LETTER

doi:10.1038/nature14317

Massive migration from the steppe was a source for Indo-European languages in Europe

Wolfgang Haak[1]*, Iosif Lazaridis[2,3]*, Nick Patterson[3], Nadin Rohland[2,3], Swapan Mallick[2,3,4], Bastien Llamas[1], Guido Brandt[5], Susanne Nordenfelt[2,3], Eadaoin Harney[2,3,4], Kristin Stewardson[2,3,4], Qiaomei Fu[2,3,6,7], Alissa Mittnik[8], Eszter Bánffy[9,10], Christos Economou[11], Michael Francken[12], Susanne Friederich[13], Rafael Garrido Pena[14], Fredrik Hallgren[15], Valery Khartanovich[16], Aleksandr Khokhlov[17], Michael Kunst[18], Pavel Kuznetsov[17], Harald Meller[13], Oleg Mochalov[17], Vayacheslav Moiseyev[16], Nicole Nicklisch[5,13,19], Sandra L. Pichler[20], Roberto Risch[21], Manuel A. Rojo Guerra[22], Christina Roth[5], Anna Szécsényi-Nagy[5,9], Joachim Wahl[23], Matthias Meyer[6], Johannes Krause[8,12,24], Dorcas Brown[25], David Anthony[25], Alan Cooper[1], Kurt Werner Alt[5,13,19,20] & David Reich[2,3,4]

We generated genome-wide data from 69 Europeans who lived between 8,000–3,000 years ago by enriching ancient DNA libraries for a target set of almost 400,000 polymorphisms. Enrichment of these positions decreases the sequencing required for genome-wide ancient DNA analysis by a median of around 250-fold, allowing us to study an order of magnitude more individuals than previous studies[1–8] and to obtain new insights about the past. We show that the populations of Western and Far Eastern Europe followed opposite trajectories between 8,000–5,000 years ago. At the beginning of the Neolithic period in Europe, ~8,000–7,000 years ago, closely related groups of early farmers appeared in Germany, Hungary and Spain, different from indigenous hunter-gatherers, whereas Russia was inhabited by a distinctive population of hunter-gatherers with high affinity to a ~24,000-year-old Siberian[6]. By ~6,000–5,000 years ago, farmers throughout much of Europe had more hunter-gatherer ancestry than their predecessors, but in Russia, the Yamnaya steppe herders of this time were descended not only from the preceding eastern European hunter-gatherers, but also from a population of Near Eastern ancestry. Western and Eastern Europe came into contact ~4,500 years ago, as the Late Neolithic Corded Ware people from Germany traced ~75% of their ancestry to the Yamnaya, documenting a massive migration into the heartland of Europe from its eastern periphery. This steppe ancestry persisted in all sampled central Europeans until at least ~3,000 years ago, and is ubiquitous in present-day Europeans. These results provide support for a steppe origin[9] of at least some of the Indo-European languages of Europe.

Genome-wide analysis of ancient DNA has emerged as a transformative technology for studying prehistory, providing information that is comparable in power to archaeology and linguistics. Realizing its promise, however, requires collecting genome-wide data from an adequate number of individuals to characterize population changes over time, which means not only sampling a succession of archaeological cultures[2], but also multiple individuals per culture. To make analysis of large numbers of ancient DNA samples practical, we used in-solution hybridization capture[10,11] to enrich next generation sequencing libraries for a target set of 394,577 single nucleotide polymorphisms (SNPs) ('390k capture'), 354,212 of which are autosomal SNPs that have also been genotyped using the Affymetrix Human Origins array in 2,345 humans from 203 populations[4,12]. This reduces the amount of sequencing required to obtain genome-wide data by a minimum of 45-fold and a median of 262-fold (Supplementary Data 1). This strategy allows us to report genomic scale data on more than twice the number of ancient Eurasians as has been presented in the entire preceding literature[1–8] (Extended Data Table 1).

We used this technology to study population transformations in Europe. We began by preparing 212 DNA libraries from 119 ancient samples in dedicated clean rooms, and testing these by light shotgun sequencing and mitochondrial genome capture (Supplementary Information section 1, Supplementary Data 1). We restricted the analysis to libraries with molecular signatures of authentic ancient DNA (elevated damage in the terminal nucleotide), negligible evidence of contamination based on mismatches to the mitochondrial consensus[13] and, where available, a mitochondrial DNA haplogroup that matched previous results using PCR[4,14,15] (Supplementary Information section 2). For 123 libraries prepared in the presence of uracil-DNA-glycosylase[16] to reduce errors due to ancient DNA damage[17], we performed 390k capture, carried out paired-end sequencing and mapped the data to the human genome. We restricted analysis to 94 libraries from 69 samples that had at least 0.06-fold average target coverage (average of 3.8-fold) and used majority rule to call an allele at each SNP covered at least once (Supplementary Data 1). After combining our data (Supplementary Information section 3) with 25 ancient samples from the literature — three Upper Paleolithic samples from Russia[1,6,7], seven people of European hunter-gatherer ancestry[2,4,5,8], and fifteen European farmers[2,3,4,8] — we had data from 94 ancient Europeans. Geographically, these came from Germany ($n = 41$), Spain ($n = 10$), Russia ($n = 14$), Sweden ($n = 12$), Hungary ($n = 15$), Italy ($n = 1$) and Luxembourg ($n = 1$) (Extended Data Table 2). Following the central European chronology, these included 19 hunter-gatherers (~43,000–2,600 BC), 28 Early Neolithic farmers (~6,000–4,000 BC), 11 Middle Neolithic farmers (~4,000–3,000 BC) including

[1]Australian Centre for Ancient DNA, School of Earth and Environmental Sciences & Environment Institute, University of Adelaide, Adelaide, South Australia 5005, Australia. [2]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. [4]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. [5]Institute of Anthropology, Johannes Gutenberg University of Mainz, D-55128 Mainz, Germany. [6]Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany. [7]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing 100049, China. [8]Institute for Archaeological Sciences, University of Tübingen, D-72070 Tübingen, Germany. [9]Institute of Archaeology, Research Centre for the Humanities, Hungarian Academy of Science, H-1014 Budapest, Hungary. [10]Römisch Germanische Kommission (RGK) Frankfurt, D-60325 Frankfurt, Germany. [11]Archaeological Research Laboratory, Stockholm University, 114 18 Stockholm, Sweden. [12]Departments of Paleoanthropology and Archaeogenetics, Senckenberg Center for Human Evolution and Paleoenvironment, University of Tübingen, D-72070 Tübingen, Germany. [13]State Office for Heritage Management and Archaeology Saxony-Anhalt and State Museum of Prehistory, D-06114 Halle, Germany. [14]Departamento de Prehistoria y Arqueología, Facultad de Filosofía y Letras, Universidad Autónoma de Madrid, E-28049 Madrid, Spain. [15]The Cultural Heritage Foundation, Västerås 722 12, Sweden. [16]Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS, St Petersburg 199034, Russia. [17]Volga State Academy of Social Sciences and Humanities, Samara 443099, Russia. [18]Deutsches Archaeologisches Institut, Abteilung Madrid, E-28002 Madrid, Spain. [19]Danube Private University, A-3500 Krems, Austria. [20]Institute for Prehistory and Archaeological Science, University of Basel, CH-4003 Basel, Switzerland. [21]Departamento de Prehistòria, Universitat Autònoma de Barcelona, E-08193 Barcelona, Spain. [22]Departamento de Prehistòria y Arqueolgia, Universidad de Valladolid, E-47002 Valladolid, Spain. [23]State Office for Cultural Heritage Management Baden-Württemberg, Osteology, D-78467 Konstanz, Germany. [24]Max Planck Institute for the Science of Human History, D-07745 Jena, Germany. [25]Anthropology Department, Hartwick College, Oneonta, New York 13820, USA.
*These authors contributed equally to this work.

©2015 Macmillan Publishers Limited. All rights reserved

the Tyrolean Iceman[3], 9 Late Copper/Early Bronze Age individuals (Yamnaya: ~3,300–2,700 BC), 15 Late Neolithic individuals (~2,500–2,200 BC), 9 Early Bronze Age individuals (~2,200–1,500 BC), two Late Bronze Age individuals (~1,200–1,100 BC) and one Iron Age individual (~900 BC). Two individuals were excluded from analyses as they were related to others from the same population. The average number of SNPs covered at least once was 212,375 and the minimum was 22,869 (Fig. 1).

We determined that 34 of the 69 newly analysed individuals were male and used 2,258 Y chromosome SNPs targets included in the capture to obtain high resolution Y chromosome haplogroup calls (Supplementary Information section 4). Outside Russia, and before the Late Neolithic period, only a single R1b individual was found (early Neolithic Spain) in the combined literature ($n = 70$). By contrast, haplogroups R1a and R1b were found in 60% of Late Neolithic/Bronze Age Europeans outside Russia ($n = 10$), and in 100% of the samples from European Russia from all periods (7,500–2,700 BC; $n = 9$). R1a and R1b are the most common haplogroups in many European populations today[18,19], and our results suggest that they spread into Europe from the East after 3,000 BC. Two hunter-gatherers from Russia included in our study belonged to R1a (Karelia) and R1b (Samara), the earliest documented ancient samples of either haplogroup discovered to date. These two hunter-gatherers did not belong to the derived lineages M417 within R1a and M269 within R1b that are predominant in Europeans today[18,19], but all 7 Yamnaya males did belong to the M269 subclade[18] of haplogroup R1b.

Principal components analysis (PCA) of all ancient individuals along with 777 present-day West Eurasians[4] (Fig. 2a, Supplementary Information section 5) replicates the positioning of present-day Europeans between the Near East and European hunter-gatherers[4,20], and the clustering of early farmers from across Europe with present day Sardinians[3,4], suggesting that farming expansions across the Mediterranean to Spain and via the Danubian route to Hungary and Germany descended from a common stock. By adding samples from later periods and additional locations, we also observe several new patterns. All samples from Russia have affinity to the ~24,000-year-old MA1 (ref. 6), the type specimen for the Ancient North Eurasians (ANE) who contributed to both Europeans[4] and Native Americans[4,6,8]. The two hunter-gatherers from Russia (Karelia in the northwest of the country and Samara on the steppe near the Urals) form an 'eastern European hunter-gatherer' (EHG) cluster at one end of a hunter-gatherer cline across Europe; people of hunter-gatherer ancestry from Luxembourg, Spain, and Hungary sit at the opposite 'western European hunter-gatherer'[4] (WHG) end, while the hunter-gatherers from Sweden[4,8] (SHG) are intermediate. Against this background of differentiated European hunter-gatherers and homogeneous early farmers, multiple population turnovers transpired in all parts of Europe included in our study. Middle Neolithic Europeans from Germany, Spain, Hungary, and Sweden from the period ~4,000–3,000 BC are intermediate between the earlier farmers and the WHG, suggesting an increase of WHG ancestry throughout much of Europe. By contrast, in Russia, the later Yamnaya steppe herders of ~3,000 BC plot between the EHG and the present-day Near East/Caucasus, suggesting a decrease of EHG ancestry during the same time period. The Late Neolithic and Bronze Age samples from Germany and Hungary[2] are distinct from the preceding Middle Neolithic and plot between them and the Yamnaya. This pattern is also seen in ADMIXTURE analysis (Fig. 2b, Supplementary Information section 6), which implies that the Yamnaya have ancestry from populations related to the Caucasus and South Asia that is largely absent in 38 Early or Middle Neolithic farmers but present in all 25 Late Neolithic or Bronze Age individuals. This ancestry appears in Central Europe for the first time in our series with the Corded Ware around 2,500 BC (Supplementary Information section 6, Fig. 2b). The Corded Ware shared elements of material culture with steppe groups such as the Yamnaya although whether this reflects movements of people has been contentious[21]. Our genetic data provide direct evidence of migration and suggest that it was relatively sudden. The Corded Ware are genetically closest to the Yamnaya ~2,600 km away, as inferred both from PCA and ADMIXTURE (Fig. 2) and F_{ST} (0.011 ± 0.002) (Extended Data Table 3). If continuous gene flow from the east, rather than migration, had occurred, we would expect successive cultures in Europe to become increasingly differentiated from the Middle Neolithic, but

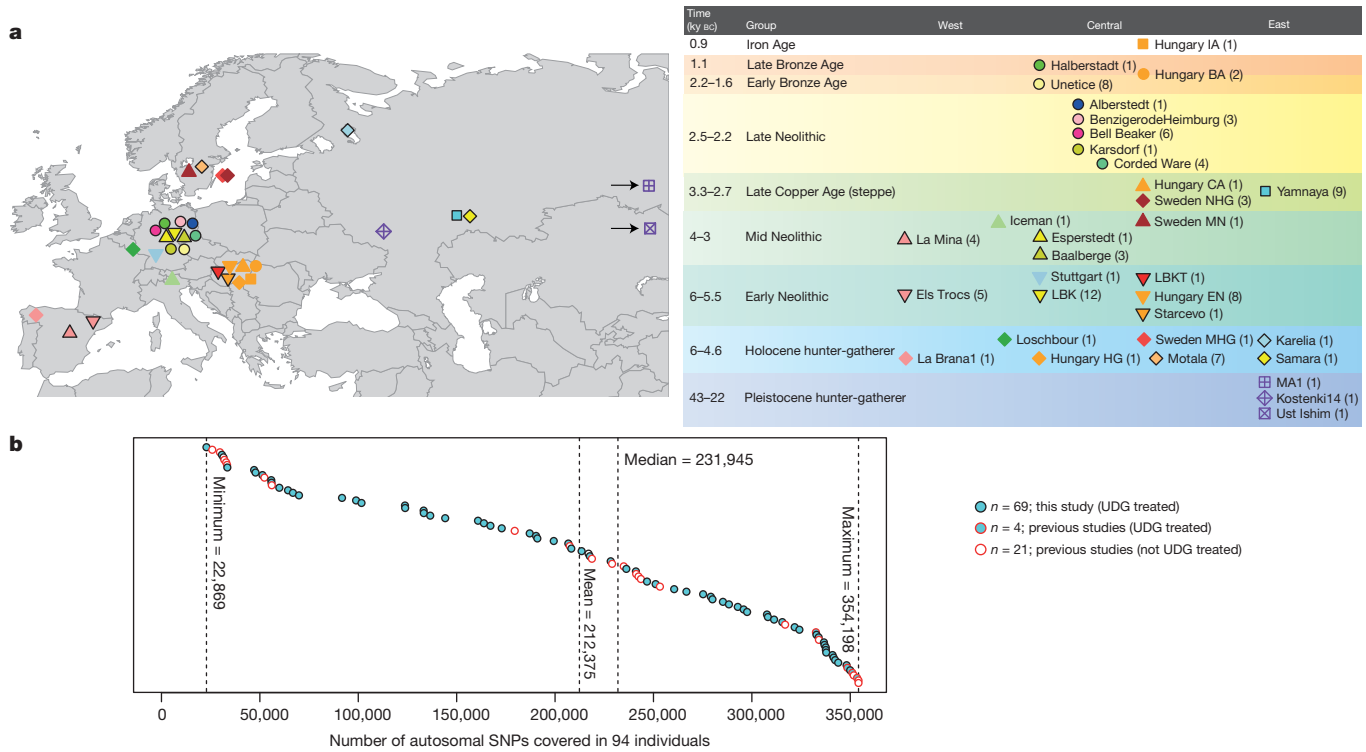


Figure 1 | Location and SNP coverage of samples included in this study. **a**, Geographic location and time-scale (central European chronology) of the 69 newly analysed ancient individuals from this study (black outline) and 25 from the literature for which shotgun sequencing data was available (no outline). **b**, Number of SNPs covered at least once in the analysis data set of 94 individuals.

©2015 Macmillan Publishers Limited. All rights reserved

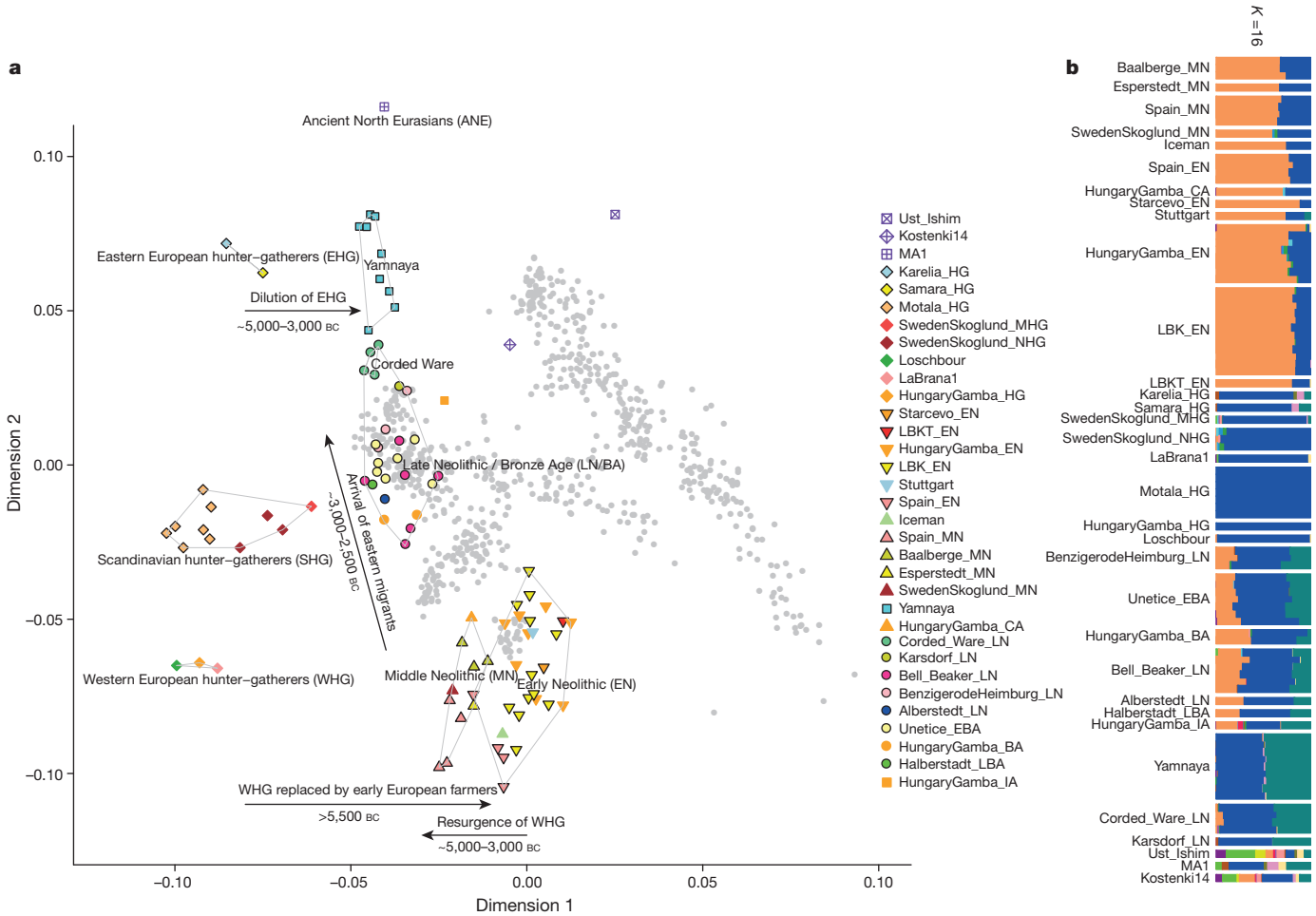


Figure 2 | Population transformations in Europe. **a**, PCA analysis. **b**, ADMIXTURE analysis. The full ADMIXTURE analysis including present-day humans is shown in Supplementary Information section 6.

instead, the Corded Ware are both the earliest and most strongly differentiated from the Middle Neolithic population.

'Outgroup' f_3 statistics[6] (Supplementary Information section 7), which measure shared genetic drift between a pair of populations (Extended Data Fig. 1), support the clustering of hunter-gatherers, Early/Middle Neolithic, and Late Neolithic/Bronze Age populations into different groups as in the PCA (Fig. 2a). We also analysed f_4 statistics, which allow us to test whether pairs of populations are consistent with descent from common ancestral populations, and to assess significance using a normally distributed Z score. Early European farmers from the Early and Middle Neolithic were closely related but not identical. This is reflected in the fact that Loschbour, a WHG individual from Luxembourg[4], shared more alleles with post-4,000 BC European farmers from Germany, Spain, Hungary, Sweden and Italy than with early farmers of Germany, Spain, and Hungary, documenting an increase of hunter-gatherer ancestry in multiple regions of Europe during the course of the Neolithic. The two EHG form a clade with respect to all other present-day and ancient populations ($|Z| < 1.9$), and MA1 shares more alleles with them ($|Z| > 4.7$) than with other ancient or modern populations, suggesting that they may be a source for the ANE ancestry in present Europeans[4,12,22] as they are geographically and temporally more proximate than Upper Paleolithic Siberians. The Yamnaya differ from the EHG by sharing fewer alleles with MA1 ($|Z| = 6.7$) suggesting a dilution of ANE ancestry between 5,000–3,000 BC on the European steppe. This was likely due to admixture of EHG with a population related to present-day Near Easterners, as the most negative f_3 statistic in the Yamnaya (giving unambiguous evidence of admixture) is observed when we model them as a mixture of EHG and present-day Near Eastern populations like Armenians ($Z = -6.3$; Supplementary Information section 7). The Late Neolithic/Bronze Age groups of central Europe share more alleles with Yamnaya than the Middle Neolithic populations do ($|Z| = 12.4$) and more alleles with the Middle Neolithic than the Yamnaya do ($|Z| = 12.5$), and have a negative f_3 statistic with the Middle Neolithic and Yamnaya as references ($Z = -20.7$), indicating that they were descended from a mixture of the local European populations and new migrants from the east. Moreover, the Yamnaya share more alleles with the Corded Ware ($|Z| \geq 3.6$) than with any other Late Neolithic/Early Bronze Age group with at least two individuals (Supplementary Information section 7), indicating that they had more eastern ancestry, consistent with the PCA and ADMIXTURE patterns (Fig. 2).

Modelling of the ancient samples shows that while Karelia is genetically intermediate between Loschbour and MA1, the topology that considers Karelia as a mixture of these two elements is not the only one that can fit the data (Supplementary Information section 8). To avoid biasing our inferences by fitting an incorrect model, we developed new statistical methods that are substantial extensions of a previously reported approach[4], which allow us to obtain precise estimates of the proportion of mixture in later Europeans without requiring a formal model for the relationship among the ancestral populations. The method (Supplementary Information section 9) is based on the idea that if a Test population has ancestry related to reference populations Ref_1, Ref_2, …, Ref_N in proportions α_1, α_2, ..., α_N, and the references are themselves differentially related to a triple of outgroup populations A, B, C, then:

$$f_4(\mathrm{Test},A;B,C) = \sum_{i=1}^{N} a_i f_4(\mathrm{Ref}_i,A;B,C)$$

©2015 Macmillan Publishers Limited. All rights reserved

By using a large number of outgroup populations we can fit the admixture coefficients α_i and estimate mixture proportions (Supplementary Information section 9, Extended Data Fig. 2). Using 15 outgroups from Africa, Asia, Oceania and the Americas, we obtain good fits as assessed by a formal test (Supplementary Information section 10), and estimate that the Middle Neolithic populations of Germany and Spain have ~18–34% more WHG-related ancestry than Early Neolithic populations and that the Late Neolithic and Early Bronze Age populations of Germany have ~22–39% more EHG-related ancestry than the Middle Neolithic ones (Supplementary Information section 9). If we model them as mixtures of Yamnaya-related and Middle Neolithic populations, the inferred degree of population turnover is doubled to 48–80% (Supplementary Information sections 9 and 10).

To distinguish whether a Yamnaya or an EHG source fits the data better, we added ancient samples as outgroups (Supplementary Information section 9). Adding any Early or Middle Neolithic farmer results in EHG-related genetic input into Late Neolithic populations being a poor fit to the data (Supplementary Information section 9); thus, Late Neolithic populations have ancestry that cannot be explained by a mixture of EHG and Middle Neolithic. When using Yamnaya instead of EHG, however, we obtain a good fit (Supplementary Information sections 9 and 10). These results can be explained if the new genetic material that arrived in Germany was a composite of two elements: EHG and a type of Near Eastern ancestry different from that which was introduced by early farmers (also suggested by PCA and ADMIXTURE; Fig. 2, Supplementary Information sections 5 and 6). We estimate that these two elements each contributed about half the ancestry of the Yamnaya (Supplementary Information sections 6 and 9), explaining why the population turnover inferred using Yamnaya as a source is about twice as high compared to the undiluted EHG. The estimate of Yamnaya-related ancestry in the Corded Ware is consistent when using either present populations or ancient Europeans as outgroups (Supplementary Information sections 9 and 10), and is 73.1 ± 2.2% when both sets are combined (Supplementary Information section 10). The best proxies for ANE ancestry in Europe[4] were initially Native Americans[12,22], and then the Siberian MA1 (ref. 6), but both are geographically and temporally too remote for what appears to be a recent migration into Europe[4]. We can now add three new pieces to the puzzle of how ANE ancestry was transmitted to Europe: first by the EHG, then the Yamnaya formed by mixture between EHG and a Near Eastern related population, and then the Corded Ware who were formed by a mixture of the Yamnaya with Middle Neolithic Europeans. We caution that the sampled Yamnaya individuals from Samara might not be directly ancestral to Corded Ware individuals from Germany. It is possible that a more western Yamnaya population, or an earlier (pre-Yamnaya) steppe population may have migrated into central Europe, and future work may uncover more missing links in the chain of transmission of steppe ancestry.

By extending our model to a three-way mixture of WHG, Early Neolithic and Yamnaya, we estimate that the ancestry of the Corded Ware was 79% Yamnaya-like, 4% WHG, and 17% Early Neolithic (Fig. 3). A small contribution of the first farmers is also consistent with uniparentally inherited DNA: for example, mitochondrial DNA haplogroup N1a and Y chromosome haplogroup G2a, common in early central European farmers[14,23], almost disappear during the Late Neolithic and Bronze Age, when they are largely replaced by Y haplogroups R1a and R1b (Supplementary Information section 4) and mtDNA haplogroups I, T1, U2, U4, U5a, W, and subtypes of H[14,23,24] (Supplementary Information section 2). The uniparental data not only confirm a link to the steppe populations but also suggest that both sexes participated in the migrations (Supplementary Information sections 2 and 4 and Extended Data Table 2). The magnitude of the population turnover that occurred becomes even more evident if one considers the fact that the steppe migrants may well have mixed with eastern European agriculturalists on their way to central Europe. Thus, we cannot exclude a scenario in which the Corded Ware arriving in today's Germany had no ancestry at all from local populations.

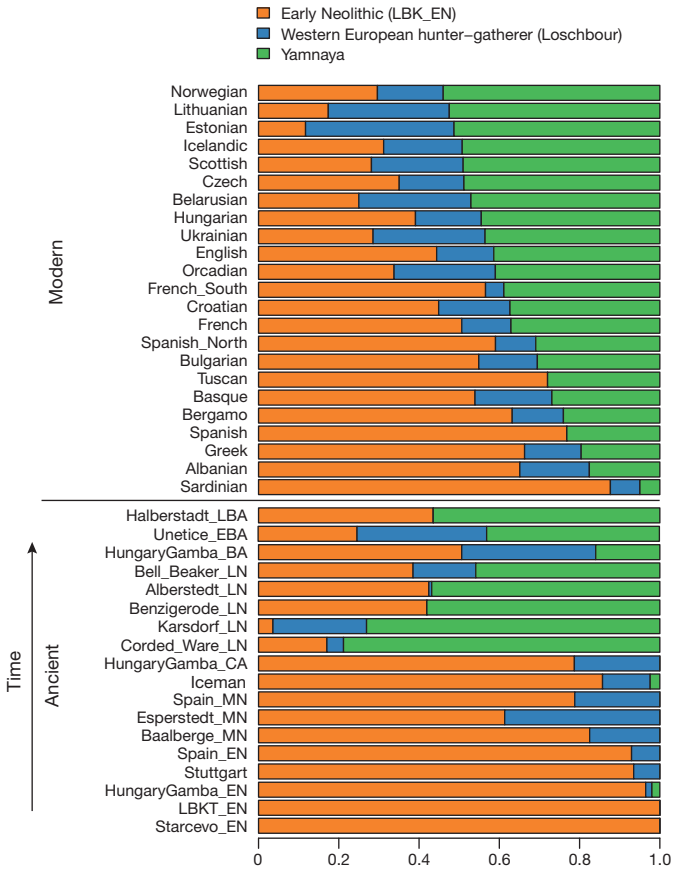


Figure 3 | Admixture proportions. We estimate mixture proportions using a method that gives unbiased estimates even without an accurate model for the relationships between the test populations and the outgroup populations (Supplementary Information section 9). Population samples are grouped according to chronology (ancient) and Yamnaya ancestry (present-day humans).

Our results support a view of European pre-history punctuated by two major migrations: first, the arrival of the first farmers during the Early Neolithic from the Near East, and second, the arrival of Yamnaya pastoralists during the Late Neolithic from the steppe. Our data further show that both migrations were followed by resurgences of the previous inhabitants: first, during the Middle Neolithic, when hunter-gatherer ancestry rose again after its Early Neolithic decline, and then between the Late Neolithic and the present, when farmer and hunter-gatherer ancestry rose after its Late Neolithic decline. This second resurgence must have started during the Late Neolithic/Bronze Age period itself, as the Bell Beaker and Unetice groups had reduced Yamnaya ancestry compared to the earlier Corded Ware, and comparable levels to that in some present-day Europeans (Fig. 3). Today, Yamnaya related ancestry is lower in southern Europe and higher in northern Europe, and all European populations can be modelled as a three-way mixture of WHG, Early Neolithic, and Yamnaya, whereas some outlier populations show evidence for additional admixture with populations from Siberia and the Near East (Extended Data Fig. 3, Supplementary Information section 9). Further data are needed to determine whether the steppe ancestry arrived in southern Europe at the time of the Late Neolithic/Bronze Age, or is due to migrations in later times from northern Europe[25,26].

Our results provide new data relevant to debates on the origin and expansion of Indo-European languages in Europe (Supplementary Information section 11). Although the findings from ancient DNA are silent on the question of the languages spoken by preliterate populations, they do carry evidence about processes of migration which are invoked by theories on Indo-European language dispersals. Such theories make predictions about movements of people to account for the spread of

©2015 Macmillan Publishers Limited. All rights reserved

languages and material culture (Extended Data Fig. 4). The technology of ancient DNA makes it possible to reject or confirm the proposed migratory movements, as well as to identify new movements that were not previously known. The best argument for the 'Anatolian hypothesis'[27] that Indo-European languages arrived in Europe from Anatolia ~8,500 years ago is that major language replacements are thought to require major migrations, and that after the Early Neolithic when farmers established themselves in Europe, the population base was likely to have been so large that later migrations would not have made much of an impact[27,28]. However, our study shows that a later major turnover did occur, and that steppe migrants replaced ~75% of the ancestry of central Europeans. An alternative theory is the 'steppe hypothesis', which proposes that early Indo-European speakers were pastoralists of the grasslands north of the Black and Caspian Seas, and that their languages spread into Europe after the invention of wheeled vehicles[9]. Our results make a compelling case for the steppe as a source of at least some of the Indo-European languages in Europe by documenting a massive migration ~4,500 years ago associated with the Yamnaya and Corded Ware cultures, which are identified by proponents of the steppe hypothesis as vectors for the spread of Indo-European languages into Europe. These results challenge the Anatolian hypothesis by showing that not all Indo-European languages in Europe can plausibly derive from the first farmer migrations thousands of years earlier (Supplementary Information section 11). We caution that the location of the proto-Indo-European[9,27,29,30] homeland that also gave rise to the Indo-European languages of Asia, as well as the Indo-European languages of southeastern Europe, cannot be determined from the data reported here (Supplementary Information section 11). Studying the mixture in the Yamnaya themselves, and understanding the genetic relationships among a broader set of ancient and present-day Indo-European speakers, may lead to new insight about the shared homeland.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 December 2014; accepted 12 February 2015.
Published online 2 March 2015.

1. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514,** 445–449 (2014).
2. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nature Commun.* **5,** 5257 (2014).
3. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature Commun.* **3,** 698 (2012).
4. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,** 409–413 (2014).
5. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507,** 225–228 (2014).
6. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505,** 87–91 (2014).
7. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346,** 1113–1118 (2014).
8. Skoglund, P. *et al.* Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344,** 747–750 (2014).
9. Anthony, D. W. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* (Princeton Univ. Press, 2007).
10. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110,** 2223–2227 (2013).
11. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. Lond. B* **370,** 20130624 (2015).
12. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192,** 1065–1093 (2012).
13. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23,** 553–559 (2013).
14. Brandt, G. *et al.* Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* **342,** 257–261 (2013).
15. Der Sarkissian, C. *et al.* Ancient DNA reveals prehistoric gene-flow from Siberia in the complex human population history of North East Europe. *PLoS Genet.* **9,** e1003296 (2013).
16. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Res.* **38,** e87 (2010).
17. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104,** 14616–14621 (2007).
18. Myres, N. M. *et al.* A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur. J. Hum. Genet.* **19,** 95–101 (2011).
19. Underhill, P. A. *et al.* The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur. J. Hum. Genet.* **23,** 124–131 (2015).
20. Skoglund, P. *et al.* Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336,** 466–469 (2012).
21. Czebreszuk, J. in *Ancient Europe, 8000 B.C. to A.D. 1000: Encyclopedia of the Barbarian World* (eds Bogucki, P. I. & Crabtree, P. J.) 467–475 (Charles Scribners & Sons, 2003).
22. Lipson, M. *et al.* Efficient moment-based inference of admixture parameters and sources of gene flow. *Mol. Biol. Evol.* **30,** 1788–1802 (2013).
23. Szécsényi-Nagy, A. *et al.* Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. Preprint at *bioRxiv* http://dx.doi.org/10.1101/008664 (2014).
24. Haak, W. *et al.* Ancient DNA from European early Neolithic farmers reveals their Near Eastern affinities. *PLoS Biol.* **8,** e1000536 (2010).
25. Hellenthal, G. *et al.* A genetic atlas of human admixture history. *Science* **343,** 747–751 (2014).
26. Ralph, P. & Coop, G. The geography of recent genetic ancestry across Europe. *PLoS Biol.* **11,** e1001555 (2013).
27. Renfrew, C. *Archaeology and Language: The Puzzle of Indo-European Origins* (Pimlico, 1987).
28. Bellwood, P. *First Farmers: The Origins of Agricultural Societies* (Wiley-Blackwell, 2004).
29. Gamkrelidze, T. V. & Ivanov, V. V. The early history of Indo-European languages. *Sci. Am.* **262,** 110–116 (1990).
30. Mallory, J. P. *In Search of the Indo-Europeans: Language, Archaeology and Myth* (Thames and Hudson, 1991).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Bellwood, J. Burger, P. Heggarty, M. Lipson, C. Renfrew, J. Diamond, S.Pääbo, R. Pinhasi and P. Skoglund for critical comments, and the Initiative for the Science of the Human Past at Harvard for organizing a workshop around the issues touched on by this paper. We thank S. Pääbo for support for establishing the ancient DNA facilities in Boston, and P. Skoglund for detecting the presence of two related individuals in our data set. We thank L. Orlando, T. S. Korneliussen, and C. Gamba for help in obtaining data. We thank Agilent Technologies and G. Frommer for help in developing the capture reagents. We thank C. Der Sarkissian, G. Valverde, L. Papac and B. Nickel for wet laboratory support. We thank archaeologists V. Dresely, R. Ganslmeier, O. Balanvosky, J. Ignacio Royo Guillén, A. Osztás, V. Majerik, T. Paluch, K. Somogyi and V.Voicsek for sharing samples and discussion about archaeological context. This research was supported by an Australian Research Council grant to W.H. and B.L. (DP130102158), and German Research Foundation grants to K.W.A. (Al 287/7-1 and 7-3, Al 287/10-1 and Al 287/14-1) and to H.M. (Me 3245/1-1 and 1-3). D.R. was supported by US National Science Foundation HOMINID grant BCS-1032255, US National Institutes of Health grant GM100233, and the Howard Hughes Medical Institute.

Author Contributions W.H., N.P., N.R., J.K., K.W.A. and D.R. supervised the study. W.H., E.B., C.E., M.F., S.F., R.G.P., F.H., V.K., A.K., M.K., P.K., H.M., O.M., V.M., N.N., S.L.P., R.R., M.A.R.G., C.R., A.S.-N., J.W., J.K., D.B., D.A., A.C., K.W.A. and D.R. assembled archaeological material, W.H., I.L., N.P., N.R., S.M., A.M. and D.R. analysed genetic data. I.L., N.P. and D.R. developed methods using *f* statistics for inferring admixture proportions. W.H., N.R., B.L., G.B., S.N., E.H., K.S. and A.M. performed wet laboratory ancient DNA work. I.L., N.R., S.M., B.L., Q.F., M.M. and D.R. developed the 390k capture reagent. W.H., I.L. and D.R. wrote the manuscript with help from all co-authors.

Author Information The aligned sequences are available through the European Nucleotide Archive under accession number PRJEB8448. The Human Origins genotype dataset including ancient individuals can be found at (http://genetics.med.harvard.edu/reichlab/Reich_Lab/Datasets.html). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.R. (reich@genetics.med.harvard.edu).

©2015 Macmillan Publishers Limited. All rights reserved

METHODS

Screening of libraries (shotgun sequencing and mitochondrial capture). The 212 libraries screened in this study (Supplementary Information section 1) from a total of 119 samples (Supplementary Information section 3) were produced at Adelaide ($n = 151$), Tübingen ($n = 16$), and Boston ($n = 45$) (Supplementary Data 1).

The libraries from Adelaide and Boston had internal barcodes directly attached to both sides of the molecules from the DNA extract so that each sequence begins with the barcode[10]. The Adelaide libraries had 5 base pair (bp) barcodes on both sides, while the Boston libraries had 7 bp barcodes. Libraries from Tübingen had no internal barcodes, but were differentiated by the sequence of the indexing primer[31].

We adapted a reported protocol for enriching for mitochondrial DNA[10], with the difference that we adjusted the blocking oligonucleotides and PCR primers to fit our libraries with shorter adapters. Over the course of this project, we also lowered the hybridization temperature from 65 °C to 60° C and performed stringent washes at 55 °C instead of 60 °C[32].

We used an aliquot of approximately 500 ng of each library for target enrichment of the complete mitochondrial genome in two consecutive rounds[32], using a bait set for human mtDNA[32]. We performed enrichment in 96-well plates with one library per well, and used a liquid handler (Evolution P3, Perkin Elmer) for the capture and washing steps[33]. We used blocking oligonucleotides in hybridization appropriate to the adapters of the truncated libraries. After either of the two enrichment rounds, we amplified the enriched library molecules with the primer pair that keeps the adapters short (PreHyb) using Herculase Fusion II PCR Polymerase. We performed an indexing PCR of the final capture product using one or two indexing primers[31]. We cleaned up all PCR reactions using SPRI technology[34] and the liquid handler. Libraries from Tübingen were amplified with the primer pair IS5/IS6[31].

For libraries from Boston and Adelaide, we used a second aliquot of each library for shotgun sequencing after performing an indexing PCR[31]. We used unique index combinations for each library and experiment, allowing us to distinguish shotgun sequencing and mitochondrial DNA capture data, even if both experiments were in the same sequencing run. We sequenced shotgun libraries and mtDNA captured libraries from Tübingen in independent sequencing runs since the index was already attached at the library preparation stage.

We quantified the sequencing pool with the BioAnalyzer (Agilent) and/or the KAPA Library Quantification kit (KAPA Biosystems) and sequenced on Illumina MiSeq, HiSeq2500 or NextSeq500 instruments for 2×75, 2×100 or 2×150 cycles along with the indexing read(s).

Enrichment for 394,577 SNP targets ('390k capture'). The protocol for enrichment for SNP targets was similar to the mitochondrial DNA capture, with the exception that we used another bait set (390k) and about twice as much library (up to 1,000 ng) compared to the mtDNA capture.

The specific capture reagent used in this study is described for the first time here. To target each SNP, we used a different oligonucleotide probe design compared to ref. 10. We used four 52 base pair probes for each SNP target. One probe ends just before the SNP, and one starts just after. Two probes are centred on the SNP, and are identical except for having the alternate alleles. This probe design avoids systematic bias towards one SNP allele or another. For the template sequence for designing the San and Yoruba panels baits, we used the sequence that was submitted for these same SNPs during the design of the Affymetrix Human Origins SNP array[12]. For SNPs that were both in the San and Yoruba panels, we used the Yoruba template sequence in preference. For all other SNPs, we used the human genome reference sequence as a template. Supplementary Data 2a–d gives the list of SNPs that we targeted, along with details of the probes used. The breakdown of SNPs into different classes is as follows.

124,106 'Yoruba SNPs': all SNPs in 'panel 5' of the Affymetrix Human Origins array (discovered as heterozygous in a Yoruba male: HGDP00927)[12] that passed the probe design criteria specified in ref. 11.

146,135 'San SNPs': all SNPs in 'panel 4' of the Affymetrix Human Origins array (discovered as heterozygous in a San male: HGDP01029)[12] that passed probe design criteria[11]. The full Affymetrix Human Origins array panel 4 contains several tens of thousands of additional SNPs overlapping those from panel 5, but we did not wish to redundantly capture panel 5 SNPs.

98,166 'compatibility SNPs': SNPs that overlap between the Affymetrix Human Origins, the Affymetrix 6.0, and the Illumina 610 Quad arrays, which are not already included in the 'Yoruba SNPs' or 'San SNPs' lists[12] and that also passed the probe design design criteria[11].

26,170 'miscellaneous SNPs': SNPs that did not overlap the Human Origins array. The subset analysed in this study were 2,258 Y chromosome SNPs (http://isogg.org/tree/ISOGG_YDNA_SNP_Index.html) that we used for Y haplogroup determination.

Processing of sequencing data. We restricted analysis to read pairs that passed quality control according to the Illumina software ('PF reads').

We assigned read pairs to libraries by searching for matches to the expected index and barcode sequences (if present, as for the Adelaide and Boston libraries). We allowed no more than 1 mismatch per index or barcode, and zero mismatches if there was ambiguity in sequence assignment or if barcodes of 5 bp length were used (Adelaide libraries).

We used Seqprep (https://github.com/jstjohn/SeqPrep) to search for overlapping sequence between the forward and reverse read, and restricted to molecules where we could identify a minimum of 15 bp of overlap. We collapsed the two reads into a single sequence, using the consensus nucleotide if both reads agreed, and the read with higher base quality in the case of disagreement. For each merged nucleotide, we assigned the base quality to be the higher of the two reads. We further used Seqprep to search for the expected adaptor sequences at either ends of the merged sequence, and to produce a trimmed sequence for alignment.

We mapped all sequences using BWA-0.6.1 (ref. 35). For mitochondrial analysis we mapped to the mitochondrial genome RSRS[36]. For whole-genome analysis we mapped to the human reference genome hg19. We restricted all analyses to sequences that had a mapping quality of MAPQ ≥ 37.

We sorted all mapped sequences by position, and used a custom script to search for mapped sequences that had the same orientation and start and stop positions. We stripped all but one of these sequences (keeping the best quality one) as duplicates.

Mitochondrial sequence analysis and assessment of ancient DNA authenticity. For each library for which we had average coverage of the mitochondrial genome of at least tenfold after removal of duplicated molecules, we built a mitochondrial consensus sequence, assigning haplogroups for each library as described in Supplementary Information section 2.

We used contamMix-1.0.9 to search for evidence of contamination in the mitochondrial DNA[13]. This software estimates the fraction of mitochondrial DNA sequences that match the consensus more closely than a comparison set of 311 worldwide mitochondrial genomes. This is done by taking the consensus sequence of reads aligning to the RSRS mitochondrial genome, and requiring a minimum coverage of 5 after filtering bases where the quality was <30. Raw reads are then realigned to this consensus. In addition, the consensus is multiply aligned with the other 311 mitochondrial genomes using kalign (2.0.4)[37] to build the necessary inputs for contamMix, trimming the first and last 5 bases of every read to mitigate against the confounding factor of ancient damage. This software had difficulty running on data sets with higher coverage, and for these data sets, we downsampled to 50,000 reads.

For all sequences mapping to the mitochondrial DNA for which the consensus mitochondrial DNA sequence had a cytosine at the terminal nucleotide, we measured the proportion of sequences with a thymine at that position. For population genetic analysis, we only used partially UDG-treated libraries with a minimum of 3% C→T substitutions as recommended by ref. 33. In cases where we used a fully UDG-treated library for 390k analysis, we examined mitochondrial capture data from a non-UDG-treated library made from the same extract, and verified that the non-UDG library had a minimum of 10% C→T at the first nucleotide as recommended by ref. 38. Metrics for the mitochondrial DNA analysis on each library are given in Supplementary Data 1.

390k capture, sequence analysis and quality control. For 390k analysis, we restricted to reads that not only mapped to the human reference genome hg19 but that also overlapped the 354,212 autosomal SNPs genotyped on the Human Origins array[4]. We trimmed the last two nucleotides from each sequence because we found that these are highly enriched in ancient DNA damage even for UDG-treated libraries. We further restricted analyses to sites with base quality ≥ 30.

We made no attempt to determine a diploid genotype at each SNP in each sample. Instead, we used a single allele—randomly drawn from the two alleles in the individual—to represent the individual at that site[20,39]. Specifically, we made an allele call at each target SNP using majority rule over all sequences overlapping the SNP. When each of the possible alleles was supported by an equal number of sequences, we picked an allele at random. We set the allele to 'no call' for SNPs at which there was no read coverage.

We restricted population genetic analysis to libraries with a minimum of 0.06-fold average coverage on the 390k SNP targets, and for which there was an unambiguous sex determination based on the ratio of X to Y chromosome reads (Supplementary Information section 4 and Supplementary Data 1). For individuals for whom there were multiple libraries per sample, we performed a series of quality control analysis. First, we used the ADMIXTURE software[40,41] in supervised mode, using Kharia, Onge, Karitiana, Han, French, Mbuti, Ulchi and Eskimo as reference populations. We visually inspected the inferred ancestry components in each individual, and removed individuals with evidence of heterogeneity in inferred ancestry components across libraries. For all possible pairs of libraries for each sample, we also computed statistics of the form $D(Library_1, Library_2; Probe, Mbuti)$, where *Probe* is any of a panel of the same set of eight reference

©2015 Macmillan Publishers Limited. All rights reserved

populations), to determine whether there was significant evidence of the *Probe* population being more closely related to one library from an ancient individual than another library from that same individual. None of the individuals that we used had strong evidence of ancestry heterogeneity across libraries. For samples passing quality control for which there were multiple libraries per sample, we merged the sequences into a single BAM.

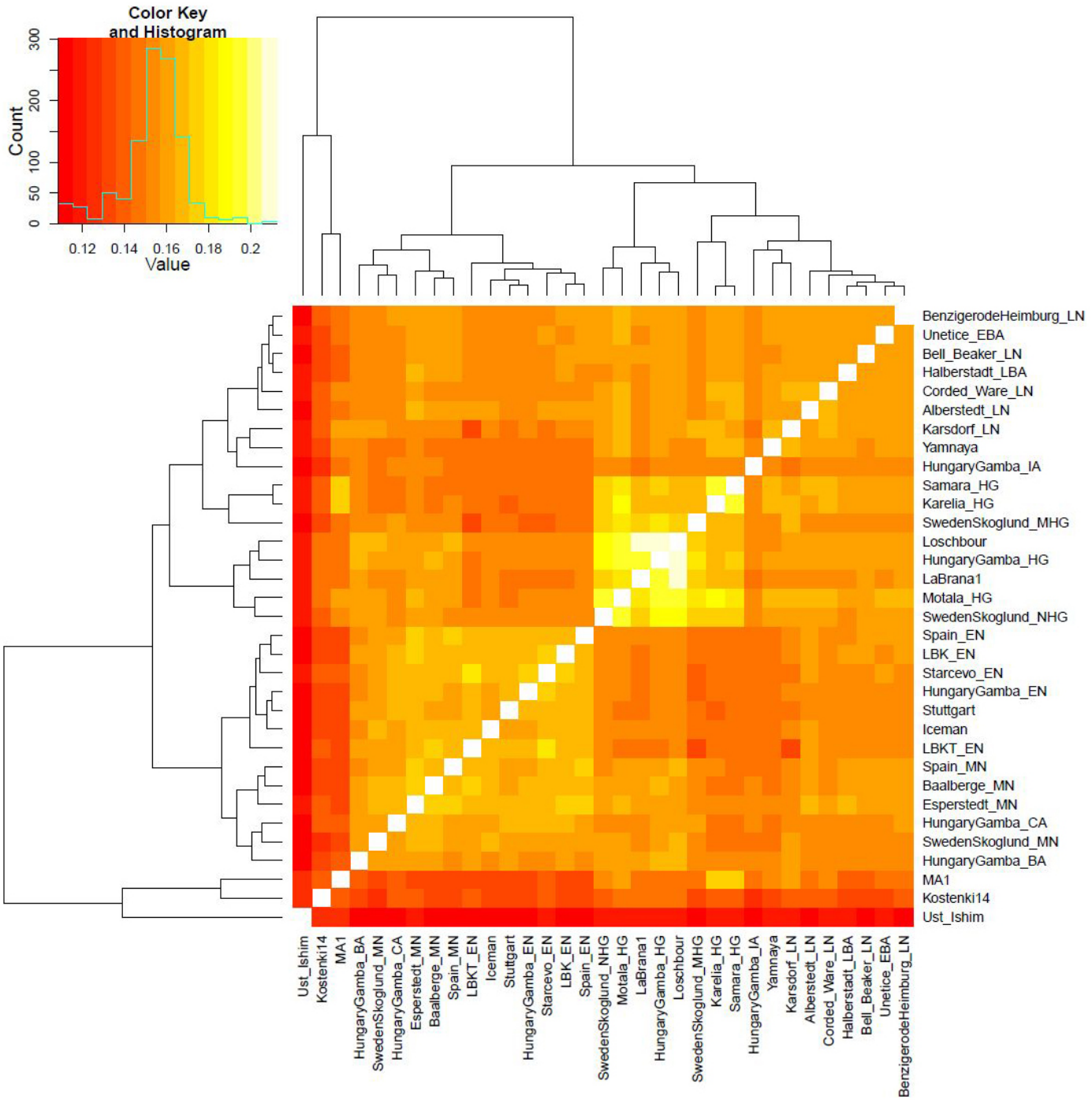
We called alleles on each merged BAM using the same procedure as for the individual libraries. We used ADMIXTURE[41] as well as PCA as implemented in EIGENSOFT[42] (using the *lsqproject: YES* option to project the ancient samples) to visualize the genetic relationships of each set of samples with the same culture label with respect to 777 diverse present-day West Eurasians[4]. We visually identified outlier individuals, and renamed them for analysis either as outliers or by the name of the site at which they were sampled (Extended Data Table 1). We also identified two pairs of related individuals based on the proportion of sites covered in pairs of ancient samples from the same population that had identical allele calls using PLINK[43]. From each pair of related individuals, we kept the one with the most SNPs.

Population genetic analyses. We determined genetic sex using the ratio of X and Y chromosome alignments[44] (Supplementary Information section 4), and Y chromosome haplogroup for the male samples (Supplementary Information section 4). We studied population structure (Supplementary Information sections 5 and 6). We used *f* statistics to carry out formal tests of population relationships (Supplementary Information section 6) and built explicit models of population history consistent with the data (Supplementary Information section 7). We estimated mixture proportions in a way that was robust to uncertainty about the exact population history that applied (Supplementary Information section 8). We estimated the minimum number of streams of migration into Europe needed to explain the data (Supplementary Information sections 9 and 10). The estimated mixture proportions shown in Fig. 3 were obtained using the *lsqlin* function of Matlab and the optimization method described in Supplementary Information section 9 with 15 world outgroups.

Sample size. No statistical methods were used to predetermine sample size.

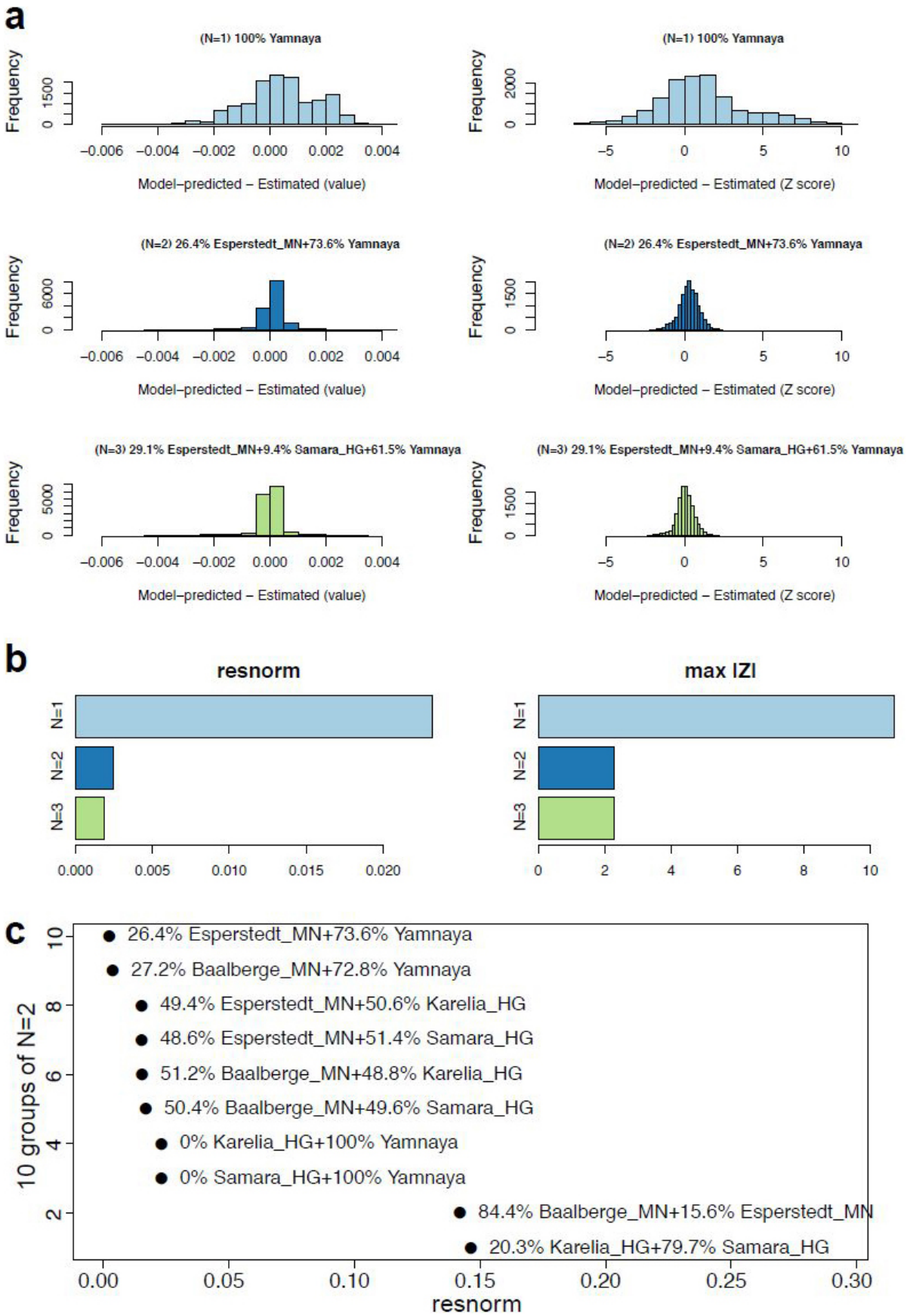
31. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40,** e3 (2012).
32. Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505,** 403–406 (2014).
33. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. Lond. B* **370,** 20130624 (2015).
34. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22,** 939–946 (2012).
35. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
36. Behar, D. M. *et al.* A ''Copernican'' reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90,** 675–684 (2012).
37. Lassmann, T. & Sonnhammer, E. L. L. Kalign—an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* **6,** 298 (2005).
38. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE* **7,** e34131 (2012).
39. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328,** 710–722 (2010).
40. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12,** 246 (2011).
41. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).
42. Reich, D., Price, A. L. & Patterson, N. Principal component analysis of genetic data. *Nature Genet.* **40,** 491–492 (2008).
43. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
44. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40,** 4477–4482 (2013).

©2015 **Macmillan Publishers Limited. All rights reserved**



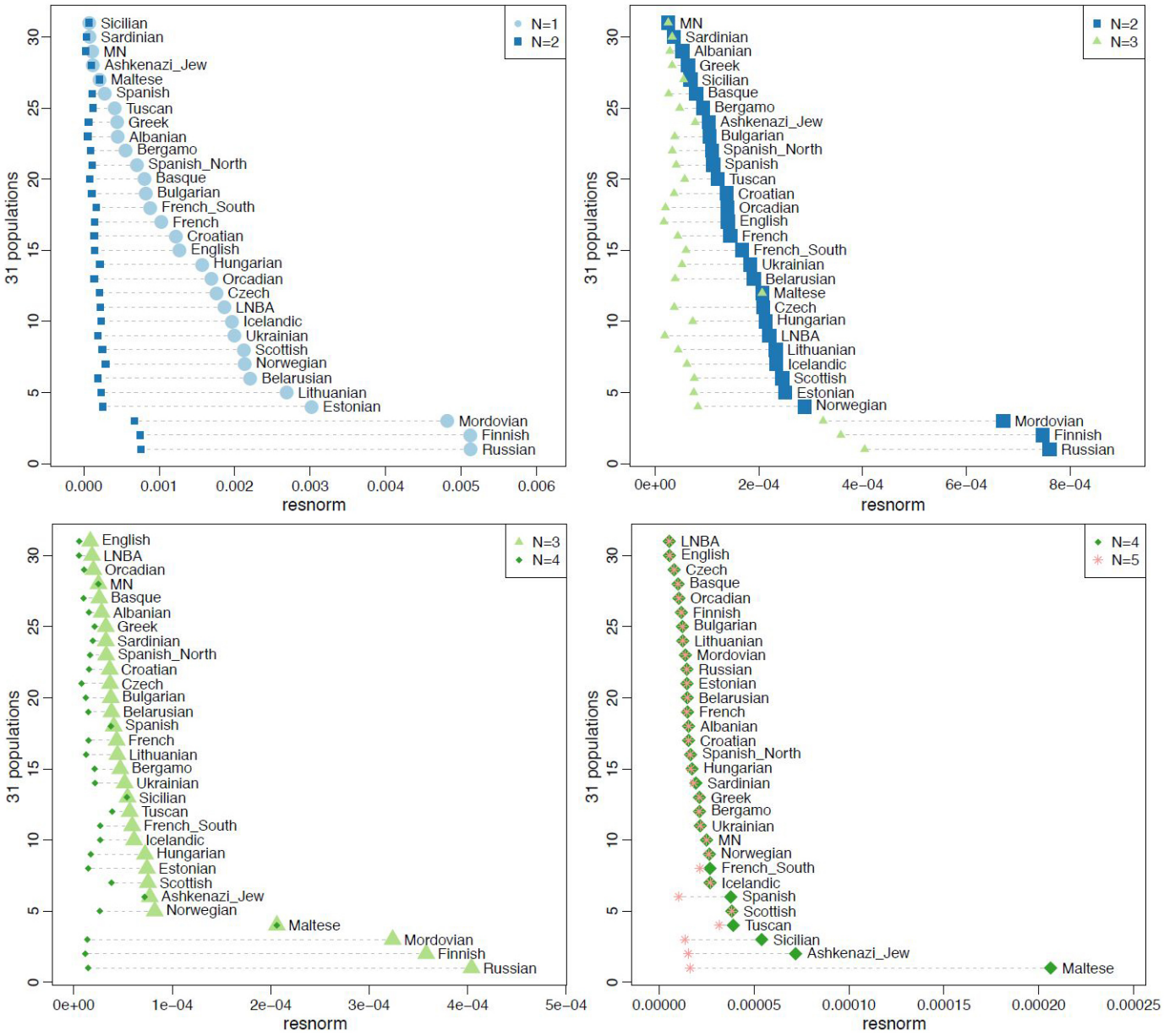
Extended Data Figure 1 | Outgroup f_3 statistic f_3(Dinka; *X*, *Y*), measuring the degree of shared drift among pairs of ancient individuals.

©2015 Macmillan Publishers Limited. All rights reserved



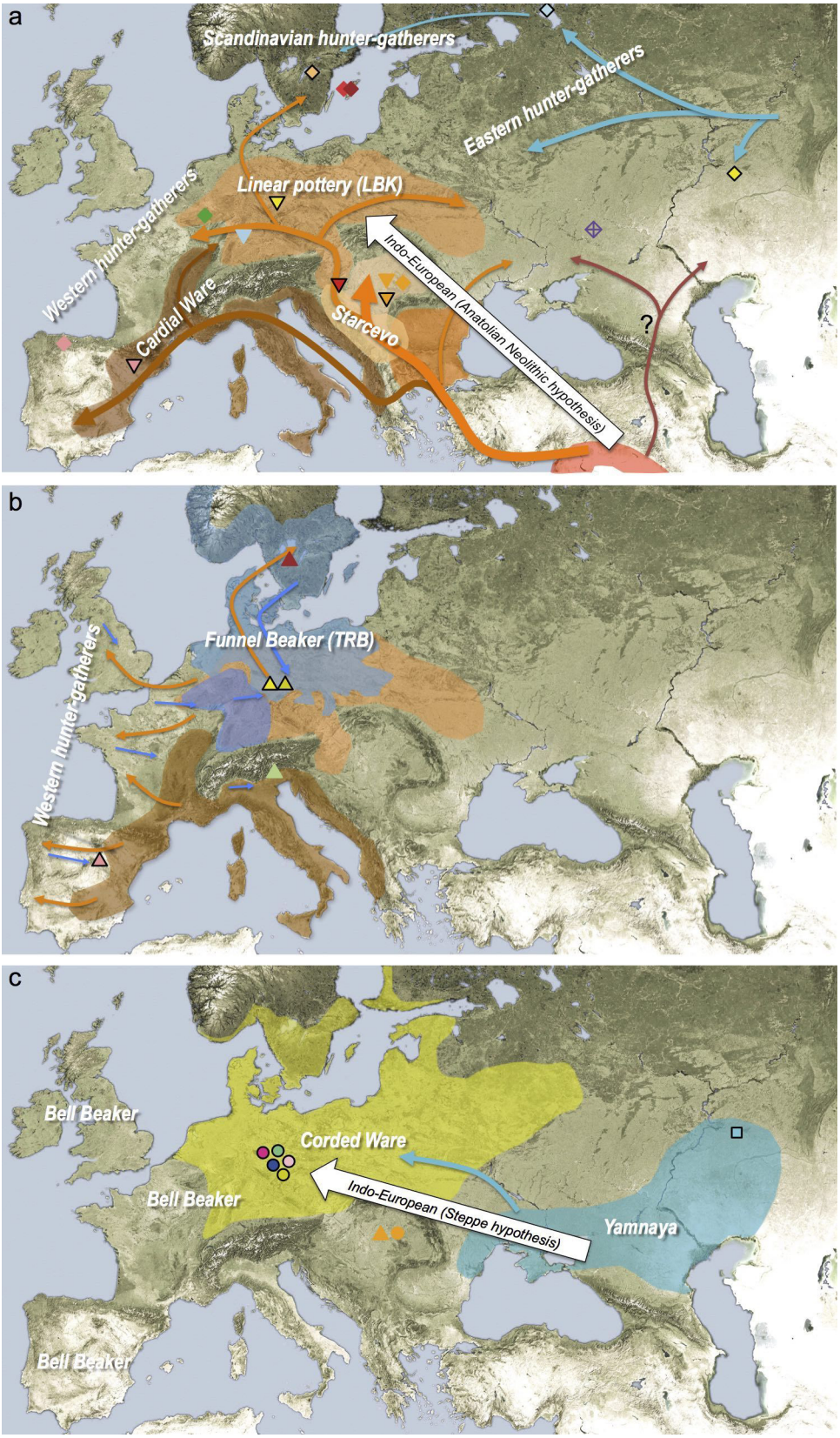
Extended Data Figure 2 | Modelling Corded Ware as a mixture of $N = 1$, 2, or 3 ancestral populations. **a**, The left column shows a histogram of raw f_4 statistic residuals and on the right Z-scores for the best-fitting (lowest squared 2-norm of the residuals, or *resnorm*) model at each N. **b**, The data on the left show *resnorm* and on the right show the maximum $|Z|$ score change for different N. **c**, *resnorm* of different $N = 2$ models. The set of outgroups used in this analysis in the terminology of Supplementary Information section 9 is 'World Foci 15 + Ancients'.

©2015 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 3 | Modelling Europeans as mixtures of increasing complexity: $N = 1$ (EN), $N = 2$ (EN, WHG), $N = 3$ (EN, WHG, Yamnaya), $N = 4$ (EN, WHG, Yamnaya, Nganasan), $N = 5$ (EN, WHG, Yamnaya, Nganasan, BedouinB). The residual norm of the fitted model (Supplementary Information section 9) and its changes are indicated.

©2015 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 4 | Geographic distribution of archaeological cultures and graphic illustration of proposed population movements / turnovers discussed in the main text. **a**, Proposed routes of migration by early farmers into Europe ~9,000−7000 years ago. **b**, Resurgence of hunter-gatherer ancestry during the Middle Neolithic 7,000−5,000 years ago. **c**, Arrival of steppe ancestry in central Europe during the Late Neolithic ~4,500 years ago. White arrows indicate the two possible scenarios of the arrival of Indo-European language groups. Symbols of samples are identical to those in Fig. 1.

©2015 Macmillan Publishers Limited. All rights reserved

Extended Data Table 1 | Number of ancient Eurasian modern human samples screened in genome-wide studies to date

First author	Description	No. samples at ≥0.05× coverage (enough for Procrustes analysis)	No. samples at >0.25× coverage (enough to analyze in pairs)
Keller[3]	Tyrolean Iceman	1	1
Raghavan[6]	Upper Paleolithic Siberians	2	1
Olalde[5]	Mesolithic Iberian from LaBrana	1	1
Skoglund[8]	Farmers and hunter-gatherers from Sweden	5	2
Lazaridis[4]	Early European farmer from Germany & Mesolithic hunter-gatherers from Luxembourg and Sweden	7	4
Gamba[2]	Neolithic, Bronze Age, Iron Age Hungary	13	9
Fu[1]	Upper Paleolithic Siberian from Ust-Ishim	1	1
Seguin-Orlando[7]	Upper Paleolithic European from Kostenki	1	1
Total before study		*31*	*20*
This study	Hunter-gatherers and pastoralists from Russia, Mesolithic hunter-gatherers from Sweden, Early Neolithic from Germany, Hungary, and Spain, Middle Neolithic from Germany & Spain, Late Neolithic / Bronze Age from Germany	69	58

Only studies that produced at least one sample at ≥ 0.05 × coverage are listed.

©2015 Macmillan Publishers Limited. All rights reserved

Extended Data Table 2 | Summary of the archaeological context for the 69 newly reported samples

Reich ID	Pop Label for Analysis	Culture	Group	Location and sample details (e.g. sample, grave and museum ID)	Date (lab no.)	Country	Sex	mt-hg	Y-hg	Autosomal SNPs
I0061	Karelia_HG	Russian Mesolithic	EHG	Yuzhnyy Oleni Ostrov, Karelia, Russia; UzOO74, grave 142, MAE RAS 5773-74	5500-5000 BC	Russia	M	C1g (formerly C1f)	R1a1	341554
I0124	Samara_HG	Russian Neolithic HG	EHG	Sok River, Samara, Russia; SVP44	5650-5555 cal BC (Beta – 392490)	Russia	M	U5a1d	R1b1	206748
I0011	Motala_HG	Swedish Mesolithic	SHG	Motala, Sweden; Motala 1	5898-5531 cal BC	Sweden	F	U5a1		228271
I0012	Motala_HG	Swedish Mesolithic	SHG	Motala, Sweden; Motala 2	5898-5531 cal BC	Sweden	M	U2e1	I2c2	292853
I0013	Motala_HG	Swedish Mesolithic	SHG	Motala, Sweden; Motala 3	5898-5531 cal BC	Sweden	M	U5a1	I2a1b	251108
I0014	Motala_HG	Swedish Mesolithic	SHG	Motala, Sweden; Motala 4	5898-5531 cal BC	Sweden	F	U5a2d		311299
I0015	Motala_HG	Swedish Mesolithic	SHG	Motala, Sweden; Motala 6	5898-5531 cal BC	Sweden	M	U5a2d	I2a1	285307
I0016	Motala_HG	Swedish Mesolithic	SHG	Motala, Sweden; Motala 9	5898-5531 cal BC	Sweden	M	U5a2	I2a1	275233
I0017	Motala_HG	Swedish Mesolithic	SHG	Motala, Sweden; Motala 12	5898-5531 cal BC	Sweden	M	U2e1	I2a1b	337794
I0174	Starcevo_EN	Starcevo	EN	Alsónyék-Bátaszék, Mérnöki telep, Hungary; BAM25a, feature 1532	5710-5550 cal BC (MAMS 11939)	Hungary	M	N1a1a1b	H2	101653
I0176	LBKT_EN	LBKT	EN	Szemely-Hegyes, Hungary: SZEH4b, feature 1001	5210-4940 cal BC (Beta - 310038)	Hungary		N1a1a1a3		30718
I0046	LBK_EN	LBK	EN	Halberstadt-Sonntagsfeld, Germany: HAL5, grave 2, feature 241.1	5206-5004 cal BC (MAMS 21479)	Germany	F	T2c1d'e'f		266764
I0048	LBK_EN	LBK	EN	Halberstadt-Sonntagsfeld, Germany: HAL25, grave 28, feature 861	5206-5052 cal BC (MAMS 21482)	Germany	M	K1a	G2a2a	123828
I0056	LBK_EN	LBK	EN	Halberstadt-Sonntagsfeld, Germany; HAL14, grave 15, feature 430	5206-5052 cal BC (MAMS 21480)	Germany	M	T2b(8)	G2a2a	136578
I0057	LBK_EN	LBK	EN	Halberstadt-Sonntagsfeld; HAL34, grave 38, feature 992	5207-5067 cal BC (MAMS 21483)	Germany	F	N1a1a1		55802
I0100	LBK_EN	LBK	EN	Halberstadt-Sonntagsfeld; HAL4, grave 1, feature 139	5032-4946 cal BC (KIA40341)	Germany	F	N1a1a1a		342342
I0659	LBK_EN	LBK	EN	Halberstadt-Sonntagsfeld, Germany; HAL2, grave 35, feature 999	5079-4997 cal BC (KIA40350) 5066-4979 cal BC (KIA30408)	Germany	M	N1a1a	G2a2a1	191007
I0821	LBK_EN	LBK	EN	Halberstadt-Sonntagsfeld, Germany; HAL24, grave 27, feature 867	5034-4942 cal BC (KIA40348)	Germany	M	Pre-X2d1	G2a2a1	55914
I0795	LBK_EN	LBK	EN	Karsdorf, Germany; KAR6a, feature 170	5207-5070 cal BC (MAMS 22823)	Germany	M	H1	T1a	47804
I0054	LBK_EN	LBK	EN	Oberwiederstedt-Unterwiederstedt, UWS4, Germany, grave 6, feature 1 14	5209-5070 cal BC (MAMS 21485)	Germany	F	J1c17		337625
I0022	LBK_EN	LBK	EN	Viesenhäuser Hof, Stuttgart-Mühlhausen, Germany; LBK1976	5500-4800 BC	Germany	F	T2e		160852
I0025	LBK_EN	LBK	EN	Viesenhäuser Hof, Stuttgart-Mühlhausen, Germany; LBK1992	5500-4800 BC	Germany	F	T2b		307686
I0026	LBK_EN	LBK	EN	Viesenhäuser Hof, Stuttgart-Mühlhausen, Germany; LBK2155	5500-4800 BC	Germany	F	T2b		315484
I0409	Spain_EN	Els_Trocs	EN	Els Trocs, Spain; Troc1	5311-5218 cal BC (MAMS 16159)	Spain	F	J1c3		172903
I0410	Spain_EN	Els_Trocs	EN	Els Trocs, Spain; Troc3	5178-5066 cal BC (MAMS 16161)	Spain	M	pre-T2c1d2	R1b1	297595
I0411	Spain_EN_relative_of_I0410	Els_Trocs	EN	Els Trocs, Spain; Troc4	5177-5068 cal BC (MAMS 16162)	Spain	M	K1a2a	F*	31507
I0412	Spain_EN	Els_Trocs	EN	Els Trocs, Spain; Troc 5	5310-5206 cal BC (MAMS 16164)	Spain	M	N1a1a1	I2a1b1	333940
I0413	Spain_EN	Els_Trocs	EN	Els Trocs, Spain; Troc7	5303-5204 cal BC (MAMS 16166)	Spain	F	V		295844
I0405	Spain_MN	La_Mina	MN	La Mina, Spain; Mina3	3900-3600 BC	Spain	M	K1a1b1	I2a1a1/H2?	133230
I0406	Spain_MN	La_Mina	MN	La Mina, Spain; Mina4	3900-3600 BC	Spain	M	H1	I2a2a1	324169
I0407	Spain_MN	La_Mina	MN	La Mina, Spain; Mina6b	3900-3600 BC	Spain	F	K1b1a1		236225
I0408	Spain_MN	La_Mina	MN	La Mina, Spain; Mina18a	3900-3600 BC	Spain	F	pre-U5b1i		321761
I0172	Esperstedt_MN	Salzmünde/Bernburg	MN	Esperstedt, Germany; ESP24, feature 1841	3360-3086 cal BC (Erl8699)	Germany	M	T2b	I2a1b1a	279147
I0559	Baalberge_MN	Baalberge	MN	Quedlinburg, Germany; QLB15D, feature 21033	3645-3537 cal BC (MAMS 22818)	Germany	M	HV6'17	R*?	64304
I0560	Baalberge_MN	Baalberge	MN	Quedlinburg, Germany; QLB18A, feature 21039	3640-3510 cal BC (Er7856)	Germany	F	T2e1		133305
I0807	Baalberge_MN	Baalberge	MN	Esperstedt, Germany; ESP30, feature 6220	3887-3797 cal BC (Er7784)	Germany	M	H1e1a	F*	33481
I0231	Yamnaya	Yamnaya	EBA	Ekaterinovka_ Southern Steppe, Samara , Russia, SVP3	2910-2875 cal BC (Beta 392487)	Russia	M	U4a1	R1b1a2a2	348142
I0357	Yamnaya	Yamnaya	EBA	Lopatino I, Sok_River, Samara, Russia; SVP5 same sample as SVP37	3090-2910 BC (Beta 392489)	Russia	F	W6		163845
I0370	Yamnaya	Yamnaya	EBA	Ishkinovka I, Eastern Orenburg, Pre-Ural steppe, Samara, Russia: SVP10	3300-2700 BC	Russia	M	H13a1a1a	R1b1a2a2	199345
I0429	Yamnaya	Yamnaya	EBA	Lopatino I, Sok River, Samara, Russia, SVP38	3339-2917 cal BC (AA47804)	Russia	M	T2c1a2	R1b1a2a2	217664
I0438	Yamnaya	Yamnaya	EBA	Luzhki I, Samara River, Samara, Russia; SVP50	3021-2635 cal BC (AA47807)	Russia	M	U5a1a1	R1b1a2a2	213493
I0439	Yamnaya	Yamnaya	EBA	Lopatino I, Sok River, Samara, Russia, SVP52	3305-2925 cal BC (Beta 392491)	Russia	M	U5a1a1	R1b1a	98900
I0441	Yamnaya	Yamnaya	EBA	Kurmanaevskii III, Buzuluk, Samara, Russia; SVP54	3010-2622 cal BC (AA47805)	Russia	F	H2b		51326
I0443	Yamnaya	Yamnaya	EBA	Lopatino II, Sok River, Samara, Russia; SVP57	3300-2700 BC	Russia	M	W3a1a	R1b1a2a	343890
I0444	Yamnaya	Yamnaya	EBA	Kutuluk I, Kutuluk River, Samara, Russia; SVP58	3300-2700 BC	Russia	M	H6a1b	R1b1a2a2	187126
I0550	Karsdorf_LN	unknown	LN	Karsdorf, Germany; KAR22a, feature 191	2564-2475 cal BC (MAMS 23344)	Germany	F	T1a1		59907
I0103	Corded_Ware_LN	Corded Ware	LN	Esperstedt, Germany; ESP16, feature 6236	2566-2477 cal BC (MAMS 21488)	Germany	F	W6a		336918
I0049	Corded_Ware_LN	Corded Ware	LN	Esperstedt, Germany; ESP22, feature 6140	2454-2291 cal BC (MAMS 21489)	Germany	F	X2b4		167170
I0106	Corded_Ware_LN	Corded Ware	LN	Esperstedt, Germany; ESP26, feature 6233.1	2454-2291 cal BC (MAMS 21490)	Germany	F	T2a1b1		69886
I0104	Corded_Ware_LN	Corded Ware	LN	Esperstedt, Germany; ESP11, feature 6216	2473-2348 cal BC (MAMS 21487)	Germany	M	U4b1a1a1	R1a1a1	336637
I0059	BenzigerodeHeimburg_LN	Bell Beaker?	LN	Benzingerode-Heimburg, Germany; BZH6, grave 2, feature/find 1287/1036	2286-2153 cal BC (MAMS 21486)	Germany	F	H1 /H1b'ad		241081
I0058	BenzigerodeHeimburg_LN	Bell Beaker	LN	Benzingerode-Heimburg, Germany; BZH4, grave 7, feature 4607	2283-2146 cal BC (MAMS 21491)	Germany	F	H1e		246728
I0171	BenzigerodeHeimburg_LN	Bell Beaker?	LN	Benzingerode-Heimburg, Germany; BZH12, grave 3 ,feature 6256	2204-2136 cal BC (KIA27952)	Germany	F	U5a1a2a		66800
I0112	Bell_Beaker_LN	Bell Beaker	LN	Quedlinburg XII, Germany; QUEXII6, feature 6256	2340-2190 cal BC (Er7038)	Germany	F	H13a1a2		341003
I0113	Bell_Beaker_LN	Bell Beaker	LN	Quedlinburg XII, Germany; QUEXII4, feature 6255.1	2290-2130 cal BC (Er7283)	Germany	F	J1c5		190352
I0108	Bell_Beaker_LN	Bell Beaker	LN	Rothenschirmbach, Germany; ROT6, feature 10044	2497-2436 cal BC (Er8710)	Germany	F	H5a3		260528
I0111	Bell_Beaker_LN	Bell Beaker	LN	Rothenschirmbach, Germany; ROT4, feature 10142	2414-2333 cal BC (Er8712)	Germany	F	H3new		208256
I0060	Bell_Beaker_LN	Bell Beaker	LN	Rothenschirmbach, Germany; ROT3, feature 10011	2294-2206 cal BC (MAMS 22819)	Germany	F	K1a2c		47085
I0806	Bell_Beaker_LN	Bell Beaker	LN	Quedlinburg VII 2, Germany; QLB28b, feature 19617	2296-2206 cal BC (MAMS 22820)	Germany	M	H1	R1b1a2a1a2	91757
I0118	Alberstedt_LN	unknown	LN	Alberstedt, Germany; ALB3, feature 7144.2	2459-2345 cal BC (MAMS 21492)	Germany	F	HV6'17		349956
I0114	Unetice_EBA_relative_of_I0117	Unetice	EBA	Esperstedt, Germany; ESP2, feature 3340.1	2131-1979 cal BC (MAMS 21493)	Germany	M	I3a	I2a2	217031
I0115	Unetice_EBA	Unetice	EBA	Esperstedt, Germany; ESP3, feature 1559.1	1931-1780 cal BC (MAMS 21494)	Germany	F	U5a1		123744
I0116	Unetice_EBA	Unetice	EBA	Esperstedt, Germany; ESP4, feature 3322/3323	2118-1961 cal BC (MAMS 21495)	Germany	M	W3a1	I2c2	308158
I0117	Unetice_EBA	Unetice	EBA	Esperstedt, Germany; ESP29, feature 3332/3333	2199-2064 cal BC (MAMS 21496)	Germany	F	I3a		279996
I0164	Unetice_EBA	Unetice	EBA	Quedlinburg VIII, Germany, QUEVII6, feature 3580	2012-1919 cal BC (MAMS 21497)	Germany	F	pre-U5b2a1b		332832
I0803	Unetice_EBA	Unetice	EBA	Eulau, Germany; EUL41A, feature 882	2115-1966 cal BC (MAMS 22822)	Germany	F	H4a1a1		144186
I0804	Unetice_EBA	Unetice	EBA	Eulau, Germany; EUL57B, feature1911.5	2131-1982 cal BC (MAMS 22821)	Germany	M	H3	I2	22969
I0047	Unetice_EBA	Unetice	EBA	Halberstadt-Sonntagsfeld, Germany; HAL16, grave 19, feature 613.1	2022-1937 cal BC (MAMS 21481)	Germany	F	V		288353
I0099	Halberstadt_LBA	Late Bronze Age	LBA	Halberstadt-Sonntagsfeld, Germany; HAL36C, grave 40, feature 1114	1113-1021 cal BC (MAMS 21484)	Germany	M	H23	R1a1a1b1a2	337566

Samples with direct radiocarbon dates are indicated by a calibrated date "cal BC" along with associated laboratory numbers. Dates that are estimated based on faunal elements associated with the samples are not indicated with 'cal' (although they are still calibrated, absolute dates).

©2015 Macmillan Publishers Limited. All rights reserved

Extended Data Table 3 | Pairwise F_{ST} for all ancient groups with ≥ 2 individuals, present-day Europeans with ≥ 10 individuals, and selected other groups

	Armenian	Baalberge_MN	Basque	BedouinB	Belarusian	Bell_Beaker_LN	BenzigerodeHeimburg_LN	Bergamo	Bulgarian	Corded_Ware_LN	Croatian	Czech	EHG	English	Estonian	French	Greek	Han	Hungarian	HungaryGamba_BA	HungaryGamba_EN	Icelandic	Iraqi_Jew	Kartliana	LBK_EN	Lezgin	Lithuanian	Mala	Mordovian	Motala_HG	Norwegian	Onge	Orcadian	Papuan	Russian	Sardinian	Sicilian	Sindhi	Spain_EN	Spain_MN	Spanish	SwedenSkoglund_NHG	Turkish	Unetice_EBA	WHG	Yamnaya	Yoruba
Armenian		0.004	0.000	0.001	0.001	0.001	0.002	0.000	0.000	0.001	0.000	0.000	0.002	0.000	0.001	0.000	0.000	0.001	0.000	0.002	0.001	0.000	0.001	0.002	0.001	0.000	0.001	0.001	0.001	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Baalberge_MN	0.023		0.003	0.004	0.004	0.005	0.006	0.004	0.004	0.005	0.004	0.003	0.007	0.004	0.004	0.003	0.003	0.004	0.003	0.008	0.005	0.004	0.004	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.004	0.003	0.004	0.003	0.003	0.003	0.003	0.005	0.005	0.003	0.013	0.003	0.004	0.005	0.004	0.004
Basque	0.017	0.019		0.001	0.000	0.001	0.002	0.000	0.000	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.000	0.002	0.001	0.000	0.001	0.002	0.001	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
BedouinB	0.025	0.047	0.041		0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.000	0.000	0.001	0.001	0.002	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Belarusian	0.014	0.026	0.012	0.041		0.001	0.002	0.000	0.000	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.000	0.002	0.001	0.000	0.001	0.002	0.001	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Bell_Beaker_LN	0.017	0.020	0.012	0.042	0.008		0.002	0.001	0.001	0.002	0.001	0.001	0.003	0.001	0.001	0.001	0.001	0.002	0.001	0.003	0.002	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.001	0.002	0.002	0.001	0.004	0.001	0.001	0.002	0.001	0.001
BenzigerodeHeimburg_LN	0.019	0.021	0.017	0.047	0.011	0.004		0.002	0.002	0.003	0.002	0.002	0.004	0.002	0.002	0.002	0.002	0.002	0.002	0.004	0.003	0.002	0.002	0.003	0.002	0.002	0.002	0.002	0.002	0.003	0.002	0.003	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.002	0.006	0.002	0.002	0.003	0.002	0.002
Bergamo	0.007	0.015	0.008	0.032	0.007	0.008	0.012		0.000	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.000	0.002	0.001	0.000	0.000	0.001	0.002	0.001	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Bulgarian	0.006	0.019	0.009	0.031	0.003	0.008	0.010	0.002		0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.000	0.002	0.001	0.000	0.000	0.001	0.002	0.001	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Corded_Ware_LN	0.023	0.032	0.025	0.052	0.015	0.008	0.007	0.019	0.017		0.001	0.001	0.003	0.001	0.001	0.001	0.001	0.002	0.001	0.003	0.002	0.001	0.002	0.002	0.002	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.001	0.002	0.002	0.001	0.004	0.001	0.002	0.002	0.002	0.002
Croatian	0.009	0.019	0.010	0.035	0.002	0.007	0.011	0.004	0.001	0.015		0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.000	0.002	0.001	0.000	0.001	0.002	0.001	0.000	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Czech	0.011	0.024	0.009	0.037	0.001	0.007	0.010	0.005	0.002	0.015	0.001		0.002	0.000	0.000	0.000	0.000	0.001	0.000	0.002	0.001	0.000	0.001	0.002	0.001	0.000	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
EHG	0.067	0.082	0.060	0.093	0.042	0.041	0.038	0.057	0.051	0.034	0.048	0.043		0.002	0.002	0.002	0.002	0.003	0.002	0.004	0.003	0.002	0.003	0.003	0.003	0.002	0.002	0.002	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.002	0.002	0.002	0.003	0.003	0.002	0.006	0.002	0.003	0.003	0.002	0.002
English	0.011	0.020	0.008	0.038	0.004	0.006	0.009	0.004	0.003	0.014	0.003	0.002	0.045		0.000	0.000	0.000	0.001	0.000	0.002	0.001	0.000	0.001	0.002	0.001	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Estonian	0.017	0.027	0.014	0.045	0.002	0.009	0.011	0.010	0.006	0.016	0.005	0.002	0.040	0.005		0.000	0.000	0.001	0.000	0.002	0.001	0.000	0.001	0.002	0.001	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
French	0.009	0.018	0.006	0.035	0.004	0.006	0.008	0.002	0.002	0.015	0.002	0.001	0.048	0.001	0.005		0.000	0.001	0.000	0.002	0.001	0.000	0.000	0.002	0.001	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Greek	0.004	0.018	0.010	0.028	0.007	0.010	0.014	0.002	0.000	0.019	0.003	0.004	0.057	0.005	0.009	0.003		0.001	0.000	0.002	0.001	0.000	0.000	0.002	0.001	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Han	0.108	0.136	0.120	0.128	0.110	0.114	0.117	0.113	0.108	0.118	0.110	0.111	0.135	0.113	0.110	0.110	0.110		0.001	0.003	0.002	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.002	0.002	0.001	0.003	0.001	0.001	0.002	0.001	0.001
Hungarian	0.009	0.019	0.009	0.036	0.001	0.007	0.009	0.003	0.001	0.014	0.002	0.000	0.046	0.001	0.003	0.001	0.003	0.108		0.002	0.001	0.000	0.001	0.002	0.001	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
HungaryGamba_BA	0.016	0.021	0.010	0.039	0.005	0.008	0.010	0.008	0.005	0.016	0.005	0.005	0.046	0.007	0.006	0.005	0.007	0.115	0.005		0.003	0.002	0.002	0.003	0.002	0.002	0.002	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.002	0.002	0.002	0.003	0.003	0.002	0.006	0.002	0.003	0.003	0.003	0.002
HungaryGamba_EN	0.016	0.009	0.018	0.037	0.021	0.020	0.025	0.009	0.012	0.037	0.015	0.018	0.081	0.018	0.026	0.014	0.010	0.129	0.016	0.014		0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.001	0.002	0.002	0.001	0.004	0.001	0.002	0.002	0.002	0.001
Icelandic	0.015	0.025	0.011	0.041	0.005	0.008	0.011	0.006	0.005	0.014	0.006	0.003	0.045	0.002	0.006	0.003	0.008	0.113	0.004	0.007	0.021		0.001	0.002	0.001	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Iraqi_Jew	0.005	0.031	0.021	0.026	0.020	0.022	0.024	0.011	0.010	0.029	0.014	0.017	0.073	0.016	0.023	0.014	0.009	0.112	0.015	0.018	0.020	0.021		0.002	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Kartliana	0.210	0.234	0.218	0.230	0.205	0.208	0.208	0.212	0.206	0.209	0.207	0.206	0.206	0.209	0.205	0.208	0.209	0.180	0.205	0.214	0.229	0.209	0.214		0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.004	0.002	0.002	0.003	0.002	0.001
LBK_EN	0.023	0.009	0.024	0.043	0.029	0.021	0.025	0.017	0.018	0.035	0.022	0.024	0.084	0.024	0.033	0.020	0.016	0.137	0.022	0.018	0.006	0.029	0.026	0.237		0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.003	0.001	0.001	0.002	0.001	0.001
Lezgin	0.005	0.033	0.018	0.034	0.012	0.015	0.015	0.010	0.007	0.017	0.009	0.010	0.052	0.011	0.014	0.009	0.007	0.103	0.009	0.017	0.025	0.013	0.011	0.201	0.032		0.001	0.001	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Lithuanian	0.019	0.030	0.014	0.046	0.001	0.010	0.010	0.011	0.006	0.016	0.005	0.003	0.043	0.006	0.002	0.006	0.010	0.114	0.004	0.007	0.026	0.007	0.024	0.209	0.034	0.015		0.001	0.000	0.001	0.000	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Mala	0.045	0.074	0.061	0.068	0.051	0.055	0.057	0.053	0.048	0.058	0.051	0.051	0.081	0.053	0.053	0.051	0.049	0.070	0.050	0.056	0.070	0.055	0.049	0.187	0.077	0.043	0.056		0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.000	0.002	0.001	0.001	0.001	0.003	0.001	0.001	0.002	0.001	0.001
Mordovian	0.014	0.028	0.014	0.041	0.002	0.010	0.011	0.009	0.005	0.013	0.005	0.003	0.040	0.005	0.003	0.005	0.008	0.100	0.003	0.007	0.025	0.006	0.020	0.196	0.032	0.011	0.004	0.047		0.001	0.000	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Motala_HG	0.088	0.083	0.072	0.113	0.060	0.057	0.053	0.073	0.069	0.060	0.067	0.062	0.048	0.063	0.058	0.064	0.075	0.166	0.062	0.058	0.091	0.062	0.095	0.254	0.093	0.080	0.058	0.112	0.060		0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.001	0.002	0.002	0.001	0.004	0.001	0.001	0.002	0.001	0.001
Norwegian	0.014	0.024	0.010	0.041	0.004	0.007	0.009	0.006	0.005	0.014	0.003	0.002	0.044	0.002	0.005	0.002	0.007	0.112	0.003	0.009	0.021	0.002	0.020	0.207	0.028	0.012	0.006	0.053	0.006	0.062		0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Onge	0.168	0.191	0.181	0.186	0.171	0.174	0.177	0.174	0.170	0.180	0.171	0.172	0.197	0.174	0.173	0.172	0.170	0.155	0.171	0.176	0.187	0.175	0.172	0.283	0.194	0.167	0.177	0.129	0.166	0.227	0.174		0.001	0.002	0.001	0.001	0.001	0.001	0.002	0.002	0.001	0.004	0.001	0.002	0.002	0.002	0.001
Orcadian	0.014	0.021	0.011	0.041	0.006	0.009	0.010	0.007	0.006	0.016	0.006	0.004	0.047	0.003	0.007	0.004	0.008	0.114	0.004	0.008	0.021	0.005	0.020	0.209	0.029	0.013	0.008	0.054	0.007	0.063	0.004	0.175		0.002	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Papuan	0.186	0.206	0.199	0.202	0.190	0.194	0.194	0.193	0.187	0.199	0.190	0.191	0.218	0.192	0.192	0.191	0.189	0.180	0.189	0.192	0.204	0.195	0.190	0.304	0.212	0.185	0.196	0.151	0.185	0.244	0.193	0.232	0.193		0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.004	0.002	0.002	0.002	0.002	0.002
Russian	0.015	0.027	0.014	0.042	0.002	0.010	0.010	0.009	0.005	0.015	0.005	0.003	0.039	0.005	0.002	0.005	0.008	0.099	0.003	0.007	0.024	0.006	0.021	0.196	0.032	0.012	0.004	0.047	0.002	0.058	0.005	0.166	0.007	0.185		0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Sardinian	0.014	0.014	0.013	0.036	0.017	0.016	0.022	0.007	0.009	0.033	0.011	0.013	0.074	0.012	0.020	0.009	0.009	0.123	0.012	0.014	0.010	0.016	0.017	0.225	0.015	0.021	0.021	0.064	0.020	0.085	0.016	0.183	0.016	0.201	0.020		0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Sicilian	0.003	0.017	0.011	0.027	0.009	0.011	0.013	0.003	0.002	0.022	0.004	0.006	0.059	0.006	0.012	0.004	0.002	0.109	0.005	0.010	0.010	0.010	0.008	0.209	0.016	0.009	0.013	0.048	0.010	0.078	0.007	0.169	0.009	0.187	0.010	0.008		0.000	0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Sindhi	0.016	0.047	0.032	0.043	0.023	0.026	0.028	0.023	0.020	0.029	0.022	0.022	0.057	0.023	0.025	0.022	0.020	0.081	0.021	0.027	0.042	0.025	0.020	0.185	0.049	0.014	0.028	0.014	0.021	0.088	0.024	0.142	0.026	0.162	0.021	0.035	0.021		0.001	0.001	0.000	0.003	0.000	0.001	0.002	0.001	0.001
Spain_EN	0.034	0.021	0.030	0.053	0.037	0.028	0.035	0.024	0.028	0.044	0.030	0.032	0.092	0.032	0.040	0.028	0.027	0.144	0.031	0.028	0.018	0.036	0.037	0.245	0.015	0.042	0.041	0.085	0.039	0.099	0.036	0.202	0.038	0.219	0.041	0.021	0.026	0.058		0.002	0.001	0.004	0.001	0.002	0.002	0.002	0.002
Spain_MN	0.033	0.014	0.025	0.054	0.032	0.022	0.026	0.023	0.026	0.035	0.027	0.029	0.077	0.027	0.036	0.024	0.025	0.138	0.028	0.028	0.023	0.031	0.037	0.241	0.016	0.040	0.035	0.079	0.035	0.081	0.030	0.196	0.031	0.214	0.035	0.021	0.025	0.052	0.009		0.001	0.004	0.001	0.002	0.002	0.002	0.002
Spanish	0.008	0.014	0.005	0.032	0.006	0.007	0.011	0.002	0.002	0.018	0.004	0.003	0.054	0.003	0.008	0.001	0.003	0.111	0.003	0.006	0.011	0.005	0.013	0.210	0.017	0.011	0.009	0.052	0.008	0.069	0.005	0.172	0.005	0.190	0.008	0.007	0.003	0.024	0.024	0.020		0.003	0.000	0.001	0.002	0.001	0.001
SwedenSkoglund_NHG	0.071	0.066	0.058	0.094	0.047	0.050	0.048	0.056	0.053	0.054	0.052	0.047	0.058	0.050	0.050	0.049	0.058	0.153	0.050	0.041	0.068	0.051	0.076	0.244	0.078	0.066	0.049	0.101	0.047	0.034	0.049	0.211	0.053	0.230	0.047	0.067	0.061	0.074	0.084	0.074	0.052		0.003	0.004	0.004	0.003	0.003
Turkish	0.001	0.023	0.014	0.026	0.010	0.013	0.016	0.005	0.004	0.021	0.006	0.008	0.058	0.009	0.013	0.006	0.003	0.095	0.007	0.013	0.016	0.012	0.008	0.198	0.022	0.005	0.015	0.040	0.010	0.081	0.011	0.160	0.012	0.178	0.011	0.013	0.003	0.013	0.032	0.030	0.006	0.084		0.001	0.002	0.001	0.001
Unetice_EBA	0.016	0.016	0.014	0.042	0.007	0.002	0.003	0.008	0.007	0.005	0.007	0.005	0.034	0.005	0.006	0.005	0.009	0.112	0.006	0.007	0.022	0.007	0.021	0.206	0.022	0.013	0.007	0.053	0.007	0.051	0.005	0.173	0.007	0.190	0.008	0.019	0.011	0.024	0.030	0.023	0.007	0.043	0.013		0.002	0.001	0.001
WHG	0.086	0.073	0.062	0.106	0.054	0.055	0.056	0.066	0.065	0.070	0.061	0.056	0.078	0.058	0.053	0.058	0.070	0.162	0.058	0.049	0.080	0.057	0.091	0.260	0.091	0.082	0.053	0.110	0.056	0.053	0.057	0.220	0.059	0.239	0.054	0.074	0.072	0.087	0.092	0.070	0.060	0.050	0.078	0.057		0.002	0.002
Yamnaya	0.030	0.054	0.034	0.060	0.020	0.016	0.013	0.028	0.023	0.011	0.022	0.020	0.028	0.021	0.020	0.022	0.026	0.118	0.020	0.023	0.050	0.020	0.037	0.204	0.052	0.019	0.022	0.056	0.018	0.063	0.021	0.178	0.022	0.197	0.019	0.044	0.030	0.028	0.060	0.053	0.027	0.062	0.026	0.012	0.076		0.001
Yoruba	0.142	0.163	0.155	0.141	0.150	0.151	0.153	0.148	0.145	0.158	0.148	0.149	0.179	0.150	0.152	0.147	0.145	0.177	0.148	0.151	0.156	0.152	0.141	0.270	0.163	0.143	0.155	0.139	0.148	0.200	0.151	0.222	0.152	0.231	0.148	0.153	0.139	0.133	0.169	0.167	0.144	0.185	0.137	0.151	0.195	0.158	

F_{ST} (below the diagonal), standard deviation (above the diagonal).

©2015 Macmillan Publishers Limited. All rights reserved