house mouse mandible. Journal of Zoological Systematics and Evolutionary Research 48: 138–150.

Renaud S., Michaux J.R., 2003. Adaptive latitudinal trends in the mandible shape of *Apodemus* wood mice. Journal of Biogeography 30: 1617–1628.

Renaud S., Michaux J., 2004. Parallel evolution in molar outline of murine rodents: the case of the extinct *Malpaisomys insularis* (Eastern Canary Islands). Zoological Journal of the Linnean Society 142: 555–572.

Renaud S., Michaux J.R., 2007. Mandibles and molars of the wood mouse, *Apodemus sylvaticus* (L.): integrated latitudinal signal and mosaic insular evolution. Journal of Biogeography 34: 339–355.

Renaud S., Auffray J.-C., de La Porte S., 2010. Epigenetic effects on the mouse mandible: common features and discrepancies in remodeling due to muscular dystrophy and response to food consistency. BMC Evolutionary Biology 10: 28.

Renaud, S. Auffray J.-C., Michaux J., 2006. Conserved phenotypic variation patterns, evolution along lines of least resistance, and departure due to selection in fossil rodents. Evolution 60: 1701–1717.

Renaud S., Benammi M., Jaeger J.-J., 1999a. Morphological evolution of the murine rodent *Paraethomys* in response to climatic variations (Mio-Pleistocene of North Africa). Paleobiology 25: 369–382.

Renaud S., Michaux J., Jaeger J.-J., Auffray J.-C., 1996. Fourier analysis applied to *Stephanomys* (Rodentia, Muridae) molars: nonprogressive evolutionary pattern in a gradual lineage. Paleobiology 22: 255–265.

Renaud S., Michaux J., Mein P., Aguilar J.-P., Auffray J.-C., 1999b. Patterns of size and shape differentiation during the evolutionary radiation of the European Miocene murine rodents. Lethaia 32: 61–71.

Renaud S., Michaux J., Schmidt D.N., Aguilar J.-P., Mein P., Auffray J.-C,. 2005. Morphological evolution, ecological diversification and climate change in rodents. Proceedings of the Royal Society of London, Biological Sciences (Series B) 272: 609–617.

Renaud S., Pantalacci S., Auffray J.-C., 2011. Differential evolvability along lines of least resistance of upper and lower molars in island mouse mice. PLoS One 6: e18951. doi: 10.1371/journal.pone.0018951

Renaud S., Pantalacci S., Quéré J.-P., Laudet V., Auffray J.-C., 2009. Developmental constraints revealed by co-variation within and among molar rows in two murine rodents. Evolution and Development 11: 590–602.

Roff D.A., 2000. The evolution of the G matrix: selection or drift? Heredity 84: 135–142.

Salazar-Ciudad I., Jernvall J., 2010. A computational model of teeth and the developmental origins of morphological variation. Nature 464: 583–586.

Satoh K., 1997. Comparative functional morphology of mandibular forward movement during mastication of two murid rodents, *Apodemus speciosus* (Murinae) and *Clethrionomys rufocanus* (Arvicolinae). Journal of Morphology 231: 131–142.

Schaub S., 1938. Tertiäre und Quartäre Murinae. Abhandlungen der Schweizerischen Paleontologischen Gesellschaft, Basel 61: 1–39. [in German]

Schluter D., 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50: 1766–1774.

Shimizu T., Oikawa H., Han J., Kurose E., Maeda T., 2004. Genetic analysis of crown size in the first molars using SMXA recombinant inbred mouse strains. Journal of Dental Research 83: 45–49.

Siahsarvie R., 2012. Comparaison de la divergence morphologique et génétique chez la souris domestique au cours de son expansion géographique. Thèse de Doctorat. Université Montpellier 2. [in French]

Siahsarvie R., Auffray J.-C., Darvish J., Bonhomme F., Claude J., 2012. Patterns of morphological evolution in the mandible of the house mouse *Mus musculus* (Rodentia: Muridae). Biological Journal of the Linnean Society 105: 635–647.

Skinner M.M., Gunz P., 2010. The presence of accessory cusps in chimpanzee lower molars is consistent with a patterning cascade model of development. Journal of Anatomy 217: 245–253.

Steppan S.J., Phillips P.C., Houle D., 2002. Comparative quantitative genetics: evolution of the G matrix. Trends in Ecology and Evolution 17: 320–327.

Wilson D.E., Reeder D.M., 2005. Mammals Species of the World: A Taxonomic and Geographic Reference. Smithsonian Institution Press, Washington.

Workman M.S., Leamy L.J., Routman E.J., Cheverud J.M., 2002. Analysis of quantitative trait locus effects on the size and shape of mandibular molars in mice. Genetics 160: 1573–1586.

Zachos J., Pagani M., Sloan L., Thomas E., Billups K., 2001. Trends, rythms, and aberrations in global climate 65 Ma to present. Science 292: 686–693.

Zelditch M.L., Lundrigan B.L., Sheets H.D., Garland T.J., 2003. Do precocial mammals develop at a faster rate? A comparison of rates of skull development in *Sigmodon fulviventer* and *Mus musculus domesticus*. Journal of Evolutionary Biology 16: 708–720.

Associate Editor: A. Cardini

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 94–102, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/6316/pdf
doi:10.4404/hystrix-24.1-6316

Research Article

Log-shape ratios, Procrustes superimposition, elliptic Fourier analysis: three worked examples in R

Julien Claude[a,*]

[a] *Institut des Sciences de l'Evolution de Montpellier, cc064, Université de Montpellier 2, 2 Place Eugène Bataillon, 34095 Montpellier cedex 5, France.*

Keywords:
rodents
morphometrics
Rattus
Asia
R

Article history:
Received: 18 June 2012
Accepted: 9 April 2013

Acknowledgements
I thank Yannick Chaval who organized body shape measurements in the CERoPath database, and who performed molecular barcoding. The landmark coordinates were obtained by Miss Duangkamol Phoophitphong (Kasetsart University), while pictures of teeth were taken by Sylvie Agret; body shape measurements were obtained from the members of the CERoPath project. Ardern Hulme-Beaman, Andrea Cardini, Anna Loy, Paolo Colangelo seriously improved the first versions of this manuscript and made valuable comments. This study is sponsored by the ANR 07BDIV012 CERoPath, 11-CEPL-002 BiodivhealthSEA, and ANR-11-BSV7-008 BigTooth projects.

Abstract

This publication uses and presents R routines that perform various morphometric analysis in the context of rodent systematics. The morphological variation of two commensal rat species, *Rattus exulans* and *Rattus tanezumi*, is analysed and the potential for discrimination between the two is assessed. Specimens were trapped in three localities of Northern and North-Eastern Thailand. Shape and size variation are analysed in regards to sex, species, and geographical effects with various morphometric methods: log-shape ratios on body measurements, elliptic Fourier analyses on teeth outlines, Procrustes superimposition on skull coordinates. Both species are significantly different; however, the discrimination seems to be better on skull Procrustes coordinates and on teeth size than on other morphometric data set. Where different allometries exist between species and where species differ in size and shape, it is shown that filtering allometry using the approach of Burnaby (1966) can improve the discrimination between species. Sex size and shape dimorphism is reduced by comparison to interindividual variation. Shape variation varies between sampled localities for *Rattus exulans*, this is not the case for *Rattus tanezumi*. This pattern is possibly related to the more commensal life of *R. exulans*.

Introduction

Exploring morphological variation and relating this variation to explanatory factors is the essential purpose of morphometrics. In systematics, morphometrics primarily helps to quantify and describe differences between taxa or populations (e.g., Loy et al. 1993; Cardini et al. 2009; Viscosi and Cardini 2011; Chiari and Claude 2012). At present, there is a very large tool box for obtaining morphometric variables that can be routinely analysed via statistical analyses (see for instance the list given at http://life.bio.sunysb.edu/morph/). Morphometric methods have undergone an important progression starting from "traditional" or "multivariate" morphometrics to "modern" or "geometric morphometrics" (Rohlf and Marcus, 1993; Adams et al., 2004). While "traditional" morphometrics mostly relies on collection of raw linear measurements, "geometric" morphometrics analyses collections of anatomical landmarks expressed as point coordinates to quantify shape and size. Depending on the diverse available methods, size and several shape variables can be extracted from the morphometric recording on a set of specimens (Claude, 2008). Among others, these can be simple transformation of linear measurements (log-shape ratios), more sophisticated variables such as Procrustes coordinates obtained trough superimposition methods for landmark data, or coefficients of shape functions fitted to curves or surfaces (e.g. elliptic Fourier analyses). All these techniques have been applied in mammal evolution and systematics (see Corti et al. 2000), and they are now routinely used by researchers.

As afore mentioned, a wide range of software is available to perform morphometric procedures and perform statistical analyses, but few offer the possibility to extract diverse parameters of shape variation for statistical shape analyses with a unique language and software. Moreover, few of these software can be run in diverse operational systems, which means that, a "Windows" or an "Apple" user may not have access to the same programs. However, many of the morphometric techniques have been recently imported into the R language and environment (Claude, 2008). The R project started in the 90's and now provides a series of advantages for obtaining several kinds of morphometric data and for analysing these morphometric data with a wide array of statistical analyses (R Core Team, 2013). R has many advantages on other software: it is free, it can be run under various operating system (Windows, Linux, Apple), it is evolving with the help of a large community of users and developers. Moreover, the R language is close to the statistical jargon, and an enormous amount of literature about R is now available, often freely on the web. In addition to the R core, a very diverse number of packages has been developed in various disciplinary fields that have strong affinities with shape analysis (for instance, packages performing phylogenetic or comparative analyses, analyses of ecological communities). There is also a tremendous number of statistical tools, that are usually used in modern systematics, ecology and evolution, like fixed effects and mixed effects linear modeling, multivariate statistics, circular data, spatialised data, phylogenetics, genetics, comparative data, and community analyses. R also has graphical interfaces that permit the production a very large array of graphics easily customisable by the users. Finally, some packages and several functions have been developed for performing morphometrics, and other are being developed. The packages available on the CRAN include `shapes` and `geomorph` for Procrustes methods (Adams and Otarola-Castillo, 2013; Dryden, 2013), `Momocs` for outline analyses (Bonhomme et al., 2012), `LOST` for missing morphometric data simulation and estimation (Arbour and Brown, 2012). In addition, the package `Morpho` (Schlager, 2013) for 3D analyses is also available at http://morpho-rpackage.sourceforge.net; and the functions developed in Claude (2008) can be downloaded or sourced at http://www.isem.univ-montp2.fr/recherche/files/2012/01/Rfunctions1.txt. It is

*Corresponding author
Email address: julien.claude@univ-montp2.fr (Julien Claude)

therefore valuable to use R and develop it further, with the possibility to ultimately adopt it as standard. R comes at first with a command line, but some more interactive graphical interfaces with buttons have been developed for users who are not familiar with R language (the R commander GUI that can be installed with the Rcmdr package (Fox, 2005), for instance). The use of a command line may be at first daunting to the new comers. However, it also gives an enormous advantage, as you can write scripts, save them and keep in memory the whole analyses in a text file. This can represent a certain gain of time to the users and is more efficient than having to play with buttons; in addition, because scripts are written in R, they can be posted to other users which review and improve them or to adapt it to other data sets. While describing an original study, the purpose of this paper is also to provide simple R commands for several kinds of morphometric analyses to be used with a number of data sets: landmark coordinates, shape measurements, or outline coordinates. The codes are supplied as supplementary material. The functions and codes that were used here principally come from Claude (2008) and are explained in that book. The supplementary material also contains some functions of the newly developed packages for obtaining graphics and performing tests. For the reader who may not be familiar with R, "R for beginners" by Paradis (2005) provides the essentials.

The scripts presented here are focused on taxonomic and simple evolutionary questions regarding rodent phenotypic characteristics. The applied part of the study aimed at depicting differences between species, relating shape variation to geographical factors and sex dimorphism, and to compare the congruency of results between different shape features using R routines. I also tried to determine which of the shape features and analyses could offer the best discrimination between the two taxa that were analysed. The main idea remains to show that one can rapidly obtain nice results and graphics with few line commands in R. I applied various morphometric methods for quantifying differences between two species of obligatory or facultatively commensal rodents, which both occurs in South-East Asia. I also tested whether sexual dimorphism or geographical differences could explain this variation. The two species of interest are *Rattus exulans* and *Rattus tanezumi*. Both species can be found indoors, sometime in the same house. They are considered as pests (Aplin et al., 2011) and potential reservoirs for several zoonotic diseases (Aplin et al., 2011; Lerdthusnee et al., 2008). Overall, the two species are very similar in shape but can be differentiated based on their size and ecology. *Rattus exulans* is a small sized species that is exclusively commensal in Thailand, found in house and farms, while *Rattus tanezumi* is larger, more opportunistic, living both in houses and diverse habitats (plantations, forests, agricultural lands) (Lekagul and McNeely, 1988; Corbet and Hill, 1992). There is, however, a considerable overlap in ecology and size variation, and small *R. tanezumi* found in human habitations can be easily confused with *Rattus exulans*.

In the literature, the two species are also differentiated by their mammae formula (Lekagul and McNeely, 1988; Corbet and Hill, 1992), but mammae formula are sometimes difficult to observe, and some variation occurs. As for many other cases, morphometrics (traditional or modern) can help identify species in the field. Traditional morphmetrics is regularly used for species recognition. For instance, in Lekagul and McNeely (1988) and in Corbet and Hill (1992), tables of measurements are given and are intended to help in identification of taxa. In rodent taxonomy, however, differences can be subtle, with only minor skull or teeth differences, and the statistical analysis of complex shapes becomes sometimes the only resource for measuring how much species differ, and to potentially offer clues for taxonomic diagnosis. Furthermore, a morphometric analysis can be complementary to molecular methods (e.g., Guillot et al. 2012; Pages et al. 2013, as it allows to quantify phenotypic variation in populations. When species niches or geographical range overlap, morphometric analyses allow to evidenciate character displacements that could explain differences between populations (see Adams and Rohlf 2000; Loy and Capanna 1998). Here I explored the variability of the two rodent species using different morphometric methods.

Figure 1 – Localities where rodents were sampled.

Materials and Methods

Three original datasets have been obtained from 200 rodents trapped during field work in Northern and North-Eastern Thailand. Specimens were captured from three localities during field trips from 2006 to 2009 (Fig. 1, Tab. 1). The geographic range for a sampling site was around 20 km × 20 km, so that several habitats and houses were surveyed for each field session. Rodents were identified based on gross morphology, and some of them were identified via molecular analyses. *Rattus tanezumi* is often confounded in the literature with *Rattus rattus*, which was absent from the sites that were sampled for this study (Aplin et al., 2011; Lack et al., 2012; Pages et al., 2013). As there are taxonomic issues regarding *R. tanezumi* (Pages et al., 2013), most individuals of that species were also molecularly identified. *Rattus exulans* were determined mostly based on their gross morphology, mammae formula, and from the capture location (indoors). Whenever direct identification was not clear, or a morphological overlap was possible with *R. tanezumi* (for instance for subadults, unusually large specimens, when mammae formula was not clear, or when specimens were trapped outdoors) individuals were molecularly controlled also for this species. Here following are explanation of the three data sets (linear measurements, landmark coordinates on skulls and tooth outline coordinates) provided as supplementary material.

Linear measurements

The first data-set "`body.csv`" is a compilation of five linear measurements (HBL: head + body length, EL: ear length, FL: hindfoot length, HL: head length, TL: tail length). Measurements were obtained using callipers (HL) and rulers (HBL, EL, FL, TL), following the protocol described in Auffray et al. (2011). The data set contains 10 variables organized in 10 columns. The first corresponds to the individual label, the three next are explanatory variables (sp: species, loc: locality ("n" for `Nan`, "l" for Loei, "p" for Phrae), sex: `M` for males, `F` for females);

Table 1 – Number of sampled individuals according to sexes, species and localities.

Locality	Loei		Nan		Phrae	
Sex	F	M	F	M	F	M
R. exulans	22	19	20	19	23	19
R. tanezumi	12	22	10	9	11	13

the five next columns are response variables (morphometric measurements), and the last column indicates whether there was a molecular control on the species identification (either based on Cytochrome Oxidase 1 or on Cytochrome b). For simplicity and portability, this data set is written in `.csv` format and can be opened by any text editor. Data were collected only once by different operators. We did not test measurement error for these data, but are confident that our protocol is sufficiently standardized to limit variation due to the different operators.

Morphometric Procedures

The body measurements were analysed using the log-shape ratios approach (Mosimann, 1970). For each individual, size was computed as the geometric mean of all measurements, and each measurement was divided by size to obtain the shape ratios. The log of this quantity was used as raw data for subsequent analyses. Shape ratios are redundant in terms of information: one degree of freedom is lost due to scaling, and therefore four dimensions instead of five are necessary to describe shape variation. A principal components analysis was performed on five shape ratios and only the first four principal component scores were kept for multivariate analyses of variance. Observations with missing measurements (13.5%) were excluded from the analysis.

Statistical Analyses

In order to visualize how shape variation was structured, multivariate ordination of individuals was plotted on the first two PC axes using different colours and symbols for species, sex, and localities. Contributions of original variables on the corresponding eigenvectors (PC loadings) were analysed to understand which shape features had more influence on each PC.

The first PC axis represents 48.3% of shape variation while the second represents 18.9% (Fig. 2). The fourth PC (not shown here) seems to be related with species differentiation (*Rattus exulans* having higher scores), but there is an important overlap between species even on this axis. The fourth PC axis opposes the variables HBLand ELwith the variables FL, HLand FL, which means that *Rattus exulans* have in general relatively smaller ears and body length, and longer head, foot and tail compared to *Rattus tanezumi*.

Differences between Species, Sexes, and Localities

Effects of sexes, species and localities were estimated on geometric size using a multiple linear model, and the significance of these factors was tested using an ANOVA with type II sums of squares. Type II sums of squares are calculated in such way that the effect of a factor, for a given order of interaction, is evaluated once all the other factors were taken into account. This approach does not violate the principle of marginality, although it is conservative (see Venables 1998; Claude et al. 2003, 2004). Similarly, a multiple and multivariate linear model was applied on principal components of shape variation with non-zero eigenvalues. The sex, species, locality factors, the size variable, as well as the interactions until the third order were considered as explanatory variables. A multivariate analysis of variance using type II sums of squares was performed on the different variances and covariances explained by the factors and covariables. Results indicate that size was significantly related to species but not to other factors (Tab. 2).

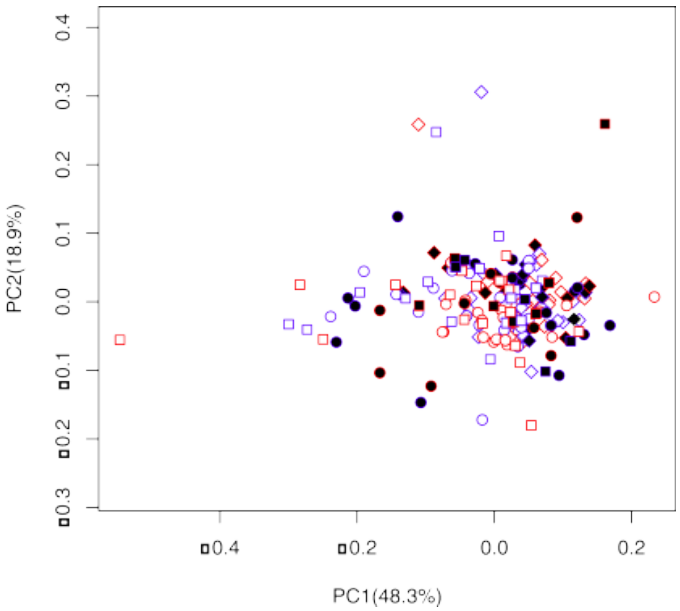


Figure 2 – PCA on log-shape ratios, localities are displayed with different symbols (circle, square, diamond), symbol outline colours represent sexes (red: females, blue: males), while symbol inner colours represent species (white: *R. exulans*, black: *R. tanezumi*).

Table 2 – ANOVA on geometric size (loc: localities, sex: sex, sp: species).

Factor	SS	df	F	*p*-value
sex	3.6	1	0.21	0.65
loc	14.1	2	0.42	0.66
sp	9914.5	1	591.29	<0.0001
sex:loc	4.9	2	0.14	0.87
sex:sp	29.3	1	1.75	0.19
loc:sp	79.5	2	2.37	0.10
sex:loc:sp	8.1	2	0.24	0.78
Residuals	2682.8	160		

Size, species and localities significantly explained shape variation, while sex shape dimorphism was not significant (Tab. 3). No interaction of second or third order were found to be significant. Also, no interaction was found between species and size, suggesting that allometries are similar among groups.

By analysing each species separately (tests not shown in the main text), it can be seen that the locality effect is highly significant for *R. exulans* ($p < 0.0001$) and only significant for *R. tanezumi* ($0.01 < p < 0.05$), and that there is a triple interaction between sex, locality and size for *R. exulans*.

Table 3 – MANOVA on the first four PCs of log-shape ratios.

Factor	df	Pillai	approx. F	num. df	den. df	*p*-value
sex	1	0.04	1.46	4	147	0.22
loc	2	0.24	5.14	8	296	<0.0001
sp	1	0.47	33.21	4	147	<0.0001
size	1	0.56	46.9	4	147	<0.0001
sex:loc	2	0.05	0.90	8	296	0.51
sex:sp	1	0.03	1.29	4	147	0.28
sex:size	1	0.04	1.66	4	147	0.16
loc:sp	2	0.07	1.39	8	296	0.20
loc:size	2	0.09	1.76	8	296	0.08
sp:size	1	0.012	0.44	4	147	0.78
sex:loc:sp	2	0.07	1.34	8	296	0.22
sex:loc:size	2	0.07	1.28	8	296	0.25
sex:sp:size	1	0.01	0.32	4	147	0.86
loc:sp:size	2	0.06	1.15	8	296	0.33

Performance of Log-Shape Ratios for species Identification

In order to evaluate whether it is possible to easily distinguish species based on shape variables, a linear discriminant analysis using the species as group factor was performed on the log-shape ratios of molecularly identified specimens. Although the two species were significantly different in shape, it was not possible to obtain a good discrimination on log-shape ratios. The predictive discriminant analysis based on molecularly identified specimens correctly assigned only 25% of the specimens that were not molecularly controlled. This extremely low percentage (less than the worse expected: 50%) probably comes from the unbalanced sampling of the reference (few *R. exulans* were molecularly identified which may distort the discriminant coefficients in favour to special features of this small group).

A linear discriminant analysis using the species grouping was also performed on the whole log-shape ratio dataset. The percentage of correctly assigned specimens was computed using a leave-out-one cross validation procedure (jackknife) and only reached 73%.

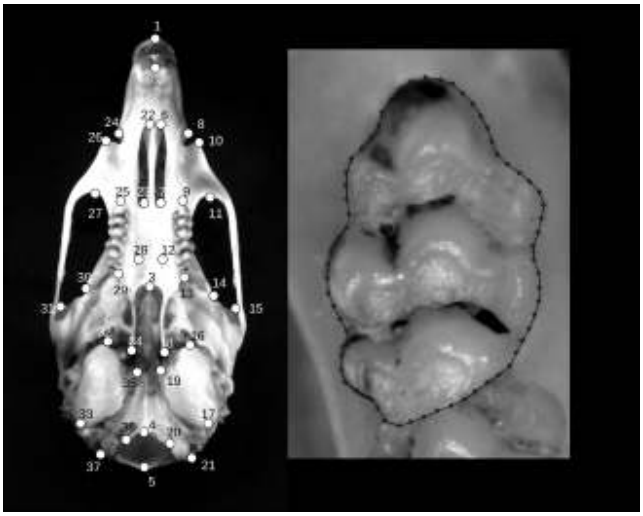


Figure 3 – Left: Landmark locations on the palatine view of the rodent skull (the two landmarks on the scale are not represented). Right: Upper first molar with digitized points around the outline.

Finally, when a discriminant analysis was directly applied on body shape measurements (therefore, when size was included), the percentage of correctly classified individuals reached 98% for the non-molecularly identified specimens. The same percentage of correct assigned individuals were obtained when using the leave-one-out cross-validation on the whole data set. By introducing this new variable, the linear combination of variables corresponding to discriminant coefficients was therefore much more efficient to discriminate between species.

Landmarks and Procrustes superimposition

A second dataset, "`skull.tps`", was produced from skulls. This data set is a collection of 39 points digitized for appraising the variation of the skull in palatine view (Fig. 3). The dataset contains the same individuals as the linear measurement dataset except one specimen. This dataset was obtained by first photographing the skulls of the specimens with a Pentax K200 camera, keeping always the same focal and distance between the specimens and the camera. Specimens were repositioned, rephotographed, and digitized a second time in order to estimate the percentage of error measurements during the digitization process. For each configuration, the 37 first pairs of coordinates correspond to the landmarks digitized on the skull palatine view (Fig. 3). Two additional points, spaced by 1 cm from each other, were measured on a millimetre scale photographed together with the skull, and used for scaling objects. For this study, landmark coordinates were originally recorded using the `TPSdig2` digitization software (Rohlf, 2013), but could have been directly digitized using `R` functions of the package `geomorph`, such as `digitize2d` or a more general function like `locator`. The `.tps` format is explained in Rohlf (2013). Each image was labelled so that the four first characters corresponded to the specimen number, the fifth to the species, the sixth to the locality, the seventh to the sex, and the eight indicated whether the specimen was identified with molecular data (`g`) or directly on the field (`f`); the last character corresponds to the session number.

Morphometric Procedures

Coordinates of landmarks were transformed from pixel into cm by dividing the raw coordinates by the Euclidean distance between the two landmarks digitized on the scale. All configurations (including replicates) were scaled to unit centroid size, translated, rotated, and optimally superimposed through the Generalized partial Procrustes Analysis (Dryden and Mardia, 1998; Claude, 2008). In this procedure, symmetric and asymmetric components were not partitioned. Details concerning Procrustes superimpositions can be found inBookstein (1990, 1991, 1996); Goodall (1991); Small (1996); Rohlf and Slice (1990); Dryden and Mardia (1998); Viscosi and Cardini (2011). Centroid size was used as an estimator for size. Procrustes coordinates were projected into the Euclidean tangent shape space using an orthogonal projection (Kendall, 1984; Goodall, 1991; Small, 1996; Dryden and Mardia, 1998; Claude, 2008). Superimposed coordinates are redundant: there are more coordinates than the number of dimensions in the shape space because the translation, scaling and rotation consume two, one and one degrees of freedom respectively (Dryden and Mardia, 1998). Therefore, a principal components analysis was performed on superimposed coordinates, and the 2 (dimensions) by 37 (landmarks) minus 4 (lost degrees of freedom) PCs were considered for multivariate analyses of variance.

Measurement Error

Percentage of error measurement was obtained following the ANOVA approach described in Yezerinac et al. (1992), directly on centroid size and it was adapted to Procrustes data. For Procrustes coordinates, I followed the Procrustes ANOVA approach described in Goodall (1991); Klingenberg and McIntyre (1998); Claude et al. (2003). The among and within variances were calculated directly from the mean squares and crossproducts corresponding to the specimen and residual sources of variation. The percentage of measurement error is less than 1% for centroid size, and 26% for shape. The specimen factor is always significant, which means that interindividual size and shape variations are stronger than variation between replicated measurements on the same individual. One can note that this percentage of measurement error could be greater if one consider a smaller stratum for our samples (e.g., one single sex, one single species, one single locality).

Principal Components of Shape Variation

A principal component analysis was performed on averaged configurations for each individual (average of the two replicates) and ordination of individuals was explored on the first two PCs. To understand which shape features were involved in the patterns of variation observed along the two axes, reconstruction of extreme morphologies along each PC was obtained for Procrustes data. The projection of individuals on eigenvectors (Fig. 4) shows that both species are well distinguished along PC 2, which involves the length of the molar row relatively shorter for *Rattus exulans* than for *Rattus tanezumi*, and the orientation of the incisor. Although PCs are computed in such way that each PC is independent and orthogonal to the others, one can see that there is a relationship between PC 1 and PC 2 if one consider each species independently (Fig. 4).

These apparent relationships certainly come from the fact that strong allometric relationships exist (cf. patterns observed on PC 1 that remind rodent skull ontogeny) and possibly differ between species (see statistical analyses below).

Differences between Species, Sexes, and Localities

As for log-shape ratios, a univariate linear model on centroid size was applied to understand the possible effects of species, locality and sex. Similarly, a multivariate linear model was applied on principal components of shape variation with non-zero eigenvalues. The sex, species, locality factors, the size variable, as well as the interactions until the third order were considered in this model as explanatory variables. Univariate (for size) and multivariate (for shape) analyses of variance using type II sums of squares were performed on the different variances explained by the factors and covariables. Skull centroid size was influenced by species, and sex size dimorphism differed between species (Tab. 4). Males of *Rattus tanezumi* appear larger than females, while sex size dimorphism was not detected in *Rattus exulans*.

Skull shape variation was significantly affected by species, size, sexes and localities. In addition, size significantly interacted with species, and species with localities (Tab. 5). The patterns of shape differences between localities differed in the two species and the allometric growth pattern differed among species.

Allometry-free approach

Because size interacted with species in the previous linear model, an allometry-free approach was conducted following the Burnaby Procedure (Burnaby, 1966). This approach permitted to filter out the effect of

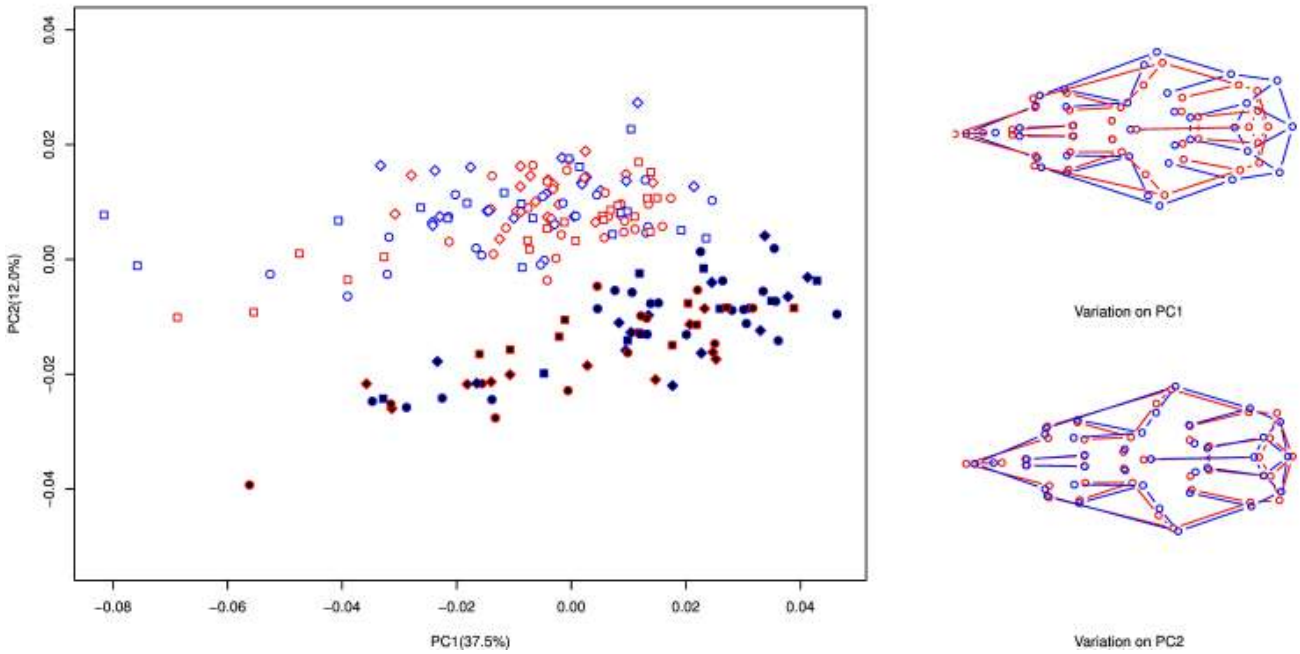


Figure 4 – PCA on Procrustes coordinates, similar symbols of Fig. 2. Patterns of variation along PC 1 and PC 2 are presented on the right side of the plot, blue corresponding to minimal scores, red to maximal ones.

Table 4 – ANOVA on skull centroid size (loc: localities, sex: sex, sp: species).

Factor	SS	df	F	*p*-value
sp	146.086	1	652.94	<0.0001
sex	0.770	1	3.44	0.07
loc	0.599	2	1.34	0.26
sp:sex	1.803	1	8.06	0.0050
sp:loc	0.001	2	0.00	0.99
sex:loc	0.226	2	0.50	0.60
sp:sex:loc	0.588	2	1.31	0.27
Residuals	41.839	187		

Table 5 – MANOVA on the $37 \times 2 - 4$ first shape principal components for Procrustes data.

Factor	df	Pillai	approx. F	num. df	den. df	*p*-value
sp	1	0.91	16.44	70	108	< 0.0001
size	1	0.92	17.77	70	108	< 0.0001
sex	1	0.51	1.63	70	108	0.0108
loc	2	1.21	2.38	140	218	< 0.0001
sp:size	1	0.58	2.10	70	108	0.0003
sp:sex	1	0.35	0.83	70	108	0.80
sp:loc	2	0.94	1.39	140	218	0.0147
size:sex	1	0.42	1.12	70	108	0.29
size:loc	2	0.87	1.20	140	218	0.11
sex:loc	2	0.73	0.89	140	218	0.78
sp:size:sex	1	0.40	1.03	70	108	0.44
sp:size:loc	2	0.83	1.11	140	218	0.24
sp:sex:loc	2	0.62	0.70	140	218	0.99
size:sex:loc	2	0.64	0.73	140	218	0.98

growth from the data. Although type II sums of squares are used for estimating effects, it could be interesting to remove all variation that could be due to growth. In order to reach this goal, I projected the data onto an space where ordination of individuals for every species would be independent of growth, rather than analyse variation once the effect of size was taken into account through regression (this is actually what is done with type II sums of squares). Since each species displayed different allometries, this approach is completely different than analysing variation after applying a linear model on individuals introducing size, and interaction between size and species. In practice, the vector of allometric coefficients for each species was obtained, and Procrustes coordinates were projected onto a an orthogonal space based on the direction of these vectors following the procedure of Burnaby (1966).

Once allometric growths within species were filtered from the shape variation, the two species appeared well discriminated on the first transformed PC axis (Fig. 5). In this allometry-free shape space, as for the previous analysis, species differentiation is related to the relative length of the tooth row and the orientation of the incisor. These characters can therefore be used independently of specimen size (ultimately age).

Species, sexes, and localities significantly differed in this new allometry-free shape space (Tab. 6). There is an interaction between the locality and species factors indicating that differences between localities are not the same for both species, once growth is filtered out from the data.

By examining each species separately (data not shown but see tests in supplementary material), *Rattus exulans* skull shape significantly differed between localities, with a marginal effect of sex, while *Rattus tanezumi* skull shape did not differ between localities and sexes.

Performance of Procrustes Data for Species Identification and Species Differences

A linear discriminant analysis using the species factor was performed on the Procrustes coordinates for the molecularly identified specimens in order to know whether one can predict the species based on Procrustes data. The prediction of the specimens that were not molecularly controlled on this analysis was weak: only 26.8% of these were assigned to the correct species. This low score can in part be explained by the unbalanced samples in the reference or by the confounding effect of growth.

A linear discriminant analysis using the species was then performed on the dataset of Procrustes coordinates including all specimens, and the percentage of correctly assigned specimens was computed using a

Table 6 – MANOVA on the $37 \times 2 - 6$ first allometry free shape principal components for Procrustes data.

Factor	df	Pillai	approx. F	num. df	den. df	*p*-value
sex	1	0.47	1.55	68	120	0.0190
sp	1	0.95	33.56	68	120	< 0.0001
loc	2	1.14	2.37	136	242	< 0.0001
sex:sp	1	0.40	1.20	68	120	0.20
sex:loc	2	0.70	0.95	136	242	0.62
sp:loc	2	0.91	1.47	136	242	0.0046
sex:sp:loc	2	0.70	0.95	136	242	0.62

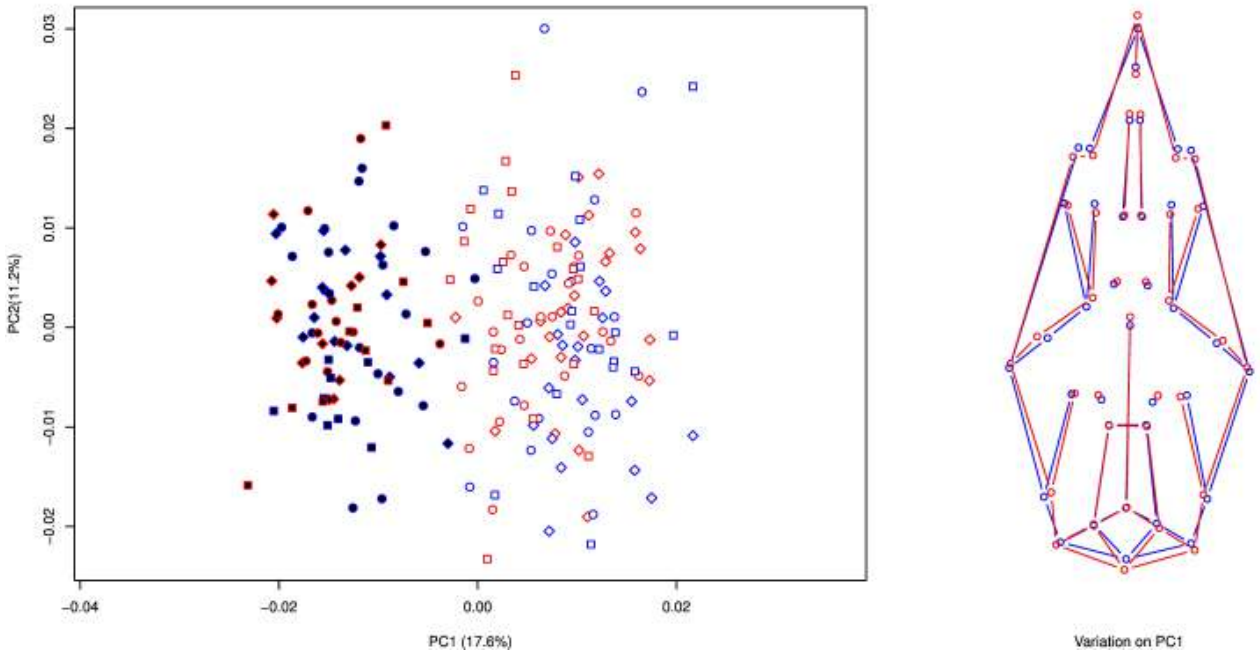


Figure 5 – PCA on Procrustes coordinates corrected for intraspecific ontogenetic allometries, similar symbols of Fig.2. Patterns of variation along PC 1 are presented on the right side of the plot, blue corresponding to minimal scores, red to maximal ones.

leave-one-out cross-validation (jackknife) procedure. In this analysis, 99.5% of correct re-assignation was obtained.

Discriminative features between species were estimated by rescaling the linear discriminant coefficients and are presented in Fig. 6. Indeed the metric of the discriminant space corresponds to the Mahalanobis distance and it is desirable to re-incorporate the pooled intra-group variance on the coefficients in order to depict shape changes in terms of original units (here displacements of landmarks in cm). As expected from the observations on PCAs, shape features that better discriminate between species concerned the relative length of the tooth row, the incisor position, as well as differences in the morphology of the posterior part of the skull (Fig. 6).

When intraspecific allometries are filtered out, 92.6% of inviduals are correctly assigned to the good species for the samples that were not identified by molecular markers; while 99.5% of individuals are correctly re-assigned using the leave-one-out cross-validation procedure on the whole data set.

Elliptic Fourier Analysis

"`teeth.tps`" is a collection of points digitized on the outline of the first right upper molar (Fig. 3) recorded a subset of 62 individuals. The teeth were photographed with a CCD camera mounted on a stereomiscroscope. For each specimen, the two first digitized coordinates correspond to two points defined by an inframillimetric scale. As for skulls, specimens were randomly repositioned, photographed and digitized a second time for later estimating measurement error due to the digitization process. Points were digitized along the outline of the teeth in a clockwise way starting from the anterior end of the first upper molar using `TPSdig2` (Rohlf, 2013). Sixty four points were resampled from these former set using `TPSdig2` (Rohlf, 2013). In `R`, such a procedure can be done using the `locator` function and/or more automatised functions such as those presented in Claude (2008). In future years, it can be expected that some improvements of the current packages will offer a GUI for easily digitizing curves. Images were labelled in a similar way as the "`skull.tps`" dataset.

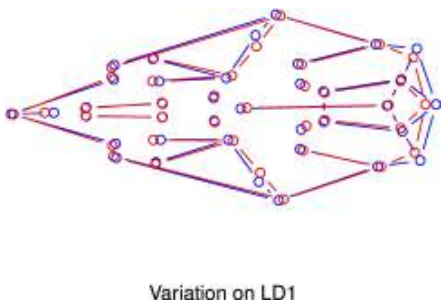


Figure 6 – Shape features that best discriminate species: "typical" *Rattus tanezumi* shape in red, while "typical" *Rattus exulans* is in blue.

Morphometric Procedures

As for skull coordinates, pixels were transformed to mm by dividing the raw coordinates by the Euclidean distance of the scale and multiplying by the scale actual length. For each configuration, the set of coordinates was then described using normalized elliptic Fourier coefficients. Details of the method can be found in Kuhl and Giardina (1982); Rohlf and Archie (1984); Claude (2008). Size was estimated as the longer radius of the first fitting ellipse. To reduce the number of variables by comparison to the number of individuals, only the first harmonics that showed a reasonable digitization error rate (< 35%) were kept. Indeed high order harmonics are more sensitive to small random variations (and therefore to noise) that could be introduced during the digitization process.

Measurement Error

Sixty-four points were sampled for elliptic Fourier analysis, but only the first 32 harmonics were retained (following the Nyquist theorem Shannon 1949). The percentage of error on harmonic coefficients is calculated with a similar approach as the Procrustes ANOVA. The mean sums of squares were first calculated for the four coefficients of each harmonics to observe the evolution of percentage of error according to the rank of harmonics. The total measurement error rate was obtained by summing the different mean squares for the four coefficients within each harmonics. Tooth size measurement error (based on the larger radius of the first ellipse) reaches 1.4%. Tooth shape measurement error depends on harmonic rank (Fig. 7).

The first normalised harmonic coefficients displayed an important digitization error rate (61%) while the coefficients of the six following reach around 30%. The high level of error found for the first harmonic certainly comes from the variation in orientation of the tooth under the steromicroscope, since the corresponding coefficients measure the width on length ratios (and this ratio depends on the orientation

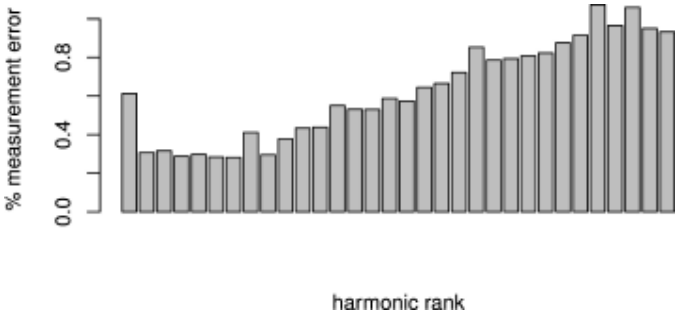


Figure 7 – Change of measurement error in regard to harmonic rank.

of the occlusal surface relative to the horizontal plan). After the seventh harmonic, the percentage of error variation increased. Shape variables that were considered for further tests were therefore summarized by the seven first harmonic coefficients by excluding the first harmonic coefficients.

Analysis of Principal Components of Shape Variation

As for Procrustes data, a principal components analysis was performed on averaged elliptic Fourier coefficients of tooth outlines for each individual. Ordination of individuals was later plotted on the first two PC axes. To understand which shape features were involved, reconstruction of extreme morphologies along each PC was obtained for elliptic Fourier data. In order to reconstruct tooth outline from Fourier coefficients, the inverse Fourier transform was used to reconstruct these theoretical teeth outlines (Rohlf and Archie, 1984; Claude, 2008). A large overlap between species is observed for tooth outline shape variation on the first two principal component axes (Fig. 8).

Differences between Species, Sexes, and Localities

The effects of sex, species and locality were estimated by a linear model on tooth size. Tooth size was expressed by the length of the major axis of the best fitting ellipse defined by the first harmonic coefficients. A multivariate linear model was applied on principal components of the normalized elliptic Fourier coefficients with non-zero eigenvalues. The sex, species, locality factors, the variable size, as well as the interactions until the third order were considered as explanatory variables. Univariate and multivariate analyses of variance using type II sums of squares were performed on the different variances explained by the factors and covariables for size and shape respectively. Species significantly differed in size and there was no overlap in tooth size between species (Tab. 7). No other factors contributed significantly as a source of size variation for teeth. Species also significantly differed in shape and there was a significant relationship between tooth shape and size (Tab. 8). However, no interaction and no sex effects were significant.

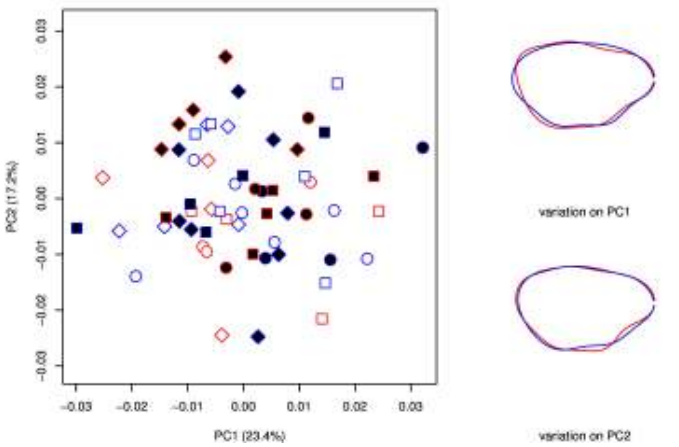


Figure 8 – PCA on the Fourier coefficients obtained from tooth outline, patterns of variation along PC 1 and PC 2 are presented on the right side of the plot, blue corresponding to minimal scores, red to maximal ones.

Table 7 – ANOVA on on first upper molar size.

Factor	SS	df	F	*p*-value
sex	0.00002	1	0.1480	0.70
sp	0.13466	1	1196.65	< 0.0001
loc	0.00046	2	2.04	0.14
sex:sp	0.00013	1	1.17	0.28
sex:loc	0.00009	2	0.41	0.67
sp:loc	0.00048	2	2.13	0.13
sex:sp:loc	0.00005	2	0.20	0.82
Residuals	0.00540	48		

Table 8 – MANOVA on the elliptic Fourier coefficients.

Factor	df	Pillai	approx. F	num. df	den. df	*p*-value
sp	1	0.88	4.47	24	15	0.0021
sex	1	0.78	2.23	24	15	0.06
loc	2	1.38	1.47	48	32	0.12
size	1	0.87	4.36	24	15	0.0024
sp:sex	1	0.45	0.50	24	15	0.94
sp:loc	2	1.10	0.82	48	32	0.74
sp:size	1	0.72	1.57	24	15	0.18
sex:loc	2	1.23	1.07	48	32	0.43
sex:size	1	0.41	0.44	24	15	0.97
loc:size	2	1.18	0.96	48	32	0.56
sp:sex:loc	2	1.17	0.94	48	32	0.58
sp:sex:size	1	0.44	0.49	24	15	0.94
sp:loc:size	2	1.30	1.24	48	32	0.27
sex:loc:size	2	1.19	0.98	48	32	0.54

When species were tested separately on the 17 first PCs (99.5% of shape variation) for size, sex and locality and the different interaction between these factors, significant differences were found between localities for *R. exulans* but not for *R. tanezumi*. Sex had a marginal effect only for *R. tanezumi*.

Performance of Elliptic Fourier Coefficients for Species Identification and Species Differences

As too few individuals of *Rattus exulans* were molecularly identified, only the predictive discriminant analysis based on the whole dataset of coefficients was performed with the leave-one-out cross-validation procedure. Sixty-six percent of individuals were correctly reclassified in this analysis. Linear discriminant coefficients were rescaled by reincorporating intragroup variance covariance to identify shape features that increased the "ratio" of inter- on intra-specific variation. The shape features that discriminate between species were very subtle and are illustrated in Fig. 9. It can be seen that the anterior part of the tooth is better demarcated in *R. tanezumi* (Fig. 9).

Discussion

Both species differed for size, linear measurements, skull and tooth shapes. However, some structures more effectively discriminated between species: tooth size as well as features of skull palatine views seem better discriminators than log-shape ratios or tooth shape parameters. Discrimination between species mostly involved tooth row length.

R. exulans is smaller than *R. tanezumi* in teeth, skull, and body measurements. The length of the tooth row relative to the skull size is smaller in *R. exulans* than in *R. tanezumi*, suggesting evolutionary allometries between species. Post-natal ontogeny strongly structure the morphological variation of the skull shape within each species, but the relationships between shape and size differ between species. Not only mean shape changes occurred in the evolution of *Rattus* but developmental features, such as the allometric relationships between shape and size were also modified. It is interesting to note that tooth outline shape is rather preserved in the two species,while tooth size differs. Tooth shape has often been found to differ between closely related species within the Murinae (e.g., Macholan 2006).

Discriminating between closely related rodent species can sometimes be difficult. The *Rattus* genus is known to be a very diverse group for which delimiting species is difficult due to an important intraspe-

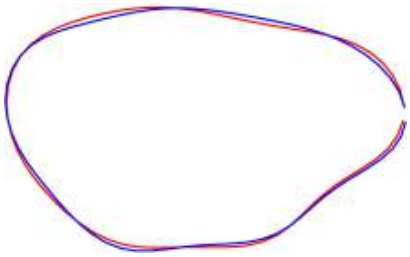


Figure 9 – Tooth shape variation that best discriminates between species, the red outline corresponds to *R. tanezumi* phenotype while the blue outline corresponds to *R. exulans*.

cific variation (Rowe et al., 2011). When one can have access to the dynamic of growth and when this effect can be filtered out from variation, landmark data on skulls performs rather well to discriminate species. As there is also a good differentiation between species in size, adding this variable in the discriminant analysis can increase the discrimination between species. Based on body measurements, form (shape + size) discriminates rather well between the two species compared to shape only. This study also shows that equal sampling between groups as well as sufficient effective size matter for obtaining better predictions (see the very low score of predictive discriminant analyses when species have unequal numbers of observations).

All the shape features that could differentiate the two species were not exhaustively explored in this study. An exploration of other skull views, osteological traits, body measurements, or teeth could possibly yield also good results. In addition, it is clear now that sampling size should be large enough for morphometrics to serve at identifying specimen using predictive discriminant analyses.

In contrast to the analysis of shape ratios, Procrustes or outline analyses provide the considerable advantage to visually display shape changes along exploratory axes (PCA) or explanatory axes (linear discriminant axes). Candidate features for discriminating the two species include tooth length, configuration of the posterior part of the skulls (relatively wider in *R. exulans*).

Sexual dimorphism was significant only for skull shape and seems to be stronger in *R. exulans* than in *R. tanezumi*. No sexual dimorphism of size could be identified. One must, however, stress that significance of effects depends not only on the magnitude of effect variation but also on the residual variation. This characteristic if tested with controlled genetic and environmental variation could become possibly significant.

Rattus exulans differs between localities for skull morphometrics and tooth shape, while *Rattus tanezumi* does not. Motokawa et al. (2004) and Pages et al. (2013) reported morphological differences for *Rattus exulans* and for *Rattus tanezumi* at a larger geographical scale. The observed differences probably demonstrate that geographical variation exists at finer geographical scale for *Rattus exulans*. Although it is speculative at this stage, one could hypothesize that differences in demographic features and kinship structures may possibly explain the differentiation between *R. exulans* populations and the homogeneity between *R. tanezumi* populations. It is also possible that due to its strong relationship with humans, some individuals of *R. exulans* have settled a distinct population helped with human activity (transportation, importation of food, etc.). It is also possible that *R. tanezumi* is more panmictic because it exploits a larger range of habitats (indoor, agricultural and forested areas) than *R. exulans*, which is always commensal in Thailand. Larger sampling and molecular studies could probably help to understand the different results obtained for both species.

Finally, scripts that were developed for this study can easily be reused and adapted for other groups, and for addressing similar questions: morphological differentiation, evaluation of sex dimorphism, geographical variation. Furthermore, since morphometrics is now going along with several studies such as quantitative genetics, morphological integration, phylogenetics, fluctuating asymmetries, etc., and because many R libraries are developed and devoted with some aspects of these problems (mixed linear modelling, matrix calculation, reconstruction of ancestral character states), the script can probably serve as a base or as a tutorial for going beyond taxonomic issues.

References

Adams D.C., Otarola-Castillo E., 2013. `geomorph`: an R package for the collection and analysis of geometric morphometric shape data. Methods in Ecology and Evolution 4: 393–399. [R package version 1.0] http://cran.r-project.org/web/packages/geomorph/index.html

Adams D., Rohlf F.J., 2000. Ecological character displacement in *Plethodon*: biomechanical differences found from a geometric morphometric study. Proceedings of the National Academy of Sciences, U.S.A. 97: 4106–4111.

Adams D.C., Rohlf F.J., Slice D.E., 2004. Geometric morphometrics: ten years of progress following the "revolution". It. J. Zool. 71: 5–16.

Aplin K.P., Suzuki H., Alejandro A., Chinen A.A., Chesser R.T., ten Have J., Donnellan S.C., Austin J., Frost A., Gonzales J.P., Herbreteau V., Catzeflis F., Soubrier J., Fang Y.-P., Robins J., Matisoo-Smith E., Bastos A.D.S., Maryanto I., Sinaga M.H., Denys C., Van Den Bussche R., Conroy C., Rowe K., Cooper A., 2011. Multiple geographic origins of commensalism and complex dispersal history of black rats. Plos ONE, 6(11): e26357. doi:10.1371/journal.pone.0026357

Arbour J., Brown C., 2012. `LOST`: Missing morphometric data simulation and estimation. [R package version 1.0] http://cran.r-project.org/web/packages/LOST/index.html

Auffray J.-C., Blasdell K., Bordes F., Chabé M., Chaisiri K., Charbonnel N., Chaval Y., Claude J., Cosson, J.-F., Dei-Cas E., Desquesnes M., Dobigny G, Douangboupha B., Galan M., Haukisalmi V., Henttonen H., Herbreteau V., Hugot J.-P., Jiyipong T., Latinne A., Michaux J., Milocco C., Morand S., Pagès M., Phoophitpong D., Pumhom P., Ribas Salvador A., Soonchan S., Suputtamongkol Y., Waengsothorn S., Waywa D., Xuéreb A., 2011. Protocols for Field and Laboratory Rodent Studies. Kasetsart University Press.

Bonhomme V., Picq S., Claude J., 2013. `Momocs`: Shape Analysis of Outlines. [R package version 0.2-02] http://vincentbonhomme.fr/momocs/

Bookstein F.L., 1990. Introduction to methods for landmark data. In: Rohlf F.J., Bookstein F.L. (Eds.) Proceedings of the Michigan Morphometric Workshop. University of Michigan Museum of Zoology Special Publication 2. 215–226.

Bookstein F.L., 1991. Morphometric tools for landmark data: Geometry and Biology. Cambridge University Press, Cambridge.

Bookstein F.L., 1996. Combining the tools of geometric morphometrics. In: Marcus L.F., Corti M., Loy A., Naylor G.J.P., Slice D., Bookstein F.L. (Eds.) Advances in Morphometrics. NATO ASI Series. Plenum Press, New York. 131–152.

Brown C., Arbour J., Jackson D., 2012. Testing of the effect of missing data estimation and distribution in morphometric multivariate data analyses. Systematic Biology 61: 941–956 .

Burnaby T.P., 1966. Growth-invariant discriminant functions and generalized distances. Biometrics 22: 96–110.

Cardini A., Nagorsen D., O'Higgins P., Polly P.D., Thorington J., Tongiorgi P., 2009. Detecting biological distinctiveness using geometric morphometrics: an example case from the Vancouver island marmot. Ethology, Ecology and Evolution 21: 209–223.

Chiari Y., Claude J., 2012. Morphometric identification of individuals when there are more shape variables than reference specimens: a case study in galapagos tortoises. Comptes rendus Biologies 335: 62–68.

Claude J., 2008. Modern Morphometrics with R. Springer, New York.

Claude J., Paradis E., Tong H., Auffray J.-C., 2003. A geometric morphometric assessment of the effects of environment and cladogenesis on the evolution of the turtle shell. Biological Journal of the Linnean Society 79: 485–501.

Claude J., Pritchard P., Tong H., Paradis E., Auffray J.-C., 2004. Ecological correlates and evolutionary divergence in the skull of turtles: a geometric morphometric assessment. Systematic Biology 53: 933–948.

Corbet G.B., Hill J.E., 1992. The Mammals of the Indomalayan Region: a systematic review. Natural History Museum Publications, Oxford University Press, Oxford.

Corti M., Marcus L., Hingst-Zaher E., 2000. Introduction to the Symposium: Geometric morphometrics in mammalogy. Hystrix 11(1): 3–7. doi:10.4404/hystrix-11.1-4133

Dryden I.E., 2013. `shapes`: Statistical shape analysis. [R package version 1.1-6] http://www.maths.nottingham.ac.uk/~ild/shapes

Dryden I.E., Mardia K.V., 1998. Statistical Shape Analysis. Wiley, Chichester.

Fox J., 2005. The R commander: A basic-statistics graphical user interface to R. Journal of Statistical Software, 14(9): 1–42.

Goodall C.R., 1991. Procrustes methods in the statistical analysis of shape. Journal of the Royal Statistical Society, series B 53: 285–339.

Guillot G., Renaud S., Ledevin R., Michaux J., Claude J., 2012. A Unifying Model for the Analysis of Phenotypic, Genetic and Geographic Data. Systematic Biology 61(5): 897–911.

Kendall D., 1984. Shape manifolds, procrustean metrics, and complex projective spaces. Bulletin of the London Mathematical Society 16: 81–121.

Klingenberg C.P., McIntyre G.S., 1998. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. Evolution 54(4): 1363–1375.

Kuhl F.P., Giardina C.R., 1982. Elliptic Fourier features of a closed contour. Computer Graphics and Image Processing 18: 236–258.

Lack J.B., Greene D.U., Conroy C.J., Hamilton M.J., Braun J.K., Mares M.A., Van Den Bussche R.A., 2012. Invasion facilitates hybridization with introgression in the *Rattus rattus* species complex. Molecular Ecology 21(14): 3545–3561.

Lekagul B., McNeely J.A., 1988. Mammals of Thailand, 2nd edition. Darnsutha Press, Thailand.

Lerdthusnee K., Nigro J., Monkanna T., Leepitakrat W., Leepitakrat S., Insuan S., Charoensongsermkit W., Khlaimanee N., Akkagraisee W., Chayapum K., Jones J., 2008. Surveys of rodent-borne disease in Thailand with a focus on scrub typhus assessment. Integrative Zoology 3: 267–273.

Loy A., Capanna E., 1998. A parapatric contact area between two species of moles (genus *Talpa*): character displacement investigated through the geometric morphometric of skull. Acta Zoologica Academiae Scientiarum Hungaricae 44(1–2): 151–164.

Loy A., Corti M., Marcus L.F., 1993. Landmark data: Size and shape analysis in systematics. A case study on old world talpidae (Mammalia, Insectivora). In: Marcus L.F., Bello E., Garcia-Valdecasas A. (Eds.) Contribution to Morphometrics. Museo Nacional de Ciencias Naturales, Madrid. 215–240.

Macholan M., 2006. A geometric morphometric analysis of the shape of the first upper molar in mice of the genus *Mus* (Muridae, Rodentia). Journal of Zoology 270: 672–681.

Mosimann J.E., 1970. Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions. Journal of the American Statistical Association 65: 930–948.

Motokawa M., Lin L.-K., Lu K.-H., 2004. Geographic variation in cranial features of the polynesian rat *Rattus exulans* (Peole, 1848) (Mammalia: Rodentia: Muridae). The Raffles Bulletin of Zoology 52(2): 653–663.

Pages M., Bazin E., Galan M., Chaval Y., Claude J., Herbreteau V., Michaux J., Piry S., Morand S., Cosson J.-F., 2013. Cytonuclear discordance among southeast asian black rats (*Rattus rattus* complex). Molecular Ecology 22: 1019–1034.

Pages M., Chaval Y., Herbreteau V., Waengsothorn S., Cosson J.-F., Hugot J.-P., Morand S., Michaux J., 2010. Revisiting the taxonomy of the Rattini tribe: a phylogeny-based delimitation of species boundaries. BMC Evolutionary Biology 10: 184.

Paradis E., (2005). R for beginners. Available online at http://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rohlf F.J., 2013. tpsDIG2: Digitize landmarks & outlines from image files, scanner, or video. Available online at http://life.bio.sunysb.edu/morph/soft-dataacq.html

Rohlf F.J., Archie A.W., 1984. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Cuculidae). Systematic Zoology 33(3): 302–317.

Rohlf F.J., Marcus L.F., 1993. A revolution in morphometrics. Trends in Ecology and Evolution 8(4): 129–132.

Rohlf F.J., Slice D.E., 1990. Extension of the Procrustes method for the optimal superimposition of landmarks. Systematic Zoology 39: 40–59.

Rowe K.C., Aplin K.P., Braverstock P.R., Moritz C., 2011. Recent and rapid speciation with limited morphological disparity in the genus *Rattus*. Systematic Biology 60: 188–203.

Schlager S., 2013. Morpho: Calculations and visualizations related to Geometric Morphometrics. [R package version 0.17] http://sourceforge.net/projects/morpho-rpackage/

Shannon C.E., 1949. Communication in the presence of noise. Proc. Institute of Radio Engineers 37(1): 10–21.

Small C.G., 1996. The statistical theory of shape. Springer, New York.

Venables W., 1998. Exegeses on linear models. Paper presented to the S-PLUS User's Conference Washington, DC, 8-9th October, 1998.

Viscosi V., Cardini A., 2011. Leaf morphology, taxonomy and geometric morphometrics: A simplified protocol for beginners. PLoS ONE 6(10): e25630. doi:10.1371/journal.pone.0025630

Yezerinac S.M., Loogheed S.C., Handford P., 1992. Measurement error and morphometric studies: statistical power and observer experience. Systematic Biology 41(4): 471–482.

Associate Editor: A. Loy

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 103–109, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/6292/pdf
doi:10.4404/hystrix-24.1-6292

Research Article

Semilandmarks: a method for quantifying curves and surfaces

Philipp Gunz[a,*], Philipp Mitteroecker[b]

[a] *Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*
[b] *Department of Theoretical Biology, University of Vienna, Vienna, Austria*

Keywords:
Procrustes
geometric morphometrics
virtual reconstruction

Article history:
Received: 31 May 2012
Accepted: 29 July 2012

Acknowledgements
We want to thank Andrea Cardini for the invitation to contribute to this issue. The data in Figs. 4 and 5 were collected by Sarah Freidline. We thank Peter Brugger for access to the human brain MRI scans used in Fig. 2 and 3. Comments and suggestions by two expert referees helped improve the manuscript.

Abstract

Quantitative shape analysis using geometric morphometrics is based on the statistical analysis of landmark coordinates. Many structures, however, cannot be quantified using traditional landmarks. Semilandmarks make it possible to quantify two or three-dimensional homologous curves and surfaces, and analyse them together with traditional landmarks. Here we first introduce the concept of sliding semilandmarks and discuss applications and limitations of this method. In a second part we show how the sliding semilandmark algorithm can be used to estimate missing data in incomplete specimens.

Introduction

Here we illustrate a geometric morphometric measurement protocol, usually referred to as semilandmarks or sliding landmarks, for analysing homologous points locations together with curves and surfaces measured on a sample of organisms in two or three dimensions. Shape analysis using geometric morphometrics (GM) is based on the statistical analysis of landmark coordinates (Bookstein, 1991; Dryden and Mardia, 1998; Adams et al., 2004; Slice, 2007; Mitteroecker and Gunz, 2009). Landmarks are point locations that are biologically homologous between specimens. Many structures, however, cannot be quantified using traditional landmarks. Muscle attachments, for example, often are visible on bones, but traditional homologous landmark are insufficient to capture their shape. Smooth two- or three-dimensional curves (outlines) or surfaces, such as the neurocranial surface of a skull, are difficult to represent by landmarks because the landmark positions along the curve or surface cannot be homologized across different individuals. Semilandmarks make it possible to quantify two- or three-dimensional homologous curves and surfaces, and to analyse them together with traditional landmarks. Here we describe the concept of sliding semilandmarks and discuss possible applications and limitations of this method. We illustrate their application by several empirical examples. For a more technical discussion of semilandmarks and algebraic details see Bookstein (1997) and Gunz et al. (2005).

The concept of sliding semilandmarks was first introduced in Bookstein (1991) and published in detail for two-dimensional curves in Bookstein (1997). The method was later extended to three-dimensional curves and surfaces in Gunz et al. (2005). There are two alternative computational approaches to sliding semilandmarks, based on the two core techniques of geometric morphometrics, the Procrustes superimposition (Rohlf and Slice, 1990) and the thin-plate spline (TPS) deformation (Bookstein, 1989, 1991): Procrustes superimposition converts the raw landmark coordinates into shape coordinates by standardizing scale, position, and orientation of the landmark configurations. The Procrustes distance between two specimens (usually approximated by the Euclidean distance between the two sets of shape coordinates) is a measure of shape difference between two objects: it is zero only when two objects have the same shape. Procrustes distance also is the natural metric in Procrustes shape space (Kendall, 1984). The thin-plate spline (TPS) is an interpolation algorithm that serves as the mathematical underpinning of intuitive visualizations of shape differences, either as transformation grids or as warped images or surfaces (Fig. 1).

A TPS deformation visualizes the shape difference between one reference form and one target form, based on a set of homologous point coordinates measured on both forms. The space in-between the measured landmarks is interpolated by the TPS function "as smoothly as possible". More technically, the TPS minimizes the integral of the squared second derivatives, a quantity referred to as bending energy, which measures the amount of local shape deformation using a mathematical model borrowed from engineering (Bookstein, 1989)s. Transformation grids tell the reader how one would have to "squeeze and stretch" the reference shape to arrive at the target shape. Hand-drawn transformation grids were introduced by D'Arcy Thompson (1917); the TPS function makes it possible to produce transformation grids according to a mathematical model (for details on the mathematics see Bookstein 1989, 1991). In Fig. 1, a modern human surface is warped into a gibbon based on a large number of landmarks and semilandmarks. The TPS transformation grid visualizes the deformation from the human shape to the gibbon shape in the midsagittal plane. Note that no prior superimposition of reference and target is required for computing a TPS – the affine transformations of translation, rotation, and scaling have zero bending energy.

TPS transformation grids and TPS surface warps are the best visual aids to present shape differences between two forms as deformations (Oxnard and O'Higgins, 2009). It is important to keep in mind that the smooth TPS interpolation of the space in-between the measured landmark and semilandmark coordinates is an elegant mathematical formalism that is not intended to model a particular biological growth process, or the bending properties of organic tissues.

*Corresponding author
Email address: gunz@eva.mpg.de (Philipp Gunz)

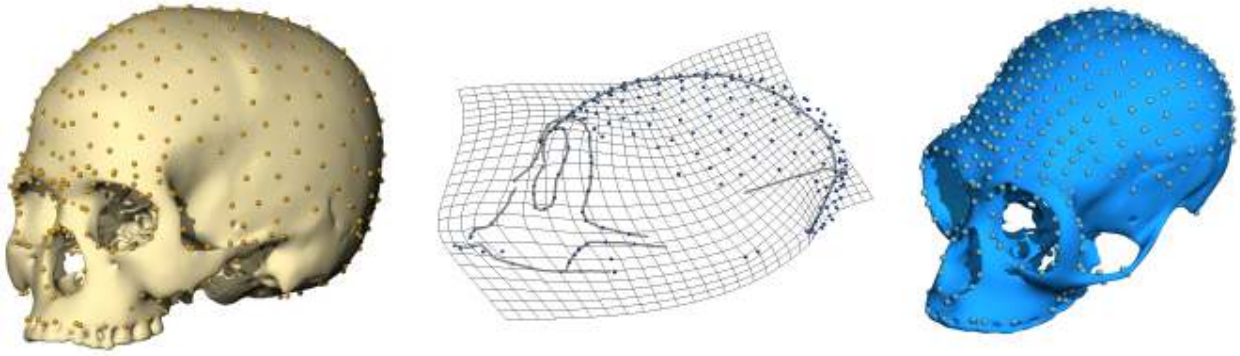


Figure 1 – Landmarks and several hundred semilandmarks measured on a computed tomographic scan of a modern human (left), and a gibbon (right). The middle panel shows the thin-plate spline (TPS) transformation grid from the human to the gibbon in the midsagittal plane. The TPS interpolation function used to draw this grid was used to warp the surface of the human cranium to the gibbon (blue surface on the right).

Point homology

In GM the measurement points are analysed with an implicit assumption about biological homology, usually based on ontogenetic or phylogenetic criteria (Bookstein, 1994; Hall, 2003; Klingenberg, 2008; Oxnard and O'Higgins, 2009). This biological homology must be expressed by a geometric homology, that is, by a point-to-point, curve-to-curve, or surface-to-surface correspondence. Furthermore, all specimens must have the same number of points. Semilandmarks are used to represent homologous curves and surfaces by sets of points, establishing a geometric homology between corresponding semilandmarks across the sample. In a first step, one distributes the same number of semilandmarks on the curves and surfaces of every specimen in roughly corresponding locations. This can either be done manually, or (semi)automatically (see below). In a second step, the spacing of the semilandmarks is optimized by allowing them to slide along the curves or surfaces (Fig. 2B). This sliding step establishes geometric correspondence of the semilandmarks by removing the effect of the arbitrary initial spacing; how the spacing is optimized differs between the two alternative computational approaches to sliding semilandmarks (Perez et al., 2006) (see below). After sliding, landmarks and semilandmarks can be treated the same way in subsequent statistical analyses. It is worth emphasizing that the rules of (semi)landmark equivalence or correspondence between forms vary according to the question at hand (Oxnard and O'Higgins 2009: 86). Oxnard and O'Higgins (2009) therefore stress that prior biological knowledge has to inform the measurement protocol, and that the choice of landmarks and semilandmarks should be driven by the biological hypotheses being tested. These authors discuss, e.g., how sagittal crests formed by the attachments of the temporal muscles in large adult male gorillas affect comparisons of neurocranial shape with female gorillas, which usually do not develop a marked crest. Below we will show that densely spaced surface semilandmarks are able to capture the shape differences between a modern human without crest, and an adult male gorilla with a marked sagittal crest (Fig. 7). In our example we treat the outer shell of the braincase in its entirety as homologous between these two specimens. While this will be helpful for many comparative purposes, inevitably this particular equivalence mapping cannot be a general solution for all possible research questions: it does not allow studying, e.g., changes in muscle attachments. Depending on the research question, one could either include additional information about the temporal muscles, e.g., by measuring curve semilandmarks along the temporal lines, or "avoid" the crests completely, e.g., by distributing the semilandmarks on the internal table of the braincase (i.e. the surface of the endocast).

Placing semilandmarks

The requirement for homology must guide any landmark and semilandmark measurement protocol. Points that are well defined by the local anatomy in all directions should be treated as traditional landmarks. Clearly observable curves on surfaces, such as ridges, should be treated as curve semilandmarks rather than surface semilandmarks. The number of semilandmarks depends on the complexity of the curve or surface and on the spatial scale of shape variation that is of interest. For many applications the semilandmarks shown in Fig. 1–8 are probably more densely spaced than necessary. However, this redundant oversampling

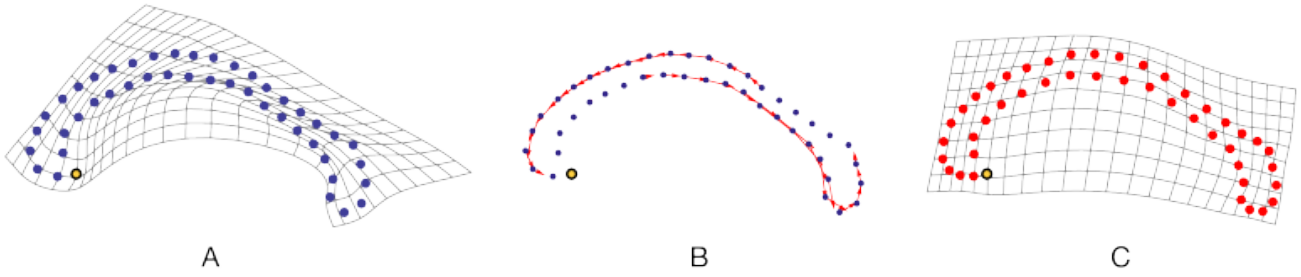


Figure 2 – One landmark (yellow) , and 87 curve semilandmarks were measured on the midline of the corpus callosum (see main text for details). TPS transformation grids between the Procrustes mean shape shape and one specimen before (A) and after semilandmark sliding (C). B: The initially equidistant curve semilandmarks (blue) slide (red arrows) along the curve until the TPS bending energy between this specimen and the Procrustes mean shape is minimal. Note that semilandmark sliding does not change the shape of the digitized curve, only the spacing of the semilandmarks along the curve changes.

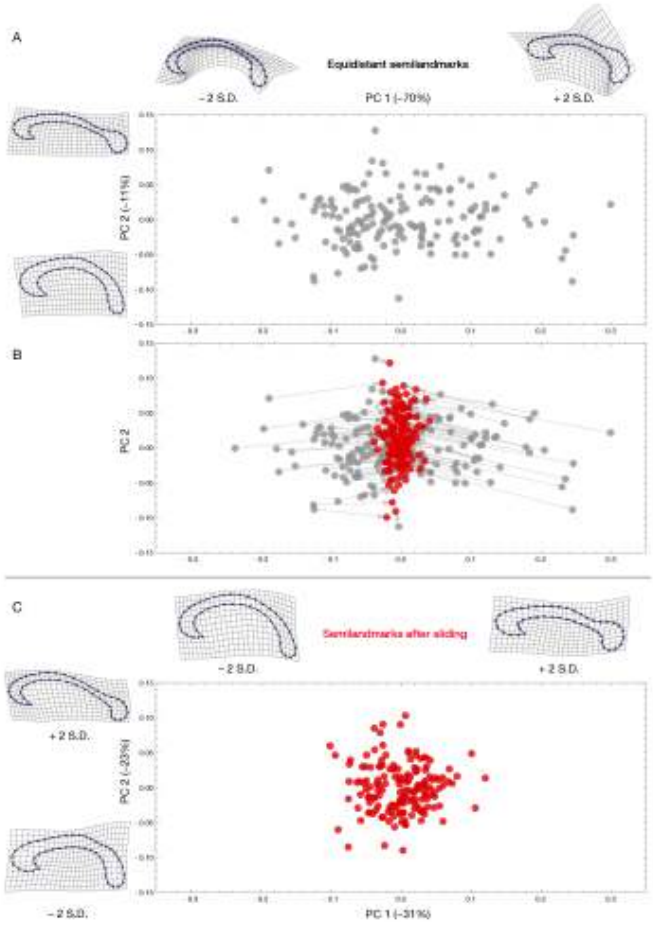


Figure 3 – Principal component (PC) scores of Procrustes shape coordinates before (A,B) and after semilandmark sliding (C). TPS transformation grids visualize the shape changes associated with the PC axes. B: the slid specimens (red) are projected into the PC space of the unslid (equidistant) specimens (gray). The first principal component of the equidistant semilandmarks is dominated by the effects of the semilandmark spacing (A).

of morphology is critical for effective visualizations and exploratory studies, as well as for estimating missing data (Mitteroecker and Gunz, 2009; Gunz et al., 2009b). In principle, it is always advantageous to capture the morphology in great detail using densely spaced semilandmarks. The only caveat is that the number of variables often exceeds the number of cases whereas many classical multivariate methods require an excess of cases over variables. It may thus be necessary to use dimension reduction techniques, such as principal component analysis, prior to other multivariate methods, or to use methods that impose no constraints on the number of variables, such as partial least squares analysis (Rohlf and Corti, 2000; Bookstein et al., 2003; Mitteroecker and Bookstein, 2007, 2008), between-group PCA (Mitteroecker and Bookstein, 2011), and permutation tests (Good, 2000); examples can be found in Mitteroecker et al. (2005) and Mitteroecker and Gunz (2009).

For curves in two or three dimensions, one usually starts by distributing a sequence of the same number of equidistantly spaced points along the curve. It is convenient (however, not algebraically necessary) to have the start and end of each curve delineated by real landmarks. While these roughly equidistant points can be placed manually, it is often practical to resample the curves to the same point count automatically.

Placing the same number of semilandmarks on homologous locations is more challenging for surfaces than it is for curves. One approach is to measure a mesh of surface semilandmarks on a single template specimen, and project this mesh onto all other forms in the sample: we start by measuring landmarks and curves on all specimens and automatically resample the curves to the same point count of equidistant curve semilandmarks. We then measure a mesh of surface semilandmarks on a template specimen and use the TPS interpolation function computed from the landmarks and curve semilandmarks to warp this mesh from the template to each specimen. These warped mesh points usually "float" in the vicinity of the actual surface of the specimen and therefore need to be lofted onto the surface before they can be used as surface semilandmarks. At this stage all specimens in a sample have the same number of landmarks, curve semilandmarks, and surface semilandmarks in approximately homologous positions.

Semilandmark sliding

In several geometric morphometric approaches to outline analysis, equidistant points along curves are analysed directly, however we will show below that equidistance can lead to serious statistical and visualisation artifacts. It is important to keep in mind that equidistance is an intuitive, yet arbitrary way of distributing semilandmarks, which does not necessarily lead to geometric or biological correspondence of the points across specimens (Bookstein, 1997; Gunz et al., 2005). Likewise, the spacing of surface semilandmarks is also arbitrary. The semilandmarks are thus allowed to slide along their curve or surface in order to remove the effects of the arbitrary spacing by "optimizing" the position of the semilandmarks with respect to the average shape of the entire sample (average of the Procrustes shape coordinates). The sliding algorithm is iterative: First the semilandmarks of each specimen are allowed to slide with respect to one (arbitrary) template specimen. Then one computes a Procrustes superimposition from these sled coordinates and a mean shape. Subsequently all semilandmarks are allowed to slide with respect to the average Procrustes shape.

Fig. 2 and 3 illustrate that equidistant semilandmarks can cause serious statistical and visualization artefacts. The semilandmarks were measured on magnetic resonance imaging (MRI) scans of human brains; they trace the midline of the corpus callosum, the structure in the brain that connects the two hemispheres. On 163 MRI scans we extracted the midsagittal slice, digitized one anatomical landmark, and distributed 87 equidistant curve-semilandmarks along the outline of the corpus callosum. We then allowed the equidistant semilandmarks on each individual to slide along the curve to minimize the TPS bending energy between this individual and the Procrustes mean shape. Fig. 2 shows the TPS transformation grid between the mean shape and one individual before (Fig. 2A) and after sliding (Fig. 2C). It is evident from Fig. 2 that semilandmark sliding has a profound effect on the TPS visualization: what looks like a strong local shape signal in the posterior part of the corpus callosum in Fig. 2A, turns out to be an artefact of the equidistant semilandmark spacing: this signal disappears when the semilandmarks are allowed to slide along the curve. We then computed two principal component analyses (PCA) in shape space to visualize the axes of largest shape variation. A comparison of the first two PC scores in shape space before (Fig. 3A) and after semilandmark sliding (Fig. 3B and 3C) reveals that the arbitrary semilandmark spacing accounts for the axis of largest variance in the dataset, the first principal component in shape space. In Fig. 3B the PC scores after semilandmark sliding (red) are projected into the PC space of the equidistant semilandmarks. This figure demonstrates that the shape variation along the first PC "disappears" almost completely when the semilandmarks are allowed to slide along the corpus callosum outline. As expected, the TPS transformation grids that visualize the shape changes along the first two principal components of shape space of the sled semilandmarks (Fig. 3C) are much smoother than the TPS transformation grids of Fig. 3A.

To make the computation of the semilandmark sliding computationally tractable, the semilandmarks slide on tangent lines to the respective curve, or on tangent planes to the respective surface (Fig. 4A). For curve semilandmarks these tangents are computed based on the two adjacent landmarks and semilandmarks; it is therefore convenient to start and end each curve with real landmarks, so reliable tangents can be computed for the first and last semilandmark along a curve. For surface semilandmarks the tangent planes are computed as the first two eigenvectors of the covariance matrix of adjacent points on the surface; for each semilandmark here we used the five closest landmarks

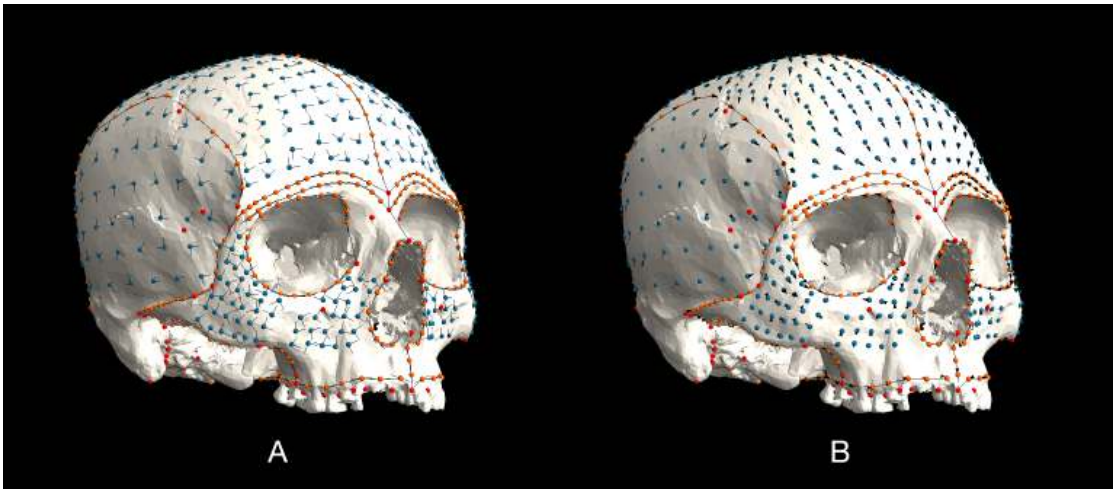


Figure 4 – Landmarks (red), curve semilandmarks (orange), and surface semilandmarks (blue) on a modern human cranium. A: Semilandmarks are allowed to slide along tangents (curves), and tangent planes (surfaces) so as to minimize the thin-plate spline bending energy between this specimen and the Procrustes average shape of the sample. B: After sliding, the semilandmarks are projected back onto the surface. Arrows connect semilandmarks before and after sliding. In this example, the positions of the semilandmarks change only subtly.

and semilandmarks to approximate the respective tangent planes. After each sliding step the slid semilandmarks can be projected back onto the curves or surfaces to ensure that they stay on the form. Whether or not this projection step is necessary depends on the complexity of the curve, the number of semilandmarks, and the amount of sliding. Fig. 4A shows the tangents for each curve semilandmark (orange spheres), and the two tangent vectors for each surface semilandmark (blue spheres). Fig. 4B visualizes the semilandmarks before and after sliding (minimizing bending energy) in a sample of *Homo sapiens* crania; it is evident that in this example the position of the semilandmarks only changes subtly.

Procrustes sliding *vs.* bending energy sliding

As mentioned above, there are two alternative computational approaches to sliding semilandmarks. In both approaches the semilandmarks slide so as to minimize shape differences between each specimen and the average shape in the sample. That is, shape variation due only to the arbitrary spacing of semilandmarks is reduced. The two approaches differ in the way shape differences are quantified, and so, in what is being minimized. In the most common approach, the one originally published by Bookstein (1997) and Gunz et al. (2005), the bending energy between all specimens and the average shape is minimized by the iterative sliding. Alternatively, it has been suggested to minimize Procrustes distance instead of bending energy (Fig. 5).

Minimizing bending energy is the optimal solution for producing transformation grids between specimens because both techniques are based on the TPS formalism. Bending energy only takes into account local shape deformation; uniform shape differences such as stretching and shearing have no effect on bending energy and the sliding process. Minimizing Procrustes distance, which is faster to compute than minimizing bending energy, is a least-squares procedure and more closely resembles the usual sum-of-squares decomposition in statistics. The most important difference, however, is the notion of homology implicit in the two approaches. Bending energy is based on all landmarks and semilandmarks and the "smoothness" of the shape deformation as a whole. All semilandmarks slide together and are influenced by the anatomical landmarks. When minimizing Procrustes distance, by contrast, each semilandmark slides separately and, apart from the common Procrustes superimposition, the sliding is not influenced by the other landmarks and semilandmarks. For instance, when minimizing Pro-

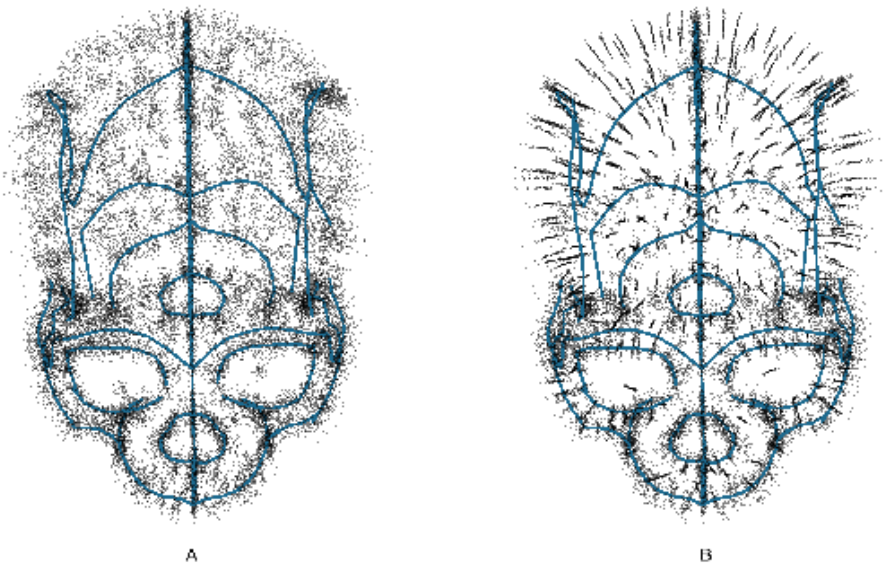


Figure 5 – Procrustes superimposition of 46 modern human crania. A: Semilandmarks were allowed to slide so as to minimize thin-plate spline bending energy between each specimen and the Procrustes average shape (blue curves). B: Semilandmarks were allowed to slide so as to mimize the Procrustes distance between each specimen and the Procrustes average shape.

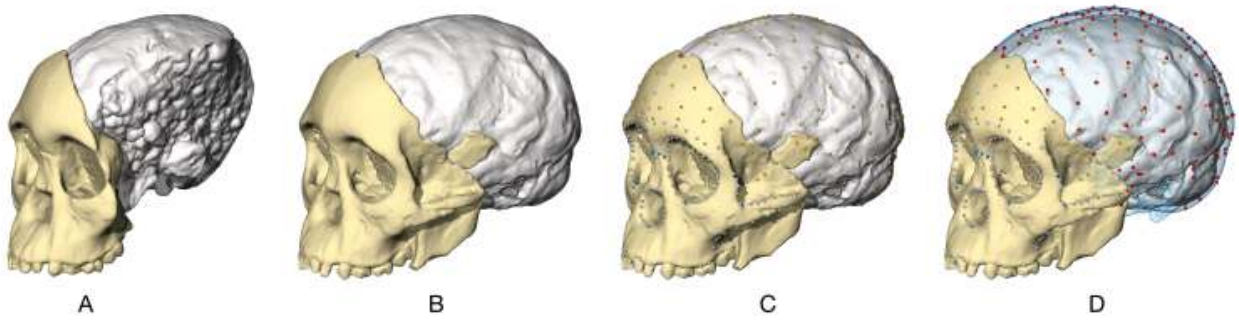


Figure 6 – Virtual reconstruction of the *Australopithecus africanus* fossil cranium Taung 1. A: Computed tomographic scan of the original fossil. B: Face and natural endocast were mirror-imaged on the computer. C: Landmarks (blue), curve semilandmarks (gray), surface semilandmarks (orange) were measured on the mirror-imaged version of the Taung child and a complete reference specimen (a subadult human). D: Landmarks and semilandmarks missing on the Taung child (red) were declared missing and estimated by allowing them to move freely so as to minimize the bending energy between Taung and the reference cranium. TPS interpolation was then used to warp the braincase of the human reference cranium to Taung (semitransparent blue surface).

crustes distance, curve semilandmarks can potentially slide beyond the endpoint of the curve or may pass another semilandmark. In contrast, this is almost impossible when minimizing bending energy. Both sliding approaches yield similar results if shape variation is small and if the semilandmarks do not need to slide much; for larger shape variation and more extensive sliding, minimizing bending energy usually leads to better results that are in line with our notion of biological homology. As mentioned above, a potential pathology is that semilandmarks end up on a different structure after sliding, e.g., semilandmarks on the frontal bone might "cross" the coronal suture and end up on the parietal bone. This is easily avoided when sliding minimizes bending energy, however, by placing a few real landmarks or curve landmarks on the coronal suture; these points will constrain the movement of the surface semilandmarks. Whether such constraints on sliding are desirable depends on the research question: for a comparison of overall braincase shape across genera the relative position of sutures might be not be particularly informative. In this case one might consider allowing semilandmarks to slide across the entire neurocranial surface, thereby ignoring sutures (Gunz et al., 2005). Incorporating landmarks or curve semilandmarks along sutures, in contrast, will be informative about individual development instead (i.e., how a particular neurocranium manages to realize its shape ontogenetically.)

Semilandmarks for missing data estimation

As GM methods require all specimens to have the same number of landmark and semilandmark coordinates, partially preserved specimens present a big challenge for any GM analysis. One can either restrict the analysis to the subset of landmarks and semilandmarks available on all specimens, or estimate the missing data. In many applications the first option is impractical, as the number of landmarks available on all specimens is often very small. In bilaterally symmetrical structures, like the skull, it usually best to start by mirror-imaging preserved parts from one side to the other (Gunz et al., 2009a, 2011, 2012; Weber and Bookstein, 2011; Weber at al., 2012; Neubauer et al., 2012; Benazzi et al., 2009; Benazzi and Senck, 2010; Benazzi et al., 2011a,b; Zollikofer et al., 1995, 1998; Ponce de León and Zollikofer, 1999, 2001; Zollikofer, 2002; Zollikofer et al., 2005; Ponce de León et al., 2008; Kalvin et al., 1995; O'Higgins et al., 2011; Zollikofer and Ponce de León, 2005). For parts that are missing on both sides, or in the symmetry axis, one can use the semilandmark algorithm to estimate the location of landmarks and semilandmarks that cannot be measured in a specimen, because, e.g., a part is broken off. At its core the concept of semilandmark sliding is already a missing data algorithm (Gunz et al., 2009a): the position of semilandmarks along a curve or surface is not well-defined and therefore not interesting; only their position in the direction normal to the curve or surface is relevant for the statistical analysis. In other words, a semilandmark's position along the curve or surface is missing: a curve semilandmark is missing one coordinate (along the tangent direction); a surface semilandmark is missing two (along the tangent plane); missing landmarks or semilandmarks are missing all three coordinates.

To estimate missing coordinates in an incomplete target form, a thin-plate spline interpolation is computed based on the landmarks and semilandmarks that are available in both a complete reference specimen and the incomplete target specimen. This interpolation function is used to map the missing landmarks and semilandmarks from the reference onto the target (Gunz et al., 2009b). This is accomplished during the sliding step, as missing landmarks and semilandmarks are allowed to move freely so as to minimize the bending energy between the incomplete specimen and a complete reference specimen.

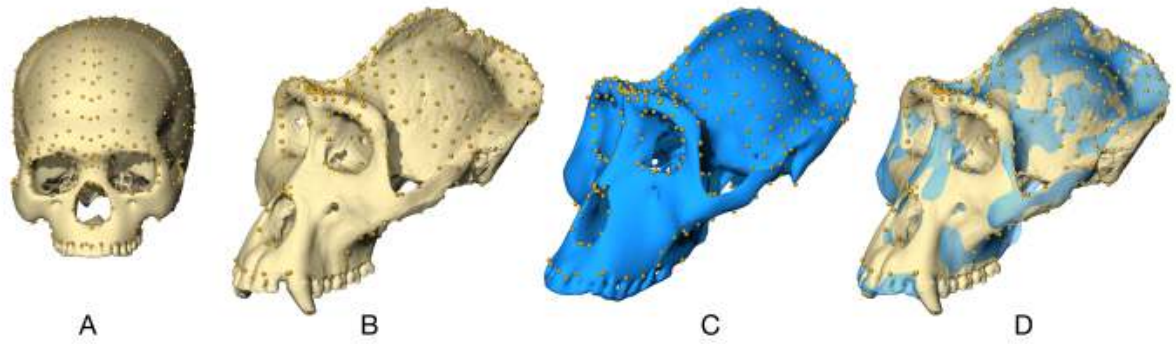


Figure 7 – Landmarks and semilandmarks on a modern human (A) and a gorilla male (B), were used to warp the modern human surface (C). D: Despite the substantial shape differences, the warped human surface (semitransparent blue) matches closely with the actual surface of the gorilla (bone colour).

We illustrate this virtual reconstruction protocol in Fig. 6, using a computed tomographic scan of a hominin fossil belonging to the species *Australopithecus africanus*. The holotype of this species, the subadult specimen Taung 1, comprises parts of the face, mandible, and a fossilized imprint of the braincase – a so-called natural endocast (Fig. 6A). After mirror-imaging across the midsagittal plane (Fig. 6B), we measured landmarks and semilandmarks on this specimen (Fig. 6C) and a complete reference cranium (a modern human subadult). We then used the TPS interpolation function to estimate the missing landmarks and semilandmarks on the exterior of the braincase in the child (Fig. 6D).

The accuracy of a virtual reconstruction depends upon the size of the defect and the number of coordinates that are recorded in the vicinity of the missing part such that reconstruction of a small defect with many adjacent coordinates will have greater accuracy (Gunz et al., 2009a; Grine et al., 2010; Weber at al., 2012; Benazzi et al., 2011a; Weber and Bookstein, 2011; Neubauer et al., 2012). A TPS interpolation can only be computed between two forms. As one might expect, the choice of the reference specimen used for estimation of the missing data will influence the final shape of the reconstruction. In most practical applications, however, it is impossible to determine what the "correct" reference form is. In Gunz et al. (2009b) we have therefore suggested creating multiple reconstructions based on as many reference forms as possible to assess the influence of the arbitrary reference choice. Shape differences among the resulting reconstructions of the same specimen provide a sense of the reconstruction uncertainty (Gunz et al., 2009a, 2010; Grine et al., 2010; Benazzi et al., 2011a; Neubauer et al., 2012; Gunz et al., 2012).

Applications and limitations

Using several hundred semilandmarks one can use a TPS interpolation to morph a human cranium into a gibbon (Fig. 1), and a gorilla (Fig. 7C). The surface warps are only used as visual aids here; however, the close correspondence between the actual gorilla surface and the warped human surface (Fig. 7D) confirms that after semilandmark sliding these point coordinates can be treated as homologous, even when the shape differences are fairly substantial, and even if there are few anatomical landmarks (like, e.g., on the neurocranium).

Applications of semilandmarks are not restricted to crania (Weber et al., 2001; Mitteroecker et al., 2004; Schaefer et al., 2004; Mitteroecker et al., 2004; Mitteroecker and Bookstein, 2008; Gunz et al., 2009a; Gonzalez et al., 2010; Heuzé et al., 2010; Benazzi et al., 2011a; Stansfield Nee Bulygina and Gunz, 2011; Weber and Bookstein, 2011; Weber at al., 2012), endocasts (Neubauer et al., 2004, 2005; Neubaer et al., 2009; Gunz et al., 2010; Neubauer et al., 2010, 2012; Gunz et al., 2012), or mandibles (Coquerelle et al., 2011; Benazzi et al., 2011b). They can also be used to quantify the enamel dentine junction (Skinner et al., 2008, 2009a; Skinner and Gunz, 2010) from high-resolution computed tomographic data of teeth (Fig. 8A), and the bony labyrinth of the inner ear (Fig. 8B). As shape is captured in such great detail using semilandmarks, it was possible to document subtle, yet statistically significant shape differences between subspecies of chimpanzees (*Pan troglodytes troglodytes* and *P. t. verus*), both for the enamel-dentine junction (Skinner et al., 2009a), and the semicircular canals of the inner ear (Gunz et al., 2012). The ability to discriminate dental shape and labyrinth shape at the subspecies level demonstrates that semilandmarks can be extremely useful in taxonomic studies of extant and fossil specimens (Gunz et al., 2012). It is not only possible to capture very subtle shape differences among groups, one can also visualize the results of the statistical analysis as shape deformations.

Semilandmarks (like all geometric morphometric techniques) are not suitable for the comparison of forms when the curves and surfaces are not homologous among specimens. The method was also developed for capturing fairly smooth surfaces and sharp ridges, so there are practical limitations regarding the complexity of surfaces. While it is, e.g., possible to quantify the relatively smooth surface of brain endocasts using semilandmarks (Neubaer et al., 2009), the gyri and sulci of the brain's surface might prove to be too complex and irregular for semilandmarks.

Semilandmarks and alternative methods

Several alternative morphometric approaches exist for quantifying curves and surfaces. Of particular interest are recent methodological advances that require less manual input than the sliding semilandmark approach discussed here. In these "homology-free" methods (for reviews see Polly 2008; Klingenberg 2008) the forms usually are aligned first – either based on a few landmarks, or completely automatically based on principal axes; subsequently a correspondence map between two shapes is computed automatically (Specht et al., 2007; Polly, 2008; Durrelman et al., 2011; Boyer et al., 2011). The obvious advantage of such a semi-automated data collection is the possibility to process large sample sizes in a fairly short amount of time (Polly, 2008; Klingenberg, 2008). The processing speed comes at a cost, however. Many automated methods are restricted to fairly small ranges of shape variation. Moreover, the point homology across specimens, which is "enforced" by the experienced morphometrician measuring semilandmarks on curves and surfaces manually, is no longer guaranteed. As a result, sample averages and variances may be meaningless and biologically not interpretable. If one aims to go beyond the mere discrimination of groups and tries to identify the biological factors underlying shape differences, the time spent digitizing curves and surfaces as semilandmarks is almost always worthwhile.

Software tools

Sliding semilandmarks for two-dimensional data (usually digitized from images) can easily be handled by the free TPS series by Jim Rohlf (http://life.bio.sunysb.edu/ee/rohlf/software.html). Three-dimensional

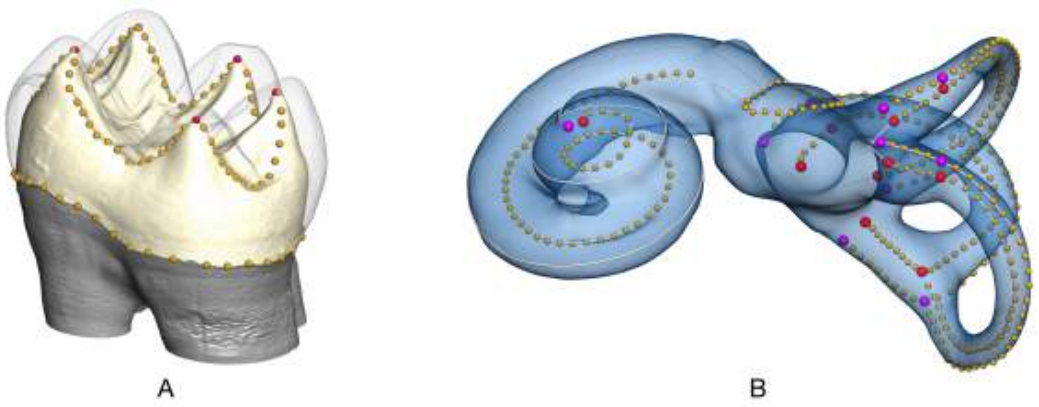


Figure 8 – A: Landmarks (red) and curve semilandmarks (orange) quantify the enamel-dentine junction of a molar based on a micro CT scan (Skinner et al., 2008, 2009a,b; Skinner and Gunz, 2010). B: Landmarks (red and magenta) and curve semilandmarks (orange) on the bony labyrinth of the inner ear (Gunz et al., 2012).

semilandmarks are supported by the EVAN toolbox by the EVAN Society (http://evan-society.org/), and the commercial software package Viewbox (http://www.dhal.com/). Scripts for sliding semilandmarks in the statistical software package R (http://www.R-project.org), developed by Stefan Schlager are available (http://sourceforge.net/projects/morpho-rpackage). The Mathematica (Wolfram Inc.) code for 2D and 3D sliding semilandmarks developed by the authors, which was used to create the figures and analyses in this paper, is available from the authors upon request.

References

Adams D.C., Rohlf F.J., Slice D.E., 2004. Geometric morphometrics: ten years of progress following the "revolution". Italian Journal of Zoology 71: 5–16.

Benazzi S., Bookstein F.L., Strait D.S., Weber G.W., 2011. A new OH5 reconstruction with an assessment of its uncertainty. J. Hum. Evol. 61(1): 75–88.

Benazzi S., Fiorenza L., Kozakowski S., Kullmer O., 2011. Comparing 3D Virtual Methods for Hemimandibular Body Reconstruction. Anat. Rec. (Hoboken) 294(7): 1116–1125.

Benazzi S., Senck S., 2010. Comparing 3-Dimensional Virtual Methods for Reconstruction in Craniomaxillofacial Surgery. J. Oral. Maxillofac. Surg. 69(4): 1184–1194.

Benazzi S., Stansfield E., Milani C., Gruppioni G., 2009. Geometric morphometric methods for three-dimensional virtual reconstruction of a fragmented cranium: the case of Angelo Poliziano. Int. J. Legal. Med. 123: 333–344.

Bookstein F.L., 1989. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE. Transactions on Pattern Analysis and Machine Intelligence V: 567–585.

Bookstein F.L., 1991. Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge University Press, Cambridge, UK.

Bookstein F.L., 1994. Can biometrical shape be a homologous character? In: Hall B.K. (Ed.). Homology: The Hierarchical Basis of Comparative Biology. Academic Press, New York. 197–227.

Bookstein F.L., 1997. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Med. Image. Anal. 1: 225–243.

Bookstein F.L., Gunz P., Mitteroecker P., Prossinger H., Schaefer K., Seidler H., 2003. Cranial integration in *Homo*: singular warps analysis of the midsagittal plane in ontogeny and evolution. J. Hum. Evol. 44: 167–187.

Boyer D.M., Lipman Y., St. Clair E., Puente J., Patel B.A., Funkhouser T., Jernvall J., Daubechies I., 2011. Algorithms to automatically quantify the geometric similarity of anatomical surfaces. Proc. Natl. Acad. Sci. U.S.A. 108: 18221–18226.

Coquerelle M., Bookstein F.L., Braga J., Halazonetis D.J., Weber G.W., Mitteroecker P., 2011. Sexual dimorphism of the human mandible and its association with dental development. Am. J. Phys. Anthropol. 145: 192–202.

Dryden I.L., Mardia K.V., 1998. Statistical Shape Analysis. John Wiley & Sons, London, Chichester.

Durrleman S., Pennec X., Trouvé A., Ayache N., Braga J., 2011. Comparison of the endocranial ontogenies between chimpanzees and bonobos via temporal regression and spatiotemporal registration. J. Hum. Evol. 62(1): 74–88.

Gonzalez P.N., Perez S.I., Bernal V., 2010. Ontogeny of robusticity of craniofacial traits in modern humans: a study of South American populations. Am. J. Phys. Anthropol. 142: 367–379.

Good P.I., 2000. Permutation tests: a practical guide to resampling methods for testing hypotheses. Springer, New York.

Grine F.E., Gunz P., Betti-Nash L., Neubauer S., Morris A.G., 2010. Reconstruction of the Late Pleistocene human skull from Hofmeyr, South Africa. J. Hum. Evol. 59: 1–15.

Gunz P., Bookstein F.L., Mitteroecker P., Stadlmayr A., Seidler H., Weber G.W., 2009. Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario. Proc. Natl. Acad. Sci. U.S.A. 106: 6094-6098.

Gunz P., Mitteroecker P., Bookstein F.L., 2005. Semilandmarks in three dimensions. In: Slice D.E. (Ed.). Modern Morphometrics in Physical Anthropology. Kluwer Academic/Plenum Publishers, New York. 73–98.

Gunz P., Mitteroecker P., Neubauer S., Weber G.W., Bookstein F.L., 2009. Principles for the virtual reconstruction of hominin crania. J. Hum. Evol. 57: 48–62.

Gunz P., Neubauer S., Golovanova L., Doronichev V., Maureille B., Hublin J.J., 2012. A uniquely modern human pattern of endocranial development. Insights from a new cranial reconstruction of the Neandertal newborn from Mezmaiskaya. J. Hum. Evol. 62: 300–313.

Gunz P., Neubauer S., Maureille B., Hublin J.J., 2010. Brain development after birth differs between Neanderthals and modern humans. Current Biology 20: R921-R922.

Gunz P., Neubauer S., Maureille B., Hublin J.J., 2011. Virtual Reconstruction of the Le Moustier 2 newborn skull. Implications for Neandertal ontogeny. PALEO 22: 155–172.

Gunz P., Ramsier M., Kuhrig M., Hublin J.J., Spoor F., 2012. The mammalian bony labyrinth reconsidered, introducing a comprehensive geometric morphometric approach. J. Anat. 220(6): 529–543.

Hall B.K., 2003. Descent with modification: the unity underlying homology and homoplasy as seen through an analysis of development and evolution. Biological Reviews 78: 409-433.

Heuzé Y., Boyadjiev S.A., Marsh J.L., Kane A.A., Cherkez E., Boggan J.E., Richtsmeier J.T., 2010. New insights into the relationship between suture closure and craniofacial dysmorphology in sagittal nonsyndromic craniosynostosis. J. Anat. 217: 85–96.

Kalvin A.D., Dean D., Hublin J.-J., 1995. Reconstruction of human fossils. IEEE Computer Graphics and. Applications 15: 12–15.

Kendall D.G., 1984. Shape manifolds, Procrustean metrics and complex projective spaces. Bull. London Math. Soc. 16: 81–121.

Klingenberg C.P., 2008. Novelty and "Homology-free´´ Morphometrics: What's in a Name? Evolutionary Biology 35: 186–190.

Mitteroecker P., Bookstein F., 2007. The conceptual and statistical relationship between modularity and morphological integration. Syst. Biol. 56: 818–836.

Mitteroecker P., Bookstein F., 2008. The evolutionary role of modularity and integration in the hominoid cranium. Evolution 62: 943–958.

Mitteroecker P., Bookstein F., 2011. Linear Discrimination, Ordination, and the Visualization of Selection Gradients in Modern Morphometrics. Evolutionary Biology 39: 100–114.

Mitteroecker P., Gunz P., 2009. Advances in Geometric Morphometrics. Evolutionary Biology 36: 235–247.

Mitteroecker P., Gunz P., Bernhard M., Schaefer K., Bookstein F.L., 2004. Comparison of cranial ontogenetic trajectories among great apes and humans. J. Hum. Evol. 46: 679–697.

Mitteroecker P., Gunz P., Bookstein F.L., 2005. Heterochrony and geometric morphometrics: A comparison of cranial growth in *Pan paniscus* versus *Pan troglodytes*. Evolution and Development 7: 244–258.

Mitteroecker P., Gunz P., Weber G.W., Bookstein F.L., 2004. Regional dissociated heterochrony in multivariate analysis. Annals of Anatomy 186: 463–470.

Neubauer S., Gunz P., Hublin J.J., 2009. The pattern of endocranial ontogenetic shape changes in humans. J. Anat. 215: 240–255.

Neubauer S., Gunz P., Hublin J.J., 2010. Endocranial shape changes during growth in chimpanzees and humans: a morphometric analysis of unique and shared aspects. J. Hum. Evol. 59: 555–566.

Neubauer S., Gunz P., Mitteroecker P., Weber G.W., 2004. Three-dimensional digital imaging of the partial *Australopithecus africanus* endocranium MLD 37/38. Can. Assoc. Radiol. J. 55: 271–278.

Neubauer S., Gunz P., Schwarz U., Hublin J.J., Boesch C., 2012. Endocranial Volumes in an Ontogenetic Sample of Chimpanzees from the Tai Forest National Park, Ivory Coast. Am. J. Phys. Anthropol. 147(2): 319–325. doi:10.1002/ajpa.21641

Neubauer S., Gunz P., Weber G.W., 2005. Digital reconstruction of *P. boisei* OH 5. Am. J. Phys. Anthropol. Suppl. 40: 161.

Neubauer S., Gunz P., Weber G.W., Hublin J.J., 2012. Endocranial volume of *Australopithecus africanus*: New CT-based estimates and the effects of missing data and small sample size. J. Hum. Evol. 62(4): 498–510.

O'Higgins P., Cobb S.N., Fitton L.C., Gröning F., Phillips R., Liu J., Fagan M.J., 2011. Combining geometric morphometrics and functional simulation: an emerging toolkit for virtual functional analyses. J. Anat. 218: 3–15.

Oxnard C., O'Higgins P., 2009. Biology Clearly Needs Morphometrics. Does Morphometrics Need Biology? Biological Theory 4: 84–89.

Perez S.I., Bernal V., Gonzalez P.N., 2006. Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. J. Anat. 208: 769–784.

Polly P.D., 2008. Developmental Dynamics and G-Matrices: Can Morphometric Spaces be Used to Model Phenotypic Evolution? Evolutionary Biology 35: 83–96.

Ponce de León M.S., Golovanova L., Doronichev V., Romanova G., Akazawa T., Kondo O., Ishida H., Zollikofer C.P., 2008. Neanderthal brain size at birth provides insights into the evolution of human life history. Proc. Natl. Acad. Sci. U.S.A. 105: 13764–13768.

Ponce de León M.S., Zollikofer C.P., 1999. New evidence from Le Moustier 1: computer-assisted reconstruction and morphometry of the skull. Anat. Rec. 254: 474–489.

Ponce de León M.S., Zollikofer C.P., 2001. Neanderthal cranial ontogeny and its implications for late hominid diversity. Nature 412: 534–538.

Rohlf F.J., Corti M., 2000. Use of two-block partial least-squares to study covariation in shape. Syst. Biol. 49: 740–753.

Rohlf F.J., Slice D., 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39: 40–59.

Schaefer K., Mitteroecker P., Gunz P., Bernhard M., Bookstein F.L., 2004. Craniofacial sexual dimorphism patterns and allometry among extant hominids. Annals of Anatomy 186: 471–478.

Skinner M.M., Gunz P., 2010. The presence of accessory cusps in chimpanzee lower molars is consistent with a patterning cascade model of development. J. Anat. 217(3): 245–253.

Skinner M.M., Gunz P., Wood B.A., Boesch C., Hublin J.J., 2009. Discrimination of extant *Pan* species and subspecies using the enamel-dentine junction morphology of lower molars. Am. J. Phys. Anthropol. 140: 234–243.

Skinner M.M., Gunz P., Wood B.A., Hublin J.J., 2008. Enamel-dentine junction (EDJ) morphology distinguishes the lower molars of *Australopithecus africanus* and *Paranthropus robustus*. J. Hum. Evol. 55: 979–988.

Skinner M.M., Gunz P., Wood B.A., Hublin J.J., 2009. How Many Landmarks? Assessing the Classification Accuracy of *Pan* Lower Molars Using a Geometric Morphometric Analysis of the Occlusal Basin as Seen at the Enamel-Dentine Junction. Front. Oral. Biol. 13: 23–29.

Slice D.E., 2007. Geometric morphometrics. Annu. Rev. Anthropol. 36: 261–281.

Specht M., Lebrun R., Zollikofer C.P.E., 2007. Visualizing shape transformation between chimpanzee and human braincases. Visual. Computer 23: 743–751.

Stansfield Nee Bulygina E., Gunz P., 2011. Skhodnya, Khvalynsk, Satanay, and Podkumok calvaria: possible Upper Paleolithic hominins from European Russia. J. Hum. Evol. 60: 129–144.

Thompson D., 1917. On Growth and Form. Cambridge University Press, Cambridge, UK.

Weber G.W., Bookstein F.L., 2011. Virtual anthropology: a guide to a new interdisciplinary field. Springer, Wien, London.

Weber G.W., Gunz P., Neubauer S., Mitteroecker P., Bookstein F.L., 2012. Digital South African fossils: morphological studies using reference-based reconstruction and electronic preparation. In: Reynolds S.C., Gallagher A. (Eds.). African Genesis: Perspectives on Hominin Evolution. Cambridge University Press, Cambridge, UK. 298–316.

Weber G.W., Schaefer K., Prossinger H., Gunz P., Mitteroecker P., Seidler H., 2001. Virtual anthropology: the digital evolution in anthropological sciences. J. Physiol. Anthropol. Appl. Human. Sci. 20: 69–80.

Zollikofer C.P., Ponce de León M.S., Lieberman D.E., Guy F., Pilbeam D., Likius A., Mackaye H.T., Vignaud P., Brunet M., 2005. Virtual cranial reconstruction of *Sahelanthropus tchadensis*. Nature 434: 755–759.

Zollikofer C.P.E., 2002. A Computational Approach to Paleoanthropology. Evolutionary Anthropology 11: 64-67.

Zollikofer C.P.E., Ponce de Leon M.S., Martin R.D., Stucki P. 1995. Neanderthal computer skulls. Nature. 375: 283–285.

Zollikofer C.P.E., Ponce de León M.S., 2005. Virtual reconstruction: a primer in computer-assisted paleontology and biomedicine. Wiley-Interscience, Hoboken, N.J.

Zollikofer C.P.E., Ponce De León M.S., Martin R.D., 1998. Computer-assisted paleoanthropology. Evolutionary Anthropology 6: 41–54.

Associate Editor: A. Loy

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 110–125, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/6299/pdf
doi:10.4404/hystrix-24.1-6299

Research Article

Geometric Morphometric Approaches to Acoustic Signal Analysis in Mammalian Biology

Norman MacLeod[a,*], Jonathan Krieger[b], Kate E. Jones[c]

[a] *Palaeontology Department, The Natural History Museum, Cromwell Road, London, SW7 5BD, UK*
[b] *Herbarium, Library, Art & Archives Directorate, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK*
[c] *Institute of Zoology, Zoological Society of London, Regents Park, London NW1 4RY, UK*

Keywords:
morphometrics
bats
echolocation
systematics
acoustics
geometry
shape analysis

Article history:
Received: 27 May 2012
Accepted: 9 July 2012

Acknowledgements
The order of authorship reflects the relative contribution of individuals to the writing of this article. We would like to thank the following have provided support and/or encouragement for the bat echolocation research reported herein: Charlotte Walters, Lucinda Kirkpatrick, Stuart Parsons, and Alanna Collen. This investigation depended critically on the EchoBank bat echolocation call archive. All contributors to this archive are acknowledged here and thanked for their efforts.

Abstract

In the quarter century since the development of geometric morphometrics the community of practitioners has largely been occupied with training issues and anatomy-based applications research in the biological sciences. However, just as the scope of geometry transcends comparative anatomy, the potential scope of morphometric analysis transcends investigations of the form and shape of organismal bodies. An important area of opportunity for morphometricians lies in the application of geometric methods to non-traditional form/shape analysis problems. To illustrate the potential of morphometric data analysis approaches to contribute to investigations outside its traditional base in (physical) morphology we report here results of an investigation into the morphometrics of bat echolocation calls. By treating Hanning windowed spectrograms of bat search echolocation calls as complex 3D surfaces, and by using a variant of eigensurface analysis to sample and compare these surfaces, it is possible to identify bat species to very high levels of accuracy (> 90% for raw cross-validated training set identifications, > 80% for jackknifed training set identifications), even for species (e.g., *Myotis*) whose spectrograms have resisted separation into species-specific clusters using traditional spectrogram descriptors. Moreover, the shape modeling capabilities of geometric morphometrics render the complex mathematical subspaces within which these spectrogram shape data reside – along with the discriminant functions used to separate training-set clusters – interpretable in a simple, intuitive, and biologically informative manner. These results demonstrate the rich source of species-specific information bioacoustic signal structures represent. They also illustrate the type of advances that can be made when morphometricians venture beyond the traditional confines of their field to address wider questions of significance in the biological and the physical sciences.

Introduction

It is commonplace to read that a revolution has taken place in morphometrics. When making such statements most authors refer to the development of what has come to be called "geometric morphometrics" (GM), a term that usually goes undefined even in review articles about it (e.g., Adams et al. 2004). A systematic evaluation of the morphometric literature reveals the presence of at least two competing definitions of GM. The larger proportion of articles, either explicitly or implicitly, identify it with a specific set of data-analysis procedures (e.g., Procrustes superposition, relative warps analysis, principal warps analysis) that were formulated originally to operate on Cartesian coordinate data directly, as multivariable data sets, without transforming them first into scalar distances angles, areas, form factors, etc. as was commonly the case prior to the 1990s. In our view this is the "weak" definition of GM; inadequate insofar as the technique lists offered are always exemplary rather than definitive and deficient in that no attempt is made to explain what unites these (and other) data analysis approaches together either mathematically or conceptually. This definition leads to confusing ambiguities and inconsistencies over what is, and what is not, a GM method.

The alternative "strong" definition of GM understands this to include only those aspects of shape analysis that are undertaken in a Kendall (or a mathematically similar) shape space (Kendall, 1984; Bookstein, 1991) or some lower dimensional derivative thereof. This is a set of hypothetical mathematical spaces – actually the surfaces of mathematical manifolds – unified by the fact that each point in the space corresponds to a possible configuration of n landmark or semilandmark points, usually after the canonical "nuisance" factors of position, size and rotation of have been eliminated from consideration. There are an infinite number of such shape spaces. These geometric spaces make no assumptions regarding the size of the landmark/semilandmark sets that fall into their domain (n can be any integer), the rules used to specify locations of the landmarks/semilandmarks, the nature of the objects on which these landmarks/semilandmarks are located, or the range of procedures used to analyze such shape coordinate data.

While this "strong" definition of GM has the advantage of enforcing conceptual consistency, it is perhaps too restrictive if it is understood to apply only to the subset of GM methods that operate in the Kendall shape space *sensu stricto* (e.g., principal warps analysis, relative warps analysis). For example, the outline data analysis methods of elliptical Fourier analysis (Kuhl and Giardina, 1982; Ferson et al., 1985) and eigenshape analysis (Lohmann, 1983; MacLeod, 1999) are employed routinely by geometric morphometricians, but neither operates in the Kendall shape space (see Bookstein 1991). However, if the concept of GM is extended to apply to all methods used to analyse data in which point in the space corresponds to a possible configuration of n landmark points however determined, elliptical Fourier analysis, eigenshape analysis, and a host of other data formulations can be used by GM practitioners to test form and shape-based hypotheses rigorously.

The conceptual synthesis responsible for geometric morphometrics can accommodate this ecumenical approach to shape space definition easily and, indeed, can reap substantial benefits from its employment. This synthesis took place some time ago now – between 1984 and 1989 – and involved three individuals primarily: Fred Bookstein, Colin

*Corresponding author
Email address: n.macleod@nhm.ac.uk (Norman MacLeod)

doi:10.4404/hystrix-24.1-6299

Goodall, and David Kendall (see Bookstein 1993). All three had interests in geometry and statistics, but only Bookstein had direct and extensive experience with shape-based data-analysis problems in biology. Subsequent to the late 1980s, a number of important conceptual additions to the corpus of GM have been made (e.g., Rohlf and Slice 1990; Rohlf 1993; Mardia and Dryden 1998; Dryden and Mardia 1998), new data analysis methods developed (e.g., the Dryden-Mardia shape test), and methods developed in other contexts applied to morphometric datasets (e.g., machine learning methods, see MacLeod 2008). Nevertheless, from the late 1990s to the present day the main efforts of morphometricians have been spent in introducing individuals to the concepts of GM and applying these concepts to various problems in comparative morphology, or anatomical, analysis (see Adams et al. 2004).

But, does this exhaust the range of contexts appropriate for morphometric intervention? After all, form, and shape are not only important attributes of non-anatomical aspects of biology, they transcend the biological sciences entirely. For example, a host of variables important for managing forests in conservationally responsible and commercially optimal ways can be inferred from image texture analysis (e.g., Roiu and Seyer 1997; Franklin et al. 2001; Kayitakire et al. 2006). To date GM-based morphometric methods have not been used to characterize or compare forest canopy textures though in principle this geometric problem does not differ substantially from the characterization of bone or shell surface textures in biological species. Mechanistically, correct geometric design of the surfaces of the receptor cells responsible for the senses of taste and smell is as important (if not more so) than the ability of these receptors to respond to the presence of various chemical species (Young, 2001; Cramer, 2004). The tools of GM could, in principle, assist with the quantification and design of molecules that match molecular receptor surfaces. And while sound often has quite a complex structure, any digitized sound can be represented as a shape and so be subjected, in principle, to GM analysis.

To date GM has not been employed either extensively or routinely in any research field outside that of comparative anatomy/morphology. But this only means that the utility of GM approaches to the study of phenomena that represent the subject matter of these fields remains unexplored. In the past extra-anatomical applications have been limited because the nature of the shapes these phenomena present differ markedly from the anatomical/morphological structures that are well understood by morphometricians from the standpoint of shape characterization. In particular, many of these non-traditional shapes are best characterized as continuous two-dimensional (2D) functions or three-dimensional (3D) surfaces that do not appear to lend themselves to appropriate characterization using a small set of topologically homologous landmarks (see Aston et al. 2012). However, as the semilandmark – once all but excluded from consideration as an adequate morphometric descriptor (e.g., Bookstein 1990, 1991) – has now been rehabilitated as a useful implement in the morphometrics toolkit, and as semilandmark-based sampling protocols have been developed to facilitate the analysis of heretofore "featureless" surfaces (see Bookstein 1997; MacLeod 1999, 2008, 2012; Adams et al. 2004; Gunz et al. 2005; Polly 2008; Polly and MacLeod 2008; Klingenberg 2008; Oxnard and O'Higgins 2008; Sievwright and MacLeod 2012), the conceptual gaps between extra-anatomical objects and the sorts of forms GM is used to analyze, routinely seems less formidable now than they once did.

In order to explore the contribution GM can make to the analysis of non-traditional form, this contribution will focus on the analysis of sound; specifically the geometric analysis of bat echolocation calls. Some bats, along with odonocetes (toothed whales & porpoises), some birds (oilbirds, swiftlets), and some terrestrial mammals (shrews, tenrecs, even humans to a limited extent – see Supa et al. 1994; Thaler et al. 2011 for a discussion) use acoustic echolocation to navigate through, and identify objects in, their environment, though this capability is better developed in bats than any other organismal group.

Bats typically emit calls in the frequency range from 14.0 kHz to over 100 kHz; well outside the sensitivity range of human audio reception (9 Hz – 212 kHz). Bat calls are known to have a significant association with habitat, sex, age and presence of con–specifics in terms of different types of calls being used for different purposes. When hunting, bats use different calls to locate, identify, track, and intercept different types of prey. Different types of calls are also used in different locations though, if possible, a bat will prefer to hunt in an area for which its call type is suited (see Schnitzler and Kalko 2001). For these functional reasons bat echolocation calls are known to be species specific as well as diverse. This latter attribute raises the possibility of designing automated systems to identify bat species remotely by analysing the characteristics of their calls. If call-based remote identification can be realized on a sufficiently large scale, and with sufficient accuracy, it would be of great importance to bat conservation efforts as bats themselves are difficult to catch, especially without causing physical harm to the individual. Bat calls are able to be recorded using microphones, but such recordings are only accurate at present if expensive audio sensors are used. There is even an app – iBat – for iPhone and Android operating systems that will allow smart phones to be used as bat call recording devices and to upload the recorded calls to a central repository at the Institute of Zoology, London where software can assist in identifying the call (http://www.ibats.org.uk). If robust automated algorithms for identifying bat species from the physical attributes of their call patterns can be realized, the efforts of a growing cadre of committed and enthusiastic citizen scientists could be enhanced to provide reliable census data in the cause of promoting bat biology/ecological research and bat conservation efforts (Jones et al., 2013).

Figure 1 – Bat species used as the primary subjects of this investigation.

Accordingly, the goals of this investigation are fourfold.

i) Description of a generalized, geometry-based strategy for analysing bat search phase echolocation calls quantitatively that takes advantage of GM concepts and tools.
ii) Comparison and contrast of results obtained using this new morphometric approach to acoustic signal analysis with more traditional approaches.
iii) Discussion of avenues and opportunities for future research that might be pursued in the context of a shape analysis-based approach to the analysis of bioacoustic signals.
iv) Encouragement to (a) morphometricians to expand the scope of their work beyond the routine analysis of physical morphology and (b) non-morphometricians to realize and appreciate the potential of geometric approaches to contribute directly to testing what have traditionally been regarded as non-morphology-based hypotheses in their field(s) of study.

Materials and methods

For the primary example dataset a sample of calls from five bat species (Fig. 1) was obtained from the EchoBank bat call archive, a bat call reference library hosted by the Zoological Society of London. These species are all known to occur in the UK. Twenty calls were selected from each species at random in order to obtain a model of within-species call variation. In addition to this a second set of EchoBank calls was collected from bats belonging to the genus *Myotis*, which is regarded as a challenging group to identify to species level from echolocation call signatures alone (Kalko and Schnitzler, 1993; Parsons and Jones,

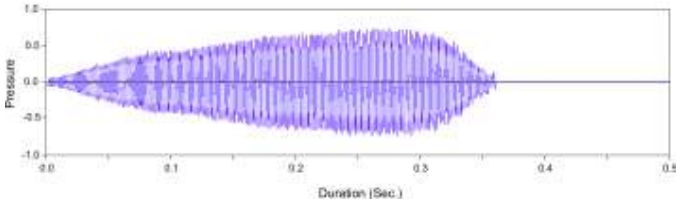


Figure 2 – Raw (non-normalized) oscillogram for a typical *Pipistrellus pipistrellus* search call. This call has a duration of 0.36 sec. and a maximum volume of 0.62 pressure units.

2000; Lundy et al., 2011). Nine *Myotis* species were used for this part of the analysis: *M. bechsteinii*, *M. blythii*, *M. brandtii*, *M. capaccinii*, *M. daubentonii*, *M. emarginatus*, *M. myotis*, *M. mystacinus*, and *M. nattereri*. Each of these species was represented by ten calls.

The calls themselves were full-spectrum, digital audio recordings of bat detection or search calls (as opposed to feeding buzzes or social calls) recorded at a sampling frequency of 44.1 kHz. These calls were obtained as primary digital audio recordings and saved to disk in the Microsoft .wav file format. Although each recording contains multiple calls from a single individual, only one was selected for analysis in this investigation. Minimal processing was applied to each call to standardize its structure. This processing procedure consisted of (1) normalizing the amplitude of each call to standardize its volume and (2) editing each call file to ensure it encompassed an equivalent total duration. The latter step was accomplished by determining the duration of the longest call in the sample and padding the ends of the shorter calls with silence taking care to make certain that each call began at the first position in the file listing. This step is necessary to ensure that each call can be compared across the sample in a reasonable manner and to verify that each call will be represented by the same number of geometric variables (see below).

Currently there exist three approaches to the quantitative analysis of bioacoustic signal data generally and bat echolocation calls in particular (Russ, 2012). The primary method of data collection quantifies the sound wave as a series of pressure readings taken at equal time intervals during the course of the call. A graph of these data that plots sound amplitude (= energy) against time is termed an oscillogram (Fig. 2). Oscillograms have been used to study many aspects of sound and are familiar to many musicians and fans of digital music, as well as acoustic researchers, from the graphical displays of digital sound editing software.

Spectrograms are constructed from oscillograms by applying a Fast Fourier Transform (FFT) to these amplitude vs. time data to represent or re-describe the complex waveform as a series of frequencies with associated frequency amplitudes. A sound file that has been re-expressed as a Fourier series of frequencies is said to have been transformed into the "frequency domain". A plot of a sound's structure on a graph of amplitude vs. frequency is often referred to as a "power spectrum".

Finally, a 3D representation of the sound can be obtained using a sampling window that breaks the sound into chunks composed of an equal number of data points; usually some power of 2 (2^2, 2^4, 2^8, 2^9, 2^{10}), a convention that derives from the mathematics of the FFT. These windows are moved down the signal by a specified amount, termed the "offset" (usually another power of 2, but one that is less than the window length), so that successive windows overlap by a constant amount. A FFT of the data included in each window is calculated and the associated amplitudes recorded as a matrix of numbers with the rows of the matrix representing the number of windows or chunks and the columns representing the number of frequency harmonics used to describe the sound included in each window. Of course, since the number of points included in each window is the same, the number of frequency harmonics extracted from each window will be the same. This procedure is referred to as a short-time or "windowed" Fourier analysis.

One complication inherent to the windowed approach is it will rarely be the case that the first and last point within each windowed dataset will have the same amplitude value. If this is not the case the Fourier decomposition procedure will artificially assign a high amplitude to a high-frequency harmonic. The most commonly employed procedure to correct this problem is to multiply every point in the raw (windowed) dataset by a continuous weighting function that, in effect, forces the ends to adopt the value 0.0 while not inducing distortion in the central section of the sound segment. A number of different weighting (or window) functions have been developed for this purpose including the Hanning, Hamming, Blackman, Bartlett, Turkey, and Laczos functions (see Harris 1978).

Typically, the set of variables resulting from a windowed Fourier analysis – call duration, set of harmonic frequencies, and amplitude values associated with each frequency for each windowed chunk of the sound – are assembled into a 3D image of sonic structure. These "spectrograms" have long been used to represent and compare both human and animal vocalization patterns. Spectrograms of this general form have been referred to variously as spectral waterfalls, sonograms, voiceprints or voicegrams. An example of a Hanning windowed call frequency spectrogram for a typical bat echolocation call is shown in Fig. 3. The 3D structure of the call is usually represented as a colour-

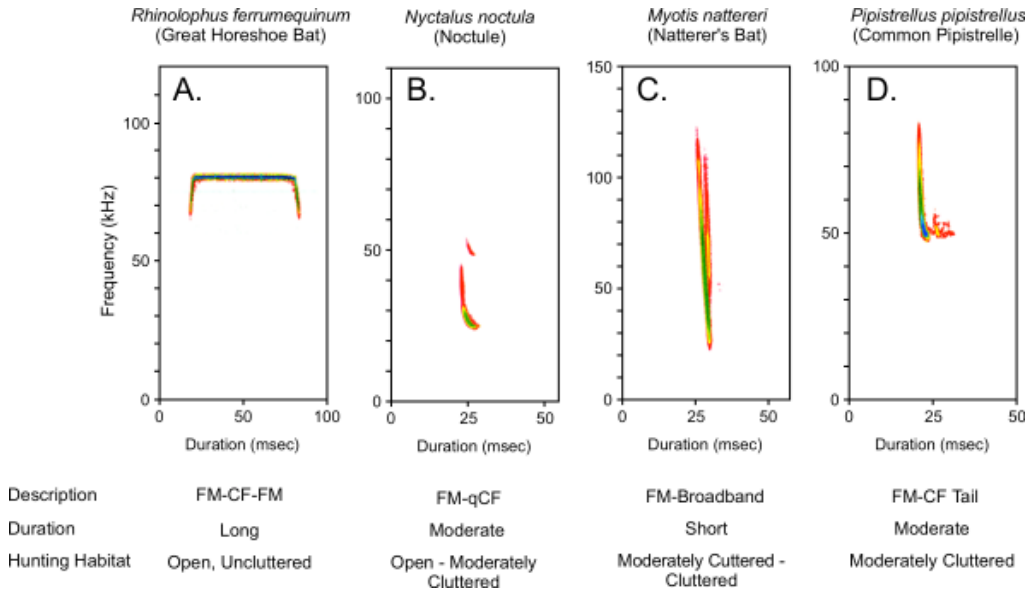


Figure 3 – Representative bat spectrogram call forms with their standard qualitative descriptors (description, duration) and typical hunting habitat. Abbreviations: FM – frequency modulated, CF – constant frequency, qCF – quasi-constant frequency. Redrawn from Russ (2012).

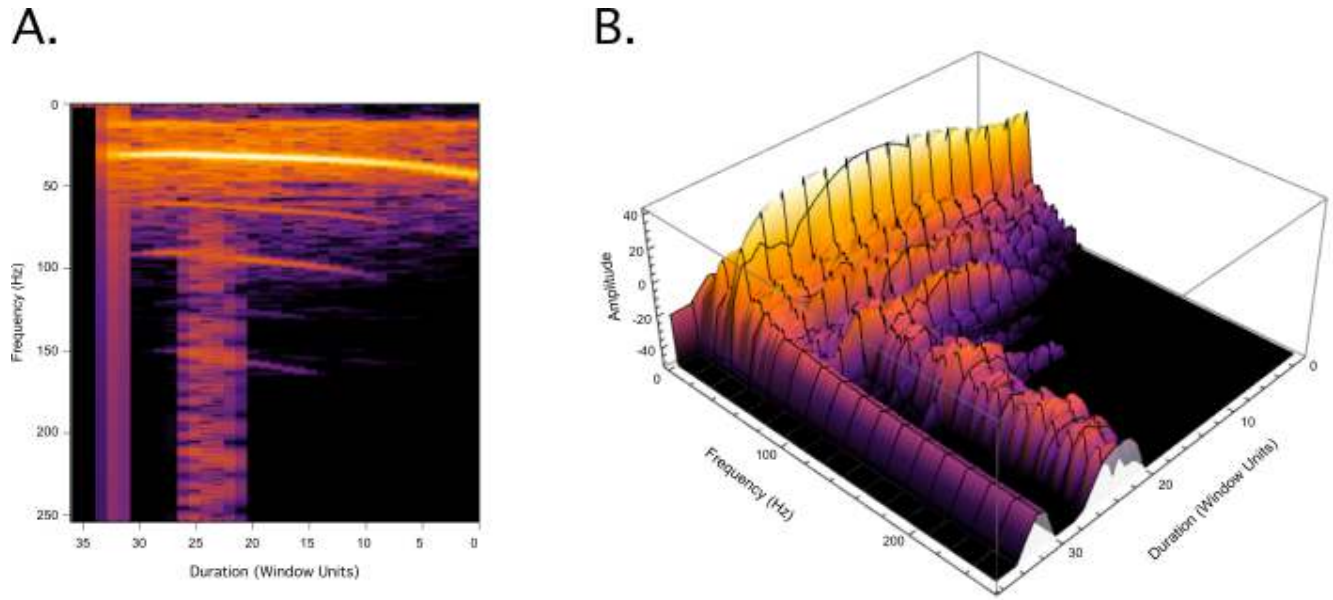


Figure 4 – Hanning windowed spectrograms for the *Pipistrellus pipistrellus* search call shown in Fig. 1 after normalization and padding to extend it to boundary of the oscillogram sampling window. The spectrogram is shown as a color coded 2D matrix (A) and an interpolated 3D surface (B). Note that the bright (yellow-white) band marking the fundamental harmonic sweep in (A) corresponds to the prominent ridge of amplitude values in the 3D surface representation (B) of the call's structure.

coded topological plot of call duration and harmonic frequency with amplitude values represented as a greyscale or color region mappings (e.g., see Schnitzler and Kalko 2001; Teeling 2009; Russ 2012; Fig. 3).

The traditional approach to using of spectrograms to achieve a quantitative description of acoustic signal structure in a manner that lends itself to qualitative analysis has been to employ a rather small number of observations and/or simple descriptive terms that capture a very limited subset of the spectrogram's overall geometry (see Russ 2012 and Fig. 3). Nonetheless, using these data, in addition to visual inspection of the spectrogram patterns themselves, it has been possible to determine that, as a group, bat species have evolved differently structured calls to take advantage of, or to compensate for, physical features of their preferred hunting environments and preferred prey items. For example, bats that hunt in open spaces utilize long duration, constant frequency (Fig. 3A) or quasi-constant frequency (Fig. 3B) calls that achieve maximum range with low atmospheric attenuation (Schnitzler and Kalko, 2001; Teeling, 2009; Russ, 2012). In contrast, bats that hunt in spatially complex, cluttered environments tend to utilize either short-duration, broadband, linear frequency modulated calls or short-duration, broadband, linear period modulated calls to sense the structure of their surroundings (Fig. 3C and 3D). Many bat species also rely on a variety of additional strategies for prey detection (e.g., auditory cues, Doppler shift, see Fenton et al. 1995; Jones 1999; Schnitzler and Kalko 2001). Finally, species that hunt in edge or mixed environments tend to utilize calls with both constant frequency and frequency modulation components with a relatively longer, narrow bandwidth, quasi-constant frequency character to achieve both localization sensitivity and high detection performance (Schnitzler and Kalko, 2001).

Over and above this generalized relation of call sonic structure to the physical aspects of different hunting environments, representation of bat search echolocation calls by means of a spectrogram can support, in a general sense, the testing of a variety of functional, ecological, and phylogenetic hypotheses. It does this in the same way that morphological features of anatomy support the same sorts of investigations, including phylogenetic analyses. For example, it is already known that, in many bats, call type is consistently associated with differences in homologous anatomical characters (e.g., length, width, and shape of the wing). This should not come as a surprise because, even through the concept of biological homology cannot be applied to acoustic structures *per se*, there is an obvious functional relation between the physical capabilities of particular (say) wing designs with regard to characteristics of the environment in which flight takes place and the type of audio signal best suited for echolocation duties in that same environment. Therefore, provided methods for representing the complex structure of these calls can be developed, it should be the case that morphometric approaches can be applied to the characterization, comparison, and analysis of these "non-morphological" structures in the same way they are applied to morphological structures.

For this investigation all calls were placed within an interval of 0.1023 seconds, which yielded call files 4512 sample values in length. This is equivalent to representing each call as a 4512-dimensional column vector. All calls were set to begin at time = 0.0 (sample = 1). Beyond normalization of the call amplitude and padding of the call duration, no filters were used to "clean" (e.g., sense or eliminate acoustic reflections from nearby surfaces) or enhance the sound. While on occasion reflections from certain surfaces can be helpful in bat identifications (e.g., bats that hunt over water, see Russ 2012), identification using acoustic signals generated by the bat itself is usually preferable.

While it has been the case to date that sonic spectrogram data have been characterized by and described on the basis of the 3D color-coded contour maps such as those shown in Fig. 3, these data can just as easily – and arguably more accurately from the standpoint of quantitative form characterization/comparison – be represented as true 3D elevation plots (see Fig. 4). Accordingly, sonic spectrograms were calculated from each call oscillogram in order to represent its form as a 3D surface. A spectrogram chunk size of 512 data points, with a chunk offset of 128 data points, was chosen to order to construct the window which also employed the Hanning function to minimize the amount of frequency leakage that occurs as a result of the chunked signal segments being non-continuous. These are standard spectrogram window settings.

This calculation resulted in a representation of each call as set of 36 chunks each of which was 512 samples long with each chunk being described by 512 Fourier harmonic amplitude values. Note this is a complete Fourier spectrum. To avoid redundancy due to aliasing all analyses were confined to the unique portion of the Fourier spectrogram (see Fig. 3). Once conversion to the windowed spectrogram had been accomplished for all calls amplitude values less than an arbitrarily chosen cut-off value of -50 were reassigned that value in order to provide a limit against which to distinguish between the call signal and background "noise". In terms of the shape of the call structure this background establishes the duration and frequency attributes of each call's form. This background normalization is a standard signal-processing technique (see Russ 2012) and, in a sense, is the sonic equivalent of placing a specimen on a black background prior to collecting a

digital image of its form. Exploratory experiments showed that above the -50 value results of the analysis changed markedly depending on which background normalization cut-off level was chosen, but below this level results were remarkably constant.

Redescription of the bat call oscillograms as Fourier-transformed sonic spectrograms increased the dimensionality of the call form from 4512 values to 8960 values, with this increase resulting from the fact that each window chunk is described by 256 unique harmonic amplitudes. This procedure resulted in a highly-detailed, but also highly redundant, representation of the each call's physical structure. Such redundancy can be minimized, and the major features of the call structure preserved for analysis, by mathematically laying a call-sampling grid over the windowed call and recording only the duration, frequency, and amplitude values that occupy the nodes of the sampling grid. This represents a surface-shape sampling system analogous to that used in eigensurface analysis (MacLeod, 2008; Sievwright and MacLeod, 2012) for which object size (here represented by call duration) has been included for an identical purpose: to represent the geometry of a highly variable, but consistent, featureless surface accurately and efficiently as a set of topologically homologous semilandmarks that bear a consistent geometric relation to each other and to the underlying "morphology" – in this case the windowed spectrogram. Gunz et al. (2005) utilized a similar semilandmark-based phenetic procedure to analyze the shape of the cranial vault in humans though the grid sampling procedure used in eigensurface analysis is both more structured and applicable to a greater diversity of forms. Specification of grid dimensions provides analysts control over the fidelity of the call's spatial – and so acoustic – representation. Coarse sampling grids will capture only the gross call form whereas finer grids will preserve greater levels of sonic detail. Using this strategy there is even scope for automating the spectrogram-sampling process so that analysts can be sure all grids sample the spectrogram to a consistent minimum quality criterion (see MacLeod 1999 and MacLeod 2008 for a discussion in the context of eigenshape analysis and eigensurface analysis respectively). Figure 5 shows results for a series of sub-samplings of the *Pipistrellus pipistrellus* spectrogram shown in Fig. 3 using square sampling grids of 10, 15, 25, and 35 cells per side. Note the rapid convergence on a reasonably detailed estimate of raw spectrograph shape both in terms of call feature shape and call feature location even at what would be considered coarse grid resolutions. In this study either a 30-cell (mixed bat genera dataset) or a 25-cell (*Myotis* species) grid was used to represent the generalized aspects of bat call structure. This level of detail was was judged (via visual inspection) to contain all the key features of the original spectrogram (compare Fig. 5 with Fig. 4). Selection of this resolution means that each spectrogram was described by 900 variables. Of course, this is still a very high-dimensional dataset. But in fact these data represent only 20 percent of the original spectrogram data; a considerable reduction in the dimensionality of the original dataset.

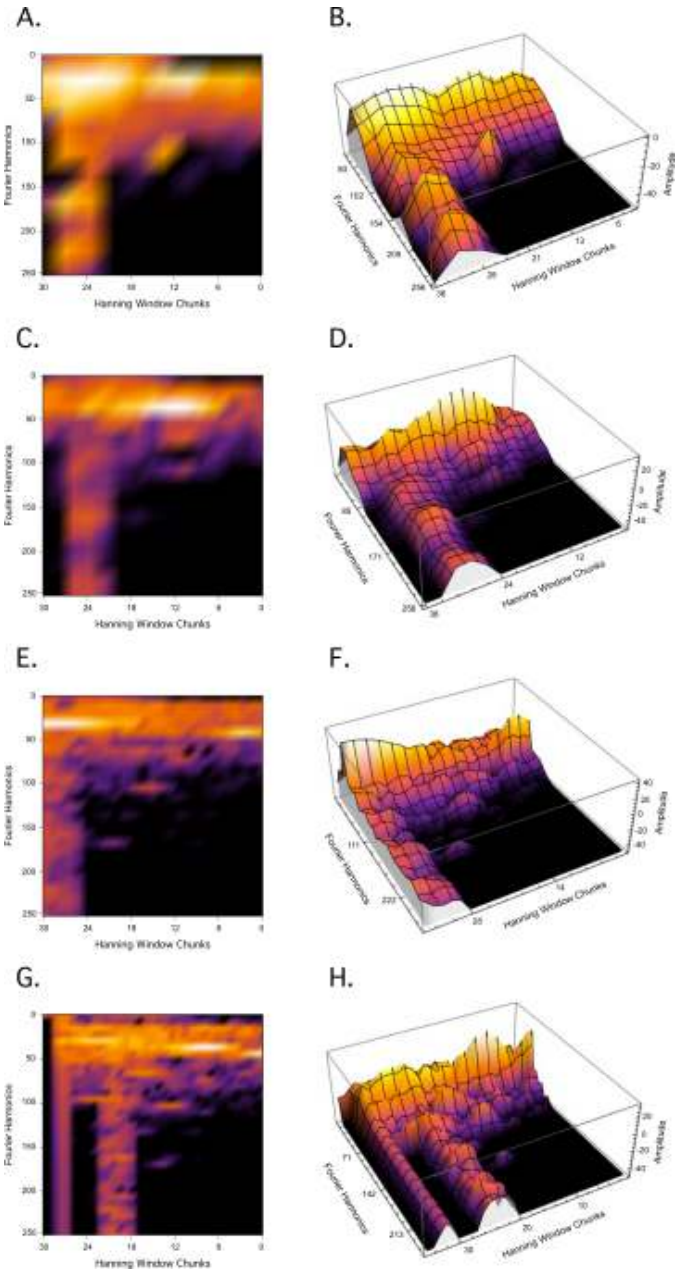


Figure 5 – The effect of choosing different subsampling grid resolutions on the character of the estimated spectrogram. Compare each down-sampled *Pipistrellus pipistrellus* spectrogram with the full-resolution *Pipistrellus pipistrellus* spectrogram shown in Fig. 3. For the purposes of the analyses of the primary bat call dataset a square grid resolution of 30 cells per side (900 cells in total) was chosen.

Sampled in this way the spectrogram data are, effectively, shape data that reside as point locations on a high-dimensional (n = 900) Kendall shape manifold. The 30-cell grids that were mathematically superimposed over the spectrogram surface are topologically homologous across the dataset in the sense that each grid point bears a consistent spatial relation to all other points on the grid. Indeed, the duration and frequency coordinates of all grid nodes are identical across all spectrograms in the subsampled dataset; only the amplitude values vary. As is standard practice in geometric morphometric investigations, these amplitude data were re-expressed as deviations from the mean spectrogram shape for the pooled sample (Fig. 6). Re-expression of the spectrogram data in this manner allows the acoustic structure of the search call sounds to be represented in a rigorous and fully quantifiable manner as shapes. Once these grid-based samplings of the original spectrogram data are in this form they can be operated on by all the procedures of geometric morphometrics.

For this investigation a preliminary covariance-based principal components analysis (PCA) was carried out on the pooled bat call spectrogram shape dataset in order to reduce the effective dimensionality of the dataset still further. This step is also important for assessing the major directions of shape (= call) variability for the sample and for serving as a basis space for call modeling procedures that will be used to interpret the placement of call groups in a linear projection space derived from the Kendall shape manifold. Results of the PCA analysis were used to decide how many latent shape variable axes to retain for subsequent discriminant analysis. The decision criterion for this phase of the investigation was to retain call configuration scores on a sufficient number of eigenvectors to ensure that at least 95 percent of the observed call-shape variability was retained for subsequent group-based analyses.

To serve the needs to rhetorical brevity, these data analysis steps – including (1) calculation of the Hanning windows, (2) subsampling of these windows based on grids of user-specified dimensions, and (3) summarization of major trends in acoustic structural variation via ordination of the positions of spectrogram surface shape coordinates in a reduced PCA subspace – will henceforth be referred to as "eigensound" analysis. This term is simply a convenience that streamlines procedural and interpretive descriptions and discussions in much the same way that

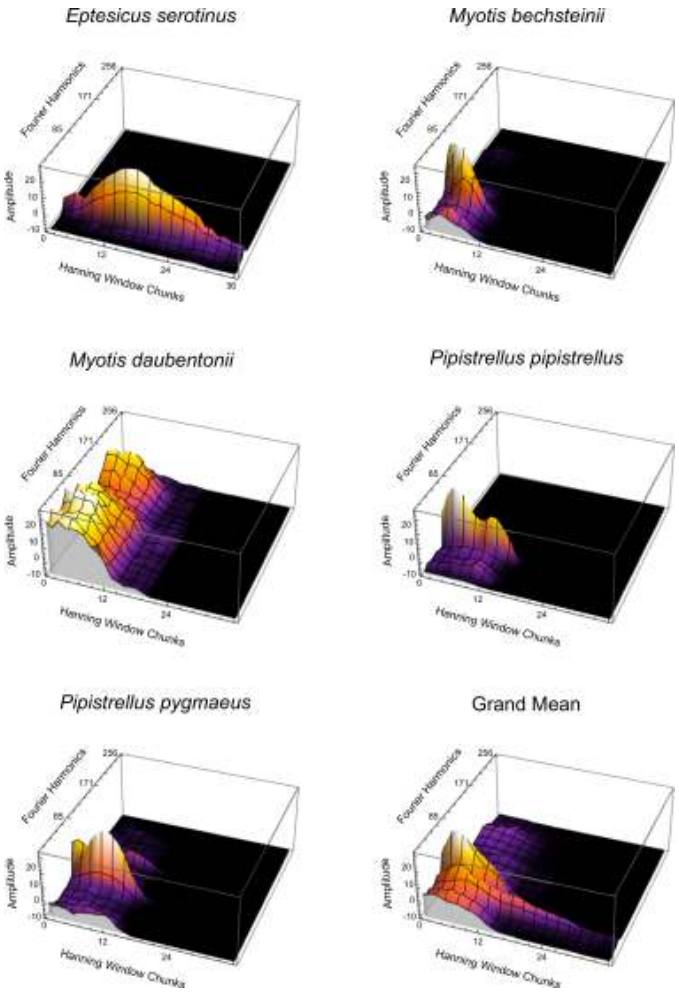


Figure 6 – Mean search call spectrograms for the five bat species included in the primary call dataset along with the grand mean for the pooled sample.

the terms "Procrustes analysis", "principle warps analysis" and "relative warps analysis" function in the standard GM literature. "Eigensound analysis" was chosen to highlight the conceptual links between this geometry-based approach to acoustic signal surface analysis and its morphological equivalents: eigenshape analysis (Lohmann, 1983; MacLeod, 1999) and eigensurface analysis (MacLeod, 2008; Polly and MacLeod, 2008; Sievwright and MacLeod, 2012).

Once the secondary data matrix of PC scores had been assembled, these data were combined with a grouping variable that associated each set of PC scores with the species name of the caller, and the dataset submitted to a canonical variates analysis (CVA, see Campbell and Atchley 1981; MacLeod 2007). Since five species were present in the primary sample test data, four discriminant functions were calculated and used in subsequent investigations. The *Myotis* dataset was treated in an identical manner which, owing to the larger number of species groups present in that dataset, resulted in the specification of eight discriminant functions. In order to obtain a robust estimate of group discrimination efficiency and address issues arising from the high dimensionality of the eigensound dataset both standard CVA and Monte Carlo CVA (see Manly 1997) procedures were employed.

Geometric interpretation of the CVA space was facilitated through the calculation of along-axis shape models using the back-projection procedure presented originally in MacLeod (2009a) and used in a number of recent articles (e.g., MacLeod 2008; Bolton et al. 2008; Sievwright and MacLeod 2012). Statistical tests of the separation between-group centroids in the CVA space relative to within-group dispersion of the data were carried out using the log-likelihood ratio (Φ) method for which the probability (ρ) of obtaining observed differences between sample mean vectors can be determined via reference to the χ^2 distribution (see Manly 1994). Both efficiency and stability of the discriminant functions calculated on the basis of call geometry were also tested using both the raw training set data and a jackknifed CVA.

Finally, in order to compare and contrast results obtained using a geometric approach to bat echolocation call analysis a set of standard call-description variables was obtained by Collen (2012) using SonoBat software from the EchoBank archive. Tab. 1 lists the variables included in that reference dataset. These variables were used by Walters et al. (2012) as the subset of possible descriptors that are most useful for quantifying between-taxonomic group distinctions between call types. To ensure strict comparability of results, these traditional spectrogram descriptor data were subjected to the same data-analysis procedures as the eigensound data.

Results

Inspection of mean spectrograms for the initial five-species dataset (Fig. 6) provides evidence for pronounced species-specific differences in call form. *Eptesicus serotinus* exhibited the most divergent call pattern, a long duration call with a narrow bandwidth focused into a prominent low-frequency fundamental harmonic. This pattern is typical of a high duty-cycling call, often employed by species hunting in open uncluttered environments. In contrast the two *Myotis* species are characterized by comparatively short, low duty-cycling calls with pronounced low-frequency energy peaks that may (*M. daubentonii*) or may not (*M. bechsteinii*) exhibit frequency modulated mean call shapes. *Myotis* species typically hunt in cluttered environments and/or over water (*M. daubentonii*). The two *Pipistrellus* species exhibit calls with their own structural differences. The *Pipistrellus pipistrellus*' mean call exhibits the short-to-intermediate duration and narrow bandwidth typical of species that hunt in marginal, semi-cluttered environments whereas the *Pipistrellus pygmaeus* mean call exhibits a form of similar duration, but longer bandwidth (especially at higher frequencies) and a marked difference in peak-amplitude profile. Both these species exhibit a prominent low-frequency fundamental harmonic, but the former is unique in its possession of a well-defined, subsidiary, higher-frequency secondary harmonic ridge.

The pooled sample mean shape (Fig. 6, lower right corner) represents a complex amalgam of these singular patterns. This mean is an abstract mathematical concept that corresponds to the call pattern of

Table 1 – Traditional bat echolocation call descriptors.

Variable Name	Description
LowFreq	Minimum frequency of the call (kHz)
FreqMaxPwr	Frequency of the call at the point of maximum amplitude (kHz)
HiFreq	Maximum frequency of the call (kHz)
Bndwdth	Bandwidth: total frequency spread of the call, calculated from the difference between maximum and minimum frequencies of the call (kHz)
CallDuration	Duration of the call (ms)
FreqCtr	Frequency at half the duration of the call (kHz)
Fc	Characteristic frequency: frequency of the instantaneous point in the final 40% of the call with lowest slope (kHz)
FreqKnee	Frequency at which the initial slope of the call most abruptly transitions to the slope of the body of the call (kHz)
FreqLedge	Frequency of the most extended flattest slope section of the call preceding the characteristic frequency (kHz)
StartSlope	Slope in the first 5% of the call duration (kHz/ms)
SteepestSlope	Steepest slope of the call: the maximum of linear regressions of any segment of 10% of the duration of the call (kHz/ms)
HiFtoKnSlope	Slope of the call calculated from the frequency and time of the point of highest frequency to the frequency and time of the knee (kHz/ms)

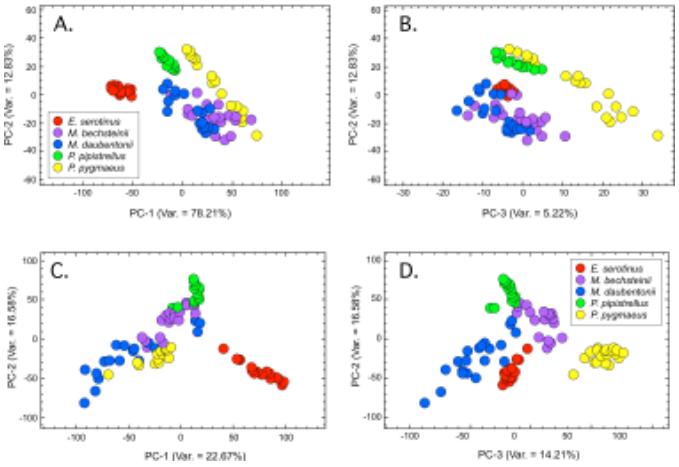


Figure 7 – A.-B. Distribution of bat species call geometries in the subspace formed by the first three principal components of correlation matrix calculated from 12 traditional spectrograph descriptors. C.-D. Distribution of bat species call geometries in the subspace formed by the first three principle components of a 30-cell sampling grid-based representation of spectrograph shape coordinates. Note differences in the axis scales which have been adjusted to save space. See text for discussion.

no known bat species, ancient or modern. Nevertheless, it plays an important role in the analysis as it specifies the semilandmark point configuration that locates the optimal set of linear planes tangent to the Kendall shape manifold on which to project the actual call configurations in order to visualize the structure of shape relations among them.

Principal Component Analysis – Traditional Spectrogram Variables

To establish a baseline against which the performance of an eigensound approach to acoustic spectrogram characterization and analysis can be evaluated, results obtained via application of this method were compared to results obtained from a mathematically comparable analysis of a series of 12 traditional spectrogram variables typically used to assess the structure of spectrogram similarities and differences in call echolocation studies across the a sample of the same five species that comprised the primary dataset (see Walters et al. 2012). These data represent observations of interest with respect to the characterization and comparison of spectrogram-based representations of acoustic data (e.g., maximum recorded frequency, bandwidth). They are not geometric in the sense of making any systematic attempt to represent the form or shape of the call spectrogram in any but its most generalized aspects. Nevertheless, it is these types of variables that are used at present to quantitatively characterize all bat echolocation calls (see Russ 2012).

As a first step in analyzing these traditional spectrogram data a PCA was performed to make a preliminary assessment of the datasets' major axes of variation and, if appropriate, reduce its dimensionality by focusing the spectrogram shape information distributed across all raw variables into a smaller set of composite, or latent, variables. Because the units associated with the traditional spectrogram variables differ from one another a correlation (rather than a covariance) matrix was used to assess the between-variable structure of these data.

Figures 7A-B shows the subspace formed by projection of the raw data values onto the first three eigenvectors of the correlation matrix calculated from this traditional spectrogram descriptor variable set. This subspace represents 96.3 percent of the observed variation described by the set of traditional variables. Although the point clouds for all species except *E. serotinus* are distributed over relatively large regions, three of the five species occupy unique domains within this subspace. The two *Myotis* species' domains overlap strongly, a result that is consistent with previous reports of difficulties separating *Myotis* species on the basis of their call patterns as assessed by traditional spectrogram descriptors (Vaughan et al., 1997; Parsons and Jones, 2000; Walters et al., 2012). Interestingly, whereas *E. serotinus* calls project to uniformly low positions along PC 1, the extremes of PC 2 and PC 3 are occupied by multiple groups. This result suggests that, with the exception of the *E. serotinus*, between-groups variation is not well aligned with the major axes of call form variation in the pooled dataset. Nevertheless, the scatter of points in the subspace these three eigenvectors indicates that these traditional variables do capture important aspects of within-species similarity and between-species differences.

Inspection of the eigenvector loadings for these axes indicates that calls plotting low on PC 1 are characterized by relatively high minimum frequencies, a high frequency that represents the transition from the initial phase and the body of the call, low maximum frequencies, and low frequency slope gradients. Calls that plot high on PC 1 exhibit the opposite characteristics and trends. Along PC 2 calls that project to low positions are characterized by low amplitudes, low bandwidths, high ledge frequencies and high frequency gradients whereas calls that project to high positions along this axis are characterized by high call amplitudes, high bandwidths, low ledge frequencies and low frequency gradients. Calls that project to positions low on PC 3 possess high ledge frequencies, high frequency gradients, low maximum frequencies, and low bandwidths whereas those that project to high positions are characterized by low ledge frequencies, low frequency gradients, high maximum frequencies, and high bandwidths. Note that, despite the apparent specificity of these spectrogram differences, it is quite difficult to form an intuitive impression of exactly what aspects of these parameters are responsible for the broad range of species-specific call variation domains, characteristic differences in species' call variation patterns, and the between-species distinctions in call form as expressed in this ordination of call structure data. In essence, these traditional variables are either too generalized or too idiosyncratic with respect to the structure of call variation to yield a detailed yet easy-to-visualize result, and the loading patterns too complex to allow for simple and clear interpretations of the PCA space geometries.

Principal Component Analysis – Spectrogram Shape Coordinates

As noted above, a covariance-based PCA of the eigensound dataset was also used to assess the major directions of call variation as the final step in an eigensound analysis. The primary purpose of this procedure was to assess dominant patterns of call structure variation and further reduce the dimensionality of the spectrogram structure dataset by focusing the information content of grid-based sampling procedure into a small number of composite, uncorrelated variables. Figures 7C-D show the ordination of call geometries within the subspace formed by the first three eigensound PC axes. Together these axes represent 75 percent of the observed spectrogram surface shape variation.

This ordination of call geometries based on the 3D spectrogram surface shape shows unexpected structure with a suggestion of a classic horseshoe pattern in the plane formed by the first two PC axes (Fig. 7C). The presence of this pattern in the PCA result indicates the existence of a non-linear gradient in these acoustic data. Since this is the first investigation (to our knowledge) that has operated on acoustic spectrogram data using sonic semilandmarks, it is unclear whether such gradients are common in these types of data or whether this is an idiosyncratic feature of this particular dataset. If non-linear gradient-like trends are common in acoustic spectrogram shape data their analysis my require methods specifically formulated to handle such data (e.g., non-linear PCA, kernel PCA, machine learning approaches; see Kramer 1991; Schölkopf et al. 1998; Friston et al. 2000; Scholz et al. 2007).

With respect to linear data analysis, specialists are of two minds regarding the "issues" posed by the horseshoe pattern. Ecologists tend to regard its presence as problematic and have developed a variety of *ad hoc* transformations to eliminate it from their datasets (e.g., de-trended correspondence analysis, see Pielou 1984; Hammer and Harper 2006). Unfortunately, employment of these algorithms runs the risk of obscuring other aspects of the data pertinent to its interpretation. In this context it should be remembered that the horseshoe pattern is always an accurate portrayal of the nonlinear pattern of the data, albeit in a linear space. Most mathematically inclined commentators advocate retention of the horseshoe pattern in the data – its removal is usually justified

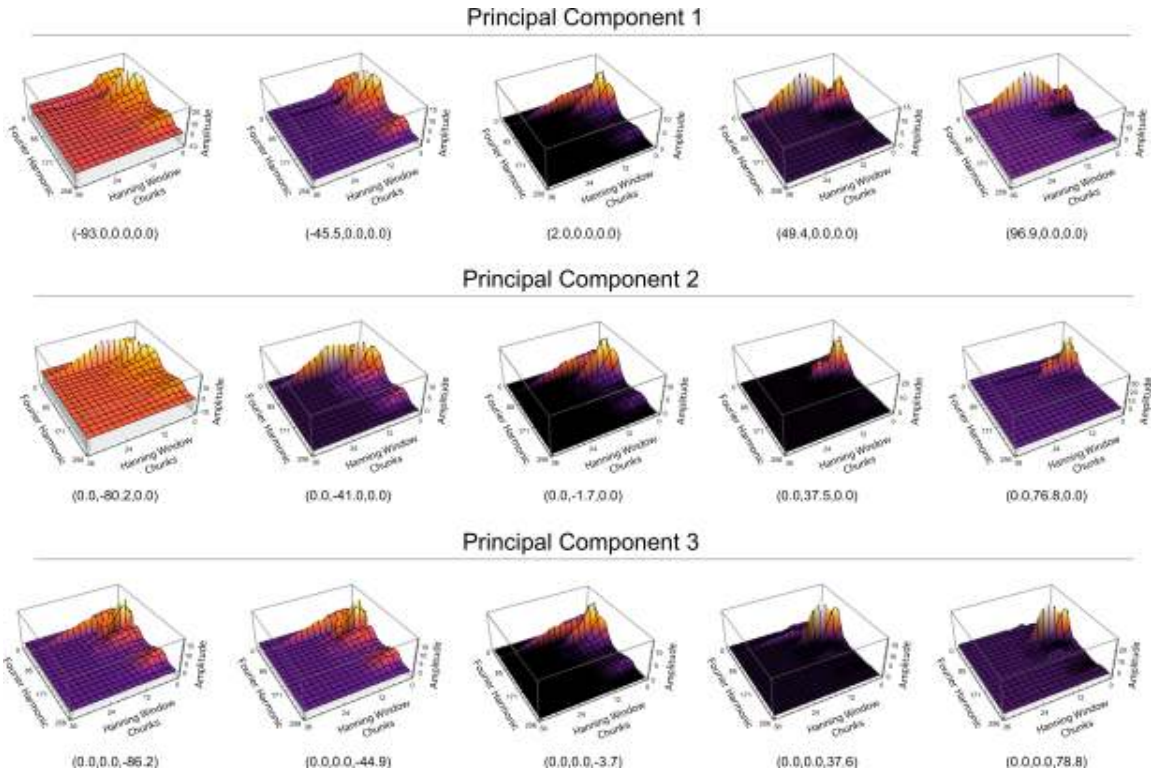


Figure 8 – Hypothetical models of search call shape at a series of equally spaced coordinate positions (listed below each plot) along the first three PC axes. These models provide a visual aid for developing interpretations of the PCA space shown in Fig. 7. See text for discussion.

primarily on aesthetic grounds – while being mindful of its proper interpretation or migrating to a non-linear data analysis procedure if the situation warrants it (see Greenacre 1984; Reyment 1991; Reyment and Jöreskog 1993; Podani and Miklós 2002; MacLeod 2006). Since the purpose of this analysis is to determine whether bat species can be identified by the geometry of their echolocation call patterns, and since the call configuration point distributions along the first two pooled-sample PC axes exhibit a high degree of species-specific clustering, recourse to de-trending algorithms or non-linear variants of PCA was deemed unnecessary.

Unlike the traditional spectrogram descriptor variable results (see Fig. 7A-B), in the eigensound PCA space the extremes of shape variation tend to be occupied by single species groups (Fig. 7C-D). For example, PC 1 represents a contrast between *M. daubentonii* (low scores) and *E. serotinus* (high scores); high scores along PC 2 tend to be dominated by *P. pipistrellus* and *M. bechsteinii*, with *P. pygmaeus* dominating the high end of PC 3. To be sure, some apparent overlap between species groups does characterize this subspace and group-level outliers are by no means uncommon. However, alignment between the major axes of pooled sample shape variation and the primary 3D surface structure-based distinctions in the species' call spectrograms is much greater for the eigensound variables than it was for the traditional set of spectrogram descriptor variables; especially for the acoustically challenging *Myotis* species.

Along PC 1 the two most divergent call geometries are those of *M. daubentonii* and *E. serotinus*. Since the former is the species with the longest mean call duration and the latter the species with the shortest, this would suggest to many that PC 1 represents a call-duration axis. Nonetheless, close inspection of the ordering of group centroids along PC 1 is inconsistent with this simplistic interpretation.

One of the advantages of choosing a geometric approach to acoustic spectrogram analysis is that it is a relatively easy matter to calculate the shapes of spectrograms for any point location on the Kendall shape manifold. This capability is in keeping with the fundamental theory that underlies all Kendall space shape analyses – that each point in the shape space corresponds to a unique configuration of landmark – or, in this case, semilandmark – points. Figure 8 shows a set of five hypothetical spectrogram point configurations that illustrate the manner in which call structure changes along each of the first three eigensound axes. This set of shape models can be used to understand the detailed geometric nature of the ordination space shown in Fig. 7C-D and refine the biological interpretation of the ordinated points therein.

Hypothetical spectrogram shape models for the first eigensound axis (PC 1 of the pooled search call dataset) indicate call semilandmark configurations that plot low on this axis represent, short duration, multimodal frequency modulated call shapes that exhibit two distinct, relatively high amplitude, but low frequency energy peaks and a low amplitude, but broader high-frequency amplitude peak. This model matches the mean representation of *M. daubentonii* well (compare with Fig. 6). With movement in a positive direction along the eigensound axis 1 the call shape changes in three ways: call duration is increased, the low-frequency amplitude ridge of high call energy splits into two (biharmonic) sharply defined, low frequency ridges with high frequency components of the call becoming progressively less well-defined and more attenuated overall.

Because of the influence of *E. serotinus* call shapes on the dataset as a whole, duration plays a strong role in the ordination of individuals along each of the first three eigensound axes, though its effect is most pronounced along axis 1. In this sense then, a naïve, qualitative interpretation of the PC space based on species located at the extremes of the various axes (see above) would be very misleading. These along-axis graphical models show quite clearly that, instead of call duration *per se*, axis 1 actually captures the contrast between low duty cycling, frequency modulated calls with a moderate and broadly defined low-frequency energy peak (low scores) and high duty cycling, narrowband, biharmonic calls with sharply defined low-frequency harmonics and attenuation of the call structure at higher frequencies (high scores).

Extending this geometric interpretation to the second and third eigensound dimensions, axis 2 captures the distinction between long-duration calls with low levels of frequency modulation and a pronounced low frequency biharmonic structure (low scores) passing along the axis to calls typified by short durations, high levels of frequency modulation and a sharp, well-defined, linear fundamental harmonic in which the maximum energy level is reached very early in

the call sequence (high scores). Similarly, axis 3 captures the distinction between long-duration calls with a well developed multi-harmonic structure that occupies (discretely) the entire frequency range (low scores) to calls characterized by short-durations and a sharply defined, low frequency, fundamental harmonic structure in which the call's maximum energy is reached early, but extended over the entire initial phase of the call.

When call spectrogram geometries are projected into this subspace species-specific clouds of points are fairly well segregated. Outlying call shapes exist for all these species; particularly so in the cases of *M. daubentonii* and *E. serotinus*. However, the eigensound ordination space is not designed to gather groups together and should not be used to evaluate hypotheses of either group membership or group distinctiveness unless such hypotheses are bound up with assessments of major directions of variation in the pooled dataset. As this is not the case in the present study, these results were used primarily to further reduce the dimensionality of the spectrogram shape characterization problem by focusing the information content of the 900 sampled amplitudes at the sampling grid nodes into a small number of orthogonal variables. Inspection of the table of associated eigenvalues indicated that the first 28 eigenvectors of the pooled-sample covariance matrix capture 95 percent of the spectrogram surface shape variation. Accordingly, the scores on these first 28 eigensound axes, along with a grouping variable specifying the positions of *a priori* groups within the dataset, were assembled and submitted to a CVA. Note this reduction from the 4512 original values in the .wav data files represents a dimensionality savings of 99.6 percent with less than 5 percent loss of geometric information content for this sample. Of course, part of this dimensionality reduction is bound up with the value of n (= number of specimens, in this case 100) which, for most datasets of this type, will always be much less than m (= number of variables, in this case 900). But even if the size of the dataset rather than the number of variables is used as the standard of comparison, a 72 percent reduction in dimensionally with less than a 5 percent loss of geometric information content remains impressive.

Canonical Variates Analysis – Traditional Spectrogram Variables

Although four discriminant axes with positive eigenvalues were specified as a result of the traditional variables CVA analysis, the character of group-optimized separations can be appreciated from an inspection of CV axes 1-3 (Fig. 9). For this variable set *E. serotinus*, *P. pipistrellus*, and *P. pygmaeus* all formed tight, well-separated domains within the CV space. However, the two *Myotis* species exhibited a much wider range of variation along with a substantial overlap in their call form distributions. Again, this result is consistent with the experience of other analysts who have employed a traditional spectrogram descriptor variables, even when these variable sets are analysed by non-linear procedures (e.g., Parsons and Jones 2000; Redgewell et al. 2009; Walters et al. 2012).

Based on results of the log-likelihood ratio test for group centroid separation relative to group dispersion (Φ = 638.00, df = 24), the null hypothesis that these call geometries can been explained by drawing

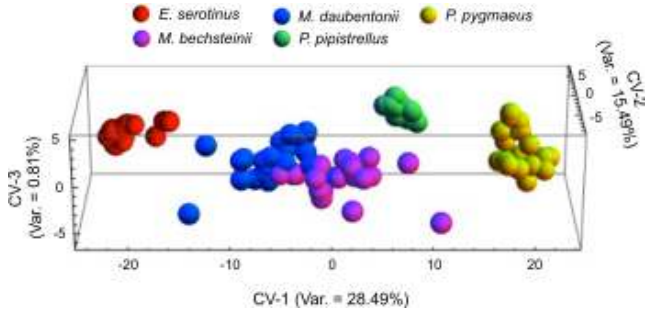


Figure 9 – Distribution of bat species call geometries in the subspace formed by the first three canonical variates of a six-dimensional PC-based representation of traditional spectrograph descriptors. Note differences in the axis scales. See text for discussion.

Table 2 – Confusion matrices for eigensound-based CVA results of call spectrogram form for mixed bat species.

Raw cross-validation identifications of the training set specimens

Species	*E. serotinus*	*M. bechsteinii*	*M. daubentonii*	*P. pipistrellus*	*P. pygmaeus*	Total	% Correct
E. serotinus	20					20	100.00
M. bechsteinii		20				20	100.00
M. daubentonii			20			20	100.00
P. pipistrellus				20		20	100.00
P. pygmaeus					20	20	100.00
Total	20	20	20	20	20	100	100.00
% Incorrect	0.00	0.00	0.00	0.00	0.00	0.00	

Jackknifed cross-validation identifications of the training set specimens

Species	*E. serotinus*	*M. bechsteinii*	*M. daubentonii*	*P. pipistrellus*	*P. pygmaeus*	Total	% Correct
E. serotinus	20					20	100.00
M. bechsteinii		15	5			15	75.00
M. daubentonii		4	16			16	80.00
P. pipistrellus				20		20	100.00
P. pygmaeus					20	20	100.00
Total	20	19	21	20	20	91	91.00
% Incorrect	0.00	4.00	5.00	0.00	0.00	9.00	

calls randomly from a single call distribution was rejected with a high degree of confidence (ρ = 0.00%). Even more importantly, results of a jackknife test of *post-hoc* identification efficiency (Tab. 2) indicated that these discriminant functions are relatively stable and might be expected to return up to 90 percent accurate results for sets of unknowns drawn from statistically similar populations. As with the analysis of post-hoc training set discrimination, results of the jackknife test identifies the two *Myotis* species as being similar to one another in terms of call structure when assessed by the traditional spectrogram descriptor dataset.

In terms of gaining insight into the aspects of the traditional call variable sets responsible for the observed between-groups distinctions, because these CVA results are based on PCA scores, interpretation of the CVA space involves using the CVA eigenvector loading coefficients to interpret degree of alignment between the CVA axes and particular the PCA variables, and then using the PCA variables loadings to interpret degree of alignment between the CVA axes and particular sets of original variables. For variables expressed in differing units and that have little conceptual relation to one another, this is a daunting interpretive task; rarely attempted by even the most experienced CV data analysts. This task could simplified to some extent by using the original data as input directly into the CVA routine. However, doing this would forego the opportunity to achieve preliminary dimensionality reduction – which could be important for dataset that employ large numbers of descriptive variables. Indeed, for this dataset a direct CVA was not possible as the magnitudes of the variable values included in the traditional descriptor dataset differed by ten orders of magnitude, thus preventing the original matrix from being inverted.

Canonical Variates Analysis – Spectrogram Shape Coordinates

While ordination of the search echolocation calls within the the subspace formed by the first three CV axes of the eigensound form data (Fig. 10A) may appear similar to that of the traditional variables superficially, there are important differences. In Fig. 9 most between-groups separation occurred in the plane defined by CV 1 and CV 2. In the case of the geometric CVA analysis each of the first three canonical variates contribute to group separation. This is a much more balanced discriminant result than was found using the traditional data. In principle this

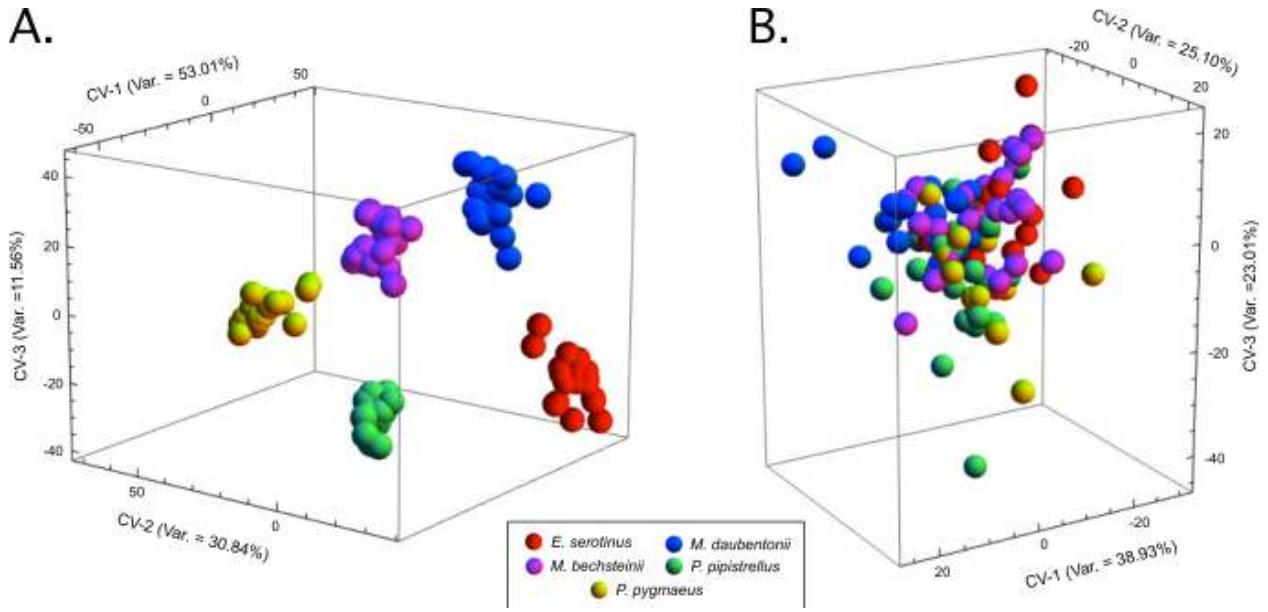


Figure 10 – Distribution of bat species call geometries in the subspace formed by the first three canonical variates of a 24 dimensional PC-based representation of spectrograph shape coordinates. A. Grouping variable set to reflect true species differences. B. Grouping variable set randomly in order to determine the degree with which the result presented in A could be consistent with the null hypothesis of no between-groups spectrogram shape differences. Note differences in the axis scales. See text for discussion.

should allow for a great degree of certainty in, and stability of, both group characterization and unknown call identification.

As before, *E. serotinus*, *P. pipistrellus*, and *P. pygmaeus* calls form tight, well-separated domains within the eigensound CV space (Fig. 10A). However using the eigensound variables the two *Myotis* species also separated cleanly into discrete and tightly clustered groups. This result is unprecedented in that *Myotis* species calls have proven to be resistant to separation based on traditional spectrogram descriptor variables even when analysed by themselves. To achieve such clear separations between the *Myotis* species' call structures when other species are present in the dataset suggests that a heretofore unexpected level of call distinctiveness exists between *Myotis* species. This interpretation is also consistent with the overall level of distinctiveness that appears to characterize the mean shape representations of *M. bechsteinii* and *M. daubentonii* in Fig. 6.

A log-likelihood ratio test for group centroid dispersion relative to within-groups variation rejects the the null hypothesis of no group-level structure to a high level of significance (Φ = 1051.00, df = 232, ρ = 0.00%), as does a 10000 pseudoreplicate Monte Carlo CVA designed to relax the distributional assumptions inherent in this parametric statistical test (Φ = 1045.42, df = 112, critical value = 134.55, ρ = 0.00%).

Despite the graphic result shown in Fig. 10A, some might question whether this grouping pattern can be used to refute the null hypothesis of no deterministic shape difference structure between groups owing to the relatively high dimensionality and relative modest number of individuals included in the dataset (see Bellman 1957; MacLeod 2007; Kovarovic et al. 2011). This issue can be resolved in two ways. First it is possible to repeat the analysis on the same set of data after the true group-level structure has been destroyed via randomized group membership assignments. The scatterplot obtained for this randomized-groups test is show in Fig. 10B. Note than randomising group membership resulted in complete destruction of group-specific spectrogram shape differences irrespective of the relatively high dimensionality of the dataset. These randomized data fail to pass a log-likelihood ratio test for the separation between group centroids (Φ = 106.10, df = 112, ρ = 63.96%) and exhibit a post-hoc efficiency of assigning members of the training set to their correct (randomized) groups as little better than would be expected due to chance alone (e.g., 50%, if there were just two groups).

The second test is to use a jackknife sampling strategy to examine the stability of the discriminant functions (Manly 1997, see Tab. 2). This test results in only nine individuals being misclassified post-hoc: five *M. bechnsteinii* calls mistaken for *M. daubentonii* calls and four *M. daubentonii* calls were mistaken for *M. bechnsteinii*. While the jackknifed cross-validation result is not perfect, it is well within the accuracy expectations of identifications based on other morphological data and far better than has been achieved by any similar analysis of bat echolocation call data previously. Further, these results suggest that, in addition to greater power of group characterization afforded by the eigensound approach to acoustic signal analysis, it may well be possible to conduct robust tests of systematic hypothesis with smaller sample sizes using geometric data than would be possible using a traditional set of spectrogram descriptors.

Perhaps best of all, use of geometric approaches to characterize spectrogram form allow for precise interpretations to be made of the geometric character of the discriminant space because it is an easy matter to project vectors the CVA space, the PCA space, and on into the space of the original variables (see MacLeod 2007, for a discussion of the projection equations). Figure 11 displays the results of using this method to illustrate the pattern of spectrogram shape variation along each of the first three geometric dataset CV axes.

Shape variation along CV axis 1 is strongly reminiscent of the patterns of shape variation captured by the first pooled groups eigensound axis (see Fig. 8). The polarity of these two axes is reversed, but eigenvector polarity is arbitrary, a by-product of the procedure used to estimate the eigenvectors. Other than this the pattern of shape change along these axes is almost identical, but with one important difference. Spectrogram surface shapes that project to positions high on CV axis 1 are characterized by long call durations and high duty cycling at low frequencies, but a sharply defined, broadband, multi-harmonic character. Given this characteristic spectrogram form it is readily understandable why the *E. serotinus* group projects to a position high on this axis. However, the opposite end of CV axis 1 in characterized by short duration, low duty cycling, frequency modulated calls accompanied by a single, low frequency, fundamental harmonic whose energy peak is realized quickly after call initiation. This is not the characteristic call form of *M. daubentonii*, which occupied the opposite end of the first eigensound axis, but rather of *M. bechsteinii*. Along this CV axis the broadband call typical of *M. daubentonii* occupies a position much closer to that of *E. serotinus* than to its congener *M. bechsteinii*.

Variation along CV axes 2 and 3 presents additional and even more subtle contrasts between spectrogram surface shapes, few of which could be understood in any detail without the graphical assistance provided by the eigensound spectrogram shape models. Along both

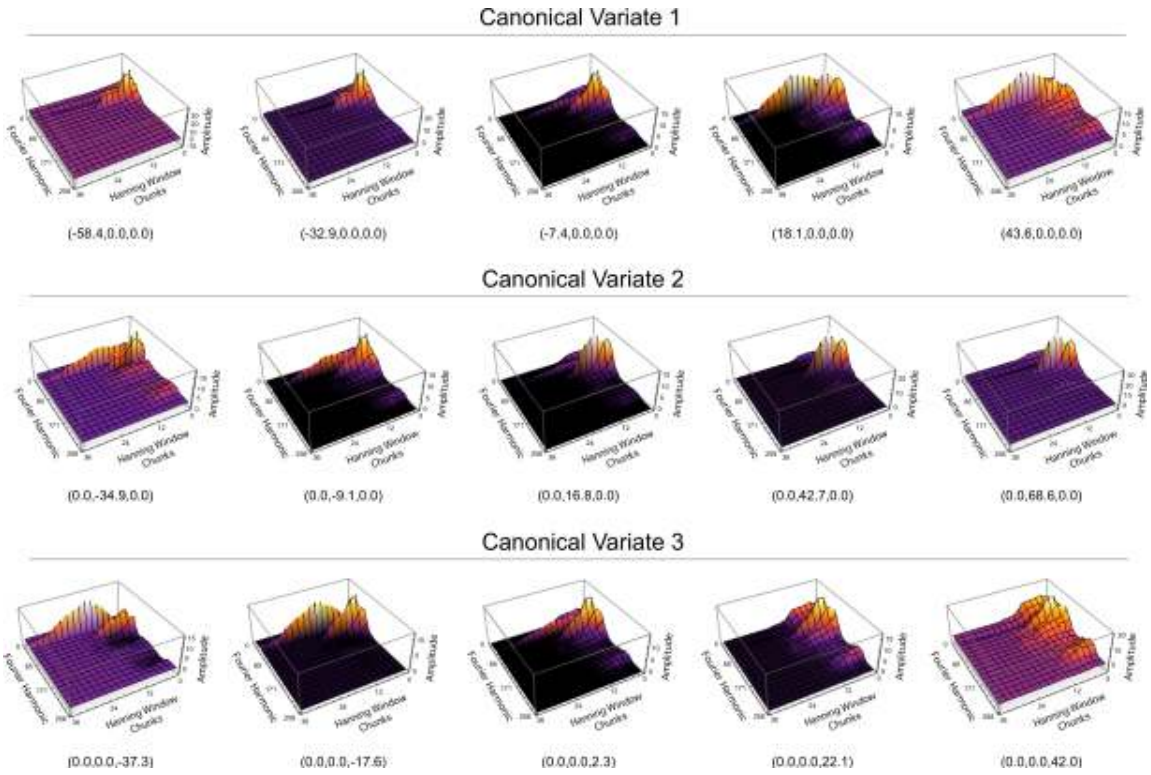


Figure 11 – Hypothetical models of search call shape at a series of equally spaced coordinate positions (listed below each plot) along the first three CV axes. These models provide a visual aid for developing interpretations of the CVA space shown in Fig. 11. See text for discussion.

of these axes there is a contrast between long-duration, high duty-cycling, low frequency, biharmonic calls (low scores) with (CV axis 3) or without (CV axis 2) subsidiary higher frequency energy components and short calls with primary harmonics in the low-frequency band exhibit and energy peak in the very earliest call stage (CV axis 2) or throughout the early portion of the call (CV axis 1). Aside from these patterns there is a clear distinction between CV axis 2 and CV axis 3 in terms of the degree of frequency modulation they represent. This ranges from weak frequency modulation with strong attenuation (CV axis 2) to strong frequency modulation, but weak attenuation (CV axis 3). In the cases of both these axes the "middle ground" of the ordination spaces is characterised by relatively long duration, multi-harmonic calls with low frequency modulation overall. Given the patterns illustrated by these CV space models, not only can the ordination of group placement relative to each other be understood quickly and easily, reliable predictions can be made about call forms in (presently) unoccupied regions of the discriminant space.

Myotis Analyses

As a second demonstration of the flexibility and power of adopting a geometric approach to acoustic signal analysis we consider the case of European *Myotis*, a group of bats regarded as being very difficult to identify on the basis of call structure alone (Walters et al., 2012). These calls were collected at a sampling rate of 312500 Hz and a total (padded) duration of 0.01 seconds, parameters that yielded a sample of 3094 digitized values per call. All calls were amplitude standardized and transformed into a spectrogram using a 512 Hanning window with an offset of 128. For eigensound analysis these calls were down-sampled using a 25 cell grid, which reduced the effective dimensionality from 12800 spectrogram values to 625, or 5 percent of the total spectrogram information. Mean 3D spectrograms for each species are shown in Fig 12.

These down-sampled spectrogram data were processed according to the eigensound protocol (as outlined above) which focuses spectrogram shape information into a small number of composite geometric shape variables. For the *Myotis* dataset 29 eigensound variables/axes were needed to represent 95 percent of the observed 3D spectrogram shape variation. Scores of each shape configuration across all 29 variables were submitted to a CVA to create a linear space that maximized between-groups separation relative to within-groups dispersion. A 3D plot of the subspace formed by the first three CV axes is shown in Fig. 13A as a way of illustrating the general character of between-groups separation that resulted from this analysis and that is resident with the spectrogram data for this sonically "difficult-to-characterize" multi-species group.

Within the low-dimensional subspace shown in Fig. 13A four species – *M. myotis*, *M. blythii*, *M. nattereri*, and *M. emarginatus* – all form tight, isolated clusters of call spectrogram geometries, well-separated

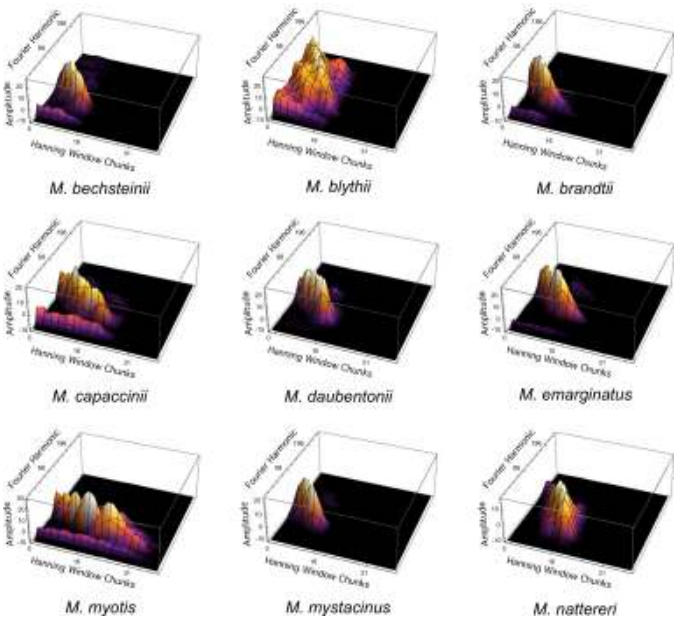


Figure 12 – Mean 3D spectrogram surface shapes that have been down-sampled to a 25 gird resolution for the nine *Myotis* species using in this investigation. These plots illustrate representative between-species call structure differences.

Table 3 – Confusion matrices for eigensound-based CVA results of call spectrogram form for *Myotis* species.

Raw cross-validation identifications of the training set specimens

Species	*M. bechsteinii*	*M. blythii*	*M. brandtii*	*M. capaccinii*	*M. daubentonii*	*M. emarginatus*	*M. myotis*	*M. mystacinus*	*M. nattereri*	**Total**	**% Correct**
M. bechsteinii	8				1			1		10	80.00
M. blythii		10								10	100.00
M. brandtii			9		1					10	90.00
M. capaccinii			1	9						10	90.00
M. daubentonii					10					10	100.00
M. emarginatus						10				10	100.00
M. myotis							10			10	100.00
M. mystacinus			1					9		10	90.00
M. nattereri									10	10	100.00
Total	8	10	11	9	12	10	10	10	10	90	94.44
% Incorrect	0.00	0.00	2.22	0.00	2.22	0.00	0.00	1.11	0.00	5.56	

Jackknifed cross-validation identifications of the training set specimens

Species	*M. bechsteinii*	*M. blythii*	*M. brandtii*	*M. capaccinii*	*M. daubentonii*	*M. emarginatus*	*M. myotis*	*M. mystacinus*	*M. nattereri*	**Total**	**% Correct**
M. bechsteinii	7	1			1			1		10	70.00
M. blythii	2	7						1		10	70.00
M. brandtii	1		8		1					1	70.00
M. daubentonii			1		9					10	90.00
M. emarginatus				1		9				10	90.00
M. myotis							10			10	100.00
M. mystacinus		1	1		1			7		10	70.00
M. nattereri									10	10	100.00
Total	26	29	35	26	39	30	30	30	30	90	82.22
% Incorrect	3.33	2.22	3.33	1.11	4.44	1.11	0.00	3.33	0.00	17.78	

from each other and from a central cluster containing the remaining species. But even within the central cluster there is evidence of strong within-groups clustering and between-groups separation based on call structure. Again, there is no expectation that separations between all groups will be represented accurately within in this low-dimensional subspace. However, to see so many overt clusters in such a low-dimensional subspace is very encouraging in terms of the ability of a geometric approach to the analysis of call form to succeed in representing species-specific call distinctions in this widely acknowledged "difficult" species group.

Myotis myotis calls exhibit the greatest level of geometric distinction within the CV 1 to CV 3 subspace. This result is consistent with a visual inspection of the average spectrogram forms in Fig. 12. *Myotis myotis* obviously has the call of greatest mean duration relative to other species in the dataset. The very tight clustering of *M. myotis* call shape ordination evident in Fig. 13A indicates that this is a consistent feature of the species. Similarly, the mean calls of *M. blythii*, *M. nattereri* and *M. emarginatus* are all more similar to each other than any are to the mean *M. myotis* call, but nevertheless retain plainly distinctive features of their own in terms of their spectrogram surface geometries. For example *M. emarginatus* is characterized by a much narrower and more well-defined fundamental harmonic than is evident in the *M. blythii* call whereas the positioning and duration-frequency orientations of the *M. emarginatus* and *M. nattereri* fundamental harmonics all differ distinctively. Again, the tight clustering of these species groups in the discriminant spectrogram shape space suggests that these – and other – geometric feature(s) are consistent and distinctive call characteristics of these species. A log-likelihood ratio test of the *Myotis* discriminant space for significant group-centroid dispersion (Φ = 1051.00, df = 232, ρ = 0.00%) resulted in rejection of the null hypothesis of no group-level structure as does a 10000 pseudoreplicate Monte Carlo CVA designed to relax the distributional assumptions inherent in parametric statistical tests for group centroid separation relative to group dispersion (Φ = 1051.02, df = 232, critical value = 236.60, ρ = 0.00%).

As was the case with the more diverse sample (see above), the possibility exists that the relative high dimensionality of the dataset and low number of specimens may conspire to enable any combination of groups to appear significantly separated (see above). To address this issue a *Myotis* group-randomized dataset was created and subjected to CVA. A plot of the first three CV axes for these data is shown in Fig. 13B. As with the mixed-genus randomized group result, amalgamation of spectrograms from different species into the same group effectively destroys the group-level structure within the discriminant space. This lack of discriminatory power is also reflected in the results of the log-likelihood test for these randomized data (Φ = 197.20, df = 232, ρ = 95.28%) and in the confusion matrix that summarizes post-hoc discrimination performance (44 incorrect spectrogram assignments or 48.9% of the total).

On the basis of these results there is little question that the species-level call shape separations seen in Fig. 13A are consistent with the recognition of heretofore unanticipated, but nonetheless profound and statistically significant, levels of distinction between *Myotis* species' echolocation call structures. The efficiency of the *Myotis* discriminant functions in assigning these 90 calls to the correct groups *post-hoc* is impressive (Tab. 3; 94.4% correct identifications). Cross-validation of these results using the more rigorous jackknife strategy provides a

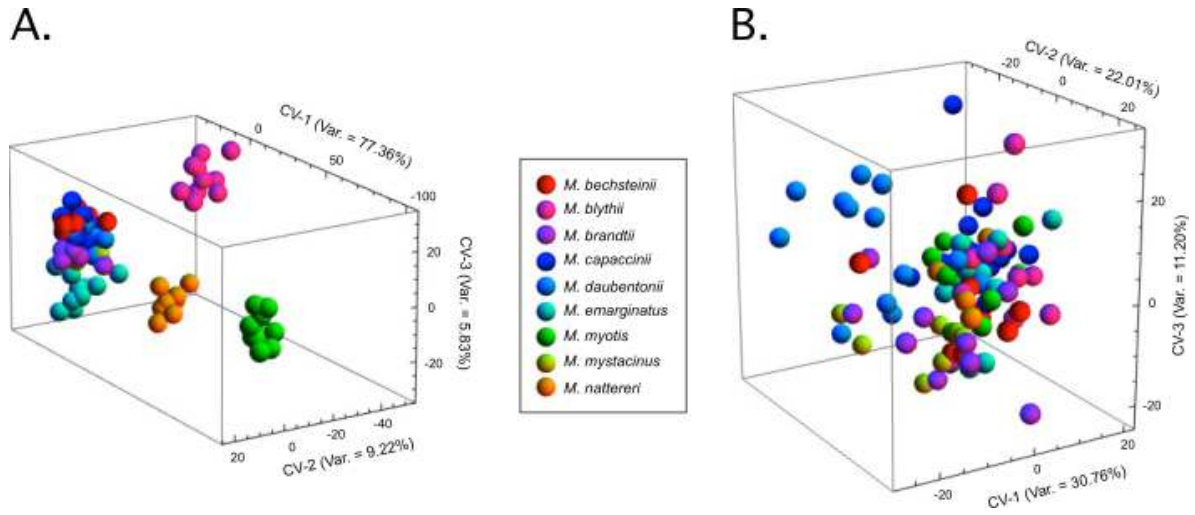


Figure 13 – Distribution of *Myotis* species call geometries in the subspace formed by the first three canonical variates of a PCA-based summarization of their spectrograph shape coordinate data. A. Grouping variable set to reflect true species differences. B. Grouping variable set randomly in order to determine the degree with which the result presented in A could be consistent with the null hypothesis of no between-groups spectrogram shape differences. Note differences in the axis scales. See text for discussion.

more robust assessment of how the discriminant functions determined from this dataset might perform in a generalized context (Tab. 3, 82.2% correct identifications). But even given the circa ten percent decrease in identification performance under the more rigorous test, the result remains favourable especially given the small sample sizes involved in this study and when compared to previous attempts to characterize these species through echolocation call data alone (see below).

Discussion

Taken at face value these results underscore the potential of geometric morphometric approaches to play a significant role in a host of non-morphological (in a strict sense of that term) contexts. They become even more so when it is recalled that these results were generated by (1) quite small datasets and (2) quite a low level of 3D spatial resolution. Yet both mixed species and *Myotis* results have revealed a wealth of useful structure; structure that, up to now, had not been recognized, much less exploited successfully, to understand the systematics and biology of bat echolocation calls. Access to additional phenomenological levels in acoustic spectrogram data can be gained easily by increasing or decreasing the spatial resolution of the eigensound sampling grid. There is also scope for windowing of the sample grid itself in order to conduct analyses on mathematically isolated spatial components of spectrogram variation. Moreover, this generalized conceptualization, sampling, and geometric data-analysis strategy can be applied to any representation of any sort of acoustic data no matter how abstract or divorced from the anatomical roots of contemporary morphometric analysis.

In a previous investigation Lundy et al. (2011) claimed that their application of elliptical Fourier analysis (EFA) to aspects of the spectrogram of *M. daubentonii*, *M. mystacinus*, and *M. nattereri* calls represented the "first attempt to classify echolocation calls using morphometrics" (p. 103). As the generic term "morphometrics" applies to any and all attempts to quantify form and form change (see Blackith and Reyment 1971; Pimentel 1979), this claim is obviously false as it fails to acknowledge all prior attempts to quantify any aspect of bat sound – including the prior investigations cited by these authors in their own text. If this statement is taken to refer to GM, it is also questionable as, under some previous definitions of that term, EFA would be excluded because it does not operate in the Kendall shape space (see Bookstein 1990, 1991, p. 48).[1] However, the more important issue is to consider carefully which approach to the analysis of any given set of data is more appropriate for answering (in this case "the biological") questions raised successfully, adequately, and reliably.

The Lundy et al. (2011) investigation used EFA to characterize the form of the fundamental harmonic in the *Myotis* call spectrogram's frequency-duration plane. In *Myotis* species this harmonic forms only part of some species' characteristic call pattern (see Fig. 12). Indeed, it could be argued this harmonic sweep is so ill-defined as to be absent entirely in some *Myotis* species (e.g., *M. blythii*). Absent also is any criterion discussed by Lundy et al. (2011) by which the sweep's lower boundary – the boundary that controls the shape of the fundamental harmonic's outline – was determined objectively. As such, the outline of the fundamental harmonic in the spectrogram's frequency-duration plane seems a decidedly limited and problematic feature of the *Myotis* call to focus on for the purpose of sound characterization.

For those species in which the fundamental harmonic is present, its form is that of a long, narrow arc in the 3D spectrogram space. As is illustrated in Kuhl and Giardina (1982), EFA does not perform well when trying to characterize such structures. Since only 20 Fourier harmonics were used by Lundy et al. (2011) to describe the shape this structure it is quite possible that distortions in the representation of these outlines were introduced by the EFA algorithm at both of the sharp ends of the fundamental harmonic's termini. Based on their Fig. 2-III the representations of these sweep patterns also seem to be of quite low resolution, with highly aliased boundaries and (so) distorted forms. Lundy et al. (2011) provided no information regarding the resolution of the semilandmark data used to quantify these outlines and/or what, if any, steps were taken to ensure both sides of the harmonic's trace were represented by an equivalent number of landmarks that occupy equivalent positions in the semilandmark sampling sequence across the sample (see MacLeod 1999 for a discussion of the problems that result for poor registration of outlines across a sample along with a simple strategy for correcting the problem). In addition, the Lundy et al. approach approach appears not to focus on locating the position of the fundamental harmonic within the context of the entire call structure.

Given these issues it is remarkable that the Lundy et al. (2011) EFA approach delivered even marginally adequate results. An overall identification accuracy of 79.6 percent was achieved using stepwise (but apparently a non-cross-validated) discriminant analysis of the EFA-based characterization of 2D sweep geometry alone. This accuracy estimate was later boosted to 96.3 percent via inclusion of the traditional spectrogram descriptive parameter "maximum frequency" in the data prior to stepwise discriminant analysis. Nevertheless, this approach to bat echolocation call analysis cannot be generalized even to all *Myotis* species – much less all bat species – insofar as multi-harmonic species cannot be represent by a single harmonic sweep outline. The eigensound approach to acoustic signal shape analysis circumvents all these issues, simply, elegantly, and effectively as well as delivering superior results.

In eigensound analysis, standardization of the spatial representation of the entire call sound structure is achieved by adopting the convention that each sound file starts at call initiation, each ends at the end of the normal signal duration (so no part of the sound is stretched or compressed artificially), and by ensuring that sound files are of equivalent duration by adding silence to the ends of the shorter-duration calls. The former is comparable in the anatomical morphometric realm to beginning outline digitization at a single landmark point that corresponds to all other call initiation landmarks across the sample. The latter is effectively analogous to ensuring all calls are set to the same "size' in the sense of being represented by the same number of Fourier harmonic amplitude variables. In this sense acoustic homology is maintained across the entire dataset in terms of the physical energy-duration "form" of the call. Re-expression of the sound's information content using a set of Fourier coefficients spatially organized into a Hanning window corresponds to the re-expression of a boundary outline curve using any radial or elliptical Fourier spectra, or indeed the the re-expression of a landmark-based shape configuration by means of principle/partial warps. Use of a grid-based representation of the sound structure also ensures equivalent spatial dimensionality across all sounds included in the sample and represents the conceptually equivalent of a Procrustes alignment (without the need to actually perform the Procrustes calculations). Finally, subsampling this spectrograph grid to reduce the effective dimensionality of the data, along with use of the baseline adjustment convention, represents the sonic equivalent of standard digital image processing procedures designed to boost the effective signal-to-noise ratios of the spectrograms.

To appreciate the similarities and the differences between this method of spectrogram processing the data processing steps considered routine in GM it is important to note that each of the steps outline above is performed on individual sound files without reference to any information contributed by the sample itself. In other words, no part of these operations is optimized via reference to any other sound or sound spectrogram in the sample. While it is tempting to use the term "homology" (in its mathematical sense of spatial correspondence) to describe the equivalence between grid cells in the down-sampled Hanning grid, in order to avoid confusion the reader may want to regard these positional equivalents simply as "windowing correspondences". Similar sampling strategies are also key parts of eigensurface

[1] Adams et al. (2004) claimed that "Outline methods were the first geometric morphometric methods to be used" (p. 6), but defined GM very simplistically, as little more than a set of multivariate procedures that operate on "Data that captured the geometry of the morphological structure" (also p. 6). By this rather general (and generous) definition, virtually any data collected from an organismal body – including linear distances between landmarks (see Strauss and Bookstein 1982), and any numerical data-analysis procedure could be regarded as being consistent with the principles of GM (see also arguments presented in the Introduction).

analysis (MacLeod, 2008; Sievwright and MacLeod, 2012) and certain approaches to machine learning (MacLeod et al., 2007a,b; MacLeod, 2012). Even after using such severe down-sampling schemes as those employed in this investigation, all of the traditional spectrogram descriptor variable concepts are represented in one form or another in the gridded eigensound dataset and so are part of the the overall eigensound analysis. In addition to this, a wealth of other geometric information not captured by either traditional sets of spectrogram scalar descriptors or the Lundy et al. (2011) EFA approach are also present in the eigensound data.

The results achieved through employment of the eigensound sampling and data characterization strategy speak for themselves regarding this technique's effectiveness. Any acoustic signal pattern, no matter how short, how long, or how complex – indeed any type of data that can be expressed as a matrix of objects and variables – can be treated in exactly the same manner and will likely deliver results of comparable sensitivity. In particular, the shape modeling capabilities of eigensound analysis represent a significant advance in the ability of mathematically complex ordination spaces to be assessed, interpreted, and used to facilitate communication with others about the nature of these spaces in a simple, informative, and intuitive visual manner (see also MacLeod 2002, 2008).

Once echolocation calls have been quantified using the eigensound approach it becomes possible to address a wide variety of questions pertinent to improving our understanding of bat systematics, ecology, functional morphology, and phylogenetics. For example, it has long been accepted that the form of bat echolocation calls has been determined by the functional needs of hunting particular prey in particular environments (e.g., Schnitzler and Kalko 2001; Jones and Teeling 2006). However, this assumption has been challenged recently by Collen (2012). Part of the problem in studying the phylogenetics of bat echolocation is the comparative lack of sufficiently detailed descriptive lexicon that can be used to identify call characters and character states (see Fig. 4 for an illustration). Treatment of call spectrograms as complex 3D morphological structures will facilitate their description using morphological terms for which there is a much richer vocabulary than is available in the traditional qualitative or semi-quantitative sound-description domains.

In performing such an analysis there is an implicit assumption that call structures of similar form are produced by biologically homologous physical structures, behaviours, physiological responses to external stimuli, functional constraints, etc. However, as has been demonstrated repeatedly in the contexts of comparative method and phylogenetic studies, the extent to which this assumption is justified cannot be decided *a priori* based on the nature of the putative character or its mode of description. Rather this is an empirical question that can be answered only by carrying out the analyses required to demonstrate (or not) the existence of phylogenetically structured patterns of variation in characters or variables derived, in this case, from the audio signal. But irrespective of the results that may be obtained from a test of this hypotheses, the point we are making here is that the eigensound approach to the representation, summarization, analysis, and comparison of the echolocation call's physical signal structure facilitates these types of analyses in a manner that makes it possible to describe sounds either as sets of continuous variables (e.g., Hanning window correspondences) or, depending on the structure of variation within a set of sampled spectrograms, a series of quantified geometric characters and character states rather than a series of imprecise categorical assessments (e.g., high duty cycling-low duty cycling, high frequency-low frequency, attenuated-non attenuated) based on the qualitative assessments of spectrogram patterns or via reference to crude descriptive indices (e.g., highest frequency, lowest frequency, bandwidth).

Along these same lines, quantification of bat call form will provide data analysts with the ability to exert direct control over the degree to which comparisons based on acoustic data are influenced by the phylogenetic component of cross-species call comparisons. The mappings of call form categories offered by Jones and Teeling (2006) and by Collen (2012) show that, to varying degrees, all bat calls should be expected to have component of phylogenetic covariance embedded within their structure. Felsenstein (1985, 1988, 2002); Harvey and Pagel (1991); Harvey et al. (1996); Martins and Hansen (1997); MacLeod (2001); Rohlf (2001, 2002, 2006) and a host of others have all made the case that the analysis of morphological, behavioral, ecological, and geographical data must take phylogenetic covariance into consideration when designing quantitative tests of biological hypotheses or run the risk of introducing substantial error in the results produced. Lack of a reliable and sufficiently detail approach to the quantification of acoustic data has, to date, kept bioacoustic signal analysis from taking advantage of improved statistical testing and data-analysis strategies that are robust to the effects of phylogenetic structure (e.g., Revell 2009). The eigensound approach to acoustic signal analysis provides a means by which the advantages of comparative method procedures can be introduced into the field of bioacoustic analysis.

Last but by no means least, the quantitative representation of acoustic structure is a prerequisite for the construction of reliable automated species identification systems for use in bat biodiversity and bat conservation studies. The bat systematics community is well ahead of other areas of biology in recognizing the important role such systems will play in twenty-first century biological research. The assembly of such systems presumes the existence a generalized approach to the identification and assessment of within-groups similarities, and between-groups differences among species. Any approach that employs one set of variables to identify one group of species, but another set to identify others, cannot be turned into a fully automated system easily (see Walters et al. 2012). However, the to the extent that sonogram data can be used to represent the information content of bat echolocation calls, the eigensound approach is fully generalizable and can be used as a complete and sufficient system for representing, partitioning, and identifying bat species on the basis of the sonic structure of their calls. Indeed, the computational overhead required by an eigensound-based system is relatively modest; well within the range of most high-end smartphone processors (e.g., Apple iPhone). Moreover, any phone with wifi capability can be used to upload a call record and control the server software that would be required to perform the necessary calculations with results being displayed as a web page.

While the eigensound results reported here by no means solve the "bat identification from echolocation call problem" in general or the "*Myoti*s problem" in particular, they are the best results that have been obtained to date and the first to reveal that such clear distinctions between different *Myotis* species calls exist. The fact that excellent between-species separations based on call structure were obtained for both *Myotis*-only datasets and mixed-species datasets is unprecedented. Indeed, the clear improvement in *Myotis* species identifications produced using the eigensound approach suggests that the *Myotis* problem may have more to do with the descriptive variables that have been used traditionally to characterize bat echolocation calls than with the fundamental structure of the calls themselves. Additional research in the area of call characterization is now needed, both in terms of testing the eigensound approach with larger species training sets, and testing alternative algorithms that are consistent with the geometric philosophy that stands behind eigensound's basic approach to acoustic signal characterization and analysis; especially those designed to cope with non-linear patterns of variation that may be present with bat echolocation call data (see MacLeod et al. 2007a,b; MacLeod 2007, 2012 and above).

Summary and conclusions

In this report we have taken up the issue of acoustic signal analysis and asked what (if anything) geometric morphometrics can contribute to the study of sounds. In particular we have employed a "eigensound analysis" – a Procrustes PCA applied to a spectrogram-based 3D characterization of sound structure – to analyze similarities and differences within two datasets of bat search calls, a mixed set of five species and four genera including both easy and challenging call types and a more uniform set of nine species from a single genus, *Myotis*, which is widely acknowledged to be difficult to identify to the species level based on traditional spectrogram descriptor variables. In both cases

the eigensound approach achieved excellent results, detecting complete between-groups separation for the training set sample in the first dataset and in the second to an overall accuracy of 94 percent. These results demonstrate the reality of species-specific distinctions between gross call structures. However, for the purpose of evaluating what level of performance might be realized as a result of the use of these discriminant functions to identify unknown bat calls the cross-validated and jackknifed results are more pertinent. As is typical in such analyses, these more robust assessments of discriminant function performances achieve c. 10 percent lower accuracy estimates when compared to results obtained from analyses of the original training set. We hasten to point out that we are not advocating the discriminant functions obtained during the course in our investigation be used for identifying bat species from their echolocation calls. Our results are indicative only of the type of results that might be realized using a larger bat call training set. Irrespective of this caveat though, so far as we are aware these are the best results that have been achieved to date for bat echolocation calls using any approach to spectrogram form characterization and/or analysis. Jackknifed cross-validation analyses of these data also indicate that the discriminant function systems specified as a result of this small example analysis are surprisingly robust and would be useful in automating bat species identifications based on echolocation call data alone.

Furthermore, the results reported above also address several other important issues. First, that species specific information is encoded in bat echolocation call structures in sufficient quantities to facilitate reliable species identification; even in the case of those species groups (e.g., *Myotis*) that have resisted such characterization in the past. Second, the strategy of successive focusing of the information content of a set of echolocation spectrograms through use of (a) 3D surface sampling grids and (b) PCA (or SVD) transformation can result in massive reductions in the dimensionality of the spectrogram analysis problem with little or no apparent loss in the ability of such reduced datasets to identify even difficult-to-identify species' echolocation calls successfully. Third, use of the shape modeling techniques (that have been pioneered by morphometricians) in the analysis and interpretation of spectrograph data opens the door to a new level of understanding of the complex geometric spaces within which a a wide variety of data reside. Such visualizations are simply not possible when using traditional sets of (largely scalar) spectrogram descriptor variables. More comprehensive, systematic, and sensitive quantification of spectrogram data may eventually allow the study of natural sounds (bioacoustics) to become a subject for evolutionary biologists to study by taking advantage of numerous procedures whose effectiveness – even when analysing complex data in which the concept of biological homology applies only in the loosest of terms (e.g., behavioural data, social data, cognition data) has already been demonstrated. These include the following:

i) comparative method studies that are necessary to improve the statistical testing of bioacoustic data by explicitly identifying phylogenetic covariation and incorporating these factors into the design of statistical procedures;
ii) methods that seek to identify the structure of covariance patterns between bioacoustic data and morphological data, genetic data, geographic data, ecological data, and indeed other acoustical data (e.g., to analyse patterns of correlation between the sounds made by animals and aspects of their phylogeny, ecology, behaviour, linguistic style);
iii) as a potentially new source of information for testing phylogenetic hypotheses and understanding the history of life on Earth.

At the most general level, the results achieved by this investigation suggest that geometric morphometrics can transcend its roots in biological morphology and address important questions in fields that, on first inspection, are not currently regarded as being morphological in nature. These include research programmes in other areas of the biological sciences, in the physical sciences, and possibly even in the social sciences and humanities (e.g., MacLeod 2009b).

Geometry is a fundamental aspect of the world in which we live. Owing to our own evolutionary history we have an affinity for conceptualizing the patterns we observe in geometric terms. The tools of shape theory and geometric analysis, forged as a result of the development of geometric morphometrics, have provided the scientific community with a set of data-analysis instruments of unlimited potential in terms of its range of conceivable applications. In order to push the morphometrics revolution forward into the twenty-first century morphometricians need to understand both the generalized nature of the tools they possess and the geometric dimensions of the interesting questions that exist in research fields far removed from morphometrics' traditional home in systematic biology. By expanding the scope of scientific problems that can be addressed by geometric morphometric methods the morphometrics community can not only make important contributions in areas far removed from its "local neighbourhood" of anatomy-based biological sciences, it can help reconceptualize problems across the physical, chemical, and humanistic sciences, demonstrate the ubiquity of morphological patterns throughout the nature, and bring some of the most sophisticated analytic approaches in the whole of applied mathematics to be bear on their resolution. They can, in a word, continue the ongoing morphometrics revolution, a continuation that will reap benefits for morphometics and morphometricians, as well as for those working in the fields to which these methods may be applied.

References

Adams D., Rohlf F.J., Slice D.E., 2004. Geometric morphometrics: ten years of progress following the "revolution": Italian Journal of Zoology 71: 5–16.

Aston J.D., Buck D., Coleman J., Cotter C.J., Jones N.S., Macaulay V., MacLeod N., Moriary J.M., Nevins A., 2012. Phylogenetic inference for function-valued traits: speech sound evolution: Trends in Ecology and Evolution 27(3): 160–166.

Blackith R.E., Reyment R.A., 1971. Multivariate morphometrics. Academic Press, London.

Bellman R.E., 1957. Dynamic programming. Princeton University Press, Princeton.

Bolton S., Edgecombe G.D., MacLeod N., 2008. Variability in female gonopods of scutigeromorph centipedes (Chilopoda): a geometric morphometric approach. Pecking 6: 97.

Bookstein F.L., 1990. Analytic Methods: Introduction and Overview, In: Rohlf F.J., Bookstein F.L. (Eds.) Proceedings of the Michigan Morphometrics Workshop. The University of Michigan Museum of Zoology, Special Publication 2, Ann Arbor, MI. 61–74.

Bookstein F.L., 1991. Morphometric tools for landmark data: geometry and biology. Cambridge University Press, Cambridge.

Bookstein F.L., 1993. A brief history of the morphometric synthesis. In: Marcus L.F., Bello E., García-Valdecasas A. (Eds.) Contributions to Morphometrics. Museo Nacional de Ciencias Naturales 8, Madrid. 18–40.

Bookstein F.L., 1997. Landmark methods for forms without landmarks: Localizing group differences in outline shape. Medical Image Analysis 1: 225–243.

Campbell N.A., Atchley W.R., 1981. The geometry of canonical variate analysis. Systematic Zoology 30: 268–280.

Collen A., 2012. Evolution of bat echolocation. PhD Dissertation. University College London, London.

Cramer C.J., 2004. Essentials of computational chemistry: theories and models, 2nd Edition. Wiley, New York.

Dryden I.L., Mardia K.V., 1998. Statistical shape analysis. Wiley, New York.

Felsenstein J., 1985. Phylogenies and the comparative method. American Naturalist 125: 1–15.

Felsenstein J., 1988. Phylogenies and quantitative characters. Annual Review of Ecology and Systematics 19: 445–471.

Felsenstein J., 2002. Quantitative characters, phylogenies, and morphometrics. In: MacLeod N., Forey P.L. (Eds.) Morphology, shape and phylogeny. Taylor & Francis, London. 27–44.

Fenton M.B., Audet D., Obrist M.K., Rydell J., 1995. Signal strength, timing and self deafening: the evolution of echolocation in bats. Paleobiology 21: 229–242.

Ferson S., Rohlf F.J., Koehn R.K., 1985. Measuring shape variation of two-dimensional outlines. Systematic Zoology 34(1): 59–68.

Franklin S.E., Maudie A.J., Lavigne M.B., 2001. Using spatial cooccurrence texture to increase forest structure and species composition classification accuracy. Photogrammetric Engineering and Remote Sensing 67(7): 849–855.

Friston K., Phillips J., Chawla D., Büchel C., 2000. Nonlinear PCA: characterizing interactions between modes of brain activity. Philosophical Transactions of the Royal Society, Series B (Biological Sciences) 355: 135–146.

Greenacre M.J., 1984. Theory and applications of correspondence analysis. Academic Press, London.

Gunz P., Mitteroecker P., Bookstein F.L., 2005. Semilandmarks in three dimensions. In: Slice D.E. (Ed.) Modern Morphometrics in Physical Anthropology. Kluwer Academic/Plenum Publishers, New York. 73–98.

Hammer Ø., Harper D., 2006. Paleontological data analysis. Blackwell Publishing, Oxford, UK.

Harris F.J., 1978. On the use of windows for harmonic analysis with the discrete Fourier transform. Proceedings of the IEEE 66(1): 51–83.

Harvey P.H., Leigh Brown A.J., Smith J. M., Nee S., 1996. New uses for new phylogenies. Oxford University Press, Oxford.

Harvey P.H., Pagel M.D., 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.

Jones G., 1999. Scaling of echolocation call parameters in bats. Journal of Experimental Biology 202: 3359–3367.

Jones K.E., Russ J.A., Bashta A.-T., Bilhari Z., Catto C., Csösz I., Gorbachev A., Györfi P., Hughes A., Ivashkiv I., Koryagina N., Kurali A., Langton S., Maltby A., Margiean G., Pandourski I., Parsons S., Prokofev I., Szodoray-Paradi A., Szodoray-Paradi F., Tilova

E., Walters C., Weatherill A., Zavarzin O. 2013. Indicator Bats Program: a system for the global acoustic monitoring of bats, In: Collen B.P., Pettorelli N., Durant S.M., Krueger L., Baillie J. (Eds.) Biodiversity monitoring and conservation: bridging the gaps between global commitment and local action. Wiley-Blackwell, London. 213–247.

Jones G., Teeling E.C., 2006. The evolution of echolocation in bats. Trends in Ecology & Evolution 21(3): 149–156.

Kalko E.K.V., Schnitzler H.U., 1993. Plasticity in echolocation signals of European pipistrelle bats in search flight: implications for habitat use and prey detection. Behavioral Ecology and Sociobiology 33(6): 415–428.

Kayitakire F., Hamel C., Defourny P., 2006. Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery. Remote Sensing of Environment 2006: 390–401.

Kendall D.G., 1984. Shape manifolds, procrustean metrics and complex projective spaces. Bulletin of the London Mathematical Society 16: 81–121.

Klingenberg C.P., 2008. Novelty and "homology-free" morphometrics: what's in a name. Evolutionary Biology 35: 186–190.

Kovarovic K., Aiello L.C., Cardini A., Lockwood C.A., 2011. Discriminant function analyses in archaeology: are classification rates too good to be true? Journal of Archaeological Science 38: 3006–3018.

Kramer M.A., 1991. Nonlinear principal component analysis using autoassociative neural networks. AIChE Journal 37(2): 233–243.

Kuhl R.K.J., Giardina C.R., 1982. Elliptic Fourier features of a closed contour. Computer Graphics and Image Processing 18: 236–258.

Lohmann G.P., 1983. Eigenshape analysis of microfossils: A general morphometric method for describing changes in shape. Mathematical Geology 15: 659–672.

Lundy M., Teeling E., Boston E., Scott D., Buckley D., Prodohu P., Marnel F., Montgomery I., 2011. The shape of sound: elliptic Fourier descriptors (EFD) discriminate the echolocation calls of *Myotis* bats (*M. daubentoni*, *M. nattereri* and *M. mystacinus*). Bioacoustics 20: 101–116.

MacLeod N., 1999. Generalizing and extending the eigenshape method of shape visualization and analysis. Paleobiology 25(1): 107–138.

MacLeod N., 2001. The role of phylogeny in quantitative paleobiological analysis. Paleobiology 27: 226–241.

MacLeod N., 2002. Phylogenetic signals in morphometric data. In: MacLeod N., Forey P.L. (Eds.) Morphology, shape and phylogeny. Taylor & Francis, London. 100–138.

MacLeod N., 2006. Rs and Qs II: correspondence analysis. Palaeontological Association Newsletter 62: 60–74.

MacLeod N., 2007. Groups II. Palaeontological Association Newsletter 65: 36–49.

MacLeod N., 2008. Understanding morphology in systematic contexts: 3D specimen ordination and 3D specimen recognition. In: Wheeler Q. (Ed.) The New Taxonomy. CRC Press, Taylor & Francis Group, London. 143–210.

MacLeod N., 2009a. Form & shape models. Palaeontological Association Newsletter 72: 14–27.

MacLeod N., 2009. Images, totems, types and memes: perspectives on an iconological mimetics. Culture, Theory and Critique 50(2–3): 185–208.

MacLeod N., 2012. The center cannot hold II: elliptic Fourier analysis. Palaeontological Association Newsletter 79: 29–42.

MacLeod N. (in press). A comparison of alternative form-characterization approaches to the automated identification of biological species. In: Hamilton A., Peirson E. (Eds.) Patterns in Nature. University of California Press, Berkeley, California.

MacLeod N., O'Neill M.A., Walsh S.A., 2007a. A comparison between morphometric and artificial neural net approaches to the automated species-recognition problem in systematics. In: Curry G., Humphries C. (Eds.) Biodiversity databases: techniques, politics, and applications. CRC Press, Taylor & Francis Group, Boca Raton, Florida. 37–62.

MacLeod N., O'Neill M.,Walsh S.A., 2007b. Automated tools for the identification of taxa from morphological data: face recognition in wasps. In: MacLeod N. (Ed.) Automated taxon recognition in systematics: theory, approaches and applications. CRC Press, Taylor & Francis Group, Boca Raton, Florida. 153–188.

Manly B.F.J., 1994. Multivariate statistical methods: a primer. Chapman & Hall, Bury, St. Edmonds, Suffolk.

Manly B.F.J., 1997. Randomization, bootstrap and Monte Carlo methods in biology. Chapman & Hall, London.

Mardia K.V., Dryden I., 1998. The statistical analysis of shape data. Biometrika 76: 271–282.

Martins E.P., Hansen T.F., 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. American Naturalist 149: 646–667.

Oxnard C., O'Higgins P., 2008. Biology clearly needs morphometrics. Does morphometrics need biology? Biological Theory 4(1): 84–97.

Parsons S., Jones G., 2000. Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. The Journal of Experimental Biology 203: 2641–2656.

Pielou E.C., 1984. The interpretation of ecological data. John Wiley & Sons, New York.

Pimentel R.A., 1979. Morphometrics: the multivariate analysis of biological data. Kendall/Hunt, Dubuque, IA.

Podani J., Miklós I., 2002. Resemblance coefficients and the Horseshoe Effect in principal coordinates analysis. Ecology 83: 3331–3343.

Polly P.D., 2008. Adaptive zones and the pinniped ankle: a 3D quantitative analysis of carnivoran tarsal evolution. In: Sargis E., Dagosto M. (Eds.) Mammalian Evolutionary Morphology: A Tribute to Frederick S. Szalay. Springer, Dordrecht, The Netherlands. 165–194.

Polly P.D., MacLeod N., 2008. Locomotion in fossil Carnivora: an application of the eigensurface method for morphometric analysis of 3D surfaces. Palaeontologia Electronica 11(2): 13.

Redgewell R.D., Szewczak J. M., Jones G., Parsons S., 2009. Classification of echolocation calls from 14 species of bat by support vector machines and ensembles of neural networks. Algorithms 2: 907–924.

Revell L.J., 2009. Size-correction and principal components for interspecific comparative studies. Evolution 63(12): 3258–3268.

Reyment R.A., 1991. Multidimensional paleobiology. Pergamon Press, Oxford.

Reyment R.A., Jöreskog K.G., 1993. Applied factor analysis in the natural sciences. Cambridge University Press, Cambridge.

Rohlf F.J., 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution 55(11): 2143–2160.

Rohlf F.J., 2002. Geometric morphometrics and phylogeny. In: MacLeod N., Forey P.L. (Eds.) Morphology, shape and phylogeny. Taylor & Francis, London. P175–193.

Rohlf F.J., 2006. A comment on phylogenetic correction. Evolution 60(7): 1509–1515.

Rohlf F.J., Slice D., 1990. Extensions of the Procrustes method for optimal superposition of landmarks. Systematic Zoology 39: 40–59.

Rohlf F.J., 1993. Relative warp analysis and an example of its application to mosquito wings. In: Marcus L.F., Bello E., García-Valdecasas A. (Eds.) Contributions to Morphometrics. Museo Nacional de Ciencias Naturales 8, Madrid. 131–160.

Roiu R., Seyler F., 1997. Texture analysis of tropical rainforest infrared satellite images: application to structural geology and soil studies. Photogrammetric Engineering and Remote Sensing 63(5): 515–521.

Russ J., 2012. British Bat Calls: A Guide to Species Identification. Pelagic Publishing, Exeter, UK.

Schnitzler H.U., Kalko E.K.V., 2001. Echolocation in insect-eating bats. Bioscience 51: 557–569.

Schölkopf B., Smola A., Müller K.-R., 1998. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10(5): 1299–1319.

Scholz M., Fraunholz M., Selbig J., 2007. Nonlinear principal component analysis: neural network models and applications. In: Alexander N., Gorban A.N., Kégl B., Wunsch D. C., Zinovyev A. (Eds.) Lecture Notes in Computational Science and Engineering. Volume 58: 44–67.

Sievwright H., MacLeod N., 2012. Eigensurface analysis, ecology, and modeling of morphological adaptation in the falconiform humerus (Falconiformes: Aves). Zoological Journal of the Linnean Society 165: 390–415.

Strauss R.E., Bookstein F.L., 1982. The truss: body form reconstruction in morphometrics. Systematic Zoology 31: 113–135.

Supa M., Cotzin M., Dallenbach K.M., 1944. "Facial vision" – the perception of obstacles by the blind. The American Journal of Psychology 57(2): 133–183.

Teeling E.C., 2009. Hear, hear: the convergent evolution of echolocation in bats? Trends in Ecology & Evolution 24(7): 351–354.

Thaler L., Arnott S.R., Goodale M. A., 2011. Neural correlates of natural human echolocation in early and late blind echolocation experts. PLOS One 6(5): 1–16.

Vaughan N., Jones G., Harris S., 1997. Identification of british bat species by multivariate analysis of echolocation call parameters. Bioacoustics 7(3): 189–207.

Walters C.L., Freeman R., Collen A., Dietz C., Fenton M.B., Jones G., Obrist M.K., Puechmaille S.J., Sattler T., Siemers B.M., Parsons S., Jones K.E., 2012. A continental-scale tool for acoustic identification of European bats. Journal of Applied Ecology 49(5): 1064–1074.

Young D., 2001. Computational chemistry: a practical guide for applying techniques to real world problems. Wiley, New York.

Associate Editor: A. Cardini

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 126–132, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/6284/pdf
doi:10.4404/hystrix-24.1-6284

Research Article

Applying geometric morphometrics to compare changes in size and shape arising from finite elements analyses

Paul O'Higgins[a,*], Nicholas Milne[b]

[a] Centre for Anatomical and Human Sciences, Hull York Medical School, University of York, York, England, YO10 5DD, UK
[b] School of Anatomy Physiology and Human Biology, University of Western Australia, Crawley 6009, Australia

Keywords:
geometric morphometrics
finite elements analysis
armadillo femur

Article history:
Received: 30 May 2012
Accepted: 4 September 2012

Acknowledgements
We are grateful to Sergio Vizcaíno and Michael Fagan who made useful suggestions in early discussions related to this work, Laura Fitton and Flora Gröning provided advice and support in our model reconstruction and FEAs. Fred Bookstein has also provided ongoing, critical and insightful feedback on our efforts to apply GM in functional studies. Any errors that remain are the authors' own responsibility. Vox-FE was developed by Michael Fagan, Roger Phillips and Paul O'Higgins with support from BBSRC (BB/E013805; BB/E009204). Nick Milne was supported by a study leave grant from The University of Western Australia.

Abstract

We consider how the methods of geometric morphometrics (GM) might combine with functional simulations using finite elements analysis (FEA). In particular we are concerned with how the deformations arising from FEA might be compared and visualized using GM. To these ends we apply these methods to a study of coronal plane forces applied to a model of an armadillo femur. We simulate the stance phase in the hind limb where the femur is subject to bending strains due to longitudinal compressive as well as abduction loads on the greater trochanter. We use this model to examine the hypothesis that muscles attaching to the third trochanter can reduce these bending strains in the loaded femur. The analysis uses standard finite element methods to produce strain maps and examine the strains at 200 point locations on the femur, but we also use the locations of the 200 points to run novel geometric morphometric analyses to assess the gross deformation of the model under different loadings. These provide insights into the application and usefulness of geometric morphometric methods in interpreting the results of finite element analyses. With further mathematical, engineering and statistical development the combination of FEA and GMM should open up new avenues of investigation of skeletal form and function in evolutionary biology.

Introduction

In this paper we consider how studies of skeletal performance using finite elements analysis (FEA) might be compared using geometric morphometric (GM) approaches. After considering what GM can and cannot do in this context, we illustrate and explore the joint use of these approaches through an example study of femoral form and function in an armadillo. GM methods alone can be used to relate variations in skeletal form to aspects of function by assessing how the form of a configuration of landmarks taken on the skeletal part covaries with functionally interesting variables. Thus, in a study of long bones, how skeletal form covaries with e.g. body mass, or limb length, or running speed etc. can be understood and compared through e.g. regression of form on relevant variables. However, this approach does not directly relate skeletal form to performance under loading. For this, simulated functional loading using finite elements analysis (FEA) with subsequent measurement and comparison of skeletal performance is required. One interesting aspect of performance is how an object deforms when loaded.

In continuum mechanics the term "deformation" means both rigid body motion (translation and rotation) together with changes in form (size and shape). Here, however we use the common definition of "deformation", which refers to changes in size and shape of an object but not rigid body motions. This is a more familiar usage for workers in GM. In engineering how the size and shape of a loaded elastic body responds to loads is commonly measured using strains (e.g. principal strains, Von Mises strains). These are used to predict failure and often in biology to assess and compare performance of either the same model under different loadings or, different models under similar loadings (e.g. in studies of the skeletal response to loads, Gupta et al. 1973; Hart et al. 1992; Korioth and Hannam 1994; Vollmer et al. 2000; Koolstra and van Eijden 2005; Moazen et al. 2009; Rayfield 2005; Ichim et al. 2007; Strait et al. 2007, 2009; Kupczik et al. 2007; Wroe et al. 2007). Strains sampled at several points may be compared singly or submitted to multivariate analysis (e.g. Gröning et al. 2012; Parr et al. 2012). However this leads to an incomplete analysis, and no proper statistical framework yet exists to compare strain fields (Bookstein, 2011; Weber et al., 2011) or to account for the effects of uncertainties in modelling on computed strains. Strains and strain maps are also frequently used as a visual guide to large scale deformations such as long bone bending or cranial "twisting" that arises from loading. Strain magnitudes and directions describe deformations at each point but are not well suited to assessment of large scale patterns of deformation (see O'Higgins et al. 2011 for more detail). An alternative is to assess large scale deformations by describing the changes in the form of a landmark configuration on a body before and after loading. This shares much in common with the application of GM approaches to kinematic analyses of motion using temporal sequences of landmark configurations, representing e.g. changing limb postures. At each temporal sampling point, the form of the configuration is recorded and the full set of configurations sampled over the period of interest is submitted to GM analysis. The analysis then focuses on comparing trajectories of form change over time. Slice (1999) and Adams and Cerney (2007) have shown how this approach can facilitate quantification and analysis of complex motions involving many joints or complex motion at few joints (e.g. the jaw in chewing). It is equally applicable to kinematic analyses of deformable surfaces or volumes such as the face during expression or speech (O'Higgins et al., 2002), or to a body deforming under loads such as is simulated in FEA. There are subtleties to such an analysis that concern: registration (i.e., if and how we scale, translate and rotate), its meaning and effects, and how motions are broken down into sequential landmark configurations (Slice, 2003).

*Corresponding author
Email address: paul.ohiggins@hyms.ac.uk (Paul O'Higgins)

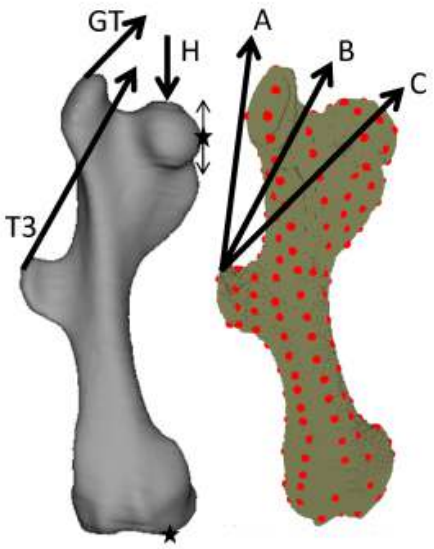


Figure 1 – The armadillo femur models showing the constraints (stars), forces applied (arrows) to the head (H), greater trochanter (GT), and third trochanter (T3). Also shown is the finite element model with; the 200 landmarks placed on the surface, and the three possible directions for the third trochanter load.

While it is unlikely that large scale deformations of skeletal parts drive either evolutionary or ontogenetic adaptations, differences can be informative from the perspective of analyses of anatomical-functional correlations (e.g Milne and O'Higgins 2012). Further, GM analyses of how a skeletal structure deforms can provide a useful adjunct to sensitivity and validation studies where subtle differences are not readily appreciated from strains sampled at a few locations or from strain maps (Gröning et al., 2011; Fitton et al., 2012). It should be noted that GM approaches to the comparison of deformations do not have the same aim as strain based comparisons; they are not directed at predicting failure. Neither can transformation grids describe the actual deformation of tissue between landmarks. This is because the method of interpolation does not reflect the actual stretching or compression of the physical material. Of course, this same issue applies to GM studies of skeletal ontogeny and evolution where grids serve as a device for visualisation of changes in size and shape of landmark configurations rather than as a representation of the biological processes underlying bone growth or evolutionary transformation; it facilitates visualisation of pattern, not process, by interpolating the changes in form of a landmark configuration to the space between and in the vicinity of the landmarks. Additionally, the landmarks themselves, unless they comprise all of the nodes of all of the finite elements, incompletely describe the deformation (O'Higgins et al., 2011). It should be borne in mind that, as with studies of e.g. growth or evolution, different landmark configurations will give rise to different distances between forms and so adequate design of the landmark configuration is important (Oxnard and O'Higgins, 2009).

We demonstrate the application of GM to the results of FEA in a study of the femur of an armadillo (*Chaetophractus villosus*). The femur presents a large third trochanter (T3; Fig 1), a common feature of xenarthrans, that varies in position along the shaft being more distal in larger animals. Many studies have highlighted how human femoral bending occurs in response to longitudinal compressive loads acting through the femoral head and abductor loads acting on the greater trochanter, and how loads simulating tension in the iliotibial tract have an "unbending" effect (using polarised light methods and beam theory, Pauwels 1980; Rybicki et al. 1972; finite element methods, Taylor et al. 1996; free body analysis, Duda et al. 1997). Similarly, Milne et al. (2011) suggested that muscles attached to T3 in xenarthrans act to reduce coronal plane bending stress in the armadillo femur. Recently this suggestion has been supported by a study (Milne and O'Higgins, 2012) that compared the "unbending" effect of muscles attached to T3 in large and small xenarthrans. The background to this work is presented here to illustrate how GM approaches can be usefully applied to interpretation of changes in skeletal form arising from simulated loading in FEA.

We carry out FEA to assess the function of the third trochanter and the impact of variations in loading and segmentation of internal architecture. We also compare the performance of a solid model with that of a model with more detailed internal structure because it is not straightforward to delineate cortical from trabecular bone in CT scans, especially in fossils which we intend to include in subsequent work. Using GM approaches, we show the effects of varying the magnitudes of simple and combined loads on the resulting strain maps and on large scale deformations of solid and hollow femoral models. These analyses allow us to assess the extent to which gluteus maximus and tensor fascia latae muscles attached to the third trochanter reduce bending.

Methods

Model building, sensitivity to modelling decisions and loading simulation. One femur from a hairy armadillo (*Chaetophractus villosus*) was CT scanned (1 mm slices with a resolution of 0.1145 mm). The CT stack was segmented (i.e the grey levels representing bone were used to isolate bone material in each CT slice in preparation for building a 3-D model of the bone) in AMIRA 4.1.1 (Mercury Computer Systems Inc., USA). The resolution of the scan meant that the cancellous bone in the epiphyses could not be segmented in any detail. In consequence the initial model was mostly solid at the epiphyses but retained the empty medullary space in the shaft. Extraneous material, including remnants of the cruciate ligaments, was removed. Following sensitivity studies that showed little difference in resulting strains or large scale deformation with voxel side length varying between 0.2 and 0.8 mm, the Amira mesh was resampled to make a model with cubic voxels of side 0.4 mm. This model was then re-segmented to fill the hollow in the shaft with solid material. The 3D volume data for the two models were exported as bitmap stacks and then converted to 8-noded linear brick finite element meshes by direct voxel conversion. The resulting models had 83627 (hollow) and 86914 (solid) elements. It has been shown in previous work that our voxel based approach achieves almost identical results to those obtained using other element types (Liu et al., 2011).

The finite element analyses (FEA) were performed using the non-commercial FEA software VOX-FE (Fagan et al. 2007; numerically validated in Liu et al. 2011; release will be announced on http://www2.hull.ac.uk./science/cmet.aspx). The models were assigned isotropic material properties within the range of published values for bone (17 GPa for Young's modulus and a Poisson's ratio of 0.3) although material properties vary from location to location (are heterogeneous, e.g. Dechow et al. 1993). This is justified on the grounds that Panagiotopoulou et al. (2012) have shown that in an elephant femur, although strains predicted by an homogeneous model less well matched experimental data than those from an heterogeneous model, the mismatch was principally in the mean magnitude of predicted strains and, to lesser degree, in the pattern of deformation (relative strain values). Thus, to predict the overall pattern, but not magnitude of deformation, the heterogeneity typical of long bones appears less important.

To approximate physiological conditions, the models were constrained in the x, y, and z directions at an area of the distal surface of the medial femoral condyle, and also in the x and y directions on the medial surface of the femoral head, thus enabling the head to move up or down under load. These constraints were chosen because, in Xenarthra the medial condyle bears most of the load at the knee (Koneval 2003; Milne et al. 2011 and references therein), and the sliding constraint on the femoral head represents the pelvis, which prevents the femoral head from being displaced medially. Simple loads (Fig. 1) were used to assess model performance, thus a compressive force representing body mass and the net action of thigh musculature was applied to the upper surface of the femoral head, and that force was directed through the centre of the constrained area of the medial femoral condyle. A force representing the lesser gluteal (abductor) muscles was applied to the superolateral part of the greater trochanter and that force was directed

superomedially, in the coronal plane at an angle of 45 degrees above the horizontal (Fig. 1). The size of the head and greater trochanter loads was 60 and 120 newtons, equivalent to the effects of approximately twice and 4 times the body weight of the armadillo. The number of constrained or loaded nodes was chosen to mirror normal anatomy while keeping the number of such nodes small, about 60 for each constraint and load. The models were iteratively solved to equilibrium on a PC.

Two hundred landmarks were placed on the surfaces of the models using Amira (Fig. 1). The landmarks were chosen to provide even and fairly dense coverage of the whole bone. There is no issue of equivalence of landmarks between load cases as the landmarks are tied to the bone mesh and move with it as it deforms. Further in a recent paper comparing the function of armadillo femora (Milne and O'Higgins, 2012) we have shown that the results of size and shape analyses using 40 "homologous" landmarks to compare femoral performance are very similar to those from the 200 landmarks used to assess deformation of single femora. This finding is consistent with a sensitivity analysis (in Fitton et al. 2012) that found, in reducing from 300 to 70 landmarks, little effect on performance analyses relating to cranial loading. This is reassuring in the context of the present study but it should be borne in mind that similar considerations apply to landmark choice in analyses of performance as apply in relation to analyses of form (Oxnard and O'Higgins, 2009) and additional internal landmarks might bring additional information into the analyses. After each load case was run, the principal strains at, and the new coordinates of, each landmark were extracted. For each model 5 load cases were run: 60 or 120 newtons at the head or greater trochanter and 60 newtons at both the head and greater trochanter (GT). Subsequently, additional loads were applied to T3. Since muscle force vectors are estimated from dissections (Koneval, 2003) and observations of articulated skeletons, three different force directions were applied to the third trochanter to assess the sensitivity of the model to force direction (Fig. 1).

Maps of either Von Mises or surface principal strains were also produced. Von Mises strains reflect the magnitude of deformation at each point and require just one diagram to map them. They simplify the presentation of results but information about the nature of strains, whether they are tensile or compressive, requires maps of the relevant principal strains. Surface Von Mises strains were mapped from the analysis of load doubling and for the comparison of hollow and solid models, because they succinctly summarise the patterns of strain density. Surface principal strains were mapped from the analysis of multiple loads because the diagrams show whether the strain at each point is tensile or compressive and so facilitate interpretation of large scale deformation of the model.

Geometric morphometric analysis of deformations. The 3D coordinates of the 200 landmarks in the unloaded model and each loadcase were submitted to geometric morphometric (GM) analysis to assess global deformations. The most common approach to GM analysis focuses on shape; it scales landmark configurations to unit centroid size; the square root of the sum of squared landmark distances from their centroid. Next, shape variables are computed by translating and rotating (registering) all configurations to minimise the sum of squared landmark distances with respect to the mean(Dryden and Mardia, 1998). Differences in shape are expressed by Procrustes distances, computed as the square root of the sum of squared differences in shape variables between configuration pairs. Differences in size are expressed by differences in centroid size, the square root of the sum of deviations of the landmarks from the centroid.

However, in this application to the comparison of results of FEA, an approach that simultaneously accounts for differences in both size and shape is required. This is because, under loading the body deforms and landmarks displace. Differences in both size and shape are consequent on the applied loads; in terms of mechanics it makes little sense to partition form changes into these components or to weight them differently. Therefore, in previous applications of GM methods (O'Higgins et al., 2011; Gröning et al., 2011; Fitton et al., 2012) principal components analysis of the shape variables plus the log of centroid size (Procrustes form analysis) was employed to assess how loadings impact on both model size and shape. However, logging of centroid size rescales this component relative to shape in such a way that size differences resulting from loading are given a relatively lower weight than would be the case if the initial scaling had not been carried out.

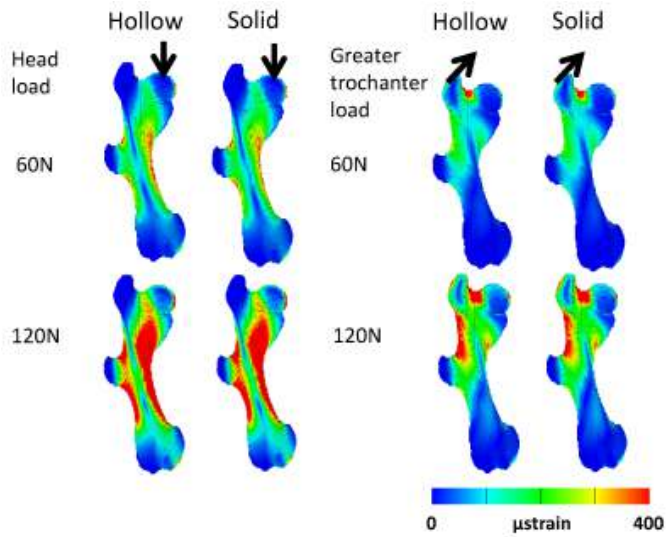


Figure 2 – Simple loadings on the head or greater in hollow and solid models. Strain maps showing the effects of head or greater trochanter loads of 60 and 120 newtons. The pairs of images allow comparison of the strain patterns in hollow and solid models. Von Mises strain 0 to 0.4 microstrains.

An alternative that preserves the weighting of both is to omit the scaling step prior to registration and carry out analyses using the resulting "size and shape" variables (Dryden and Mardia 1998; Dryden et al. 2007. Note the resulting variables combine size and shape information and in contrast to Procrustes Form analysis do not comprise shape variables plus a size variable). While omission of the scaling step does not lead to specimens being represented in the well-behaved Kendall's shape space that results from Procrustes superimposition, the fact that the deformations resulting from FEA are very small mitigates the impact of variations in size on such things as the estimation of means and covariances. Further, a consequence of omitting the scaling step is that resulting distances between loaded and unloaded forms might reflect pure shape or pure size, or some mix of size and shape depending on the directions of the vectors connecting the models in the resulting size and shape space. While this may be undesirable in some applications it is consistent with the idea that loadings can produce changes in size and/or shape. Thus, conceivably, but impractically, loads could be applied everywhere within and over an object to make it isometrically smaller or larger, but more often loads have some effect on size and some on shape. In order therefore to assess the effects of loads, size and shape changes need to be considered together.

Here, we use this approach because it is better justified from considerations of the mechanics, but the consequences on the eventual results are imperceptible with regard to relative distances among load cases and visualisations of aspects of deformation captured by PCs when compared to the results of Procrustes form (shape plus log centroid size) or shape analyses. This lack of difference between approaches is attributable to the very small differences in size relative to shape that arise from our FEAs. Thus, we carry out a "size and shape" analysis by translating and rotating but not scaling landmark configurations to minimise the sum of squared distances among landmarks. This rigid body fitting of landmark configurations from unloaded and loaded forms (see O'Higgins et al. 2012) produces "size and shape" variables. Size and shape distances among unloaded and loaded models are computed and PCA of size and shape variables is carried out to complement the strain based analyses. The aspects of size and shape variation described by each PC can be visualised, facilitating interpretation of PC plots in terms of deformations of the models. We visualise these deformations as warped rendered surface models together with a trans-

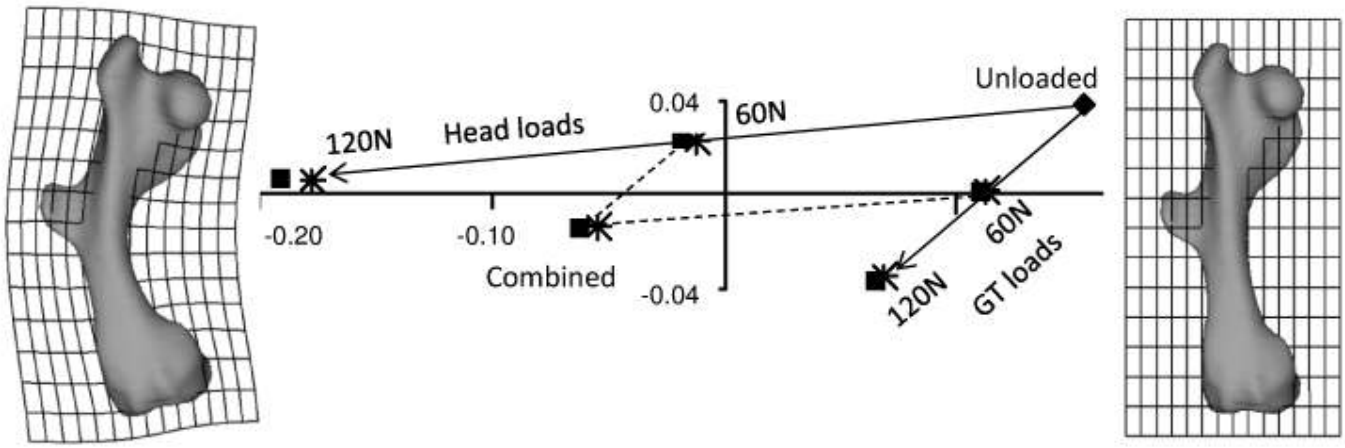


Figure 3 – GM analysis of simple and combined loads. PCI (95.9%) and PC2 (4.1%) from a size and shape space analysis of the coordinates of the 200 landmarks. The shapes to the left and right show the shape change in the model associated with PCI, from the unloaded state on the right to the loaded state with the deformed grid on the left. These deformations are exaggerated by a factor approximately 100 to aid visualisation. Diamond = unloaded, square = hollow, star = solid models. Note that two PCs are sufficient to describe the aspects of large scale deformation in this analysis because only two forces are applied, This results in two modes of deformation, albeit with varying relative magnitude among load cases, which are described by two PCs.

formation grid computed using a triplet of thin plate splines (one for displacements in each of x, y and z; Dryden and Mardia 1998). As noted earlier, the deformation of the grid is interpolated from the deformation of the landmark configuration and so only approximates the actual deformation of the model between the landmarks; in this sense it is quite unsuited to studying the elastic behaviour of the bone material. Rather, the grid is a device to facilitate visual interpretation of the large scale aspects of deformation of the landmark configuration.

Specific analyses using both strains and GM methods assessed: 1. Differences between the deformations of the solid and hollow models. 2. The effects of doubling of loads. 3. The effects of combining loads. 4. The impact of muscles acting on the third trochanter on "unbending" of the femoral shaft.

Results

Differences between the deformations of the solid and hollow models

Strain maps for the two magnitudes of head and greater trochanter (GT) loads (Fig. 2) show very similar results for the hollow and solid models. They are hard to distinguish by eye. Quantitatively, the magnitudes of tensile and compressive strains at the 200 sampled nodes are slightly lower on average for the solid than the hollow model. The ratios of the mean strains at these 200 point locations in solid compared to the hollow models are all slightly less than one (60 N loads, tensile: head 0.992, GT 0.955; compressive: head 0.959, GT 0.970; for 120 N loads these proportions are exactly the same as for the 60 N loads). Further, the strain ratios between models at each sampled node closely approximate the ratios of mean strains (e.g. 60 N loads, mean solid/hollow strain ratio: tensile: head 0.990, GT 0.948; compressive: head 0.984, GT 0.962). These findings support the similarities seen in the strain maps (Fig. 2) and indicate that the principal difference between hollow and solid models is that the solid is a little stiffer but deforms like the hollow; a finding that echoes a similar result in lizard crania (Parr et al., 2012).

When the 200 landmark co-ordinates for the unloaded and the five load conditions are submitted to geometric morphometric analysis the resulting plot of PCs 1 and 2 (accounting for 95.9% and 4.1% of the shape variation respectively) shows that: the solid models are slightly stiffer than the hollow in that they are less distant from the unloaded model (Fig. 3). Size and shape distances indicate that the solid model deforms less than the hollow (ratio solid/hollow 0.965 under the same head load, and 0.964 for the GT load).

The effects of doubling of loads

As expected, the strains at the 200 landmarks resulting from the 120 newtons loads are almost exactly twice the value of those that result from 60 newtons loads (all within 0.18%). Likewise, when the size and shape distances are compared between models with identical constraints and load vectors, these double as load magnitudes are doubled. Thus for the hollow model with head load alone the size and shape distance between the unloaded and 60 N (0.1729894666) models is half of that between the unloaded and 120 N loaded model (0.3459819139; ratio is 2.00001723). The same applies to centroid size (cs) with the change between the unloaded and the 60 N loaded models being very close to half that of the 120 N (hollow model ratios of 120 N to 60 N loadcases: head load = 1.998; GT 2.002). From Fig. 3, it is also apparent that doubling either load results in doubling of the lengths of the vectors between the unloaded and loaded models.

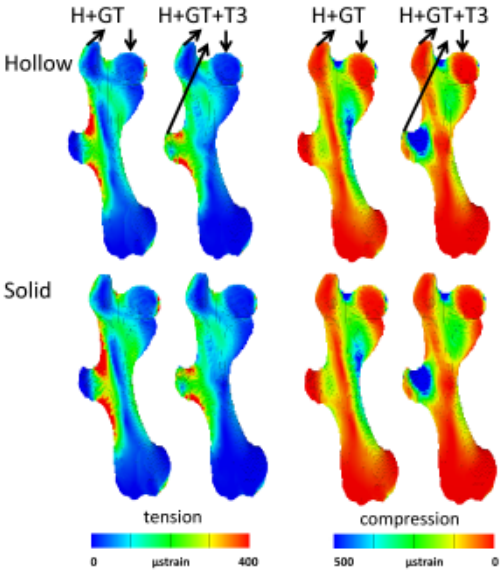


Figure 4 – The effect of third trochanter loads. Strain maps of the armadillo femur under combined head and greater trochanter (H+GT) loads, and with an additional load on the third trochanter (H+GT+T3). All the loads applied are 60 N. Both compressive (-0.5 to 0 microstrains) and tensile (0 to 0.4 microstrains) strains are shown for the hollow and solid models.

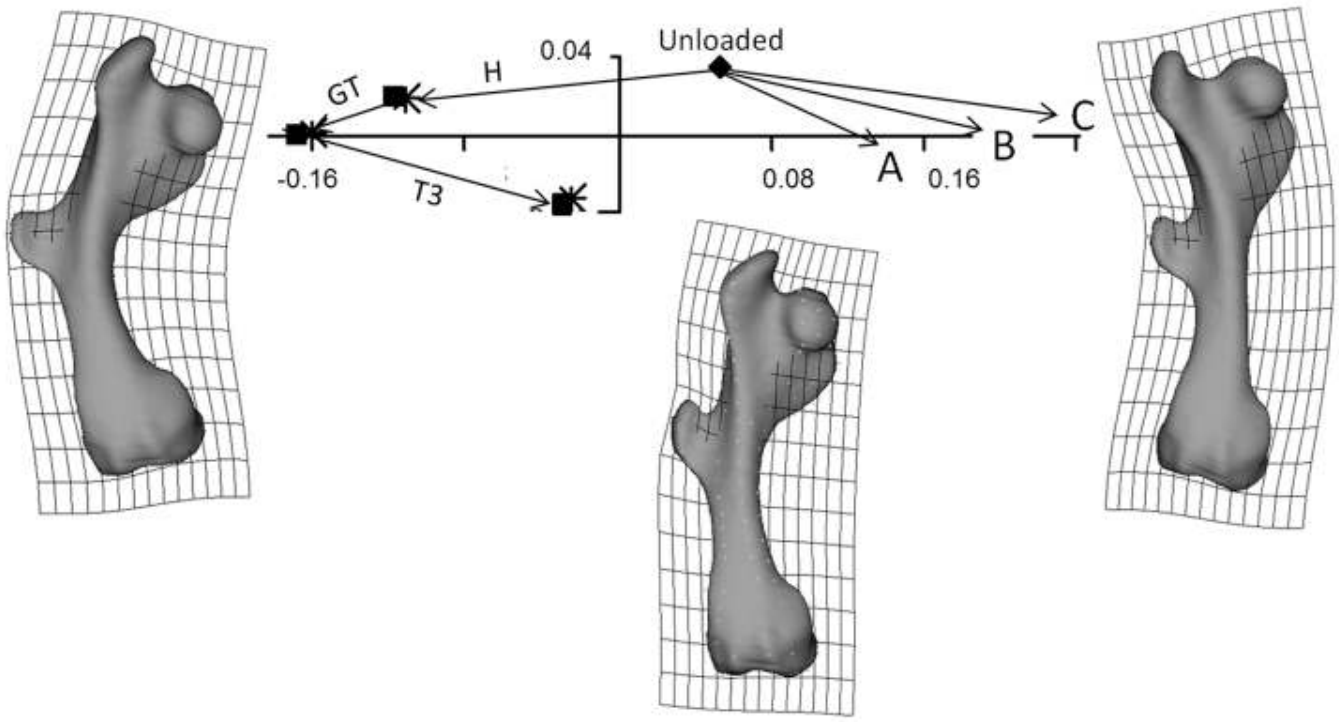


Figure 5 – The addition of third trochanter loads. PCs 1 & 2 (97.1% & 2.3%) from a GM analysis of hollow (squares) and solid (stars) models. First a head load is applied (Vector H), then a greater trochanter load is added (GT), and then the third trochanter load is added (B, labelled as T3). The points A, B and C represent the deformations due to the three different force directions applied to the third trochanter of the unloaded model (diamond). The deformed models with grids t indicate the changes in size and shape represented by PCl (left and right), and PC2 (below the plot). These deformations are exaggerated by a factor of approximately 100 to aid visualisation.

The effects of combining loads

The deformation caused by the combined load of 60 newtons on both the head and the greater trochanter is the same as the vector sum of the deformations caused by the two loads independently. This is illustrated by the parallelogram evident in Fig. 3. It is made up of the solid lines connecting the unloaded model and each of the 60 N loadcases and the dotted line connecting each of these and the combined load model. The ratios of the lengths of the short and long sides are respectively; (GT load) 0.999964, 0.999978 for the hollow model and 1.000013, 0.999993 for the solid model.

The impact of muscles acting on the third trochanter on "un-bending" of the femoral shaft

From Fig. 2 and 3, it is evident that both head and greater trochanter loads result in medial bending of the femur. From a comparison of Fig. 2 and 3 it is clear that head and greater trochanter loads combine to produce more large scale bending than either force alone. When additional forces are applied to T3 the bending strains in the femoral shaft are reduced. Fig. 4 shows the compressive and tensile strains in models with head and greater trochanter loads compared with models with additional third trochanter loads. The addition of T3 loads results in reductions in tensile strains on the lateral side, and compressive strains on the medial side of the model. The plot of PCs 1 and 2 of an analysis of all three loadcases (Fig. 5) shows that while the head and GT loads cause increased bending in the model, the T3 loads reduce that bending (Fig. 5). Fig. 5 also shows that the three different force directions for the T3 load (A, B and C) all reduce bending in the femoral models. The transformation grids in Fig. 5 show that the aspects of deformation represented by PC1 predominantly consist of large scale bending of the femoral shaft, while those represented by PC2 are principally deformations localised to the greater and third trochanters, with no apparent bending of the femoral shaft.

Discussion

We have simulated femoral loading, assessing the impact of loads, singly and then in combination, on deformation. The key underpinning technologies included imaging, image reconstruction and finite elements analysis which are commonly employed in functional analyses of skeletal elements (e.g. Hart et al. 1992; Kupczik et al. 2007; Wroe et al. 2007; Strait et al. 2009). Additionally we have applied some methods from geometric morphometrics to consider and compare large scale deformations (O'Higgins et al., 2011).

This study provides the opportunity to show how geometric morphometric methods relate to the more conventional engineering approach for quantitatively describing and comparing deformations; how they perform and inform. The situation whereby doubling of loads results in doubling of deformation is well known for principal strains (Fig. 2) and this linear relationship between load and deformation is also apparent using GM approaches (size shape distances from the unloaded model). Like strains, the empirically derived ratios of distances arising from doubling of loads are not exactly 2 but are very close, the error being most likely attributable to computational imprecisions in FEA and subsequent GM analyses. This similarity between the scaling of size and shape distances and strains and is of course to be expected, since both methods assess changes in size and shape, albeit at very different scales. It is worth noting that size and shape distances and strains also scale linearly with Young's elastic modulus (ε; a measure of model stiffness) and length. Thus halving ε doubles the size and shape distance between unloaded and loaded forms as does halving length (e.g. by isometrically scaling the form according to centroid size). These scaling relationships have a practical application in that a single load case can be used to visualise the range of deformations or estimate new deformations when constraints and load application points are kept constant but load magnitude, ε or length ("size") are varied. This avoids the need to run multiple alternative FEAs.

Since doubling of loads is effectively the combination of a load with itself, when two or more different loads are applied to the model, the resulting deformation is represented in the size and shape space as the vector sum of the deformations caused by those loads applied individu-

ally. Thus, in Fig. 3, the deformations caused by the individual loads form the sides of a parallelogram and the deformation of the combined load is represented by the diagonal of the parallelogram. This demonstrates that, as with strains, (and as expected with) size and shape distances, large scale deformations due to multiple loads can be estimated by combining the simpler constituent ones. Bear in mind that in our analysis few load cases are compared. As such the plane of PC1 vs PC2 in Fig. 3 is a perfectly adequate space in which to view results and combine deformations. As more complex analyses are undertaken, combining many load cases, it is likely that more dimensions of the size and shape space will be needed to represent the range of deformations they produce, and, as such, vector additions to combine loads will need to take account of the full dimensionality, by using all PCs.

Our analyses provide some insights into modelling and the function of the third trochanter in the armadillo. In particular it was not possible to work with the full 3D geometry of the femur because imaging resolution was inadequate to accurately segment cancellous bone detail, although the medullary cavity could be reconstructed. This situation is likely to be much worse in fossils where matrix will limit image segmentation with the result that solid models will be required. We therefore assessed the impact of working with a solid model rather than one with a hollow representing the medullary cavity. The results are encouraging in that the solid model behaves very like the hollow; the overall pattern of strains is very similar (Fig. 2 and 4). In the GM analyses (Fig. 3 and 5), the solid model deforms along the same trajectory but is a little stiffer than the hollow model. This is likely because the deformations that have arisen from our loading scenarios consist in the main of pure bending of the shaft. Because bending resistance depends on the second moment of area, which in the case of cylinders depends on the square of the distance of material from the neutral axis (see Lieberman et al. 2004 for a straightforward account), adding more material inside the bone does not have a significant impact on bending resistance. The situation would likely be different for pure compressive loads, where cross sectional area is relevant. With regard to the impact of muscles acting on the third trochanter on "unbending" of the femoral shaft, the results show that the head force produces bending in the femoral shaft, and that the action of abductor muscles acting on the greater trochanter increases this bending (Fig. 2, 3 and 4). We have demonstrated that muscles pulling on the third trochanter can counter this bending (Fig. 4 and 5). Fig. 5 also demonstrates that third trochanter loads (A-C) alone bend the femoral model in the opposite direction to head and greater trochanter loads, and that this still occurs to varying degree over the full range of possible T3 muscle force directions.

Using GM methods to analyse deformations of landmark configurations arising from FEA, provides an account of deformation that is complementary, but by no means substitutes for strain based analyses. It leads to understanding and visualisation of larger scale aspects of deformation, but does not inform in relation to likely sites of failure of the bony tissue.

We have only considered how size and shape analyses might be applied to different loadings of the same bone but a common situation involves comparison of the effects of applying the "same" loadings to different bones. We have recently published such a comparison between the armadillo (body mass $\tilde{3}$ kg) femur described in this study and the femur of its giant extinct relative the 300 kg glyptodont (Milne and O'Higgins, 2012). Our approach was to use forces that produced similar strains and large scale degrees of bending in each and then to combine the coordinate data in a single size and shape analysis in order to compare large scale bending and "unbending". This was done by first scaling translating and rotating (GPA) equivalent landmark configurations from each bone to register the load cases for each model. The differences between the coordinates of the landmarks in each loaded state and the unloaded model, were then added to the mean unloaded model shape for visualisation and the results were rescaled according to the ratio of centroid sizes between loaded and unloaded to "restore" size changes due to loading. The resulting coordinates were then submitted to a size and shape PCA of both femora. The finding was that in both animals a similar unbending effect is observed but this is greater when the third trochanter is more distal. Such analyses open up the possibility of comparing the effects of loading among different specimens and of investigating covariations between loading response and other variables (e.g. skeletal form, ecology, phylogeny, etc).

There is a need for proper mathematical, engineering and statistical development of the approaches outlined in this paper but with this, and combined with the use of warping approaches to model building (O'Higgins et al., 2011, 2012; Parr et al., 2012; Pierce et al., 2008; Sigal et al., 2008; Sigal et al, 2010; Stayton, 2009), combinations of FEA, GM and modern imaging techniques should eventually lead to new analyses that experiment with and take account of variations in form and loading and so provide novel insights into how skeletal form relates to function, ecology and evolution.

References

Adams D.C., Cerney M.M., 2007. Quantifying biomechanical motion using Procrustes motion analysis. Journal of Biomechanics 40: 437–444.

Bookstein F.L., 2011. Speculations on the next statistical toolkit for complex organized systems. Proceedings of the 29th Leeds Annual Statistical Research Workshop. http://wwwl.maths.leeds.ac.uk/statistics/workshop/lasr2011/Proceedings/Bookstein.pdf

Dechow P.C., Nail G.A., Schwartz-Dabney C.L., Ashman R.B., 1993. Elastic properties of human supraorbital and mandibular bone. American Journal of Physical Anthropology 90(3): 291–306.

Dryden I.L., Mardia K.V., 1998. Statistical shape analysis. John Wiley, London.

Dryden I.L., Hirst J.D., Melville J.L., 2007. Statistical analysis of unlabeled point sets: comparing molecules in chemoinformatics. Biometrics 63: 237–251.

Duda G.N., Schneider E., Chao E.Y.S., 1997. Internal forces and moment in the femur during walking. Journal of Biomechanics 30: 933–941.

Fagan M.J., Curtis N., Dobson C.A., Karunanayake J.H., Kupczik K., Moazen M., Page L., Phillips R., O'Higgins P., 2007. Voxel-based Finite Analysis – Working Directly with MicroCT Scan Data. Journal of Morphology 268: 1071.

Fitton L.C., Shi J.F., Fagan M.J., O'Higgins P., 2012. Masticatory loadings and cranial deformation in *Macaca fascicularis*: a finite element analysis sensitivity study. Journal of Anatomy 221: 55–68.

Gröning F., Fagan M.J., O'Higgins P., 2011 The effects of the periodontal ligament on mandibular stiffness: a study combining finite element analysis and geometric morphometrics. Journal of Biomechanics 44: 1304–1312.

Gröning F., Fagan M., O'Higgins P., 2012. Modeling the Human Mandible Under Masticatory Loads: Which Input Variables are Important? The Anatomical Record 295(5): 853–863. doi:10.1002/ar.22455

Gupta K.K., Knoell A.C., Grenoble D.E., 1973. Mathematical modeling and structural analysis of the mandible. Biomaterials, Medical Devices and Artificial Organs 1: 469–479.

Hart R.T., Hennebel V.V., Thongpreda N., van Buskirk W., Anderson R.C., 1992 Modeling the biomechanics of the mandible: a three-dimensional finite element study. Journal of Biomechanics 25: 261–286.

Ichim I., Kieser J.A., Swain M.V., 2007. Functional significance of strain distribution in the human mandible under masticatory load: Numerical predictions. Archives of Oral Biology 52: 465–473.

Koneval T.O., 2003. Comparative hindlimb anatomy and fossoriality of three armadillos: *Dasypus novemcinctus*, *Tolypeutes matacus*, and *Chaetophractus vellerosus* (Mammalia, Xenarthra Cingulata, Dasypodidae). PhD thesis, University of Massachusetts.

Koolstra J.M., van Eijden T.M.G.J., 2005. Combined finite-element and rigid-body analysis of human jaw joint dynamics. Journal of Biomechanics 38: 2431–2439.

Korioth T.W., Hannam A.G., 1994. Deformation of the human mandible during simulated tooth clenching. Journal of Dental Research 73: 56–66.

Kupczik K., Dobson C.A., Fagan M.J., Crompton R.H., Oxnard C.E., O'Higgins P., 2007. Assessing mechanical function of the zygomatic region in macaques: validation and sensitivity testing of finite element models. Journal of Anatomy 210: 41–53.

Lieberman D.E., Polk J.D., Demes B., 2004. Predicting Long Bone Loading From Cross-Sectional Geometry. American Journal of Physical Anthropology 123: 156–171.

Liu J., Shi L., Fitton L.C., Phillips R., O'Higgins P., Fagan M.J., 2011. The application of muscle wrapping to voxel-based finite element models of skeletal structures. Biomechanics and Modeling in Mechanobiology 11: 35–37.

Milne N., Toledo N., Vizcaíno S.F., 2011. Allometric and group differences in the xenarthran femur. Journal of Mammalian Evolution 19(3): 199-208. doi:10.1007/s10914-011-9171-0

Milne N., O'Higgins P., 2012. Scaling of form and function in the xenarthran femur: a 100 fold increase in body mass is mitigated by repositioning of the third trochanter. Proc. R. Soc. B 279(1742): 3449–3456. doi:10.1098/rspb.2012.0593

Moazen M., Curtis N., Evans S.E., O'Higgins P., Fagan M.J., 2009. Biomechanical assessment of evolutionary changes in the lepidosaurian skull. Proceedings of the National Acadamy of Sciences, USA 20: 8273–8277.

O'Higgins P., Jones N., Ghattaura A., Hammond P., Hutton T., Carr M., 2002. Geometric morphometric approaches to the study of soft tissue growth and expression in the human face. American Journal of Physical Anthropology Supplement 34: 119.

O'Higgins P., Cobb S.N., Fitton L.C., Gröning F., Phillips R., Liu J., Fagan M.J., 2011. Combining geometric morphometrics and functional simulation: an emerging toolkit for virtual functional analyses. Journal of Anatomy 218: 3–15.

O'Higgins P., Fitton L., Phillips R., Shi J.F., Liu J., Groening F., Cobb S.N., Fagan M.J., 2012. Virtual functional morphology: novel approaches to the study of craniofacial form and function, Evolutionary Biology 39(4): 521–535. doi:10.1007/s11692-012-9173-8

Oxnard C.E., O'Higgins P., 2009. Biology Clearly Needs Morphometrics. Does Morphometrics Need Biology? Biological Theory 4(1): 1–14.

Panagiotopoulou O., Wilshin S.D., Rayfield E.J., Shefelbine S.J., Hutchinson J.R., 2012. What makes an accurate and reliable subject-specific finite element model? A case study of an elephant femur. J. R. Soc. Interface 9(67): 351–361. doi:10.1098/rsif.2011.0323

Parr W.C.H., Wroe S., Chamoli U., Richards H.S., McCurry M.R., Clausen P.D., McHenry C., 2012. Toward integration of geometric morphometrics and computational biomech-

anics: New methods for 3D virtual reconstruction and quantitative analysis of Finite Element Models. Journal of Theoretical Biology 301: 1–14.
Pauwels. 1980. Biomechanics of the locomotor apparatus: Contributions on the functional anatomy of the locomotor apparatus. Spinger-Verlag, Berlin.
Pierce S.E., Angielczyk K.D., Rayfield E.J., 2008. Patterns of morphospace occupation and mechanical performance in extant crocodilian skulls: a combined geometric morphometric and finite element modeling approach. Journal of Morphology 269: 840–865.
Rayfield E.J., 2005. Using finite-element analysis to investigate suture morphology: a case study using large carnivorous dinosaurs. Anatomical Record A283: 349–365.
Rybicki E.F., Simonen F.A., Weis E.B., 1972. On the mathematical analysis of stress in the femur. Journal of Biomechanics 5: 203–215.
Sigal I.A., Hardisty M.R., Whyne C.M., 2008. Mesh-morphing algorithms for specimen-specific finite element modelling. Journal of Biomechanics 41: 1381–1389.
Sigal I.A., Yang H., Roberts M.D., Downs J.C., 2010. Morphing methods to parameterize specimen-specific finite element model geometries. Journal of Biomechanics 43: 254–262.
Slice D.E., 1999. Geometric motion analysis. American Journal of Physical Anthropology Supplement 28: 253–254.
Slice D.E., 2003. The analysis of shape sequences. American Journal of Physical Anthropology Supplement 36: 194–195.
Stayton C.T., 2009. Application of Thin-Plate Spline Transformations to Finite Element Models, or, How to Turn a Bog Turtle into a Spotted Turtle to Analyze Both. Evolution 63: 1348–1355.
Strait D.S, Richmond B.G., Spencer M.A., Ross C.F., Dechow P.C., Wood B.A., 2007. Masticatory biomechanics and its relevance to early hominid phylogeny: An examination of palatal thickness using finite-element analysis. Journal of Human Evolution 52: 585–599.
Strait D.S., Weber G.W., Neubauer S., Chalk J., Richmond B.G., Lucas P.W., Spencer M.A., Schrein C., Dechow P.C., Ross C.F., Grosse I.R., Wright B.W., Constantino P., Wood B.A., Lawn B., Hylander W.L., Wang Q., Byron C.D., Slice D.E., Smith A.L., 2009. The feeding biomechanics and dietary ecology of *Australopithecus africanus*. Proceedings of the National Academy of Sciences USA 106: 2124–2129.
Taylor M.E., Tanner K.E., Freeman M.A.R., Yettram A.L., 1996 Stress and strain distribution within the intact femur: compression or bending? Medical Engineering and Physics 18: 122–131.
Vollmer D., Meyer U., Joos U., Vegh A., Piffko J., 2000. Experimental and finite element study of a human mandible. Journal of Cranio-Maxillofacial Surgery 28: 91–96.
Weber G.W., Bookstein F.L., Strait D.S., 2011. Virtual anthropology meets biomechanics. Journal of Biomechanics 44: 1429–1432.
Wroe S., Moreno K., Clausen P., McHenry C., Curnoe D., 2007. High-Resolution Three-Dimensional Computer Simulation of Hominid Cranial Mechanics. The Anatomical Record 290A: 1248–1255.

Associate Editor: A. Cardini

Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Volume 24 (1): 133–140, 2013
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/6363/pdf
doi:10.4404/hystrix-24.1-6363

Research Article

Shape descriptors as ecometrics in dental ecology

Alistair R. Evans[a,*]

[a] School of Biological Sciences, Monash University, VIC 3800, Australia and Mammalogy and Geosciences, Museum Victoria, Melbourne, VIC 3001, Australia

Keywords:
teeth
mammals
shape
morphology
complexity
OPC
SSFA
hypsodonty

Article history:
Received: 20 June 2012
Accepted: 6 February 2013

Acknowledgements
I would like to thank Andrea Cardini for inviting me to contribute to this special issue. I also thank the members of the Evans lab group for discussions on this topic, particularly Peter Smits, Mikael Fortelius for comments, and Andrea Cardini and Pasquale Raia for thoughtful reviews (and the latter for reminding me of the Gould quote). This work was supported by the Australian Research Council and Monash University.

Abstract

The revolution in morphometrics over the last 20 years has largely been in shape analysis methods that explicitly encode shape. These methods, which include Fourier outline shape analysis, Procrustes-based geometric morphometrics and eigenshape analysis, can be termed "shape specifiers". Despite their tremendous power in comparisons of shape, they do not give information about more general characteristics of shape that may be useful in interpreting function or ecology of an organism. "Shape descriptors" are computational representations of shape that can summarise high-level characteristics, such as overall shape or complexity. This paper describes a number of shape descriptors that have been used to capture specific morphological features of mammal teeth. Many of these dental shape descriptors have been valuable as "ecometrics", characteristics of organisms that reflect a species' ecology and can be used to reconstruct past environments. Shape descriptors can relate to the gross morphology or to the microwear texture of the tooth surface, as each of these have different characteristics and information regarding function and ecology. While this review concentrates on shape descriptors for teeth, it is hoped that they will give inspiration and stimulation to use and discover additional descriptors for other morphological systems.

Introduction

Shape is a fundamental attribute of organisms. We can think of shape as all that is left once size, translation and rotation are removed (Kendall, 1984). While this may sound simple, the scientific study and quantification of shape has been fraught with difficulties, both theoretical and practical. In the past, analysis of shape was often carried out using ratios of linear measures, or by angles. By themselves these capture relatively limited information about the morphology of interest, and ratios often do not remove the effects of scaling as expected (Atchley et al., 1976). With the greater availability of computers in the second half of the 20th century, multivariate morphometrics became the tool of choice for the study of shape by using multivariate statistical methods such as principal component analysis (PCA) to summarise variation in a large number of linear measures (Blackith and Reyment, 1971; Dryden and Mardia, 1998). This approach is now called "traditional morphometrics" (Rohlf and Marcus, 1993).

Great leaps forward have been made in the statistical analysis of shape in the last few decades. In modern morphometric methods, the geometry of the object is captured as outlines, landmarks, semi-landmarks, or a combination of these. The suite of methods includes Fourier outline shape analysis (Christopher and Waters, 1974; Haines and Crampton, 2000), eigenshape analysis (Lohmann, 1983; MacLeod, 1999; Figueirido et al., 2011), Procrustes-based geometric morphometrics (Rohlf and Marcus, 1993; Richtsmeier et al., 2002; Adams et al., 2004; Slice, 2007), spherical harmonics (Funkhouser et al., 2003; Shen et al., 2009), eigensurface analysis (Polly, 2008; MacLeod, 2008; Polly and MacLeod, 2008; Sievwright and MacLeod, 2012), 3D semi-landmark methods (Wood, 2011) and geometric similarity based on conformational geometry and optimal mass transportation (Boyer et al., 2011). Geometric morphometrics and eigenshape analyses utilise Euclidean projections of Kendall's shape space (Kendall, 1977, 1989; Dryden and Mardia, 1998). The intention of all of these methods is to preserve the geometry of the object and explicitly compare shape among objects, and so the shape of the original form can be recovered through these methods, at least as it is represented by the outline or landmarks. I will describe all of the above methods as "shape specifiers" where the intention is to specifically represent or encode the shape.

Despite the tremendous power that comes with these new methods, most of them are limited in the range of shapes that can be compared. This is particularly the case for those based on landmarks, in part because they require the same number of landmarks on all objects. This means that they cannot compare very dissimilar objects, or objects of different classes or types. It is not trivial to represent major differences between objects, such as the appearance or disappearance of structures, as these require changes in the number of landmarks. These limitations can partially be overcome (Klingenberg, 2008; Oxnard and O'Higgins, 2009), but this remains a major challenge in statistical shape analysis.

The analysis and comparison of shape can go beyond the specification of its geometry. In comparing shapes, there may be high-level characteristics other than the specific shape itself that we would like to quantify and compare. The desire to compare apples with chairs has led to a broader suite of methods for comparing shape. Shape can be categorised or quantified using transformations that do not retain position information but instead capture aspects of shape. We can call these computational representations of shapes "shape descriptors" (Christopher and Waters, 1974; Funkhouser et al., 2003). They are statistics about the shape without trying to encode the shape itself.

Shape descriptors have been used in computer science to index shapes based on their statistical properties. These can be for either 2D images or 3D models; I will largely concentrate on 3D shape descriptors here. They have been used to aid 3D shape matching and

*Corresponding author
Email address: arevans@fastmail.fm (Alistair R. Evans)

searching and the automatic shaped-based retrieval of 3D models from large databases or web searches (Funkhouser et al., 2003).

3D shape descriptors used in computer science include moments of inertia (Elad et al., 2001), Extended Gaussian Image (EGI, a spherical function giving the distribution of surface normals; Horn 1984), spherical extent function (EXT; Saupe and Vrani 2001), shape histograms (histogram of how much surface resides at different radii from centre of mass; Ankerst et al. 1999), Light Field Descriptor (LFD, a representation of a model as a collection of images rendered from uniformly sampled positions on a view sphere; Chen et al. 2003), Depth Buffer Descriptor (DBD, a collection of depth buffer images captured from orthogonal parallel projections; Heczko et al. 2002) and D2 shape distributions (probability distributions of geometric properties computer for points randomly sampled on an object's surface; Osada et al. 2002). Spectral shape analysis based on Laplace-Beltrami spectra includes Shape-DNA (Reuter et al., 2005, 2006), Global Point Signature (GPS; Rustamov 2007) and Heat Kernel Signature (HKS; Sun et al. 2009). I mention only global shape descriptors here, which consider the shape as a whole; most recent search methods use local shape descriptors as well, which only refer to parts or elements of the shape. These types of methods are used in the annual Shape Retrieval Contest (SHREC), held since 2006, e.g. SHREC'12 (Li et al., 2012). The usefulness of these types of descriptions can be seen in the creation of MPEG-7 standards for 2D- and 3D-shape descriptors, which could enable searching of video for specific shapes. See van Kaick et al. (2011) for a recent review of the field.

Shape descriptors allow for holistic and higher-level measures of morphology, including overall shape and complexity. Unlike shape specifiers, the original shape usually cannot be reconstructed from the shape descriptors. Individual or groups of shape descriptors can be a form of bar-coding or fingerprinting, and can potentially be used for copyright protection of 3D models (Reuter et al., 2005). The main contrast between specifiers and descriptors is that the former is a representation of shape, while the latter is an abstraction. In investigating shape descriptors for biological morphology, it is likely that the criteria for useful or good shape descriptors will be very different to those used in computer science mentioned above. This will very much depend on the type of questions being addressed.

Dental ecology

A lot can be gleaned about a mammal just from its teeth. They are key to food acquisition and processing in most mammals, and their shape has substantial influence on their ability to carry out these functions. As well as being useful for taxonomic identification, often showing species-level variation, they are the most frequently preserved component in the fossil record. Steven Jay Gould, well-versed in teeth, once quipped that "mammalian evolution is a tale told by teeth mating to produce slightly altered descendant teeth" (Gould 1989: 60). Teeth are so important to our understanding of various aspects that the term "dental ecology" has been coined, referring to the study of how teeth respond to the environment (Cuozzo and Sauther, 2012). Analysis of tooth shape should therefore give information on food types that the species has adapted to consume, as well as the effect of wear over the lifetime of the animal.

Mammal teeth exhibit a surprising diversity of shapes (Ungar, 2010), which is both a blessing and a curse for shape analysis. The variability in number and position of cusps, crests and basins, as well as the effects of wear, mean that the use of homologous landmarks to specify the positions of these features on the surface is not possible when comparing a functionally and/or phylogenetically wide range of teeth. This led to the use of shape descriptors to quantify and compare tooth shapes, from the intraspecific (Zuccotti et al., 1998; Ungar and Williamson, 2000) to the order level (Evans et al., 2007b). Mammal teeth are now an important model system for investigating shape descriptors in biological morphology.

Certain characteristics of teeth increase the difficulty of assessing shape compared to other systems. During the life of a mammal, its teeth will generally wear and change shape in the process. The initially-erupted primary occlusal morphology, where the entire crown is covered in enamel, can be worn to produce a secondary occlusal morphology, resulting in a series of dentine basins surrounded by enamel. The resulting change in shape can have a significant effect on the function of the teeth, and means that assessment of morphology should consider the effect of wear.

Landmark-based geometric morphometrics has been applied to some questions of tooth morphology (e.g. Ungar et al. 1994; Hlusko 2002; Skinner et al. 2008; Piras et al. 2010; Singleton et al. 2011), but in all cases the comparisons are very limited phylogenetically and morphologically.

Ecometrics

Due to their functional importance, teeth often reflect critical characteristics of a species' ecology. Because of this, they can be used to interpret past climates given modern associations between characteristics and modern climates. Taxon-free functional trait analysis, or "ecometrics" (Eronen et al., 2010c; Polly et al., 2011), is the association of such characteristics with climate or environment. Ideally such traits are largely independent of taxonomy, such that they can be measured on a wide group of species within a larger taxonomic group. Ecometrics has been used to give a deep-time perspective on climate change using traits such as leaf shape (Wolfe, 1995), body size (Legendre, 1986) and mammal locomotion (Polly, 2010). Several aspects of tooth shape have or could be used as ecometrics.

Dental Shape Descriptors

As teeth are such a favourite of mammal palaeontologists, there has been a veritable explosion of proposed methods to quantify their shape. Here I briefly survey a range of shape descriptors that have been useful in functional interpretations of tooth shape, in comparative study across a broad taxonomic range, and as ecometrics for dietary and paleoenvironmental reconstruction.

These descriptors vary greatly in their ease of measurement, such that some can be measured using callipers while others need full 3D surface data. However, measurement simplicity is not necessarily related to their usefulness in functional or ecometric analysis. While simple shape descriptors such as hypsodonty capture very little information of the shape of the tooth, they are still powerful ecometrics.

Some of the shape descriptors here will be only applicable to teeth (e.g. shear ratio and mesowear), while others will be more generally applicable to a wide range of surfaces (e.g. OPC and microtexture analysis). One purpose of this review is to show the great variety of approaches for assessing shape in teeth; another is to show biologists working on other morphological systems the types of aspects that may be applicable to their work – to give inspiration for new ways of looking at morphology.

Gross topography

The chewing, or occlusal, surfaces of mammal teeth vary dramatically, from a few simple bumps to washboards to blades. These measures of topography intend to quantify aspects from overall form to the shape of specific regions of the teeth.

Ratios and Angles

Several simple measures have been extensively used to gauge overall tooth shape. Molar length-width ratios are common, particularly in physical anthropology, and can help in assigning a specimen to a taxon. Relative crown height, or hypsodonty, is measured as height:length or height:width ratios (Simpson, 1953; Van Valen, 1960; Janis, 1988; Fortelius et al., 2002; Damuth and Janis, 2011), and is an important ecometric correlated with precipitation (Fortelius et al., 2002; Eronen et al., 2010a,b; Liu et al., 2012; Raia et al., 2011). In carnivores, carnassial tooth shape has been quantified by the angle α, which relates the height of the protoconid relative to the length of the talonid (Crusafont-Pairó and Truyols-Santonja, 1956). This character is indicative of the level of carnivory (Wesley-Hunt, 2005) and can reveal the evolution of trophic position in carnivores (Meloro and Raia, 2010).

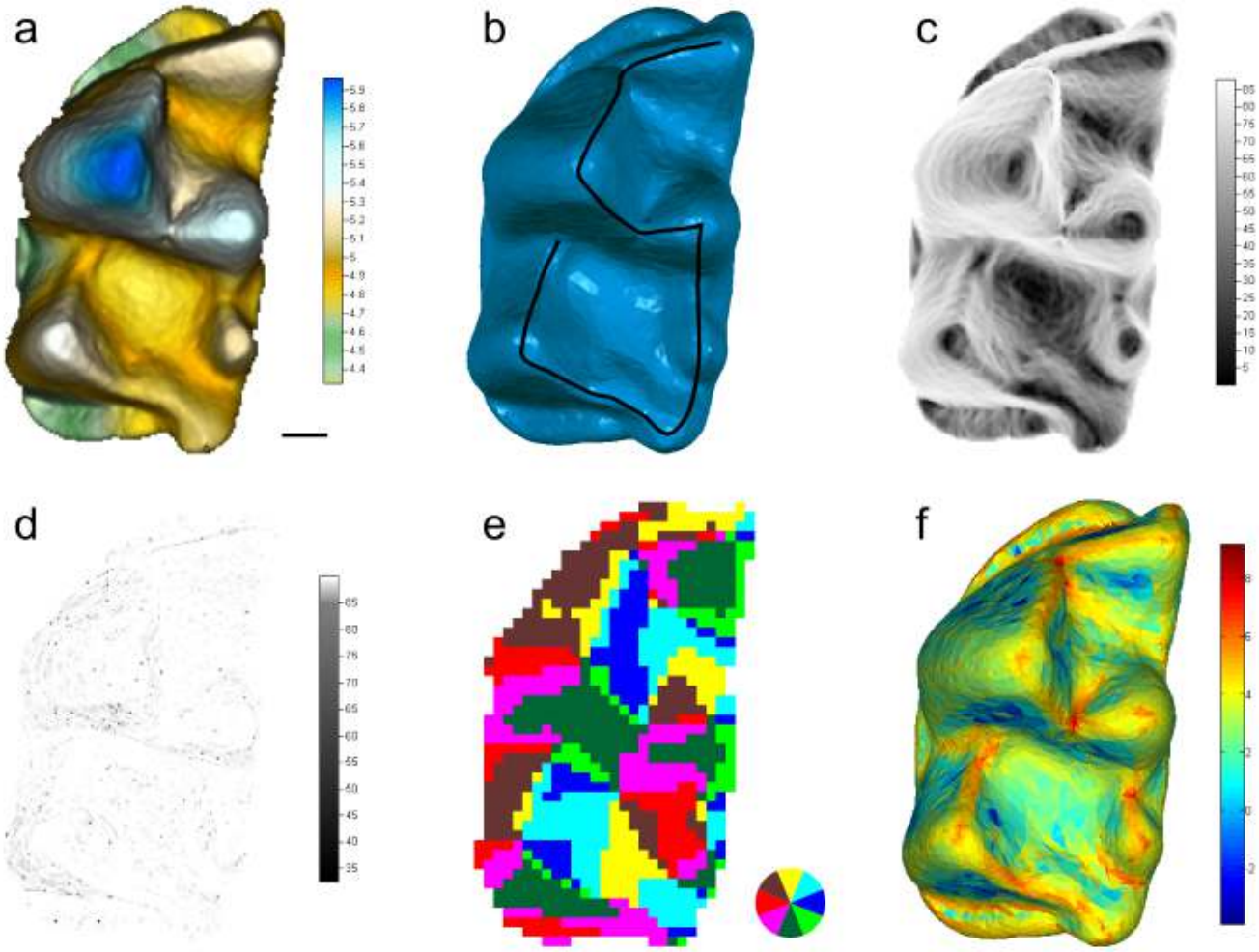


Figure 1 – Comparison of a number of gross morphology shape descriptors outlined in this review, illustrated using the dasyurid marsupial *Antechinus agilis* NMV C12676 (Museum Victoria) left lower second molar in occlusal view. a) Height-encoded map (height in mm); b) shearing edges measured for calculation of shear ratio; c) slope (in degrees); d) angularity (in degrees); e) aspect map for OPC calculation (colour wheel shows orientation); f) Dirichlet normal energy. Mean slope: 55.94; mean angularity: 89.720; Relief index 2D (M'Kirera and Ungar, 2003): 2.75; Relief index 3D (Boyer, 2008): 0.66; OPCR: 50.25; DNE: 508.39. Number of surface points: 18853 (c), 18326 (d), 980 (e) and 9995 (f). Number of anterior-posterior data rows: 50 (e). Illustrations made using Surfer (a, c, d), Geomagic (b), Surfer Manipulator (e) and Teether (f). Scale bar = 0.2 mm.

While ratios can reveal some aspects of morphology, they give no information on the outline or topography of the tooth. More specific measures of the shape of tooth components can reveal relative function of cusps and crests (Evans and Sanson, 1998, 2003). Rake angle, approach angle and edge sharpness are key shape descriptors for crest function that describe the orientation of the cutting edge and its surrounding surface relative to the direction of movement (Evans and Sanson, 2003, 2005). Changes in these factors will affect the performance of the teeth for fracturing viscoelastic foods, and can therefore quantify relative performance in unworn and worn teeth.

Shear Ratio

Kay (1975, 1978) developed one of the earliest shape descriptors for dental topography that is useful in assessing diet and ecology of primates. "Shearing quotient" and "shearing ratio" measure the relative lengths of shearing crests on the surface of a tooth (e.g. Fig. 1b). They have been extensively used in interpreting diets of placentals (see Bunn et al. 2011 for review) and marsupials (Hogue and ZiaShakeri, 2010). Disadvantages of shear ratio include the requirement of specifying the crests to measure, and its inability to deal with variable wear states (Evans and Sanson, 1998).

Mesowear

Wear on teeth during the lifetime of an animal often has a substantial effect on their shape. In many instances this has a detrimental effect on the function of the teeth (Evans, 2005; King et al., 2005). But the way in which tooth shape changes during the wear sequence is greatly influenced by the relative amounts of attrition (tooth-tooth) and abrasion (tooth-food) wear. The former tends to produce planar facets with sharp edges, while the latter causes rounding of tooth surfaces. This difference has been useful when examining wear in animals with secondary occlusal morphology, such as artiodactyls and perissodactyls, where the functional surface results from and is maintained by wear (Fortelius, 1985). If attrition dominates, as it tends to do in browsers, the resulting crests are sharp with high relief, while with increasing amounts of abrasion, found in grazers, relief is lower and the crests are rounded. With sufficient sample size for a species or population, this signature of mesowear shows a good association with dietary classification (Fortelius and Solounias, 2000).

Kaiser et al. (2013) comprehensively reviewed the associations between hypsodonty and mesowear. They found that mesowear largely indicates diet (percentage of grass in natural diet), while hypsodonty also includes effects from the environment (mean annual precipitation and openness of habitat) as well as diet.

Crown Types

To enable a broad-scale morphological comparison of teeth, Jernvall (1995; Jernvall et al. 1996) developed a topological system for categorising dental shape called "crown types". The crown type is a four-digit code counting the number of buccal and lingual cusps, and the number of longitudinal and transverse lophs. A significant feature of this method is that it assessed both developmental and functional aspects of shape, as cusps result from the folding during development of the mesenchyme-epithelium interface.

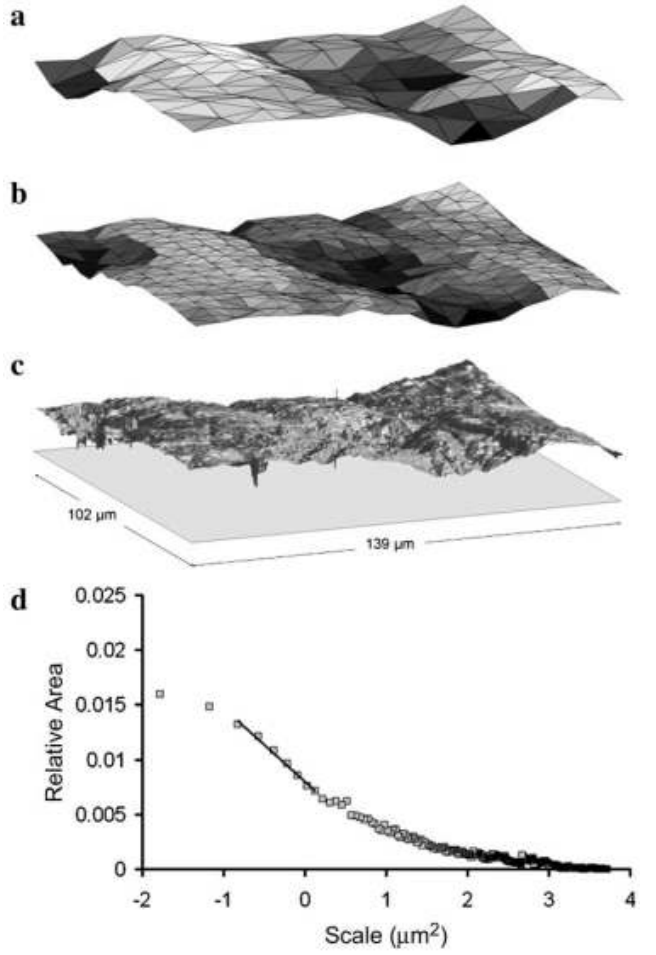


Figure 2 – Area-scale analysis in SSFA. A virtual tiling algorithm using triangles of different sizes can be used to measure surface roughness (compare a, b, and c). Complexity is represented by the steepest part of a curve fitted to the plot of relative area over scale (d). From Scott et al. (2006).

Jernvall et al. (1996) used crown types to show the difference between the Eocene and Miocene radiations in ungulate communities, with the former having higher diversity in tooth shape but lower disparity, while the latter displayed an increase in the number of lophs to presumably deal with lower quality vegetation. Together with hypsodonty, the number of longitudinal lophs highly correlates with net primary productivity (Liu et al., 2012).

Relief Index

Real-world objects are always a three-dimensional volume surrounded on all sides by a surface. However, the functional surface of teeth can largely be viewed from a single direction, as in a photograph. This type of projection of a 3D surface onto a 2D plane we can call a 2.5D surface, where for every position on an xy grid there is only one surface, at height z. Undercuts, where one part of the surface curves underneath another part, cannot be represented in 2.5D. As the crown of the tooth passes into the gingiva, the tooth tapers underneath the crown into the root or roots. However, the majority of the crown above the cervical region can be represented by 2.5D. There are many advantages to this simpler representation. Computation of surface characteristics is much simpler, including Geographic Information Systems (GIS) algorithms, which are often based on raster (gridded) 2.5D data. The use of GIS and tooth digital elevation models (DEMs) began in the late 1990s (Reed, 1997; Hunter and Jernvall, 1998; Zuccotti et al., 1998; Jernvall and Selänne, 1999; Evans et al., 2001) and has been termed "dental topographic analysis" (Ungar and Williamson, 2000).

A number of measures were first developed by Ungar and Williamson (2000) to describe shape. Relief index is the ratio of the 3D surface of a tooth to its projected 2D area. In the initial formulations, 3D surface was calculated for a 2.5D crown surface, which may only extend down the sides of the tooth to the same level as the bottom of the lowest basin (Ungar and Williamson, 2000). However, it is also possible to measure 3D surface for the full crown of the tooth (Boyer, 2008; Bunn et al., 2011). Relief index is usually considered to measure overall height of the crown or component cusps, but there is the potential for height and complexity to be confounded in the measurement of relief index (Plyusnin et al., 2008). Tooth wear reduces relief index in most species, as the height of cusps is worn down, and so wear state should be taken into account when examining relief index.

M'Kirera and Ungar (2003) calculated relief index as SA/PA, where SA is 3D surface area and PA is 2D planar area. Boyer (2008) measured relief index as $\ln(\sqrt{SA}/\sqrt{PA})$.

Relief index has been shown to differentiate diets in primates (Boyer et al., 2010; Godfrey et al., 2012; Bunn et al., 2011). Ulhaas et al. (2004, 2007), Dennis et al. (2004) and Klukkert et al. (2012) also use relief index to assess differences in relief among species and with wear.

Average Slope and Angularity

The slope can be measured at each point on a surface using GIS algorithms (Fig. 1c). An overall measure of tooth shape can be made using the average slope (Ungar and Williamson, 2000; M'Kirera and Ungar, 2003; Klukkert et al., 2012).

Angularity is the average rate of change of surface slope, calculated as the slope of the slope map (Ungar and Williamson, 2000), and so is a measure of sharpness of the edges (Fig. 1d). Despite changes in slope following wear, angularity is seemingly robust to wear (Ungar and M'Kirera, 2003).

Orientation Patch Count (OPC)

GIS can also be used to calculate the orientation or aspect of the surface at all grid points of a 2.5D surface, either as an angle from a fixed direction (e.g. y-axis or north), or as cardinal and ordinal directions (e.g. north, south-west; Fig. 1e). Adjacent grid points that are facing the same direction can be grouped together using GIS clumping procedures into a "patch". The number of these patches over the surface gives the "orientation patch count" or OPC (Evans et al., 2007b). To account for differences in size and scanning resolutions, all specimens within an analysis are standardised to a given grid length, such as 150 grid rows for a tooth row, or 50 grid rows per tooth. The effect of the positioning of the tooth on the grid can also be mitigated by rotating the orientation boundaries and recalculating OPC. The resulting value can be termed OPCR (Wilson et al., 2012).

This method was designed to give an automated quantification of dental complexity along the lines of crown typing by removing the subjective classification of which cusps or features should be counted. For many tooth forms, OPC measurement is robust to wear (Evans et al., 2007a), likely indicating a maintenance of function throughout much of the wear sequence.

Dental complexity has been used in carnivorans, rodents (Evans et al., 2007b), primates (Boyer et al., 2010; Godfrey et al., 2012; Bunn et al., 2011), bats (Santana et al., 2011), multituberculates (Wilson et al., 2012), dasyurids (Smits and Evans, 2012) and in developing teeth (Harjunmaa et al., 2012).

Section Area and Convolution

The topography of the tooth can be quantified by examining the relative area and degree of folding of contour lines in the xy plane (Plyusnin et al., 2008). The tooth is sectioned 10 times in the xy plane, and the area and perimeter of the resulting cut surface are calculated. The area measurements are standardised by the total xy area, and convolution is the length of the perimeter of all parts of the section divided by the square root of the area of that section. Convolution measures the perimeter:area ratio and quantifies the degree of folding in the shape, such that it increases as the contour is increasingly folded. Both area and convolution can be used to give a profile of the shape of the tooth from crown to base, or each section can be used separately as a shape descriptor.

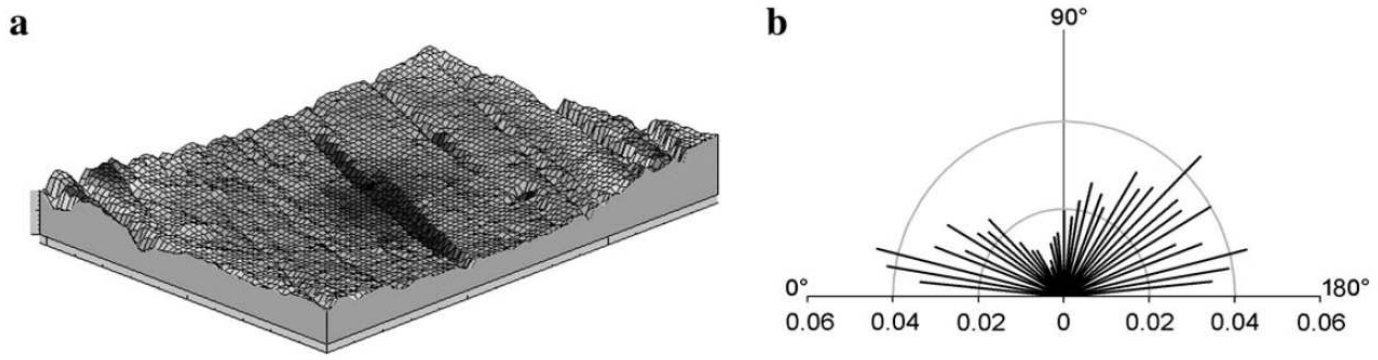


Figure 3 – Three-dimensional rendering of a striated surface (a) and the corresponding rosette plot of relative lengths taken at 36 different orientations (b). From Scott et al. (2006).

D2dist

Osada et al. (2002) conceived the descriptor D2, which measures the distance between two randomly chosen points on the surface of a model. The distribution of D2 distances was shown to be effective to distinguish a wide range of object classes. The descriptor was modified as D2dist by Plyusnin et al. (2008) for 2.5D surfaces such that the two points are chosen at random with respect to the xy plane. Plyusnin et al. (2008) used the mean and standard deviation of D2dist as shape descriptors.

Dirichlet normal energy (DNE)

Dirichlet normal energy was first used in dental topographic analysis by Bunn et al. (2011). It is based on the Dirichlet energy of the normal map of a surface, and quantifies the deviation of a surface from a plane (Fig. 1f). As a continuous function it is equivalent to measuring the sum of squares of the principal curvatures over the surface.

DNE has the advantages of being independent of position, orientation and scale. It gives an overall measure of curvature at crests and flatness of faces. Higher DNE may be the result of taller tooth features (giving larger flat faces), and is highly correlated with relief index. It reflects a change in both the height and the curvature of cusps/crests.

Microwear Surface

Surface shape extends in scale down to nanometre-level variations in surface height. Activities such as food acquisition, processing and grooming create fine use-wear features on the tooth surface called microwear. Analysis of microwear began in the late 1970s (Walker et al., 1978; Rensberger, 1978; Teaford, 1988), and showed that it was useful in distinguishing diet and/or environment. Until around 10 years ago, microwear analysis was largely been carried out by assessing the relative density and size of wear features such as pits and scratches from 2D SEM micrographs or light microscopy. Although Boyde and Fortelius (1991) suggested the use of 3D methods to examine microwear surfaces, 3D quantification did not commence until Ungar et al. (2003; Scott et al. 2006), who termed it dental microwear texture analysis. The two main sets of topographical measures have been used are described below.

Scale-sensitive fractal analysis

The initial methods to quantify dental microwear were based on scale-sensitive fractal analysis (Ungar et al., 2003; Scott et al., 2005). SSFA relies on the fractal geometry of natural objects, where the measurement of quantities such as length and area depend on the scale of measurement (Mandelbrot, 1967). Several variables have been defined that quantify different aspects of this relationship (Scott et al., 2006).

A surface can be represented by large or small triangles, giving the "scale" at which the surface is measured. When the area of a surface is measured at increasingly small scales, the measured surface area increases as smaller and smaller features are included in the measurement. Area-scale fractal complexity (Asfc) measures the steepest section of a log-log plot of scale vs. measured area (Fig. 2). The scale at which this steepest relationship occurs is called the scale of maximum complexity (Smc). Heterogeneity of the surface can be quantified by measuring Asfc for varying subregions of the surface and calculating the relative variation from the median (HAsfc).

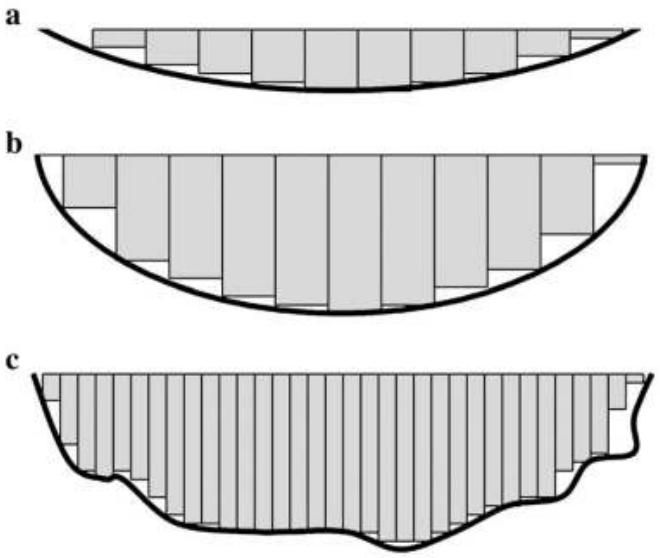


Figure 4 – Schematic comparing surfaces with (a) lower and (b) higher structural fill volumes. Finer scale prisms (c) yield structural and textural fill volumes. Textural fill volume is calculated by subtracting (b) from (c). From Scott et al. (2006).

The texture of a surface can have directionality, such as scratches in one direction. Profiles taken across the surface will vary in their length depending on whether the profile runs parallel to these scratches (and so will be flat) or perpendicular (and so will be zig-zagged). The normalized or "exact proportion" of the relative lengths can be used in a rosette diagram to visualise the degree of anisotropy (Fig. 3), and the mean vector length is defined as the exact proportion length-scale anisotropy of relief (epLsar). A higher value indicates that more wear features are perpendicular to that direction.

Variation in surface shape can also be measured by filling the volume with square cuboids of different sizes (Fig. 4). The general shape of the surface can be captured by larger cuboids, with a base length of 10 μm, measured as structural fill volume (Sfv). Using smaller cuboids (2 μm base length) will fill the finer texture of the surface, and is called the texture fill volume (Tfv).

There has been an explosion of papers using SSFA in the last three years, particularly on primates (Scott et al., 2009; Calandra et al., 2012; Merceron et al., 2009; Pontzer et al., 2011; Scott et al., 2012; Ungar et al., 2012a). SSFA has been used in other groups ranging from, bovids (Scott, 2012), cervids (Merceron et al., 2010), tragulids (Ungar et al., 2012b), macropodids (Prideaux et al., 2009) and carnivorans (Schubert et al., 2010; Stynder et al., 2012).

ISO surface texture parameters

Dental surface texture has also been investigated using two different ISO standards. The first set, from ISO 4287 (1997), quantifies roughness based on 2D surface profiles. It includes variables such as maximum (Rp), minimum (Rv), roughness average (arithmetic mean deviation; Ra), fractal dimension (Rfd), asymmetry (Rsk), kurtosis (Ruk), density of peaks (RHSC), and root mean square deviation (Rq) of each profile. Kaiser and Wolff (2005) were able to distinguish folivorous primates from gramnivores based on subsets of these variables, and Kaiser and Brinkmann (2006) found a separation of grazers and browsers in bovids.

The second is a collection of standards for the measurement of 3D surface texture (ISO 25178-2 2007). These relate to aspects such as statistics of the height of the surface (maximum Sz, skewness Ssk, kurtosis Sku), the bearing area curve (cumulative probability density function; e.g. volume Vmp, Vmc), spatial parameters such as autocorrelation (Sal) and texture aspect ratio (Str) and direction (Std, Stdi).

Schulz et al. (2010) and Calandra et al. (2012) compared the ISO standards with SSFA, illustrating that the general results concur between the two methods, and they are able to distinguish diets of differing properties. Purnell et al. (2012) were able to use the ISO standards to discriminate diets in cichlid fish.

Software

Much of the software for measuring the 3D shape descriptors discussed here was written by the researchers for the explicit purpose of quantifying dental morphology, while the remaining is commercial software that carries out some or all required functions. Once a surface has been digitised (gross morphology by computed tomography (CT), laser surface or structured light scanning; microwear surfaces by confocal microscopy or interferometry), file format conversions are usually required from the raw data (which may be as point clouds or polygons) to a format accepted by the analysis software.

Lengths, angles, 2D and 3D areas and volumes (to calculate descriptors such as relief index) can be measured from polygon models using commercial software such as Geomagic (Geomagic USA, Morrisville, NC, USA), RapidForm (3D Systems Corp., Rock Hill, SC, USA), PolyWorks (InnovMetric Software Inc., Quebec, QC, Canada) and SolidWorks (Dassault Systèmes SolidWorks Corp., Waltham, MA, USA). Some measurements can be made using the open source software Meshlab (http://meshlab.sourceforge.net). Avizo and Amira (Visualization Sciences Group, Burlington, MA, USA) can calculate these measurements from CT data (e.g. image stacks). GIS software appropriate for some shape descriptors includes Surfer (Golden Software, Inc., Golden, CO, USA) and ArcView (ESRI Corp., Redlands, CA, USA). Surfer Manipulator is custom Visual Basic software for OPC analysis that integrates with Surfer (http://users.monash.edu.au/~arevans/software.html). ToothKit is Java software to calculate descriptors such as section area, convolution and D2dist (http://www.biocenter.helsinki.fi/bi/evodevo/toothkit/index.shtml). Teether is a MATLAB package for calculating DNE (available from Julia Winchester julia.m.winchester@gmail.com). SSFA is carried out using Toothfrax and Sfrax software (Surfract Corp.). The ISO 25178-2 2007 surface standards can be measured by the software accompanying the 3D scanner, such as µsoft analysis premium software (NanoFocus AG, Oberhausen, Germany; a derivative of Mountains Analysis software by Digital Surf, Besançon, France) and Alicona IFM software (Alicona UK Ltd, Kent, UK).

Discussion

For the large number of shape descriptors covered above, several have shown great usefulness as ecometrics in mammalian biology. In terms of gross morphology, those most currently in use are relief index and dental complexity, but it is likely that new descriptors such as Dirichlet normal energy will be more broadly applied in the near future.

As a method for quantifying morphology, shape descriptors will find a much broader use in biology and palaeontology, ranging from phylogenetics, population-level variation and correlation with genetic and developmental architecture. This is due to their flexibility in being able to represent a very wide range of morphologies, and so being able to compare within and even among disparate systems. The challenge of applying these types of techniques in the future will be to ensure that any descriptors or application of them keeps in mind relevance to a biological question or hypothesis.

One recent example of the application of dental ecometrics is that of Wilson et al. (2012) in exploring the patterns of evolutionary and ecological change in the Multituberculata. The multituberculates were the most diverse and long-lived Mesozoic mammals, ranging in body mass from about 6 g to 20 kg. Because there are no modern members of the group, it has been difficult to find an appropriate analogue to assess their palaeoecology. Using OPC as a measure of dental complexity, Wilson et al. (2012) were able to show that early multituberculates were carnivorous or omnivorous in their dietary habits, but underwent an adaptive radiation around 85 million years ago into a herbivorous niche. This was corroborated by an increase in body mass, likely to be advantageous in plant feeding, and coincided with the ecological rise of angiosperms. The shape descriptor of dental complexity provided a scale- and phylogeny-independent measure of trophic position that revealed the broad-scale evolutionary patterns in this important group of mammals.

At the moment there is no overlap in the shape descriptors applied to gross morphology and microwear surfaces. There are fundamental differences between these two: the former tend to have higher relief and fewer features, while the latter are often highly repetitive and vary over the entire tooth surface. It is therefore not surprising that different sets of shape descriptors have been useful for each. The shape descriptors for gross topology are largely not applicable to microsurface textures due to much lower relief and larger number of features in the latter. It is interesting to consider, however, the degree of variability in the two types of surfaces depending on the wear state of the tooth, and to what degree the various shape descriptors are insensitive or robust to it. Depending on the application, either of these may be required.

Another feature of surface microwear is that it is an acquired "morphology" as opposed to the primary morphology resulting from folding of the mesenchyme-epithelium interface and subsequent deposition of an enamel layer. Once wear has commenced, gross topology can also be primary-derived (Evans et al., 2005) or secondary (Fortelius, 1985). Genetic determination of these morphologies therefore varies from full (primary morphology), partial (worn gross morphology, as illustrated by mesowear signature and other patterns of differential wear, such as the different carnassial forms of canids and felids; Evans et al. 2005), or none (microwear). Each of these, therefore, gives different levels of ecometric information, from days or weeks (microwear), years (mesowear) to generations (gross morphology). Only through integration of these varying time scales will we have confidence in evolutionary and environmental signals we may detect.

Despite the name "shape descriptors", many do not strictly remove size in the same way as centroid size standardization used in Procrustes-based geometric morphometrics. Generally, size is controlled for in the topographic measures based on 3D surface data by the use of a fixed number of polygons or data rows to represent the surface (e.g. relief index, OPC). However, the ISO microwear texture variables are often absolute, giving measurements in micrometers so size has not been removed at all. Several of the SSFA variables also include some aspects of scale. In principle, fractal dimension should be independent of size, but biological surfaces are not self-similar at all scales and so the range of measurement lengths will influence the measurement of fractal dimension.

As well as those with already established usefulness as ecometrics, such as hypsodonty, OPC and SSFA, I anticipate that a number of others described above, or to yet be discovered, will give further insights into the evolution of mammalian morphology and ecology.

References

Adams D.C., Rohlf F.J., Slice D.E., 2004. Geometric morphometrics: ten years of progress following the "revolution". It. J. Zool. 71: 5–16.

Ankerst M.K.G., Kriegel H.-P., Seidl T., 1999. Nearest neighbor classification in 3D protein databases. Proc. 7th International Conference on Intelligent Systems for Molecular Biology: 34–43.

Atchley W.R., Gaskins C.T., Anderson D., 1976. Statistical properties of ratios. 1. Empirical results. Syst. Zool. 25(2): 137–148. doi:10.2307/2412740

Blackith R.E., Reyment R.A., 1971. Multivariate Morphometrics. Academic Press.

Boyde A., Fortelius M., 1991. New confocal LM method for studying local relative microrelief with special reference to wear studies. Scanning 13: 429–430.

Boyer D.M., 2008. Relief index of second mandibular molars is a correlate of diet among prosimian primates and other euarchontan mammals. J. Hum. Evol. 55(6): 1118–1137. doi:10.1016/j.jhevol.2008.08.002

Boyer D.M., Evans A.R., Jernvall J., 2010. Evidence of dietary differentiation among late Paleocene–early Eocene plesiadapids (Mammalia, Primates). Am. J. Phys. Anthropol. 142(2): 194–210. doi:10.1002/ajpa.21211

Boyer D.M., Lipman Y., St Clair E., Puente J., Patel B.A., Funkhouser T., Jernvall J., Daubechies I., 2011. Algorithms to automatically quantify the geometric similarity of anatomical surfaces. Proc. Natl Acad. Sci. USA 108(45): 18221–18226. doi:10.1073/pnas.1112822108

Bunn J.M., Boyer D.M., Lipman Y., St Clair E.M., Jernvall J., Daubechies I., 2011. Comparing Dirichlet normal surface energy of tooth crowns, a new technique of molar shape quantification for dietary inference, with previous methods in isolation and in combination. Am. J. Phys. Anthropol. 145(2): 247–61. doi:10.1002/ajpa.21489

Calandra I., Schulz E., Pinnow M., Krohn S., Kaiser T.M., 2012. Teasing apart the contributions of hard dietary items on 3D dental microtextures in primates. J. Hum. Evol. 63: 85–98. doi:10.1016/j.jhevol.2012.05.001

Chen D.-Y., Tian X.-P., Shen Y.-T., Ouhyoung M., 2003. On visual similarity based 3D model retrieval. Computer Graphics Forum 22(3): 223–232. doi:10.1111/1467-8659.00669

Christopher R.A., Waters J.A., 1974. Fourier-series as a quantitative descriptor of miospore shape. J. Paleontol. 48(4): 697–709.

Crusafont-Pairó M., Truyols-Santonja J., 1956. A biometric study of the evolution of fissiped carnivores. Evolution 10: 314–332.

Cuozzo F.P., Sauther M.L., 2012. What is dental ecology? Am. J. Phys. Anthropol. 148(2): 163–170. doi:10.1002/ajpa.21656

Damuth J., Janis C.M., 2011. On the relationship between hypsodonty and feeding ecology in ungulate mammals, and its utility in palaeoecology. Biol. Rev. 86(3): 733–58. doi: 10.1111/j.1469-185X.2011.00176.x

Dennis J.C., Ungar P.S., Teaford M.F., Glander K.E., 2004. Dental topography and molar wear in *Alouatta palliata* from Costa Rica. Am. J. Phys. Anthropol. 125(2): 152-61. doi: 10.1002/ajpa.10379

Dryden I.L., Mardia K.V., 1998. Statistical Shape Analysis. John Wiley & Sons.

Elad M., Tal A., Ar S., 2001. Content based retrieval of VRML objects – An iterative and interactive approach. Eurographics Multimedia Workshop. 97–108.

Eronen J.T., Puolamaki K., Liu L., Lintulaakso K., Damuth J., Janis C., Fortelius M., 2010a. Precipitation and large herbivorous mammals II: application to fossil data. Evol. Ecol. Res. 12(2): 235–248.

Eronen J.T., Puolamaki K., Liu L., Lintulaakso K., Damuth J., Janis C., Fortelius M., 2010b. Precipitation and large herbivorous mammals I: estimates from present-day communities. Evol. Ecol. Res. 12(2): 217–233.

Eronen J.T., Polly P.D., Fred M., Damuth J., Frank D.C., Mosbrugger V., Scheidegger C., Stenseth N.C., Fortelius M., 2010c. Ecometrics: the traits that bind the past and present together. Integr. Zool. 5(2): 88–101. doi:10.1111/j.1749-4877.2010.00192.x

Evans A., Fortelius M., Jernvall J., 2007a. How does tooth wear affect dental complexity? Implications for tooth function and dietary reconstruction. J. Vert. Paleontol. 27(3): 72A.

Evans A.R., 2005. Connecting morphology, function and tooth wear in microchiropterans. Biol. J. Linn. Soc. 85(1): 81–96.

Evans A.R., Sanson G.D., 1998. The effect of tooth shape on the breakdown of insects. Journal of Zoology 246(4): 391–400.

Evans A.R., Sanson G.D., 2003. The tooth of perfection: functional and spatial constraints on mammalian tooth shape. Biol. J. Linn. Soc. 78(2): 173–191. doi:10.1046/j.1095-8312.2003.00146.x

Evans A.R., Sanson G.D., 2005. Correspondence between tooth shape and dietary biomechanical properties in insectivorous microchiropterans. Evol. Ecol. Res. 7(3): 453–478.

Evans A.R., Harper I.S., Sanson G.D., 2001. Confocal imaging, visualization and 3-D surface measurement of small mammalian teeth. J. Microsc. 204(2): 108–118. doi: 10.1046/j.1365-2818.2001.00939.x

Evans A.R., Hunter J., Fortelius M., Sanson G.D., 2005. The scaling of tooth sharpness in mammals. Ann. Zool. Fenn. 42(6): 603–613.

Evans A.R., Wilson G.P., Fortelius M., Jernvall J., 2007b. High-level similarity of dentitions in carnivorans and rodents. Nature 445(7123): 78–81. doi:10.1038/nature05433

Figueirido B., MacLeod N., Krieger J., De Renzi M., Pérez-Claros J.A., Palmqvist P., 2011. Constraint and adaptation in the evolution of carnivoran skull shape. Paleobiol. 37(3): 490–518. doi:10.1666/09062.1

Fortelius M., 1985. Ungulate cheek teeth: developmental, functional, and evolutionary interrelations. Acta Zool. Fenn. 180: 1–76.

Fortelius M., Solounias N., 2000. Functional characterisation of ungulate molars using the abrasion-attrition wear gradient: a new method for reconstructing paleodiets. Amer. Mus. Nov. 3301: 1–36.

Fortelius M., Eronen J., Jernvall J., Liu L.P., Pushkina D., Rinne J., Tesakov A., Vislobokova I., Zhang Z.Q., Zhou L.P., 2002. Fossil mammals resolve regional patterns of Eurasian climate change over 20 million years. Evol. Ecol. Res. 4(7): 1005–1016.

Funkhouser T., Min P., Kazhdan M., Chen J., Halderman A., Dobkin D., 2003. A search engine for 3D models. ACM Transactions on Graphics 22(1): 83–105.

Godfrey L.R., Winchester J.M., King S.J., Boyer D.M., Jernvall J., 2012. Dental topography indicates ecological contraction of lemur communities. Am. J. Phys. Anthropol. 148(2): 215–27. doi:10.1002/ajpa.21615

Gould S.J., 1989. Wonderful Life. W.W. Norton & Co., New York.

Haines A.J., Crampton J.S., 2000. Improvements to the method of Fourier shape analysis as applied in morphometric studies. Palaeontology 43: 765–783. doi:10.1111/1475-4983.00148

Harjunmaa E., Kallonen A., Voutilainen M., Hamalainen K., Mikkola M.L., Jernvall J., 2012. On the difficulty of increasing dental complexity. Nature 483(7389): 324–327. doi:10.1038/nature10876

Heczko M., Keim D., Saupe D., Vranić D., 2002. Methods for similarity search on 3D databases. Datenbank-Spektrum 2: 54–63. [In German]

Hlusko L.J., 2002. Identifying metameric variation in extant hominoid and fossil hominid mandibular molars. Am. J. Phys. Anthropol. 118(1): 86–97. doi:10.1002/ajpa.10051

Hogue A.S., ZiaShakeri S., 2010. Molar crests and body mass as dietary indicators in marsupials. Aust. J. Zool. 58(1): 56–68. doi:10.1071/zo09084

Horn B., 1984. Extended Gaussian images. Proc. IEEE 72(12): 1671–1686.

Hunter J.P., Jernvall J., 1998. Decomposing early morphological diversity in ungulates: analysis of early Paleocene arctocyonid teeth. J. Vert. Paleontol. 18(3): 52A–53A.

Janis C.M., 1988. An estimation of tooth volume and hypsodonty indices in ungulate mammals, and the correlation of these factors with dietary preference. Teeth Revisited: Proceedings of the VIIth International Symposium on Dental Morphology, Paris, 1986. 367–387.

Jernvall J., 1995. Mammalian molar cusp patterns: developmental mechanisms of diversity. Acta Zool. Fenn. 198: 1–61.

Jernvall J., Selänne L., 1999. Laser confocal microscopy and geographic information systems in the study of dental morphology. Palaeontologia Electronica 1: 12 pp., 219 kb. http://palaeo-electronica.org/1999_1/confocal/issue1_99.htm

Jernvall J., Hunter J.P., Fortelius M., 1996. Molar tooth diversity, disparity, and ecology in Cenozoic ungulate radiations. Science 274(5292): 1489–1492.

Kaiser T.M., Wolff C., 2005. A new approach to dental microwear in primates – a pilot study. In: Żądzińska E. (Ed.). Current Trends in Dental Morphology Research: Full Refereed Papers from 13th International Symposium on Dental Morphology. University of Łódź Press, Łódź. 513–537.

Kaiser T.M., Brinkmann G., 2006. Measuring dental wear equilibriums – the use of industrial surface texture parameters to infer the diets of fossil mammals. Palaeogeogr. Palaeoclimatol. Palaeoecol. 239(3–4): 221–240. doi:10.1016/j.palaeo.2006.01.013

Kaiser T.M., Müller D.W.H., Fortelius M., Schulz E., Codron D., Clauss M., 2013. Hypsodonty and tooth facet development in relation to diet and habitat in herbivorous ungulates: implications for understanding tooth wear. Mammal Review 43(1): 34–46. doi:10.1111/j.1365-2907.2011.00203.x

Kay R.F., 1975. The functional adaptations of primate molar teeth. Am. J. Phys. Anthropol. 43: 195–215.

Kay R.F., 1978. Molar structure and diet in extant Cercopithecidae. In: Butler P.M., Joysey K.A. (Eds.) Development, Function and Evolution of Teeth. Academic Press, London. 309–339.

Kendall D.G., 1977. Diffusion of shape. Advances in Applied Probability 9(3): 428–430. doi:10.2307/1426091

Kendall D.G., 1984. Shape manifolds, Procrustean metrics, and complex projective spaces. Bulletin of the London Mathematical Society 16(MAR): 81–121. doi:10.1112/blms/16.2.81

Kendall D.G., 1989. A survey of the statistical theory of shape. Statistical Science 4(2): 87–120.

King S.J., Arrigo-Nelson S.J., Pochron S.T., Semprebon G.M., Godfrey L.R., Wright P.C., Jernvall J., 2005. Dental senescence in a long-lived primate links infant survival to rainfall. Proc. Natl Acad. Sci. USA 102(46): 16579–16583. doi:10.1073/pnas.0508377102

Klingenberg C.P., 2008. Novelty and "homology-free" morphometrics: What's in a name? Evol. Biol. 35(3): 186–190. doi:10.1007/s11692-008-9029-4

Klukkert Z.S., Teaford M.F., Ungar P.S., 2012. A dental topographic analysis of chimpanzees. Am. J. Phys. Anthropol. 148(2): 276–84. doi:10.1002/ajpa.21592

Legendre S., 1986. Analysis of mammalian communities from the late Eocene and Oligocene of southern France. Palaeovertebrata 16: 191–212.

Li B., Godil A., Aono M., Bai X., Furuya T., Li L., López-Sastre R., Johan H., Ohbuchi R., Redondo-Cabrera C., Tatsuma A., Yanagimachi T., Zhang S., 2012. SHREC'12 Track: Generic 3D Shape Retrieval. Eurographics Workshop on 3D Object Retrieval: 119–126.

Liu L., Puolamaki K., Eronen J.T., Ataabadi M.M., Hernesniemi E., Fortelius M., 2012. Dental functional traits of mammals resolve productivity in terrestrial ecosystems past and present. Proc. Roy. Soc. B 279(1739): 2793–2799.

Lohmann G.P., 1983. Eigenshape analysis of micro-fossils – a general morphometric procedure for describing changes in shape. Journal of the International Association for Mathematical Geology 15(6): 659–672. doi:10.1007/bf01033230

M'Kirera F., Ungar P.S., 2003. Occlusal relief changes with molar wear in *Pan troglodytes troglodytes* and *Gorilla gorilla gorilla*. American Journal of Primatology 60(2): 31–41. doi:10.1002/ajp.10077

MacLeod N., 1999. Generalizing and extending the eigenshape method of shape space visualisation and analysis. Paleobiology 25(1): 107–138.

MacLeod N., 2008. Understanding morphology in systematic contexts: three-dimensional specimen ordination and recognition. In: Wheeler Q. (Ed.) The New Taxonomy. CRC Press, Taylor & Francis Group, London. 143–210.

Mandelbrot B.B. 1967. How long is the coast of Britain? Science 156: 636–638.

Meloro C., Raia P., 2010. Cats and dogs down the tree: the tempo and mode of evolution in the lower carnassial of fossil and living Carnivora. Evol. Biol. 37(4): 177–186. doi: 10.1007/s11692-010-9094-3

Merceron G., Escarguel G., Angibault J.-M., Verheyden-Tixier H., 2010. Can dental microwear textures record inter-individual dietary variations? PLoS ONE 5(3): e9542. doi:10.1371/journal.pone.0009542

Merceron G., Scott J., Scott R.S., Geraads D., Spassov N., Ungar P.S., 2009. Folivory or fruit/seed predation for *Mesopithecus*, an earliest colobine from the late Miocene of Eurasia? J. Hum. Evol. 57(6): 732–8. doi:10.1016/j.jhevol.2009.06.009

Osada R., Funkhouser T., Chazelle B., Dobkin D., 2002. Shape distributions. ACM Transactions on Graphics 21(4): 807–832.

Oxnard C., O'Higgins P., 2009. Biology clearly needs morphometrics. Does morphometrics need biology? Biological Theory 4(1): 84–97.

Piras P., Maiorino L., Raia P., Marcolini F., Salvi D., Vignoli L., Kotsakis T., 2010. Functional and phylogenetic constraints in Rhinocerotinae craniodental morphology. Evol. Ecol. Res. 12(8): 897–928.

Plyusnin I., Evans A.R., Karme A., Gionis A., Jernvall J., 2008. Automated 3D phenotype analysis using data mining. PLoS ONE 3(3): e1742. doi:10.1371/journal.pone.0001742

Polly P.D., 2008. Adaptive zones and the pinniped ankle: a 3D quantitative analysis of carnivoran tarsal evolution. In: Sargis E., Dagosto M. (Eds.) Mammalian Evolutionary Morphology: A Tribute to Frederick S. Szalay. Kluwer/Plenum, New York.

Polly P.D., 2010. Tiptoeing through the trophics: geographic variation in carnivoran locomotor ecomorphology in relation to environment. In: Goswami A., Friscia A. (Eds.)

Carnivoran Evolution: New Views on Phylogeny, Form, and Function. Cambridge University Press, Cambridge, UK. 347–410.
Polly P.D., MacLeod N., 2008. Locomotion in fossil Carnivora: an application of eigensurface analysis for morphometric comparison of 3D surfaces. Palaeontologia Electronica 11(2): 10A, 13p. http://palaeo-electronica.org/2008_2/135/index.html
Polly P.D., Eronen J.T., Fred M., Dietl G.P., Mosbrugger V., Scheidegger C., Frank D.C., Damuth J., Stenseth N.C., Fortelius M., 2011. History matters: ecometrics and integrative climate change biology. Proc. Roy. Soc. B 278(1709): 1131–1140. doi:10.1098/rspb.2010.2233
Pontzer H., Scott J.R., Lordkipanidze D., Ungar P.S., 2011. Dental microwear texture analysis and diet in the Dmanisi hominins. J. Hum. Evol. 61(6): 683–687. doi:10.1016/j.jhevol.2011.08.006
Prideaux G.J., Ayliffe L.K., DeSantis L.R., Schubert B.W., Murray P.F., Gagan M.K., Cerling T.E., 2009. Extinction implications of a chenopod browse diet for a giant Pleistocene kangaroo. Proc. Natl Acad. Sci. USA 106(28): 11646–11650. doi:10.1073/pnas.0900956106
Purnell M., Seehausen O., Galis F., 2012. Quantitative three-dimensional microtextural analyses of tooth wear as a tool for dietary discrimination in fishes. J. R. Soc. Interface 9(74): 2225–2233 doi:10.1098/rsif.2012.0140
Raia P., Carotenuto F., Eronen J.T., Fortelius M., 2011. Longer in the tooth, shorter in the record? The evolutionary correlates of hypsodonty in Neogene ruminants. Proc. Roy. Soc. B 278(1724): 3474–3481. doi:10.1098/rspb.2011.0273
Reed D.N.O., 1997. Contour mapping as a new method for interpreting diet from tooth morphology. Am. J. Phys. Anthropol. Suppl 24: 194.
Rensberger J.M., 1978. Scanning electron microscopy of wear and occlusal events in some small herbivores. In: Butler P.M., Joysey K.A. (Eds.) Development, Function and Evolution of Teeth. Academic Press, London. 415–438.
Reuter M., Wolter F.-E., Peinecke N., 2005. Laplace-Spectra as fingerprints for shape matching. SPM'05: Proceedings of the 2005 ACM symposium on solid and physical modeling, ACM Press. 101–106.
Reuter M., Wolter F.-E., Peinecke N., 2006. Laplace–Beltrami spectra as "Shape-DNA" of surfaces and solids. Computer-Aided Design 38(4): 342–366. doi:10.1016/j.cad.2005.10.011
Richtsmeier J.T., Deleon V.B., Lele S.R., 2002. The promise of geometric morphometrics. Yearbook of Physical Anthropology 45: 63–91. doi:10.1002/ajpa.10174
Rohlf F.J., Marcus L.F., 1993. A revolution in morphometrics. Trends Ecol. Evol. 8(4): 129–32. doi:10.1016/0169-5347(93)90024-j
Rustamov R.M., 2007. Laplace-Beltrami eigenfunctions for deformation invariant shape representation. Symposium on Geometry Processing 2007: 225–233.
Santana S.E., Strait S., Dumont E.R., 2011. The better to eat you with: functional correlates of tooth structure in bats. Funct. Ecol. 25(4): 839–847. doi:10.1111/j.1365-2435.2011.01832.x
Saupe D., Vrani D.V., 2001. 3D model retrieval with spherical harmonics and moments. DAGM 2001: 392–397.
Schubert B.W., Ungar P.S., DeSantis L.R.G., 2010. Carnassial microwear and dietary behaviour in large carnivorans. J. Zool. 280(3): 257–263. doi:10.1111/j.1469-7998.2009.00656.x
Schulz E., Calandra I., Kaiser T.M., 2010. Applying tribology to teeth of hoofed mammals. Scanning 32(4): 162–182. doi:10.1002/sca.20181
Scott J.R., 2012. Dental microwear texture analysis of extant African Bovidae. Mammalia 76(2) 157–174 doi:10.1515/mammalia-2011-0083
Scott J.R., Godfrey L.R., Jungers W.L., Scott R.S., Simons E.L., Teaford M.F., Ungar P.S., Walker A., 2009. Dental microwear texture analysis of two families of subfossil lemurs from Madagascar. J. Hum. Evol. 56(4): 405-416. doi:10.1016/j.jhevol.2008.11.003
Scott R.S., Teaford M.F., Ungar P.S., 2012. Dental microwear texture and anthropoid diets. Am. J. Phys. Anthropol. 147(4): 551–579. doi:10.1002/ajpa.22007
Scott R.S., Ungar P.S., Bergstrom T.S., Brown C.A., Grine F.E., Teaford M.F., Walker A., 2005. Dental microwear texture analysis shows within-species diet variability in fossil hominins. Nature 436(7051): 693–695. doi:10.1038/nature03822
Scott R.S., Ungar P.S., Bergstrom T.S., Brown C.A., Childs B.E., Teaford M.F., Walker A., 2006. Dental microwear texture analysis: technical considerations. J. Hum. Evol. 51(4): 339–349. doi:10.1016/j.jhevol.2006.04.006
Shen L., Farid H., McPeek M.A., 2009. Modeling three-dimensional morphological structures using spherical harmonics. Evolution 63(4): 1003–1016. doi:10.1111/j.1558-5646.2008.00557.x
Sievwright H., MacLeod N., 2012. Eigensurface analysis, ecology, and modelling of morphological adaptation in the falconiform humerus (Falconiformes: Aves). Zool. J. Linn. Soc. 165(2): 390–419. doi:10.1111/j.1096-3642.2012.00818.x
Simpson G.G., 1953. The Major Features of Evolution. Simon and Schuster, New York.
Singleton M., Rosenberger A.L., Robinson C., O'Neill R,. 2011. Allometric and metameric shape variation in *Pan* mandibular molars: a digital morphometric analysis. Anat. Rec. 294(2): 322–334. doi:10.1002/ar.21315
Skinner M.M., Gunz P., Wood B.A., Hublin J.J., 2008. Enamel-dentine junction (EDJ) morphology distinguishes the lower molars of *Australopithecus africanus* and *Paranthropus robustus*. J. Hum. Evol. 55(6): 979–988. doi:10.1016/j.jhevol.2008.08.013
Slice D.E., 2007. Geometric morphometrics. Ann. Rev. Anthropol. 36(1): 261–281. doi:10.1146/annurev.anthro.34.081804.120613
Smits P.D., Evans A.R., 2012. Functional constraints on tooth morphology in carnivorous mammals. BMC Evol. Biol. 12: 146. doi:10.1186/1471-2148-12-146
Stynder D.D., Ungar P.S., Scott J.R., Schubert B.W., 2012. A dental microwear texture analysis of the Langebaanweg (South African Mio-Pliocene) fossil Hyaenidae. Acta Palaeont. Pol. 57(3): 485–496. doi:10.4202/app.2011.0053
Sun J., Ovsjanikov M., Guibas L., 2009. A concise and provably informative multi-scale signature-based on heat diffusion. Computer Graphics Forum 28: 1383–1392.
Teaford M.F., 1988. A review of dental microwear and diet in modern mammals. Scanning Microscopy 2(2): 1149–1166.
Ulhaas L., Kullmer O., Schrenk F., 2007. Tooth wear and diversity in early hominid molars: A case study. In: Bailey S.E.H.J.J. (Ed.) Dental Perspectives on Human Evolution: State of the Art Research in Dental Paleoanthropology. 369–390.
Ulhaas L., Kullmer O., Schrenk F., Henke W., 2004. A new 3-d approach to determine functional morphology of cercopithecoid molars. Annals of Anatomy – Anatomischer Anzeiger 186(5–6): 487–493. doi:10.1016/s0940-9602(04)80090-6
Ungar P., Williamson M., 2000. Exploring the effects of tooth wear on functional morphology: a preliminary study using dental topographic analysis. Palaeontologia Electronica 3(1): 18 pp., 752 kb. http://palaeo-electronica.org/2000_1/gorilla/issue1_00.htm
Ungar P.S., 2010. Mammal Teeth: Origin, Evolution, and Diversity. Johns Hopkins University Press, Baltimore.
Ungar P.S., M'Kirera F., 2003. A solution to the worn tooth conundrum in primate functional anatomy. Proc. Natl. Acad. Sci. USA 100(7): 3874–3877. doi:10.1073/pnas.0637016100
Ungar P.S., Walker A., Coffing K., 1994. Reanalysis of the Lukeino molar (KNM-LU-335). Am. J. Phys. Anthropol. 94(2): 165–173. doi:10.1002/ajpa.1330940202
Ungar P.S., Brown C.A., Bergstrom T.S., Walker A., 2003. Quantification of dental microwear by tandem scanning confocal microscopy and scale-sensitive fractal analyses. Scanning 25: 185–193.
Ungar P.S., Krueger K.L., Blumenschine R.J., Njau J.K., Scott R.S., 2012a. Dental microwear texture analysis of hominins recovered by the Olduvai Landscape Paleoanthropology Project, 1995–2007. Am. J. Phys. Anthropol. 63(2): 429–437. doi:10.1016/j.jhevol.2011.04.006
Ungar P.S., Scott J.R., Curran S.C., Dunsworth H.M., Harcourt-Smith W.E.H., Lehmann T., Manthi F.K., McNulty K.P., 2012b. Early Neogene environments in East Africa: Evidence from dental microwear of tragulids. Palaeogeogr. Palaeoclimatol. Palaeoecol. 342–343: 84–96. doi:10.1016/j.palaeo.2012.05.005
van Kaick O., Zhang H., Hamarneh G., Cohen-Or D., 2011. A survey on shape correspondence. Computer Graphics Forum 30(6): 1681–1707. doi:10.1111/j.1467-8659.2011.01884.x
Van Valen L., 1960. A functional index of hypsodonty. Evolution 14: 531–532.
Walker A., Hoeck H.N., Perez L., 1978. Microwear in mammalian teeth as an indicator of diet. Science 201: 908–910.
Wesley-Hunt G.D., 2005. The morphological diversification of carnivores in North America. Paleobiol. 31(1): 35–55. doi:10.1666/0094-8373(2005)031<0035:tmdoci>2.0.co;2
Wilson G.P., Evans A.R., Corfe I.J., Smits P.D., Fortelius M., Jernvall J., 2012. Adaptive radiation of multituberculates before the extinction of dinosaurs. Nature 483(7390): 457–460. doi:10.1038/nature10880
Wolfe J., 1995. Paleoclimatic estimates from Tertiary leaf assemblages. Ann. Rev. Earth Planet. Sci. 23: 119–142.
Wood A., 2011. Regional differences in tarsal morphotypes among Late Miocene-Pliocene equids: a 3D geometric morphometric study. Journal of Vertebrate Paleontology, SVP Program and Abstracts Book 2011: 216.
Zuccotti L.F., Williamson M.D., Limp W.F., Ungar P.S., 1998. Modeling primate occlusal topography using geographic information systems technology. Am. J. Phys. Anthropol. 107: 137–142.

Associate Editor: A. Loy

HYSTRIX

the Italian Journal of Mammalogy

Volume 24(1) • 2013

Edited and published by Associazione Teriologica Italiana

Aims and scope

Hystrix, the Italian Journal of Mammalogy accepts papers on original research in basic and applied mammalogy on fossil and living mammals. The Journal is published both in paper and electronic "online first" format. Manuscripts can be published as full papers or short notes, as well as reviews on methods or theoretical issues related to mammals. Commentaries can also be occasionally accepted, under the approval by the Editor in Chief. Investigations of local or regional interest, new data about species distribution and range extensions or confirmatory research can be considered only when they have significant implications. Such studies should preferably be submitted as short notes. Manuscripts bearing only a local interest will not be accepted.

Full papers have no limits in length as well as in figure and table number and are abstracted in English. Authors are encouraged to add supplemental material in form of colour figures, original datasets and/or computer program source code. Supplemental material and colour figures will appear only on the electronic edition.

Short notes must be about 16000 characters long (including title, author names and affiliations, abstract and references), and do not include supplemental material. They are abstracted in English.

Proceedings of symposia, meetings and/or workshops, and technical reports can be published as special supplements to regular issues, under the approval by the Editor in Chief and the Associate Editors.

There are no page charges.

Manuscript submission

Manuscripts must be submitted electronically registering to the on-line submission system at the Journal web site (http://www.italian-journal-of-mammalogy.it). A comprehensive Electronic Publication Guide can be downloaded from the Journal web site: Part II of that document contains a detailed step-by-step description of the electronic submission process. Authors must submit at least a manuscript file; a cover letter and a copyright transfer form are not necessary since the electronic submission process provides both for manuscript presentation and copyright transfer acceptance. Tables and figures must be included in the manuscript file, whilst other supplemental material (if any) must be uploaded separately.

Manuscript structure

Full papers: manuscript must be divided into sections in the following sequence: title page (page 1), abstract and keywords, (page 2), introduction (from page 3 onwards), materials and methods, results and discussion, acknowledgements, list of symbols (if any), references. Tables, legends of figures and figures should be on separate pages as specified above. If necessary and useful to improve manuscript readability, a single section could be divided into subsections or paragraphs. If necessary, conclusions and/or any final consideration can be stated as a last paragraph of results and discussion.

Short notes do not have Introduction, Material and methods, Results and Discussion, and are organised in a single section. Authors are advised to structure Short notes without subdivision of the text, with an Abstract in English. The whole length of the manuscript must not exceed 16000 characters (spaces included), comprehensive of title, author names and affiliations, abstract, text body and references. In a short note references should be kept to a minimum.

Publication process

The Technical Editor checks all submitted manuscripts for compliance with the Instructions to Authors. The Editor in Chief then assigns the manuscript to an Associate Editor for the peer-review process. Once accepted, the manuscript will be typeset and a final galley will be sent to Authors for their approval. Once approved by the Authors, the manuscript will be published "online first" and will be printed in the next available issue.

Privacy statement

The names and email addresses appearing in this journal will be used exclusively for the stated journal's purposes and will not be made available for any other purpose or to any other party, as provided by the Italian Law no. 675, 31/12/1996. No notification to the Warrant is needed, as provided in art. 7, sec. 5ter, a), f), Italian Law no. 675, 31/12/1996.

Open Access Policy

This journal provides open access to all of its content on the principle that making research freely available to the public supports a greater global exchange of knowledge. For more information on this approach, see the Public Knowledge Project (http://pkp.sfu.ca), which has designed this system to improve the scholarly and public quality of research, and which freely distributes the journal system as well as other software to support the open access publishing of scholarly resources.

HYSTRIX
the Italian Journal of Mammalogy
Volume 24(1) • 2013
Edited and published by Associazione Teriologica Italiana

Contents