LGN 5822 - Biometrical Genetics

L01a – Introduction

Michele Jorge Silva Siqueira

2023

LGN 5822 - Biometrical Genetics

Outline

- Who am I?
- Course content
- Reproducibility
- Approach for the course
- Exams

Who am I?

Professor Ph.D Michele Jorge Silva Siqueira

Department of Genetics, ESALQ/USP

Schedule

- 16/08 - Introduction to Biometrics and Matrix Algebra
- 23/08 – No class - Brazilian Congress of Plant Breeding
- 30/08 - Introduction to R
- 06/09 – No class
- 13/09 - Hypothesis Testing
- 20/09 - Linear Models
- 27/09 - Linear Mixed Models
- 04/10 - Principles of Experimental Design
- 11/10 – No class
- 18/10 - First Exam

LGN 5822 - Biometrical Genetics

- 25/10 - Completely Randomized Design / Randomized Complete Block and Factorial Designs
- 01/11 - No class
- 08/11 - Latin Square and Split Plot Designs / Incomplete Block
- 15/11 - No class
- 22/11 – Lattice Designs / Augmented Designs
- 29/11 – Modeling of Genetic Correlations (Multi-Environment and Multi-Trait Analysis)
- 06/12 - Classroom Presentation: Practical Assignment – LGN5822

LGN 5822 - Biometrical Genetics

I will evaluate you:

- Exams;
- Weekly exercises with deadlines: Everything on the Forum (open to the web).

Reproducibility

- We will use tools allowing reproducibility;
- We will share data, codes, etc (videos,...)
- Try R and RStudio.

This course

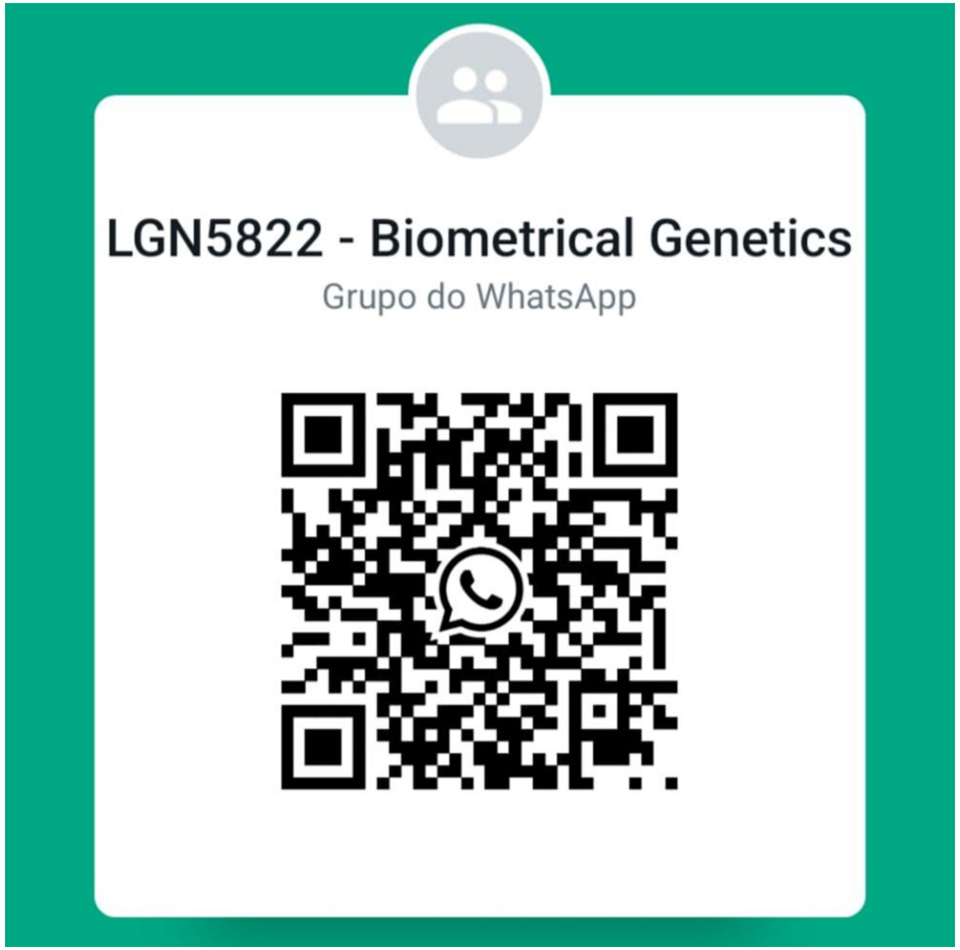
Deductive approach: General rule → Specific examples → Practice

> *"Learn continually - there's always "one more thing" to learn!"*
>
> Steve Jobs

LGN 5822 - Biometrical Genetics

We will be in touch!

- WhatsApp group



LGN5822 - Biometrical Genetics

Grupo do WhatsApp

https://chat.whatsapp.com/JDRqoyuNHWw29hb3KbUtfT

LGN 5822 - Biometrical Genetics

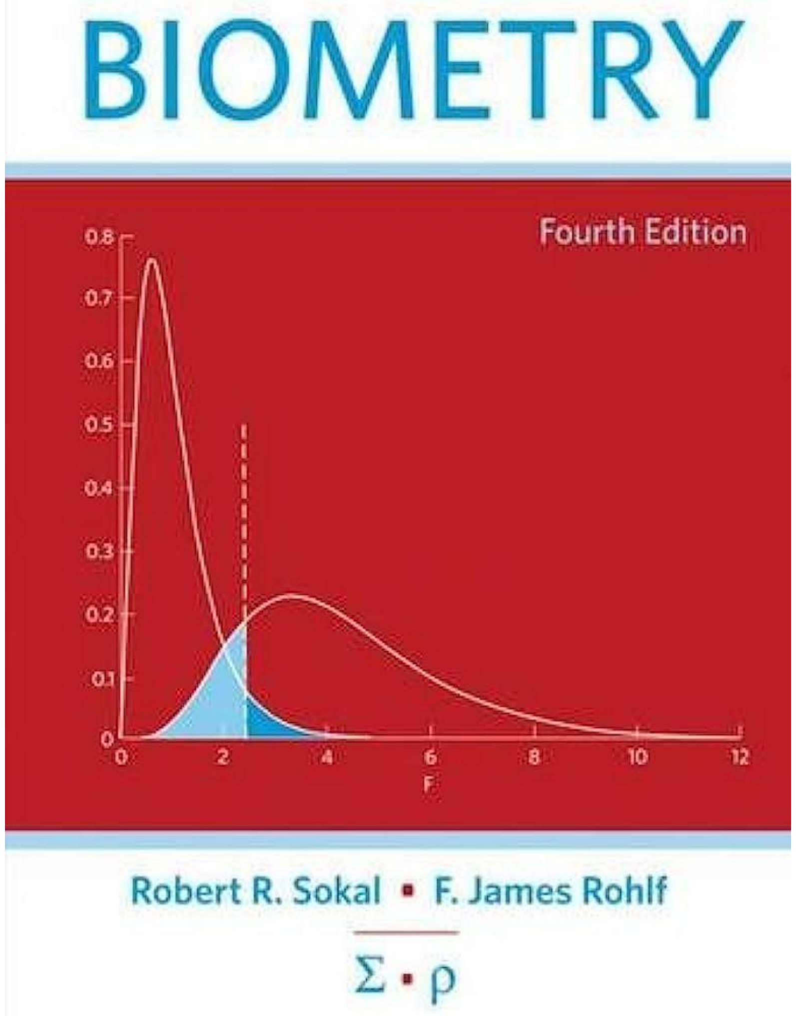
We will be in touch!

- Moodle USP: LGN5822

Check! I need USP number to add you!

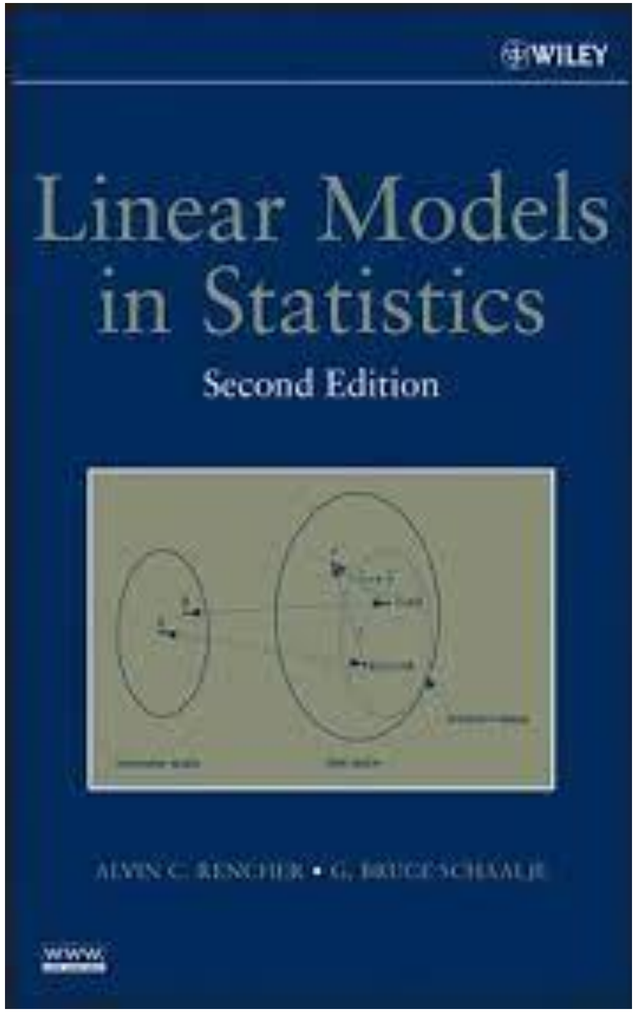
You must participate!

Introduction to Biometrics



Sokal, Robert R. "The principles and practice of statistics in biological research." Biometry (2012): 451-554.

Introduction to Biometrics



WILEY

Linear Models in Statistics

Second Edition

ALVIN C. RENCHER • G. BRUCE SCHAALJE

Rencher, Alvin C., and G. Bruce Schaalje. Linear models in statistics. John Wiley & Sons, 2008.

LGN 5822 - Biometrical Genetics

Purpose of discipline

To undestant the use of **biometrics** with particular emphasis on agricultural and biological design, analysis, and interpretation experiments.

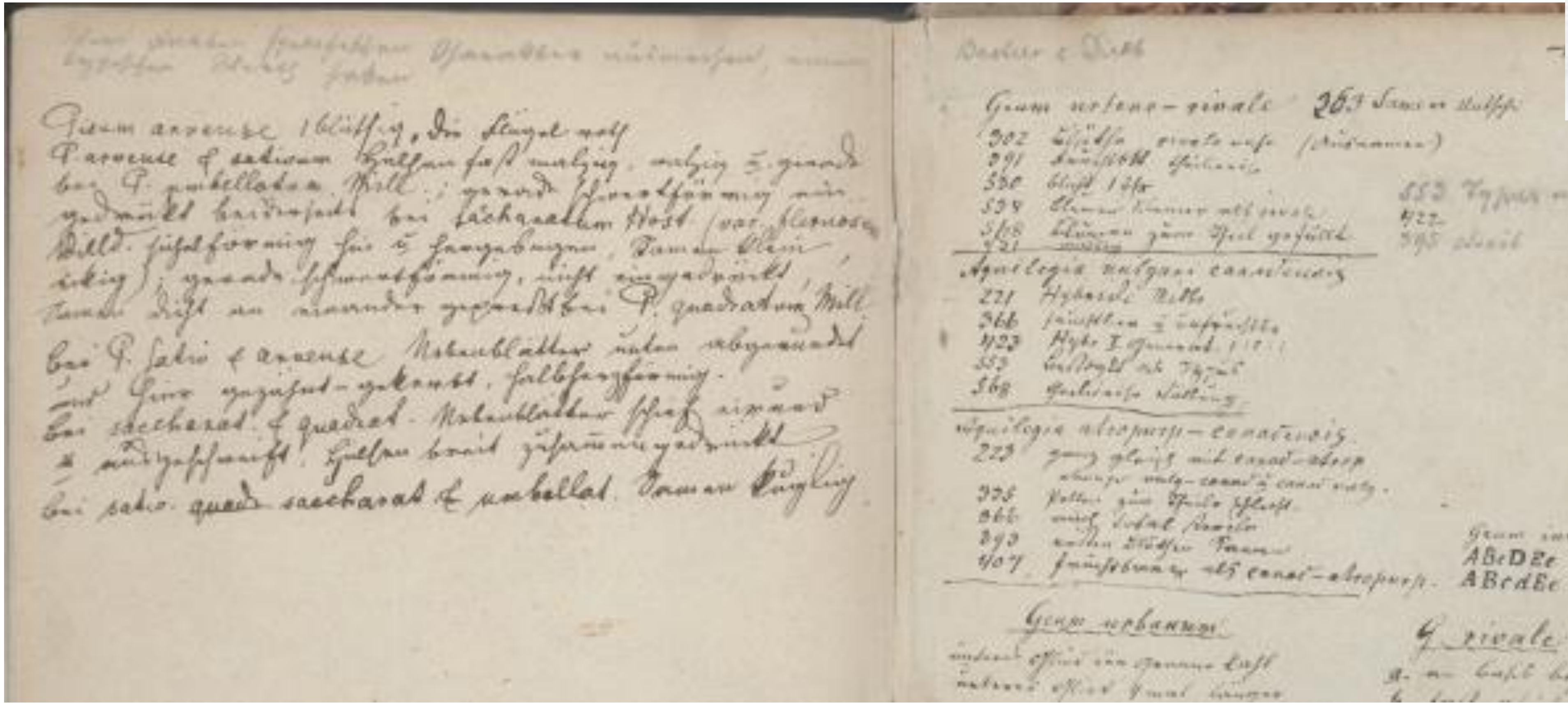
LGN 5822 - Biometrical Genetics

Introduction to Biometrics

Introduction to Biometrics

1865 – Gregor Mendel

1822-1884



Notes on the end page of Mendel's copy

Introduction to Biometrics

1865 – Gregor Mendel

From Mendel’s laws, we discovered how the factors responsible for inheritance of the simplest traits were transmitted;

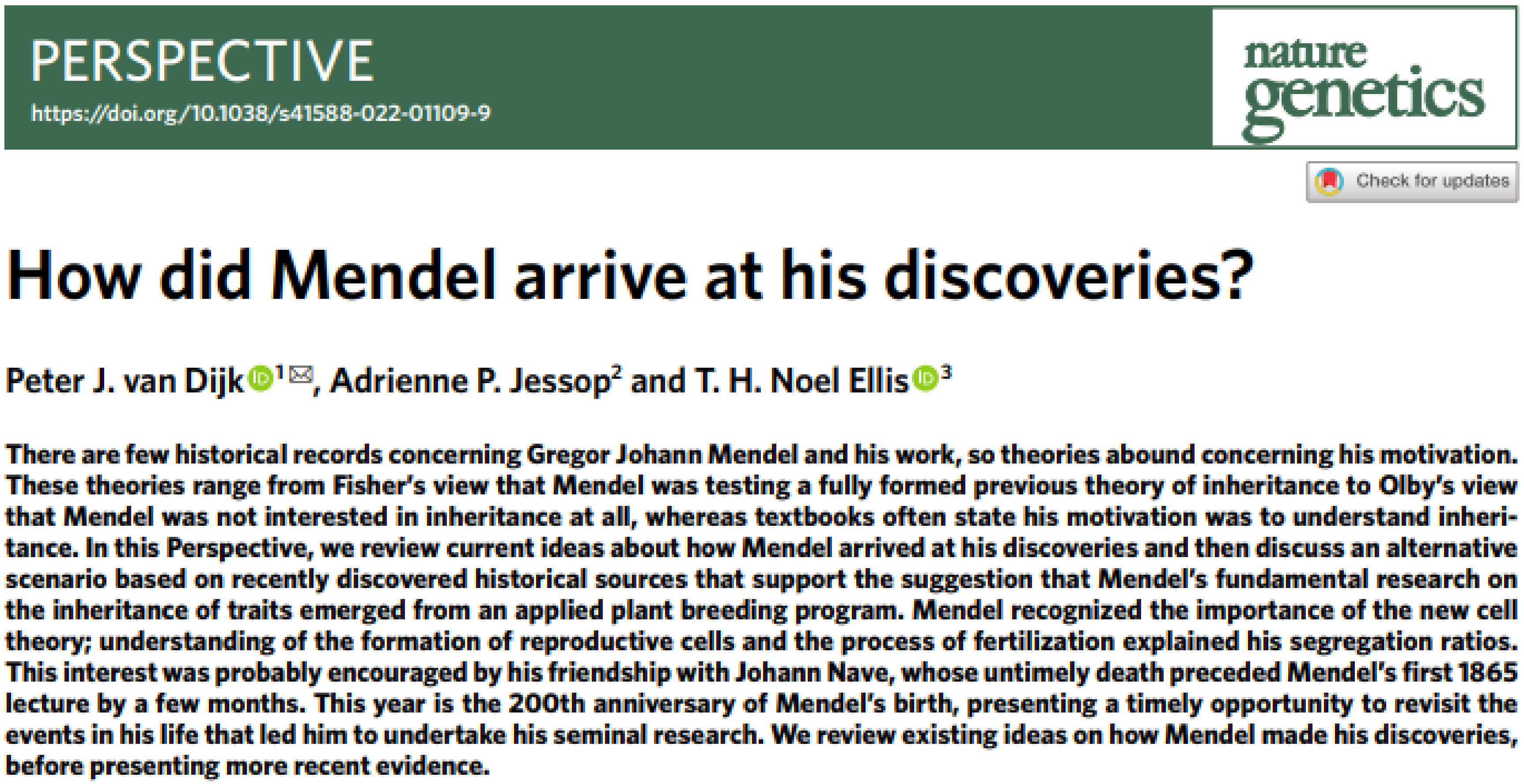
Introduction to Biometrics

1865 – Gregor Mendel

- **Mendel’s First Law**, states that during gamete formation, the two alleles at a gene locus segregate from each other; each gamete has an equal probability of containing either allele

- **Mendel's 2nd law** states that during gamete formation the segregation of each gene pair is independent of other pairs

Introduction to Biometrics



PERSPECTIVE

https://doi.org/10.1038/s41588-022-01109-9

nature genetics

Check for updates

How did Mendel arrive at his discoveries?

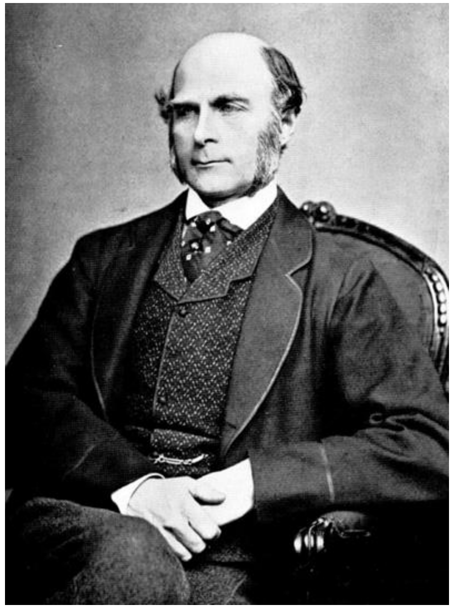
Peter J. van Dijk[1], Adrienne P. Jessop[2] and T. H. Noel Ellis[3]

There are few historical records concerning Gregor Johann Mendel and his work, so theories abound concerning his motivation. These theories range from Fisher's view that Mendel was testing a fully formed previous theory of inheritance to Olby's view that Mendel was not interested in inheritance at all, whereas textbooks often state his motivation was to understand inheritance. In this Perspective, we review current ideas about how Mendel arrived at his discoveries and then discuss an alternative scenario based on recently discovered historical sources that support the suggestion that Mendel's fundamental research on the inheritance of traits emerged from an applied plant breeding program. Mendel recognized the importance of the new cell theory; understanding of the formation of reproductive cells and the process of fertilization explained his segregation ratios. This interest was probably encouraged by his friendship with Johann Nave, whose untimely death preceded Mendel's first 1865 lecture by a few months. This year is the 200th anniversary of Mendel's birth, presenting a timely opportunity to revisit the events in his life that led him to undertake his seminal research. We review existing ideas on how Mendel made his discoveries, before presenting more recent evidence.

van Dijk, Peter J., Adrienne P. Jessop, and TH Noel Ellis. Nature Genetics 54.7 (2022): 926-933

Introduction to Biometrics

1886 - Francis Galton

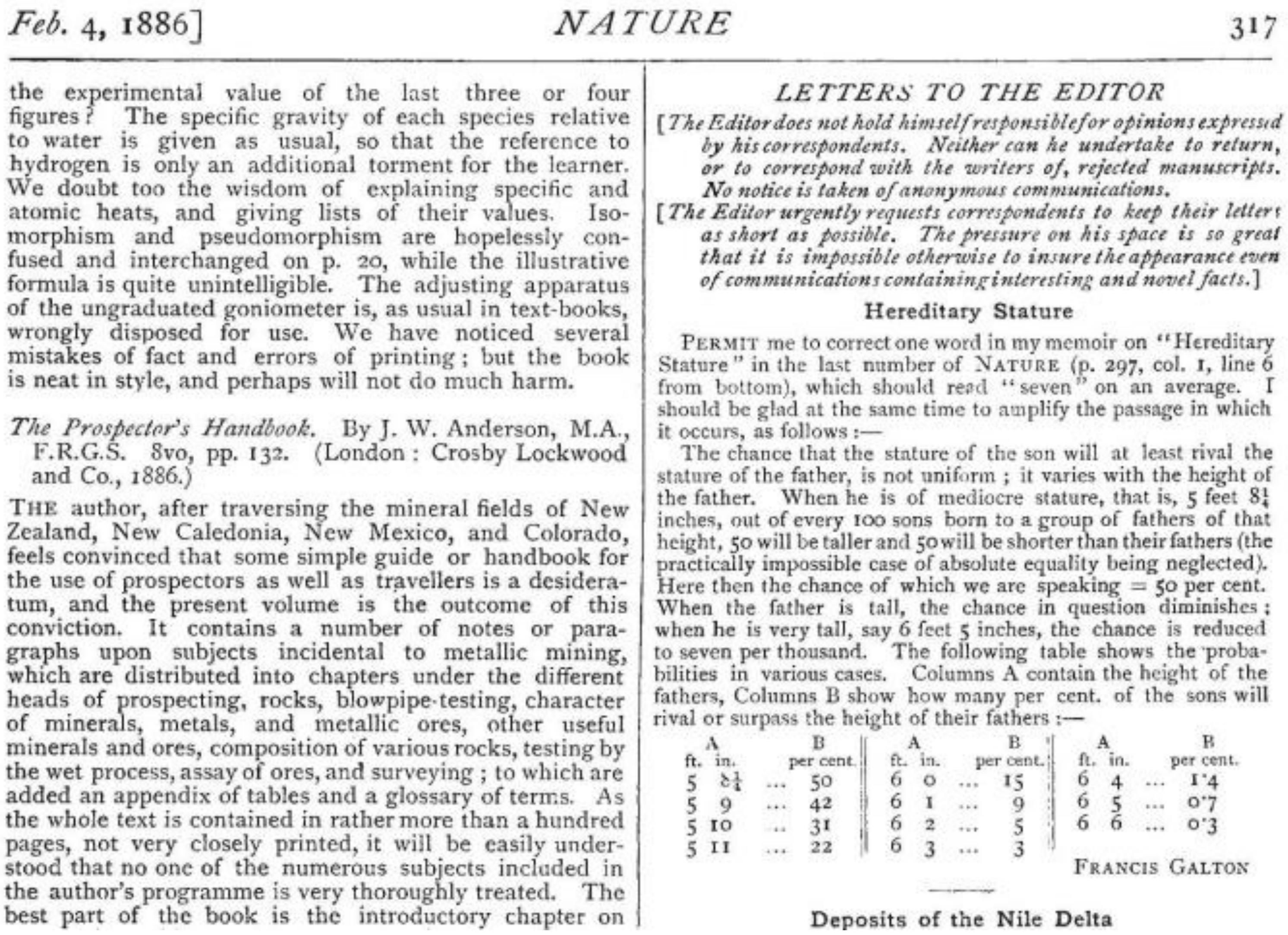


1822-1911

- The field of biometry began with Francis Galton’s work on how human characteristics, particularly height, were passed from parents to children.

- Galton's ancestral law states that the **two parent** contribute between them **one-half** of the total heritage of children, the **four grandparents one-quarter**, and so on.

Introduction to Biometrics



Feb. 4, 1886] *NATURE* 317

the experimental value of the last three or four figures? The specific gravity of each species relative to water is given as usual, so that the reference to hydrogen is only an additional torment for the learner. We doubt too the wisdom of explaining specific and atomic heats, and giving lists of their values. Isomorphism and pseudomorphism are hopelessly confused and interchanged on p. 20, while the illustrative formula is quite unintelligible. The adjusting apparatus of the ungraduated goniometer is, as usual in text-books, wrongly disposed for use. We have noticed several mistakes of fact and errors of printing; but the book is neat in style, and perhaps will not do much harm.

The Prospector's Handbook. By J. W. Anderson, M.A., F.R.G.S. 8vo, pp. 132. (London: Crosby Lockwood and Co., 1886.)

THE author, after traversing the mineral fields of New Zealand, New Caledonia, New Mexico, and Colorado, feels convinced that some simple guide or handbook for the use of prospectors as well as travellers is a desideratum, and the present volume is the outcome of this conviction. It contains a number of notes or paragraphs upon subjects incidental to metallic mining, which are distributed into chapters under the different heads of prospecting, rocks, blowpipe-testing, character of minerals, metals, and metallic ores, other useful minerals and ores, composition of various rocks, testing by the wet process, assay of ores, and surveying; to which are added an appendix of tables and a glossary of terms. As the whole text is contained in rather more than a hundred pages, not very closely printed, it will be easily understood that no one of the numerous subjects included in the author's programme is very thoroughly treated. The best part of the book is the introductory chapter on

LETTERS TO THE EDITOR

[The Editor does not hold himself responsible for opinions expressed by his correspondents. Neither can he undertake to return, or to correspond with the writers of, rejected manuscripts. No notice is taken of anonymous communications.

[The Editor urgently requests correspondents to keep their letters as short as possible. The pressure on his space is so great that it is impossible otherwise to insure the appearance even of communications containing interesting and novel facts.]

Hereditary Stature

PERMIT me to correct one word in my memoir on "Hereditary Stature" in the last number of NATURE (p. 297, col. 1, line 6 from bottom), which should read "seven" on an average. I should be glad at the same time to amplify the passage in which it occurs, as follows:—

The chance that the stature of the son will at least rival the stature of the father, is not uniform; it varies with the height of the father. When he is of mediocre stature, that is, 5 feet 8¼ inches, out of every 100 sons born to a group of fathers of that height, 50 will be taller and 50 will be shorter than their fathers (the practically impossible case of absolute equality being neglected). Here then the chance of which we are speaking = 50 per cent. When the father is tall, the chance in question diminishes; when he is very tall, say 6 feet 5 inches, the chance is reduced to seven per thousand. The following table shows the probabilities in various cases. Columns A contain the height of the fathers, Columns B show how many per cent. of the sons will rival or surpass the height of their fathers:—

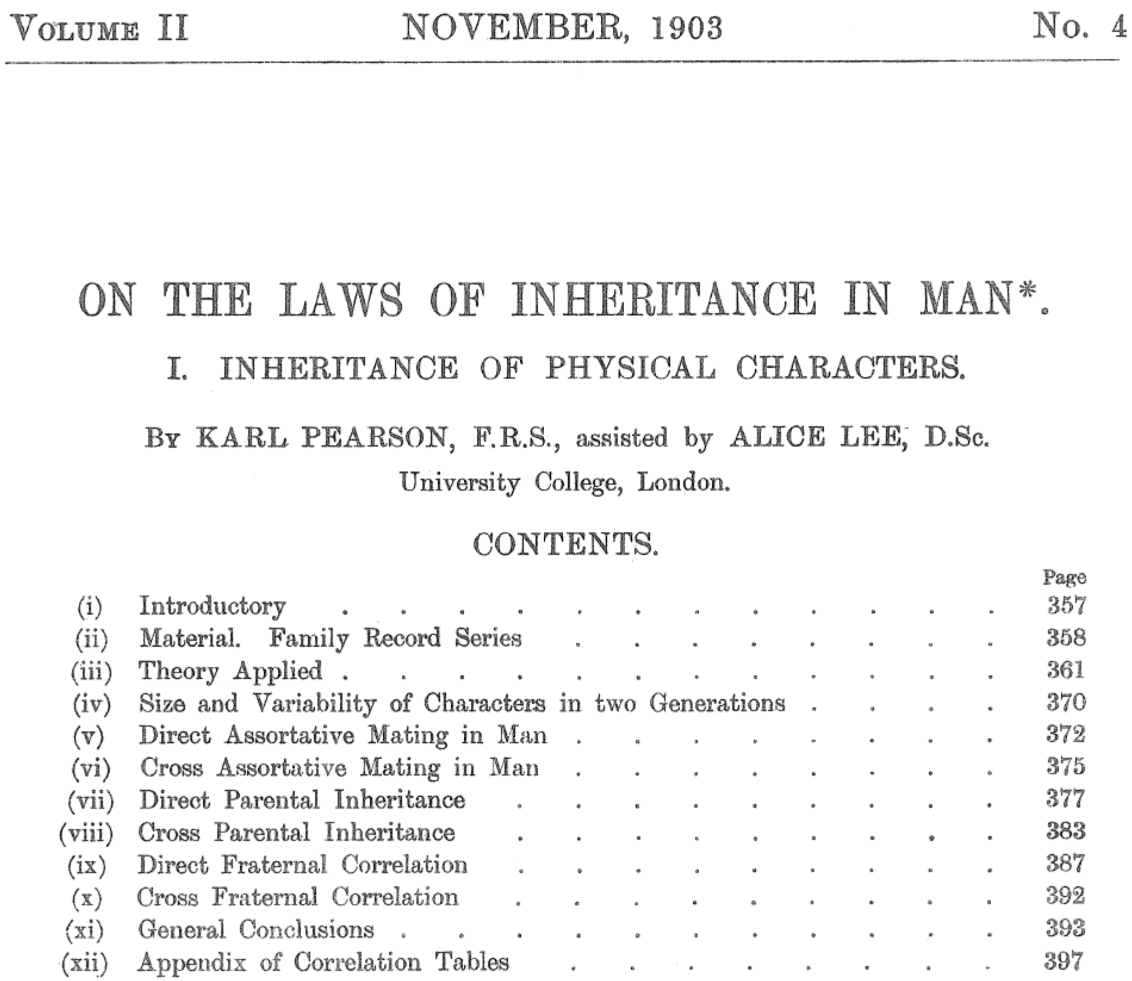
A		B	A		B	A		B
ft.	in.	per cent.	ft.	in.	per cent.	ft.	in.	per cent.
5	8¼	50	6	0	15	6	4	1·4
5	9	42	6	1	9	6	5	0·7
5	10	31	6	2	5	6	6	0·3
5	11	22	6	3	3			

FRANCIS GALTON

Deposits of the Nile Delta

Galton, F. (1886a). Hereditary stature. Nature 33, 295–298

1903 - Pearson and Lee



VOLUME II NOVEMBER, 1903 No. 4

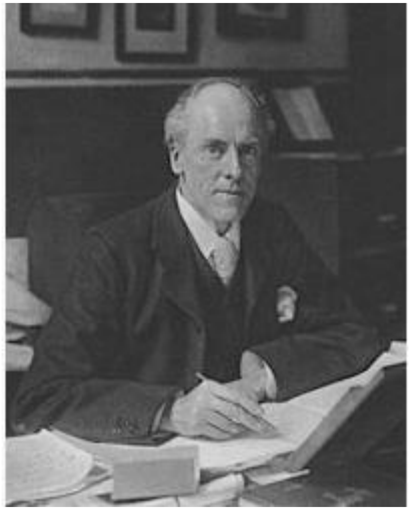
ON THE LAWS OF INHERITANCE IN MAN*.

I. INHERITANCE OF PHYSICAL CHARACTERS.

BY KARL PEARSON, F.R.S., assisted by ALICE LEE, D.Sc.
University College, London.

CONTENTS.

Page
(i) Introductory . . . 357
(ii) Material. Family Record Series . . . 358
(iii) Theory Applied . . . 361
(iv) Size and Variability of Characters in two Generations . . . 370
(v) Direct Assortative Mating in Man . . . 372
(vi) Cross Assortative Mating in Man . . . 375
(vii) Direct Parental Inheritance . . . 377
(viii) Cross Parental Inheritance . . . 383
(ix) Direct Fraternal Correlation . . . 387
(x) Cross Fraternal Correlation . . . 392
(xi) General Conclusions . . . 393
(xii) Appendix of Correlation Tables . . . 397



1857-1936

Pearson later analyzed Galton’s data as well as subsequent data and found that the **correlation of the height of fathers and mothers with the height of their sons and daughters had a mean of 0.51.**

Pearson K, Lee A (1903) On the laws of inheritance in man: I. inheritance of physical characters. Biometrika 2:357–462

Introduction to Biometrics

- These findings by the biometricians disagreed with Mendel's findings

- Mendel's results indicated a particulate nature of inheritance that led to distinct classes, rather than a continuum of observations for a given trait

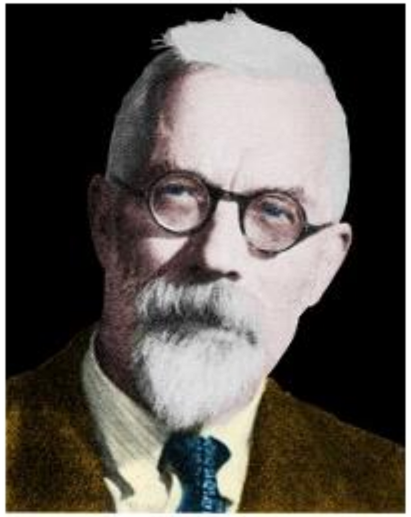
- The absence of distinct phenotypic classes then suggested to the Mendelists that continuous traits are not heritable

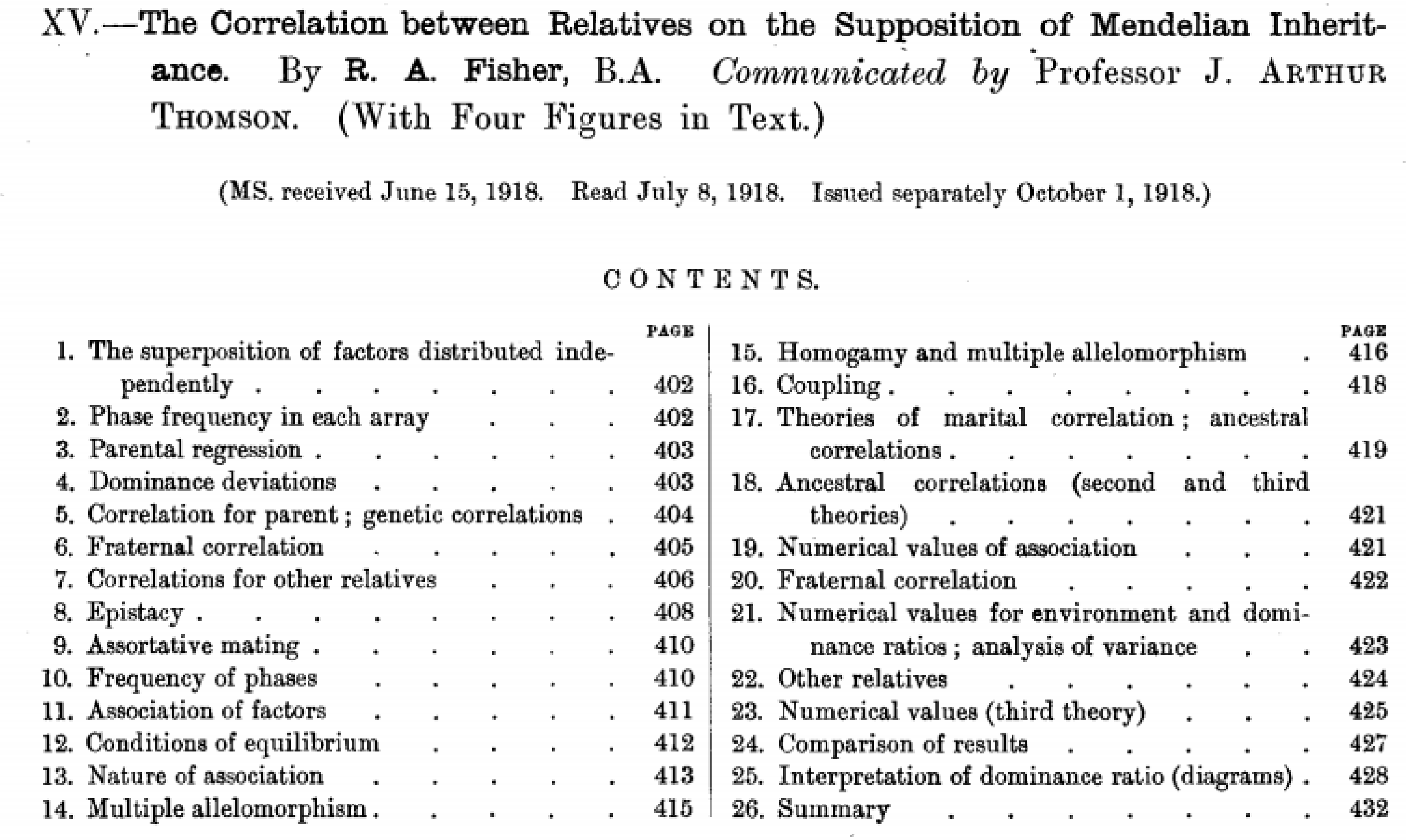
Introduction to Biometrics

- Overall, these studies by Galton and Pearson **suggested that continuous variation in humans is at least partially inherited**

Introduction to Biometrics

1918 - Fischer: To reconcile the conflicting views of the biometricians and Mendelists


Ronald Fisher in the 1950s


XV.—**The Correlation between Relatives on the Supposition of Mendelian Inheritance.** By **R. A. Fisher,** B.A. *Communicated by* Professor J. ARTHUR THOMSON. (With Four Figures in Text.)

(MS. received June 15, 1918. Read July 8, 1918. Issued separately October 1, 1918.)

CONTENTS.

	PAGE
1. The superposition of factors distributed independently	402
2. Phase frequency in each array	402
3. Parental regression	403
4. Dominance deviations	403
5. Correlation for parent; genetic correlations	404
6. Fraternal correlation	405
7. Correlations for other relatives	406
8. Epistacy	408
9. Assortative mating	410
10. Frequency of phases	410
11. Association of factors	411
12. Conditions of equilibrium	412
13. Nature of association	413
14. Multiple allelomorphism	415
15. Homogamy and multiple allelomorphism	416
16. Coupling	418
17. Theories of marital correlation; ancestral correlations	419
18. Ancestral correlations (second and third theories)	421
19. Numerical values of association	421
20. Fraternal correlation	422
21. Numerical values for environment and dominance ratios; analysis of variance	423
22. Other relatives	424
23. Numerical values (third theory)	425
24. Comparison of results	427
25. Interpretation of dominance ratio (diagrams)	428
26. Summary	432

Fisher, Ronald A. 1919 - Transactions of the Royal Society of Edinburg.

Introduction to Biometrics

1918 - Fischer

- Fisher presents the conceptual model of genetics that shows that continuous variation between phenotypic traits may be the result of Mendelian inheritance

- The paper also contains the first use of the statistical term **variance**

Fisher, Ronald A. 1919 - Transactions of the Royal Society of Edinburg

Introduction to Biometrics

Quantitative Genetics

A specific area in genetics that deals with inheritance and variation of quantitative traits, which are strongly affected by the environment

(Hallauer and Miranda Filho 1988, Vencovsky 1987).

- Has made an enormous contribution, generating valuable information placed at the service of breeders to obtain superior plants.

Introduction to Biometrics

Welcome to the universe of biometrics!

The large amount of **data** and the need for more **appropriate analyses** required a new area of science.

This new area is called "**biometry**";

Professionals acting in this area are called "**biostatisticians**"

Introduction to Biometrics

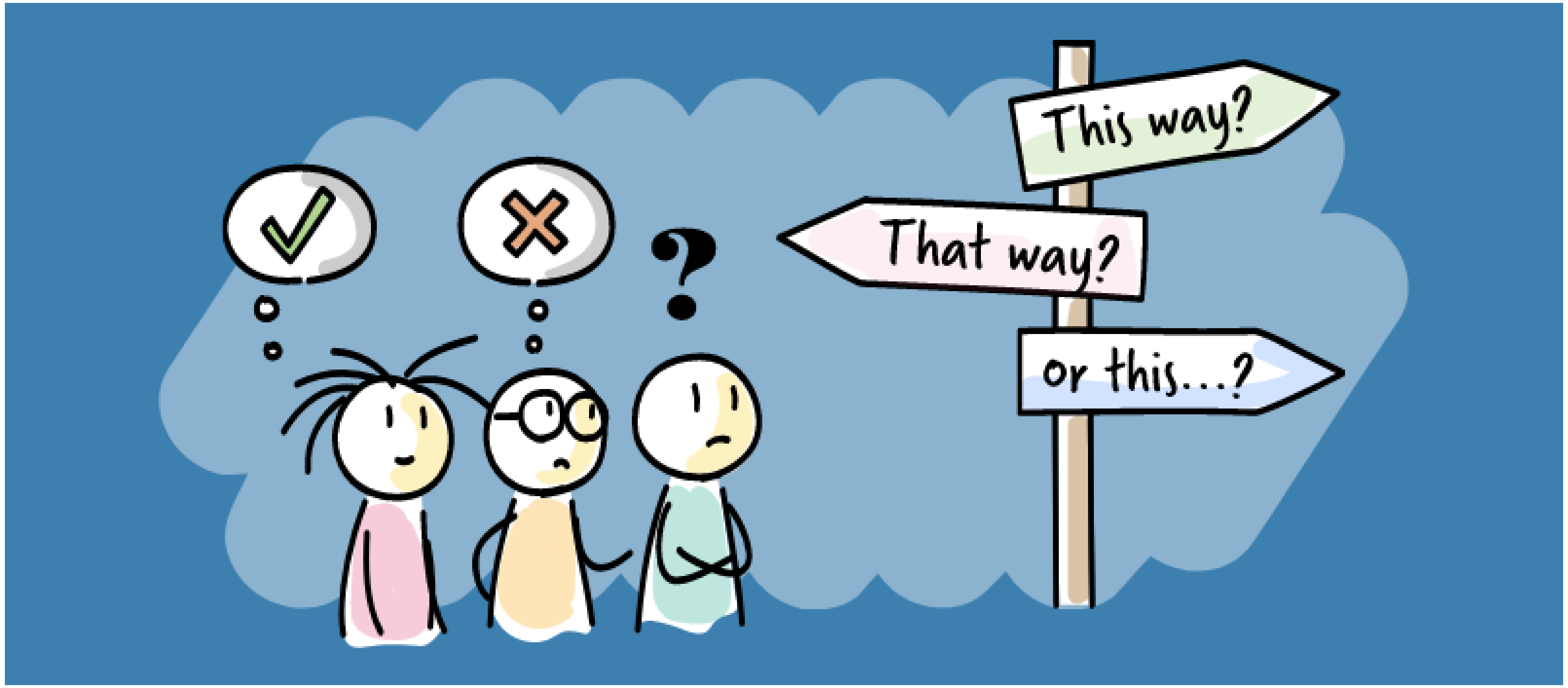
What is Biometry?

The application of methods statistics to the solution of biological problems (Sokal and Rohlf, 1969.)

Bios (“life”) and *metron* (“measure”).

Introduction to Biometrics

Decision Making!



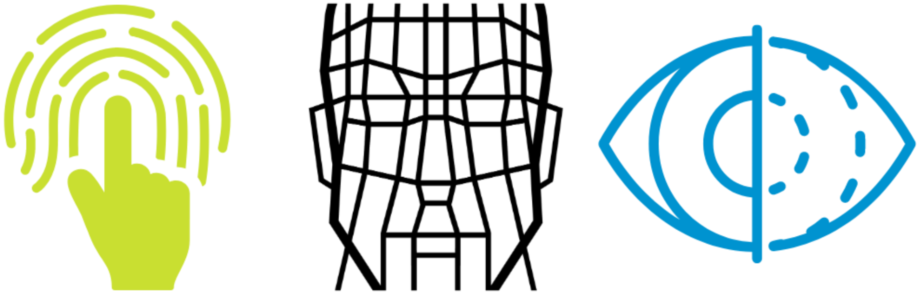
The Best Jobs in America for 2022

	Job Title	Median Base Salary	Job Satisfaction	Job Openings	
#1	Enterprise Architect	$144,997	4.1/5	14,021	View Jobs
#2	Full Stack Engineer	$101,794	4.3/5	11,252	View Jobs
#3	Data Scientist	$120,000	4.1/5	10,071	View Jobs
#4	Devops Engineer	$120,095	4.2/5	8,548	View Jobs
#5	Strategy Manager	$140,000	4.2/5	6,977	View Jobs
#6	Machine Learning Engineer	$130,489	4.3/5	6,801	View Jobs
#7	Data Engineer	$113,960	4.0/5	11,821	View Jobs
#8	Software Engineer	$116,638	3.9/5	64,155	View Jobs
#9	Java Developer	$107,099	4.1/5	10,201	View Jobs
#10	Product Manager	$125,317	4.0/5	17,725	View Jobs

https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm

Introduction to Biometrics

The term “**Biometrics**” has also been used to refer to the field of technology devoted to the **identification of individuals using biological traits**, such as those based on retinal or iris scanning, fingerprints, or face recognition.



Introduction to Biometrics

What is needed to study biometrics?

- Genetics;
- Experimental Statistics;
- Quantitative;
- Algebra;
- Programming;
- And others...

Biometry – Knowledge guiding the selection strategy

The basic principles of **experimentation**, such as the need to conduct trials with **replication**, **randomness**, and, if necessary, **local control**, led to a revolution by providing more accurate mean values, information regarding experimental accuracy, and estimates of genetic and environmental parameters useful in guiding breeding strategies.

This is complementation between **genetic** and **statistical principles** for a common purpose.

Introduction to Biometrics

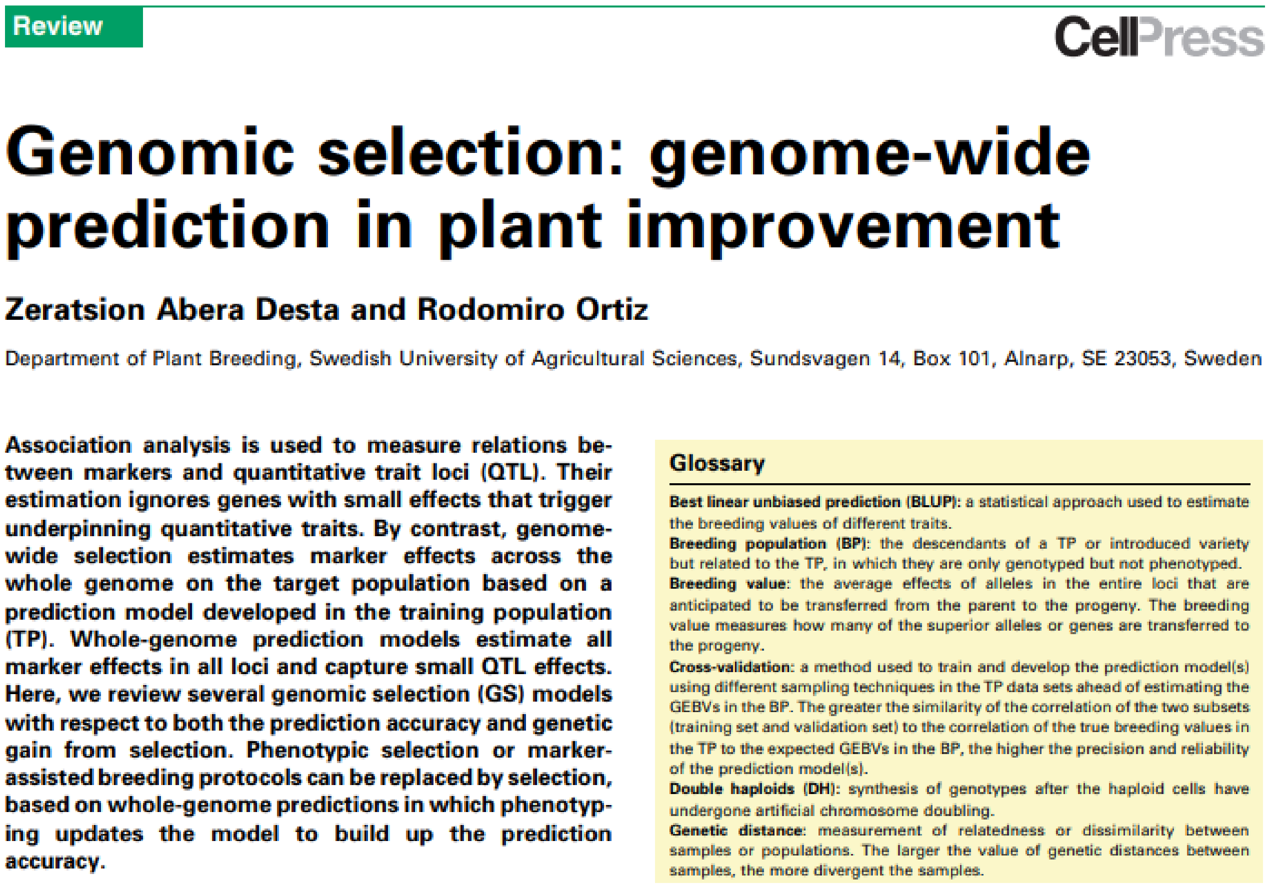
Biometry – Knowledge guiding the selection strategy

Breeding programs require intensive experimentation so that accurate genetic values can be obtained and used as selection criteria.

Introduction to Biometrics

The **genetics, experimental statistics and biometry** areas have contributed a large number of models, with the purpose of ensuring the **good choice** of **genetic material** and of **establishing a predicted value** of performance of the individual or of its descendant.

Biometry and advances: Molecular Genetics



Review

CellPress

Genomic selection: genome-wide prediction in plant improvement

Zeratsion Abera Desta and Rodomiro Ortiz

Department of Plant Breeding, Swedish University of Agricultural Sciences, Sundsvagen 14, Box 101, Alnarp, SE 23053, Sweden

Association analysis is used to measure relations between markers and quantitative trait loci (QTL). Their estimation ignores genes with small effects that trigger underpinning quantitative traits. By contrast, genome-wide selection estimates marker effects across the whole genome on the target population based on a prediction model developed in the training population (TP). Whole-genome prediction models estimate all marker effects in all loci and capture small QTL effects. Here, we review several genomic selection (GS) models with respect to both the prediction accuracy and genetic gain from selection. Phenotypic selection or marker-assisted breeding protocols can be replaced by selection, based on whole-genome predictions in which phenotyping updates the model to build up the prediction accuracy.

Glossary

Best linear unbiased prediction (BLUP): a statistical approach used to estimate the breeding values of different traits.

Breeding population (BP): the descendants of a TP or introduced variety but related to the TP, in which they are only genotyped but not phenotyped.

Breeding value: the average effects of alleles in the entire loci that are anticipated to be transferred from the parent to the progeny. The breeding value measures how many of the superior alleles or genes are transferred to the progeny.

Cross-validation: a method used to train and develop the prediction model(s) using different sampling techniques in the TP data sets ahead of estimating the GEBVs in the BP. The greater the similarity of the correlation of the two subsets (training set and validation set) to the correlation of the true breeding values in the TP to the expected GEBVs in the BP, the higher the precision and reliability of the prediction model(s).

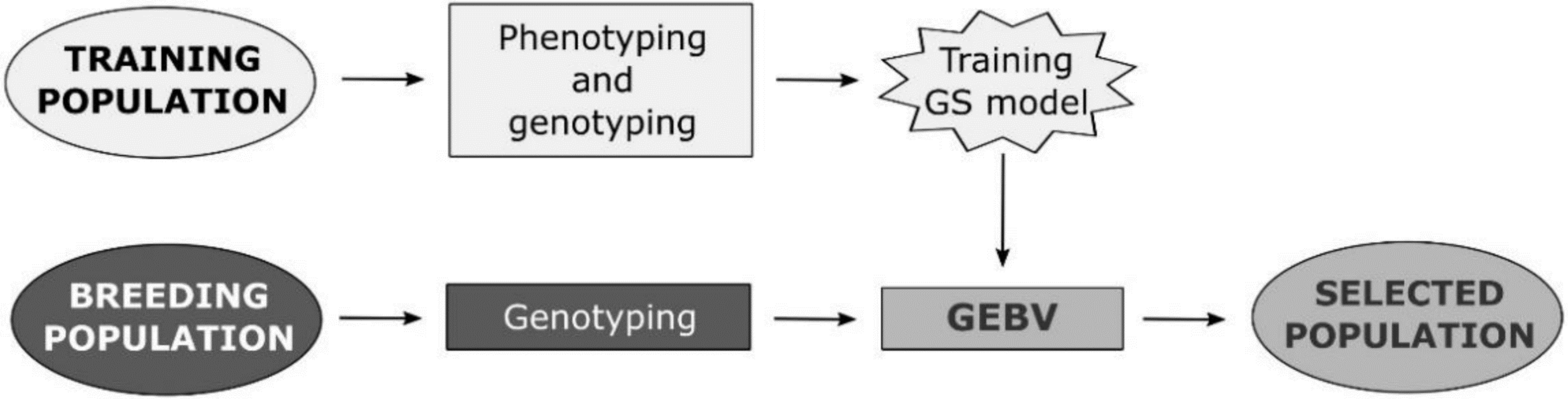
Double haploids (DH): synthesis of genotypes after the haploid cells have undergone artificial chromosome doubling.

Genetic distance: measurement of relatedness or dissimilarity between samples or populations. The larger the value of genetic distances between samples, the more divergent the samples.

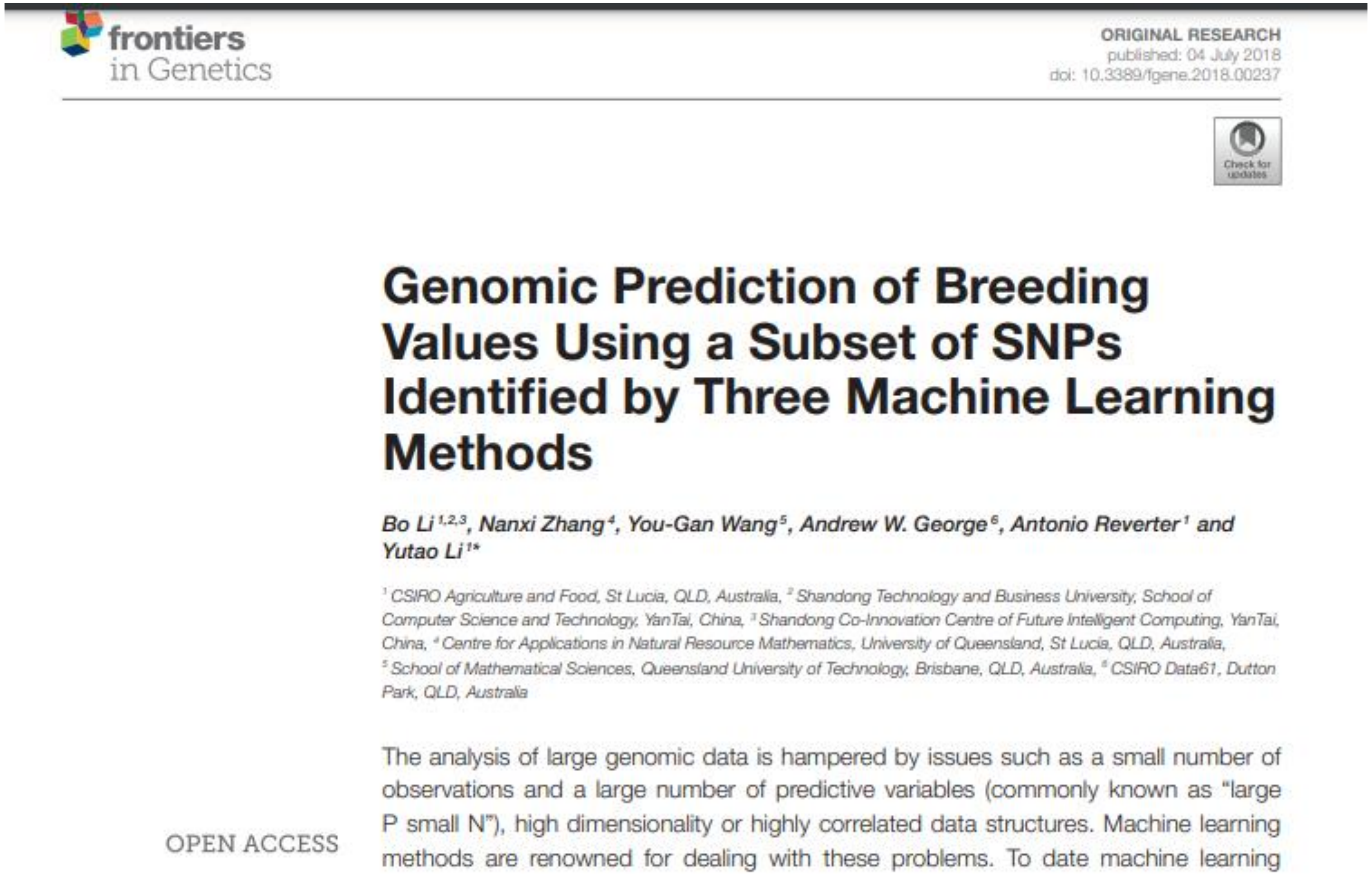
https://doi.org/10.1016/j.tplants.2014.05.006

Biometry and advances: Molecular Genetics

Genomic Selection (GS)



Biometry and advances: Machine Learning



frontiers in Genetics

ORIGINAL RESEARCH
published: 04 July 2018
doi: 10.3389/fgene.2018.00237

Check for updates

Genomic Prediction of Breeding Values Using a Subset of SNPs Identified by Three Machine Learning Methods

Bo Li [1,2,3]***, Nanxi Zhang*** [4]***, You-Gan Wang*** [5]***, Andrew W. George*** [6]***, Antonio Reverter*** [1] ***and Yutao Li*** [1*]

[1] CSIRO Agriculture and Food, St Lucia, QLD, Australia, [2] Shandong Technology and Business University, School of Computer Science and Technology, YanTai, China, [3] Shandong Co-Innovation Centre of Future Intelligent Computing, YanTai, China, [4] Centre for Applications in Natural Resource Mathematics, University of Queensland, St Lucia, QLD, Australia, [5] School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD, Australia, [6] CSIRO Data61, Dutton Park, QLD, Australia

OPEN ACCESS

The analysis of large genomic data is hampered by issues such as a small number of observations and a large number of predictive variables (commonly known as "large P small N"), high dimensionality or highly correlated data structures. Machine learning methods are renowned for dealing with these problems. To date machine learning

https://doi.org/10.3389/fgene.2018.00237

Introduction to Biometrics

Motivation

Biometry is an area of genetics in **continuous evolution**, which has contributed to generating information and finding solutions to diverse questions in **genetics and breeding**.

Introduction to Biometrics

Motivation

Biometry depends on the **critical and attentive eye of methods of data analysis** and on the algorithms that allow data processing supported by the biometric area.

Motivation

There is an urgent need for further advances, now supported by **biometric genetics**, in the context of data processing.

Introduction to Biometrics

References

Sokal, Robert R. "The principles and practice of statistics in biological research." Biometry (2012): 451-554.

Rencher, Alvin C., and G. Bruce Schaalje. Linear models in statistics. John Wiley & Sons, 2008.