ARTICLE

doi:10.1038/nature14507

Population genomics of Bronze Age Eurasia

Morten E. Allentoft[1]*, Martin Sikora[1]*, Karl-Göran Sjögren[2], Simon Rasmussen[3], Morten Rasmussen[1], Jesper Stenderup[1], Peter B. Damgaard[1], Hannes Schroeder[1,4], Torbjörn Ahlström[5], Lasse Vinner[1], Anna-Sapfo Malaspinas[1], Ashot Margaryan[1], Tom Higham[6], David Chivall[6], Niels Lynnerup[7], Lise Harvig[7], Justyna Baron[8], Philippe Della Casa[9], Paweł Dąbrowski[10], Paul R. Duffy[11], Alexander V. Ebel[12], Andrey Epimakhov[13], Karin Frei[14], Mirosław Furmanek[8], Tomasz Gralak[8], Andrey Gromov[15], Stanisław Gronkiewicz[16], Gisela Grupe[17], Tamás Hajdu[18,19], Radosław Jarysz[20], Valeri Khartanovich[15], Alexandr Khokhlov[21], Viktória Kiss[22], Jan Kolář[23,24], Aivar Kriiska[25], Irena Lasak[8], Cristina Longhi[26], George McGlynn[17], Algimantas Merkevicius[27], Inga Merkyte[28], Mait Metspalu[29], Ruzan Mkrtchyan[30], Vyacheslav Moiseyev[15], László Paja[31,32], György Pálfi[32], Dalia Pokutta[2], Łukasz Pospieszny[33], T. Douglas Price[34], Lehti Saag[29], Mikhail Sablin[35], Natalia Shishlina[36], Václav Smrčka[37], Vasilii I. Soenov[38], Vajk Szeverényi[22], Gusztáv Tóth[39], Synaru V. Trifanova[38], Liivi Varul[25], Magdolna Vicze[40], Levon Yepiskoposyan[41], Vladislav Zhitenev[42], Ludovic Orlando[1], Thomas Sicheritz-Pontén[3], Søren Brunak[3,43], Rasmus Nielsen[44], Kristian Kristiansen[2] & Eske Willerslev[1]

The Bronze Age of Eurasia (around 3000–1000 BC) was a period of major cultural changes. However, there is debate about whether these changes resulted from the circulation of ideas or from human migrations, potentially also facilitating the spread of languages and certain phenotypic traits. We investigated this by using new, improved methods to sequence low-coverage genomes from 101 ancient humans from across Eurasia. We show that the Bronze Age was a highly dynamic period involving large-scale population migrations and replacements, responsible for shaping major parts of present-day demographic structure in both Europe and Asia. Our findings are consistent with the hypothesized spread of Indo-European languages during the Early Bronze Age. We also demonstrate that light skin pigmentation in Europeans was already present at high frequency in the Bronze Age, but not lactose tolerance, indicating a more recent onset of positive selection on lactose tolerance than previously thought.

The processes that created the genetic landscape of contemporary human populations of Europe and Asia remain contentious. Recent studies have revealed that western Eurasians and East Asians diverged outside Africa between 45 and 36.2 thousand years before present (45 and 36.2 kyr BP)[1,2] and that East Asians, but not Europeans, received subsequent gene flow from remnants of an earlier migration into Asia of Aboriginal Australian ancestors at some point before 20 kyr BP[3]. There is evidence that the western Eurasian branch constituted a meta-population stretching from Europe to Central Asia[2,4] and that it contributed genes to both modern-day western Eurasians[4] and early

indigenous Americans[4–6]. The early Europeans received gene flow from the Middle East during the Neolithisation (transition from hunting-gathering to farming) around 8–5 kyr BP[7–12] and possibly also from northern Asia[10]. However, what happened hereafter, during the Bronze Age, is much less clear.

The archaeological record testifies to major cultural changes in Europe and Asia after the Neolithic period. By 3000 BC, the Neolithic farming cultures in temperate Eastern Europe appear to be largely replaced by the Early Bronze Age Yamnaya culture, which is associated with a completely new perception of family, property and

personhood[13,14], rapidly stretching from Hungary to the Urals[15]. By 2800 BC a new social and economic formation, variously named Corded Ware, Single Grave or Battle Axe cultures developed in temperate Europe, possibly deriving from the Yamnaya background, and culturally replacing the remaining Neolithic farmers[16,17] (Fig. 1). In western and Central Asia, hunter-gatherers still dominated in Early Bronze Age, except in the Altai Mountains and Minusinsk Basin where the Afanasievo culture existed with a close cultural affinity to Yamnaya[15] (Fig. 1). From the beginning of 2000 BC, a new class of master artisans known as the Sintashta culture emerged in the Urals, building chariots, breeding and training horses (Fig. 1), and producing sophisticated new weapons[18]. These innovations quickly spread across Europe and into Asia where they appeared to give rise to the Andronovo culture[19,20] (Fig. 1). In the Late Bronze Age around 1500 BC, the Andronovo culture was gradually replaced by the Mezhovskaya, Karasuk, and Koryakova cultures[21]. It remains debated if these major cultural shifts during the Bronze Age in Europe and Asia resulted from the migration of people or through cultural diffusion among settled groups[15–17], and if the spread of the Indo-European languages was linked to these events or predates them[15].

Archaeological samples and DNA retrieval

Genomes obtained from ancient biological remains can provide information on past population histories that is not retrievable from contemporary individuals[4,22]. However, ancient genomic studies have so far been restricted to single or a few individuals because of the degraded nature of ancient DNA making sequencing costly and time consuming[23]. To overcome this, we increased the average output of authentic endogenous DNA fourfold by: (1) targeting the outer cementum layer in teeth rather than the inner dentine layer[24,25], (2) adding a 'pre-digestion' step to remove surface contaminants[24,26], and (3) developing a new binding buffer for ancient DNA extraction (Supplementary Information, section 3). This allowed us to obtain low-coverage genome sequences (0.01–7.4× average depth, overall average equal to 0.7×) of 101 Eurasian individuals spanning the entire Bronze Age, including some Late Neolithic and Iron Age individuals (Fig. 1, Supplementary Information, sections 1 and 2). Our data set includes 19 genomes, between 1.1–7.4× average depth, thereby doubling the number of existing Eurasian ancient genomes above 1× coverage (ref. 27).

Bronze Age Europe

By analysing our genomic data in relation to previously published ancient and modern data (Supplementary Information, section 6), we find evidence for a genetically structured Europe during the Bronze Age (Fig. 2; Extended Data Fig. 1; and Supplementary Figs 5 and 6). Populations in northern and central Europe were composed of a mixture of the earlier hunter-gatherer and Neolithic farmer[10] groups, but received 'Caucasian' genetic input at the onset of the Bronze Age (Fig. 2). This coincides with the archaeologically well-defined expansion of the Yamnaya culture from the Pontic-Caspian steppe into Europe (Figs 1 and 2). This admixture event resulted in the formation of peoples of the Corded Ware and related cultures, as supported by negative 'admixture' f_3 statistics when using Yamnaya as a source population (Extended Data Table 2, Supplementary Table 12). Although European Late Neolithic and Bronze Age cultures such
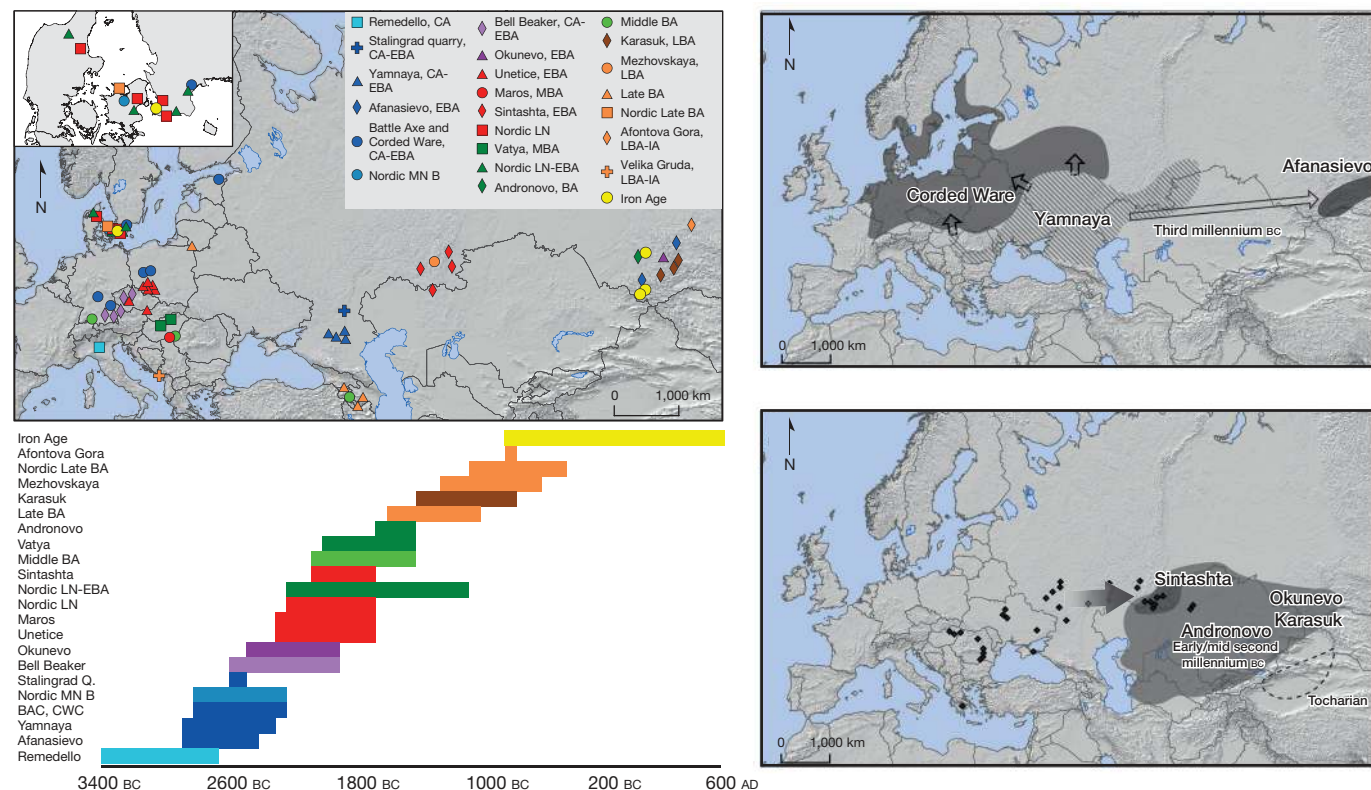
Figure 1 | Distribution maps of ancient samples. Localities, cultural associations, and approximate timeline of 101 sampled ancient individuals from Europe and Central Asia (left). Distribution of Early Bronze Age cultures Yamnaya, Corded Ware, and Afanasievo with arrows showing the Yamnaya expansions (top right). Middle and Late Bronze Age cultures Sintashta, Andronovo, Okunevo, and Karasuk with the eastward migration indicated (bottom right). Black markers represent chariot burials (2000–1800 BC) with similar horse cheek pieces, as evidence of expanding cultures. Tocharian is the second-oldest branch of Indo-European languages, preserved in Western China. CA, Copper Age; MN, Middle Neolithic; LN, Late Neolithic; EBA, Early Bronze Age; MBA, Middle Bronze Age; LBA, Late Bronze Age; IA, Iron Age; BAC, Battle Axe culture; CWC, Corded Ware culture.
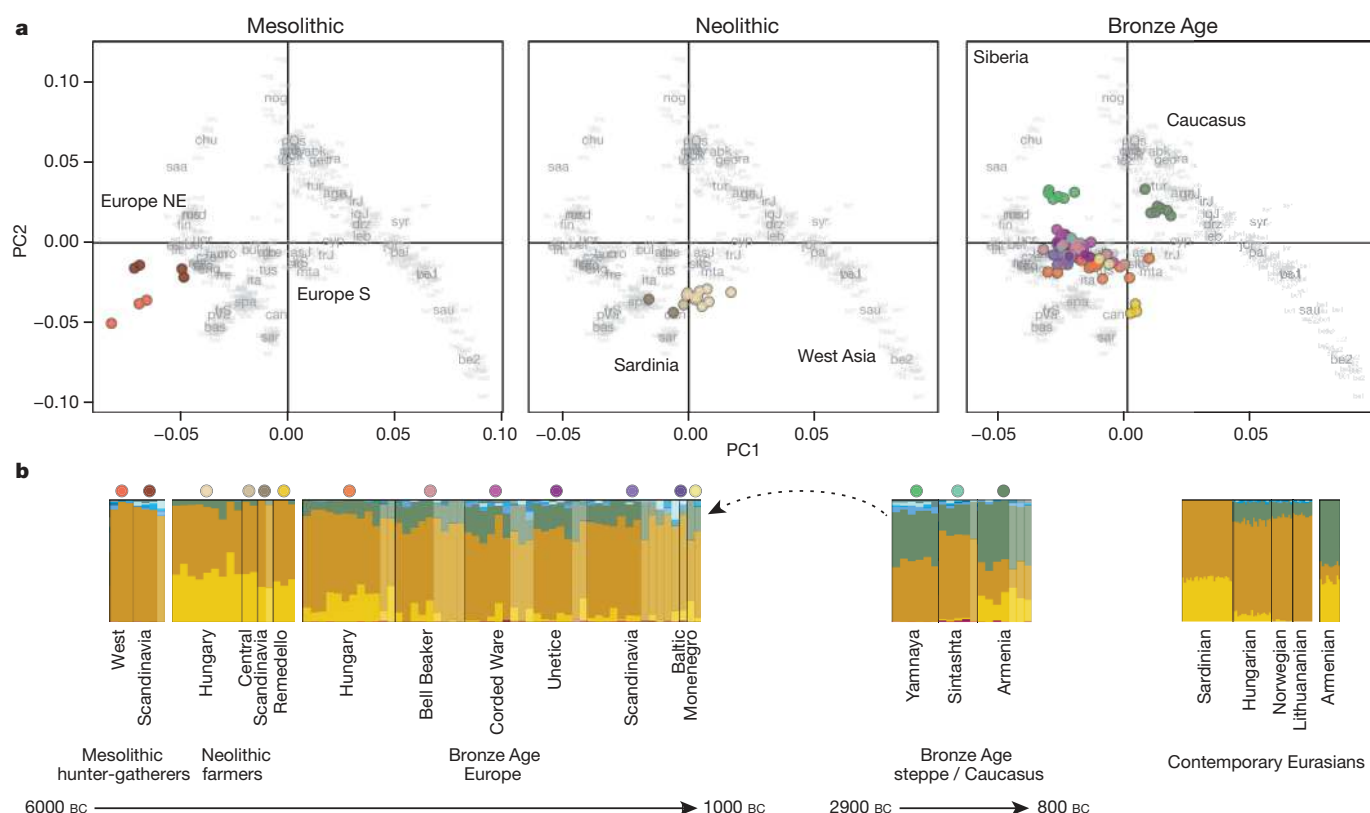
Figure 2 | Genetic structure of ancient Europe and the Pontic-Caspian steppe. a, Principal component analysis (PCA) of ancient individuals ($n = 93$) from different periods projected onto contemporary individuals from Europe, West Asia, and Caucasus. Grey labels represent population codes showing coordinates for individuals (small) and population median (large). Coloured circles indicate ancient individuals **b**, ADMIXTURE ancestry components ($K = 16$) for ancient ($n = 93$) and selected contemporary individuals. The width of the bars representing ancient individuals is increased to aid visualization. Individuals with less than 20,000 SNPs have lighter colours. Coloured circles indicate corresponding group in the PCA. Probable Yamnaya-related admixture is indicated by the dashed arrow.

as Corded Ware, Bell Beakers, Unetice, and the Scandinavian cultures are genetically very similar to each other (Fig. 2), they still display a cline of genetic affinity with Yamnaya, with highest levels in Corded Ware, lowest in Hungary, and central European Bell Beakers being intermediate (Fig. 2b and Extended Data Table 1). Using D-statistics, we find that Corded Ware and Yamnaya individuals form a clade to the exclusion of Bronze Age Armenians (Extended Data Table 1) showing that the genetic 'Caucasus component' present in Bronze Age Europe has a steppe origin rather than a southern Caucasus origin. Earlier studies have shown that southern Europeans received substantial gene flow from Neolithic farmers during the Neolithic[9]. Despite being slightly later, we find that the Copper Age Remedello culture in Italy does not have the 'Caucasian' genetic component and is still clustering genetically with Neolithic farmers (Fig. 2; Extended Data Fig. 1 and Supplementary Fig. 6). Hence this region was either unaffected by the Yamnaya expansion or the Remedello pre-dates such an expansion into southern Europe. The 'Caucasian' component is clearly present during Late Bronze Age in Montenegro (Fig. 2b). The close affinity we observe between peoples of Corded Ware and Sintashta cultures (Extended Data Fig. 2a) suggests similar genetic sources of the two, which contrasts with previous hypotheses placing the origin of Sintastha in Asia or the Middle East[28]. Although we cannot formally test whether the Sintashta derives directly from an eastward migration of Corded Ware peoples or if they share common ancestry with an earlier steppe population, the presence of European Neolithic farmer ancestry in both the Corded Ware and the Sintashta, combined with the absence of Neolithic farmer ancestry in the earlier Yamnaya, would suggest the former being more probable (Fig. 2b and Extended Data Table 1).

Bronze Age Asia

We find that the Bronze Age in Asia is equally dynamic and characterized by large-scale migrations and population replacements. The Early Bronze Age Afanasievo culture in the Altai-Sayan region is genetically indistinguishable from Yamnaya, confirming an eastward expansion across the steppe (Figs 1 and 3b; Extended Data Fig. 2b and Extended Data Table 1), in addition to the westward expansion into Europe. Thus, the Yamnaya migrations resulted in gene flow across vast distances, essentially connecting Altai in Siberia with Scandinavia in the Early Bronze Age (Fig. 1). The Andronovo culture, which arose in Central Asia during the later Bronze Age (Fig. 1), is genetically closely related to the Sintashta peoples (Extended Data Fig. 2c), and clearly distinct from both Yamnaya and Afanasievo (Fig. 3b and Extended Data Table 1). Therefore, Andronovo represents a temporal and geographical extension of the Sintashta gene pool. Towards the end of the Bronze Age in Asia, Andronovo was replaced by the Karasuk, Mezhovskaya, and Iron Age cultures which appear multiethnic and show gradual admixture with East Asians (Fig. 3b and Extended Data Table 2), corresponding with anthropological and biological research[29]. However, Iron Age individuals from Central Asia still show higher levels of West Eurasian ancestry than contemporary populations from the same region (Fig. 3b). Intriguingly, individuals of the Bronze Age Okunevo culture from the Sayano-Altai region (Fig. 1) are related to present-day Native Americans (Extended Data Fig. 2d), which confirms previous craniometric studies[30]. This finding implies that Okunevo could represent a remnant population related to the Upper Palaeolithic Mal'ta hunter-gatherer population from Lake Baikal that contributed genetic material to Native Americans[4].

Figure 3 | Genetic structure of Bronze Age Asia. a, Principal component analysis (PCA) of ancient individuals ($n = 40$) from different periods projected onto contemporary non-Africans. Grey labels represent population codes showing coordinates for individuals (small) and population median (large). Coloured circles indicate ancient individuals. **b*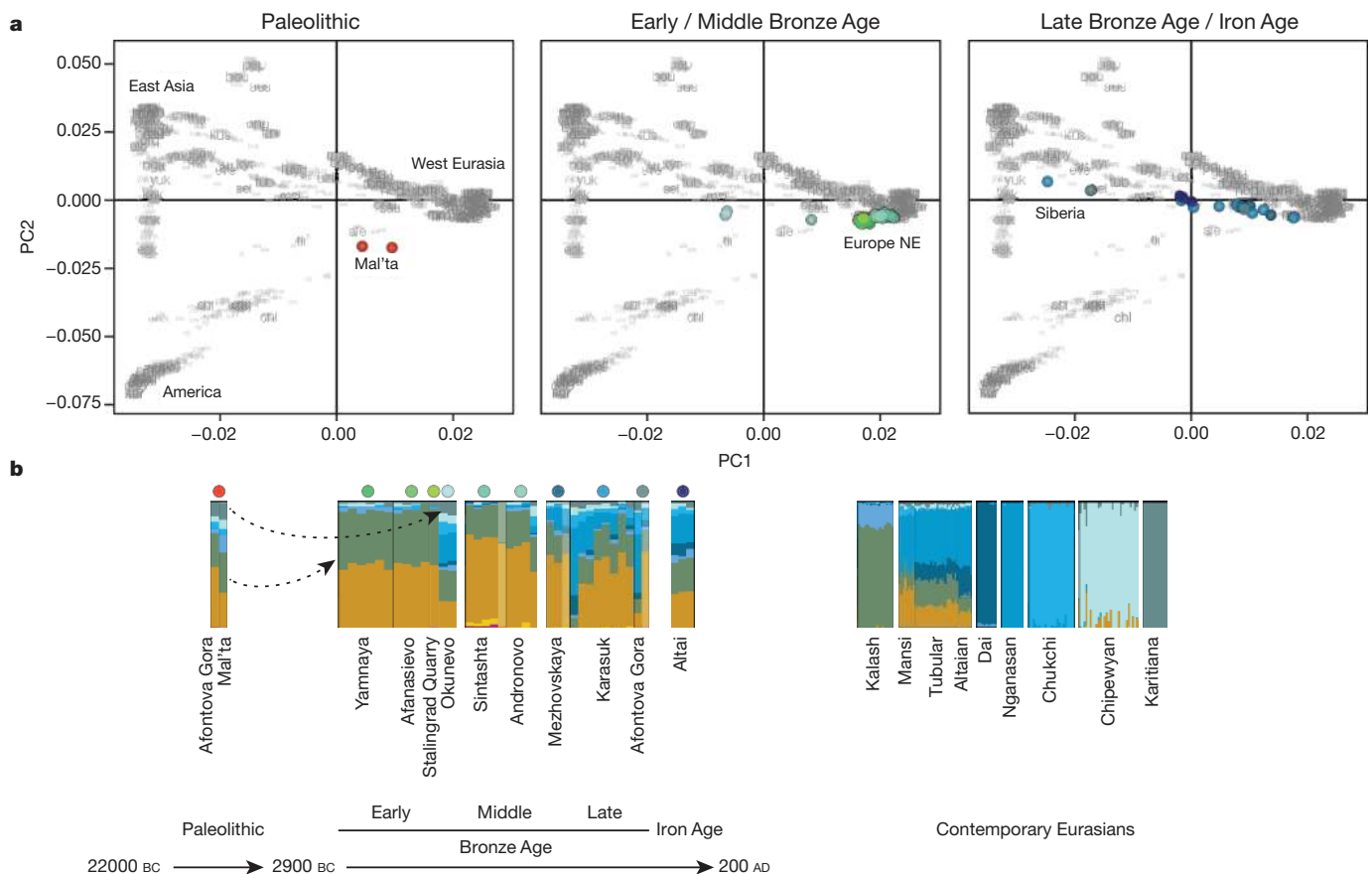*, ADMIXTURE ancestry components ($K = 16$) for ancient ($n = 40$) and selected contemporary individuals. The width of the bars representing ancient individuals is increased to aid visualization. Individuals with less than 20,000 SNPs have lighter colours. Coloured circles indicate corresponding group in the PCA. Shared ancestry of Mal'ta with Yamnaya (green component) and Okunevo (grey component) is indicated by dashed arrows.

Spread of the Indo–European languages

Historical linguists have argued that the spread of the Indo-European languages must have required migration combined with social or demographic dominance, and this expansion has been supported by archaeologists pointing to striking similarities in the archaeological record across western Eurasia during the third millennium BC[15,18,31]. Our genomic evidence for the spread of Yamnaya people from the Pontic-Caspian steppe to both northern Europe and Central Asia during the Early Bronze Age (Fig. 1) corresponds well with the hypothesized expansion of the Indo-European languages. In contrast to recent genetic findings[32], however, we only find weak evidence for admixture in Yamnaya, and only when using Bronze Age Armenians and the Upper Palaeolithic Mal'ta as potential source populations ($Z = -2.39$; Supplementary Table 12). This could be due to the absence of eastern hunter-gatherers as potential source population for admixture in our data set. Modern Europeans show some genetic links to Mal'ta[4] that has been suggested to form a third European ancestral component (Ancestral North Eurasians (ANE))[10]. Rather than a hypothetical ancient northern Eurasian group, our results reveal that ANE ancestry in Europe probably derives from the spread of the Yamnaya culture that distantly shares ancestry with Mal'ta (Figs 2b and 3b and Extended Data Fig. 3).

Formation of Eurasian genetic structure

It is clear from our autosomal, mitochondrial DNA and Y chromosome data (Extended Data Fig. 6) that the European and Central Asian gene pools towards the end of the Bronze Age mirror present-day Eurasian genetic structure to an extent not seen in the previous periods (Figs 2 and 3; Extended Data Fig. 1 and Supplementary Fig. 6). Our results imply that much of the basis of the Eurasian genetic landscape of today was formed during the complex patterns of expansions, admixture and replacements during this period. We find that many contemporary Eurasians show lower genetic differentiation (F_{ST}) with local Bronze Age groups than with earlier Mesolithic and Neolithic groups (Extended Data Figs 4 and 5). Notable exceptions are contemporary populations from southern Europe such as Sardinians and Sicilians, which show the lowest F_{ST} with Neolithic farmers. In general, the levels of differentiation between ancient groups from different temporal and cultural contexts are greater than those between contemporary Europeans. For example, we find pairwise $F_{ST} = 0.08$ between Mesolithic hunter-gatherers and Bronze Age individuals from Corded Ware, which is nearly as high as F_{ST} between contemporary East Asians and Europeans (Extended Data Fig. 5). These results are indicative of significant temporal shifts in the gene pools and also reveal that the ancient groups of Eurasia were genetically more structured than contemporary populations. The diverged ancestral genomic components must then have diffused further after the Bronze Age through population growth, combined with continuing gene flow between populations, to generate the low differentiation observed in contemporary west Eurasians.
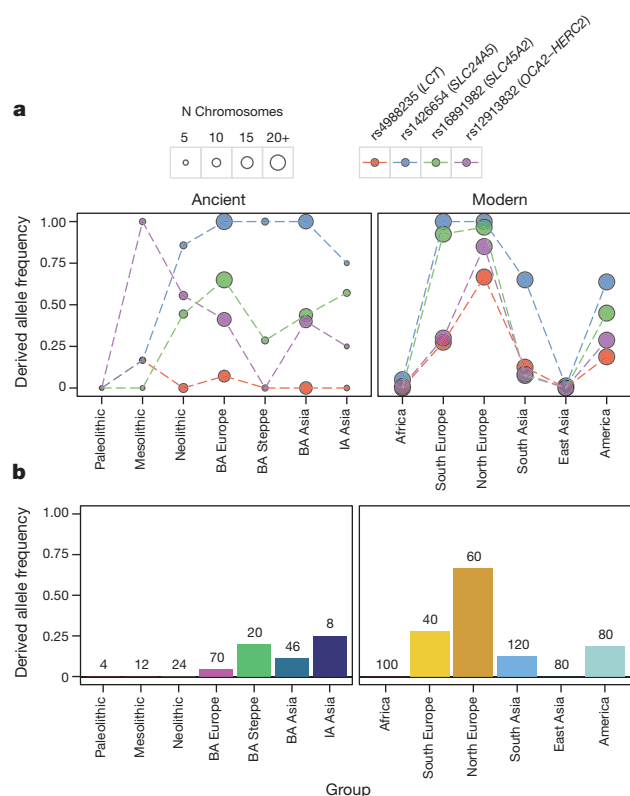
Figure 4 | Allele frequencies for putatively positively selected SNPs.
a, Coloured circles indicate the observed frequency of the respective SNP in ancient and modern groups (1000 Genomes panel). The size of the circle is proportional to the number of samples for each SNP and population. **b**, Allele frequency of rs4988235 in the *LCT* (lactase) gene inferred from imputation of ancient individuals. Numbers indicate the total number of chromosomes for each group. BA, Bronze Age; IA, Iron Age.

Temporal dynamics of selected SNPs

The size of our data set allows us to investigate the temporal dynamics of 104 genetic variants associated with important phenotypic traits or putatively undergoing positive selection[33] (Supplementary Table 13). Focusing on four well-studied polymorphisms, we find that two single nucleotide polymorphisms (SNPs) associated with light skin pigmentation in Europeans exhibit a rapid increase in allele frequency (Fig. 4). For rs1426654, the frequency of the derived allele increases from very low to fixation within a period of approximately 3,000 years between the Mesolithic and Bronze Age in Europe. For rs12913832, a major determinant of blue versus brown eyes in humans, our results indicate the presence of blue eyes already in Mesolithic hunter-gatherers as previously described[33]. We find it at intermediate frequency in Bronze Age Europeans, but it is notably absent from the Pontic-Caspian steppe populations, suggesting a high prevalence of brown eyes in these individuals (Fig. 4). The results for rs4988235, which is associated with lactose tolerance, were surprising. Although tolerance is high in present-day northern Europeans, we find it at most at low frequency in the Bronze Age (10% in Bronze Age Europeans; Fig. 4), indicating a more recent onset of positive selection than previously estimated[34]. To further investigate its distribution, we imputed all SNPs in a 2 megabase (Mb) region around rs4988235 in all ancient individuals using the 1000 Genomes phase 3 data set as a reference panel, as previously described[12]. Our results confirm a low frequency of rs4988235 in Europeans, with a derived allele frequency of 5% in the combined Bronze Age Europeans (genotype probability>0.85) (Fig. 4b). Among Bronze Age Europeans, the highest tolerance frequency was found in Corded Ware and the closely-related Scandinavian Bronze Age cultures (Extended Data Fig. 7).

Interestingly, the Bronze Age steppe cultures showed the highest derived allele frequency among ancient groups, in particular the Yamnaya (Extended Data Fig. 7), indicating a possible steppe origin of lactase tolerance.

Implications

It has been debated for decades if the major cultural changes that occurred during the Bronze Age resulted from the circulation of people or ideas and whether the expansion of Indo-European languages was concomitant with these shifts or occurred with the earlier spread of agriculture[13,15,35,36]. Our findings show that these transformations involved migrations, but of a different nature than previously suggested: the Yamnaya/Afanasievo movement was directional into Central Asia and the Altai-Sayan region and probably without much local infiltration, whereas the resulting Corded Ware culture in Europe was the result of admixture with the local Neolithic people. The enigmatic Sintashta culture near the Urals bears genetic resemblance to Corded Ware and was therefore likely to be an eastward migration into Asia. As this culture spread towards Altai it evolved into the Andronovo culture (Fig. 1), which was then gradually admixed and replaced by East Asian peoples that appear in the later cultures (Mezhovskaya and Karasuk). Our analyses support that migrations during the Early Bronze Age is a probable scenario for the spread of Indo-European languages, in line with reconstructions based on some archaeological and historical linguistic data[15,31]. In the light of our results, the existence of the Afanasievo culture near Altai around 3000 BC could also provide an explanation for the mysterious presence of one of the oldest Indo-European languages, Tocharian in the Tarim basin in China[37]. It seems plausible that Afanasievo, with their genetic western (Yamnaya) origin, spoke an Indo-European language and could have introduced this southward to Xinjang and Tarim[38]. Importantly, however, although our results support a correspondence between cultural changes, migrations, and linguistic patterns, we caution that such relationships cannot always be expected but must be demonstrated case by case.

1. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514,** 445–449 (2014).
2. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346,** 1113–1118 (2014).
3. Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334,** 94–98 (2011).
4. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505,** 87–91 (2014).
5. Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345,** 1255832 (2014).
6. Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial site in western Montana *Nature* **506,** 225–229 (2014).
7. Bramanti, B. *et al.* Genetic discontinuity between local hunter-gatherers and Central Europe's first farmers. *Science* **326,** 137–140 (2009).
8. Malmström, H. *et al.* Ancient DNA reveals lack of continuity between Neolithic hunter-gatherers and contemporary Scandinavians. *Curr. Biol.* **19,** 1758–1762 (2009).
9. Skoglund, P. *et al.* Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336,** 466–469 (2012).
10. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,** 409–413 (2014).
11. Haak, W. *et al.* Ancient DNA from European early Neolithic farmers reveals their Near Eastern affinities. *PLoS Biol.* **8,** e1000536 (2010).
12. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nature Commun.* **5,** 5257 (2014).
13. Kristiansen, K. in *The World System and the Earth System. Global Socioenvironmental Change and Sustainability Since the Neolithic* (eds Hornborg, B. & Crumley, C.) (Left Coast Press, 2007).
14. Shishlina, N. *Reconstruction of the Bronze Age of the Caspian Steppes. Life Styles and Life Ways of Pastoral Nomads.* Vol. 1876 (Archaeopress, 2008).
15. Anthony, D. *The Horse, The Wheel and Language. How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* (Princeton Univ. Press, 2007).

16. Harrison, R. & Heyd, V. The Transformation of Europe in the third millennium BC: the example of 'Le Petit-Chasseur I + III' (Sion, Valais, Switzerland). *Praehistorische Zeitschrift.* **82,** 129–214 (2007).
17. Vandkilde, H. *Culture and Change in the Central European Prehistory, 6th to 1st millennium BC* (Aarhus Univ. Press, 2007).
18. Kristiansen, K. & Larsson, T. *The Rise of Bronze Age Society. Travels, Transmissions and Transformations* (Cambridge Univ. Press, 2005).
19. Hanks, B. K., Epimakhov, A. V. & Renfrew, A. C. Towards a refined chronology for the Bronze Age of the southern Urals, Russia. *Antiquity* **81,** 353–367 (2007).
20. Kuznetsov, P. F. The emergence of Bronze Age chariots in Eastern Europe. *Antiquity* **80,** 638–645 (2006).
21. Koryakova, L. & Epimakhov, A. V. *The Urals and Western Siberia in the Bronze and Iron Ages* (Cambridge Univ. Press, 2007).
22. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463,** 757–762 (2010).
23. Carpenter, M. L. *et al.* Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *Am. J. Hum. Genet.* **93,** 852–864 (2013).
24. Barros Damgaard, P. d. *et al.* Improving access to endogenous DNA in ancient bones and teeth. Preprint at *bioRxiv* http://dx.doi.org/10.1101/014985 (2015).
25. Adler, C. J., Haak, W., Donlon, D., Cooper, A. & The Genographic Consortium. Survival and recovery of DNA from ancient teeth and bones. *J. Archaeol. Sci.* **38,** 956–964 (2011).
26. Orlando, L. *et al.* True single-molecule DNA sequencing of a Pleistocene horse bone. *Genome Res.* **21,** 1705–1719 (2011).
27. Olalde, I. & Lalueza-Fox, C. Modern humans' paleogenomics and the new evidences on the European prehistory. *Science and Technology of Archaeological Research* **1,** http://dx.doi.org/10.1179/2054892315Y.0000000002 (2015).
28. Grigoriev, S. *Ancient Indo-Europeans* (Charoid, 2002).
29. Bendezu-Sarmiento, J. *De l'Âge du Bronze et l'Âge du Fer au Kazakkstan, gestes funéraires et paramètres biologiques. Identités culturelles des population Andronovo et Saka* (De Boccard, 2007).
30. Kozintsev, A. G., Gromov, A. V. & Moiseyev, V. G. Collateral relatives of American Indians among the Bronze Age populations of Siberia? *Am. J. Phys. Anthropol.* **108,** 193–204 (1999).
31. Kristiansen, K. in *Becoming European. The transformation of third millennium Northern and Western Europe* (eds Prescott, C. & Glørstad, H.) (Oxbow Books, 2012).
32. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* http://dx.doi.org/10.1038/nature14317 (this issue).
33. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507,** 225–228 (2014).
34. Itan, Y., Powell, A., Beaumont, M. A., Burger, J. & Thomas, M. G. The origins of lactase persistence in Europe. *PLoS Computational Biol.* **5,** e1000491 (2009).
35. Mallory, J. *In Search of the Indo-Europeans. Language, Archaeology and Myth* (Thames & Hudson, 1987).
36. Renfrew, A. C. *Archaeology and Language. The Puzzle of Indo-European Origins* (Penguin, 1987).
37. Mallory, J. & Mair, V. *The Tarim Mummies. Ancient China and the Mystery of the Earliest People from the West* (Thames & Hudson, 2000).
38. Keyser, C. *et al.* Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum. Genet.* **126,** 395–410 (2009).

METHODS

DNA extraction and library preparation. A total of 603 human Bronze Age samples from across Eurasia were selected for initial molecular 'screening' to assess DNA preservation and hence the potential for genome-scale analyses. The samples consisted almost exclusively of teeth, but also a few bone and hair samples were included. All the molecular work (pre-library amplification) was conducted in dedicated ancient DNA clean laboratory facilities at the Centre for GeoGenetics, Natural History Museum, University of Copenhagen.

Preferentially targeting the outer cementum layer in teeth rather than the dentine allowed us to maximize access to endogenous DNA[24,25] (Supplementary Information, section 3). The amount of starting material varied, but was generally 100–600 mg. We also added a 'pre-digestion' step to the extraction protocol, where the drilled bone or tooth powder is incubated in an EDTA-based buffer before complete digestion to facilitate the removal of surface contaminants[24,26] (Supplementary Information, section 3). Additionally, we developed a new DNA binding buffer for extraction that proved more efficient in recovering short DNA fragments compared to previous protocols (Supplementary Information, section 3). DNA libraries for sequencing were prepared using NEBNext DNA Sample Prep Master Mix Set 2 (E6070) and Illumina-specific adapters[39] following established protocols[39–41]. The libraries were 'shot-gun' sequenced in pools using Illumina HiSeq2500 platforms and 100-bp single-read chemistry (Supplementary Information, section 3).

Molecular screening. For the molecular screening phase we generally generated between 5 and 20 million reads per library and these were used to evaluate the state of molecular preservation. Candidate samples were selected for further sequencing if they displayed a >10% C–T misincorporation damage signal in the 5′ ends as an indication of authentic ancient DNA[42,43], and a human DNA content >0.5% (Supplementary Information, section 3).

Genomic capture. We selected 24 samples with relatively low human DNA content (0.5–1.1%) for a whole-genome capture experiment[23] to enrich for the low human DNA fraction in these samples. The capture was performed using the MYbait Human Whole Genome Capture Kit (MYcroarray, Ann Arbor, MI), following the manufacturer's instructions (http://www.mycroarray.com/pdf/MYbaits-manual.pdf). After amplification, the libraries were purified using Agencourt AMPure XP beads, quantified using an Agilent 2100 bioanalyzer, pooled in equimolar amounts, and sequenced on Illumina HiSeq 2500, as described above. Methods and results are found in Supplementary Information, section 3.

Bioinformatics. The Illumina data was basecalled using Illumina software CASAVA 1.8.2 and sequences were de-multiplexed with a requirement of full match of the 6 nucleotide index that was used for library preparation. Adaptor sequences and leading/trailing stretches of Ns were trimmed from the reads and additionally bases with quality 2 or less were removed using AdapterRemoval-1.5.4. Trimmed reads of at least 30 bp were mapped to the human reference genome build 37 using bwa-0.6.2 (ref. 44) with the seed disabled to allow for higher sensitivity[45]. Mapped reads were filtered for mapping quality 30 and sorted using Picard (http://picard.sourceforge.net) and SAMtools[46]. Data was merged to library level and duplicates removed using Picard MarkDuplicates (http://picard.sourceforge.net) and hereafter merged to sample level. Sample level BAMs were re-aligned using GATK-2.2-3 and hereafter had the md-tag updated and extended BAQs calculated using SAMtools calmd[46]. Read depth and coverage were determined using pysam (http://code.google.com/p/pysam/) and BEDtools[47]. Statistics of the read data processing are shown in Supplementary Table 6.

DNA authentication. DNA contamination can be problematic in samples from museum collections that may have been handled extensively. To secure authenticity, we used the Bayesian approach implemented in mapDamage 2.0 (ref. 48) and recorded the following three key damage parameters for each sample: (1) the frequency of C→T transitions at the first position at the 5′ end of reads, (2) λ, the fraction of bases positioned in single-stranded overhangs, and (3) δs, the estimated C→T transition rate in the single-stranded overhangs (Supplementary Information, section 5). For further sequencing and downstream analyses we only considered individuals displaying at least 10% C→T damage transitions at position 1. MapDamage outputs are summarized in Supplementary Table 7.

We also estimated the levels of mitochondrial DNA contamination. We used contamMix 1.0–10 (ref. 49) that generates a moment-based estimate of the error rate and a Bayesian-based estimate of the posterior probability of the contamination fraction. We conservatively removed individuals with indications of contamination >5% (Supplementary Information, section 5). For males with sufficient depth of coverage we also estimated contamination based on the X chromosome[3] as implemented in ANGSD[50] (Supplementary Information, section 5). Results are shown in Supplementary Table 8. After implementing the 0.5% cut-off for human DNA content, combined with these ancient DNA

authentication criteria, our final sample consisted of 101 individuals (Supplementary Information, section 1).

Data sets. We constructed two data sets for population genetic analysis by merging ancient DNA data generated in this as well as previous studies with two reference panels of modern individual genotype data (Supplementary Information, section 6). For both data sets, genotypes for all ancient individuals were obtained at all variant positions in the reference panel, discarding variants where alleles for the ancient individuals did not match either of the alleles observed in the panel. Genotypes for low-coverage samples (including all data generated in this study) were obtained by randomly sampling a single read with both mapping and base quality ≥ 30. Genotypes for high-coverage samples were called using the 'call' command of bcftools (https://github.com/samtools/bcftools) and filtering for quality score (QUAL) ≥ 30. Error rates and inclusion thresholds for low coverage samples were obtained by performing PCA and model-based clustering (described below) on subsampled data sets of higher coverage individuals. For population genetic analyses (D and f statistics, F_{ST}) we obtained sample allele frequencies for the ancient groups (Supplementary Table 9) at each SNP by counting the total number of alleles observed, treating the low coverage individuals as haploid. See Supplementary Information, section 6 for more details.

PCA and model-based clustering. We performed principal component analysis with EIGENSOFT[51], projecting ancient individuals onto the components inferred from sets of modern individuals by using the 'lsqproject' option of *smartpca*. The data set was converted to all homozygous genotypes before the analysis, by randomly sampling an allele at each heterozygote genotype of modern and high-coverage ancient individuals. See Supplementary Information, section 6 for more details.

Model-based clustering analysis was carried out using the maximum-likelihood approach implemented in ADMIXTURE[52]. We used an approach where we first infer the ancestral components using modern samples only, and then 'project' the ancient samples onto the inferred components using the ancestral allele frequencies inferred by ADMIXTURE (the 'P' matrix). We ran ADMIXTURE on an LD-pruned data set of all 2,345 modern individuals in the Human Origins SNP array data set, assuming $K = 2$ to $K = 20$ ancestral components, selecting the best of 50 replicate runs for each value of K. See Supplementary Information, section 6 for more details. Genotypes where the ancient individuals showed the damage allele at C > T and G > A SNPs were excluded for each low coverage ancient individual.

D- and f-statistics and population differentiation. We used the D and f statistic framework[53] to investigate patterns of admixture and shared ancestry in our data set. All statistics were calculated from allele frequencies using the estimators described previously[53], with standard errors obtained from a block jackknife with 5 Mb block size. We investigated population differentiation by estimating F_{ST} for all pairs of ancient and modern groups from allele frequencies using the sample-size corrected moment estimator of Weir and Hill[54], restricting the analysis to SNPs where a minimum two alleles were observed in each population of the pair. See Supplementary Information, section 6 for more details.

Phenotypes and positive selection. To investigate the temporal dynamics of SNPs associated with phenotypes or putatively under positive selection, we estimated allele frequencies for a catalogue of 104 SNPs[33] in all ancient and modern groups in the 1000 Genomes data set. Genotypes for the *LCT* region were imputed from genotype likelihoods with the 1000 Genomes Phase 3 reference panel[55] using BEAGLE[56]. See Supplementary Information, section 6 for more details.

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Code availability. Source code with R functions used in the analysis for this study is available as an R package at GitHub https://github.com/martinsikora/admixr.

39. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protocols* (2010).
40. Orlando, L. *et al.* Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499,** 74–78 (2013).
41. Malaspinas, A.-S. *et al.* Two ancient human genomes reveal Polynesian ancestry among the indigenous Botocudos of Brazil. *Curr. Biol.* **24,** R1035–R1037 (2014).
42. Willerslev, E. & Cooper, A. Ancient DNA. *Proc. Royal Soc. B* **272,** 3–16 (2005).
43. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104,** 14616–14621 (2007).
44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
45. Schubert, M. *et al.* Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13,** 178 (2012).
46. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

47. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
48. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* (2013).
49. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110,** 2223–2227 (2013).
50. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15,** (2014).
51. Patterson, N., Price, A. L. & Reich, D. Population structure and Eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
52. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).
53. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192,** 1065–1093 (2012).
54. Weir, B. S. & Hill, W. Estimating F-statistics. *Annu. Rev. Genet.* **36,** 721–750 (2002).
55. Nyström, V. *et al.* Microsatellite genotyping reveals end-Pleistocene decline in mammoth autosomal genetic variation. *Mol. Ecol.* **21,** 3391–3402 (2012).
56. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81,** 1084–1097 (2007).
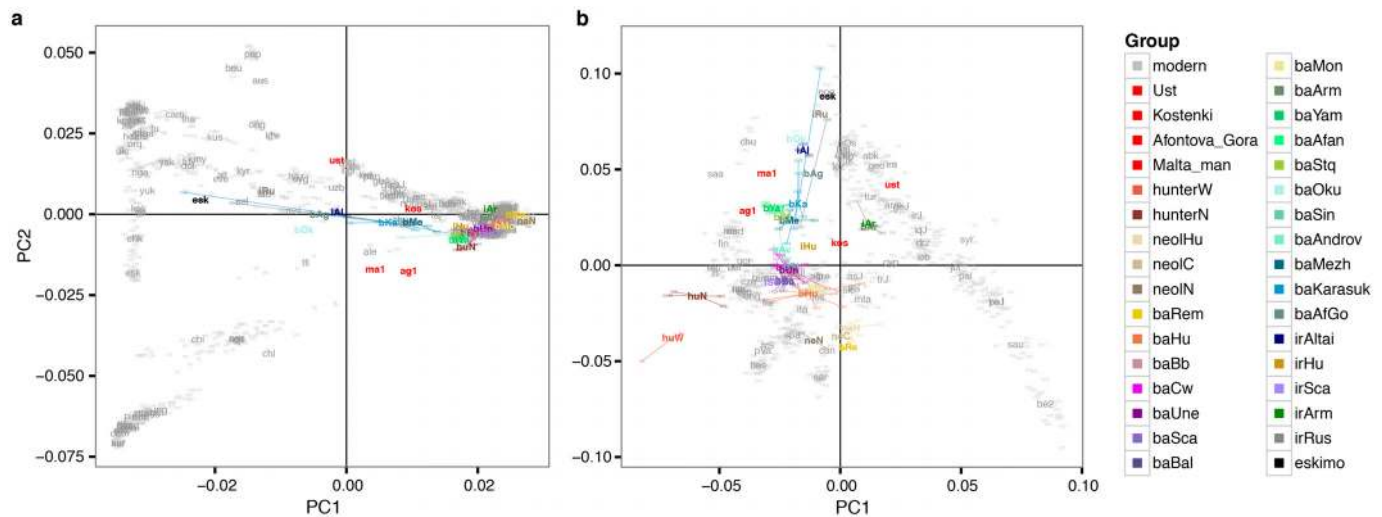
Extended Data Figure 1 | Principal component analysis of ancient genomes.
a, **b**, Principal component analysis of ancient individuals projected onto
contemporary individuals from non-African populations (**a**), Europe, West
Asia and the Caucasus (**b**). Grey labels represent population codes indicating

coordinates for individuals (small) and median of the population (large).
Coloured labels indicate positions for ancient individuals (small) and median
for ancient groups (large). Ancient individuals within a group are connected to
the respective median position by coloured lines.
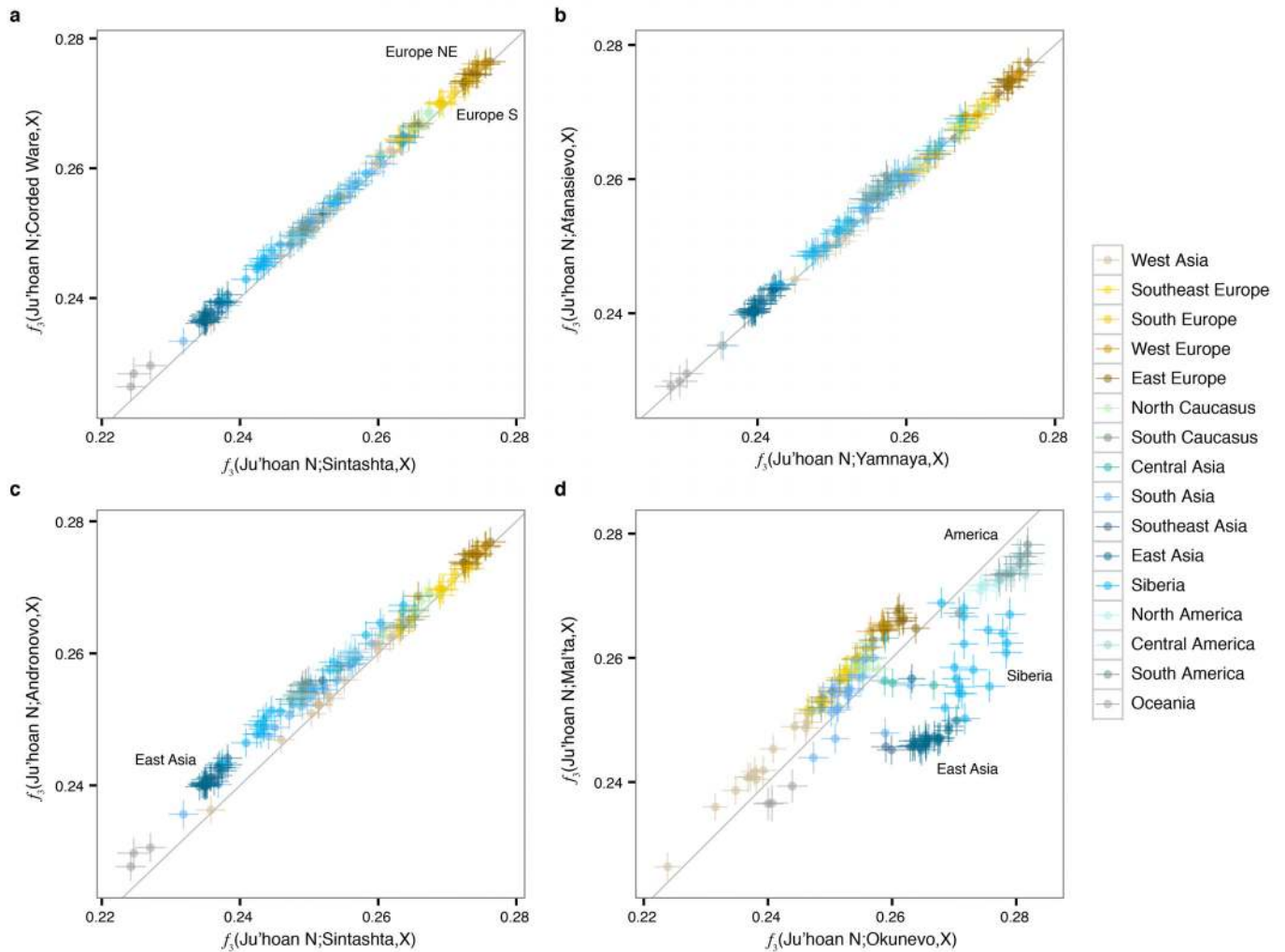
Extended Data Figure 2 | Pairwise outgroup f_3 statistics. Panels depict pairwise plots of outgroup f_3 statistics of the form f_3(Ju'hoan North;Population$_1$, Population$_2$), showing the correlation of the amount of shared genetic drift for a pair of ancient groups (Population$_1$) with all modern populations (Population$_2$) in the Human Origins data set (panel A). Closely related ancient groups are expected to show highly correlated statistics. **a**, Sintashta/Corded Ware. **b**, Yamnaya/Afanasievo. **c**, Sintashta/Andronovo. **d**, Okunevo/Mal'ta. Coloured circles indicate modern populations; error bars indicate ± 1 standard error from the block jackknife.
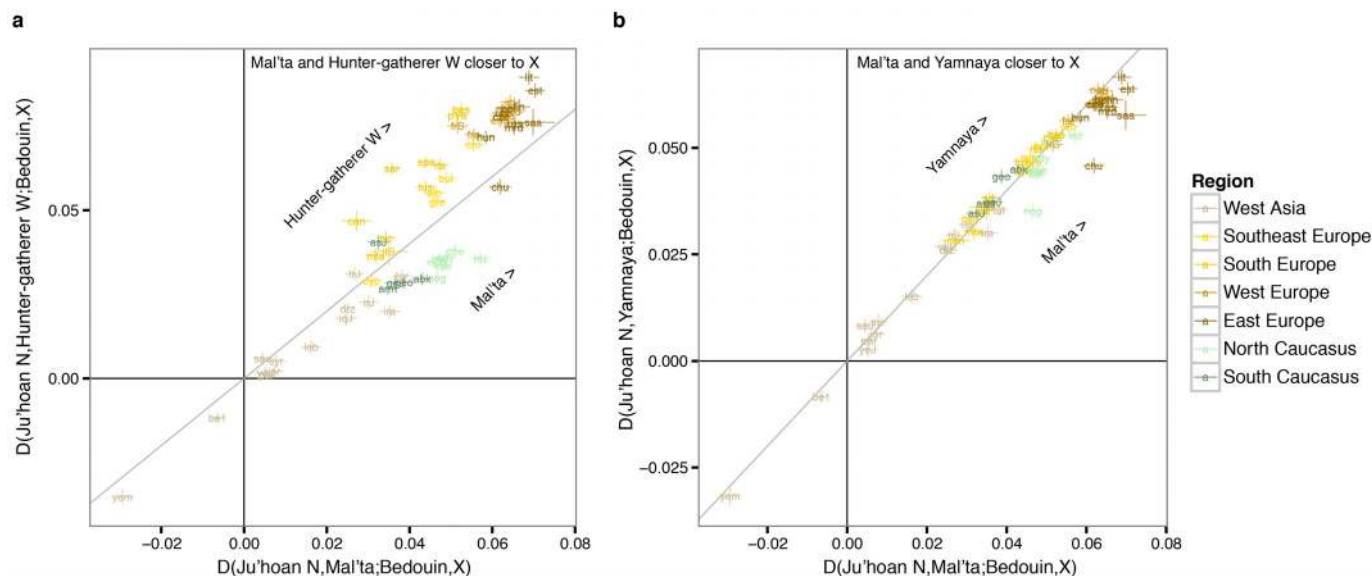
Extended Data Figure 3 | Yamnaya ancestry mirrors Mal'ta ancestry in present-day Europeans and Caucasians. Panels show pairwise plots of D-statistics D(Outgroup, Ancient)(Bedouin, Modern), contrasting Mal'ta (MA1) and Hunter-gatherers (**a**), and MA1 and Yamnaya (**b**). Coloured labels indicate modern populations, with lines corresponding to ± 1 standard error of the respective D-statistic from block jacknife. Text away from the diagonal line indicates an ancient group with relative increase in allele sharing with the respective modern populations.
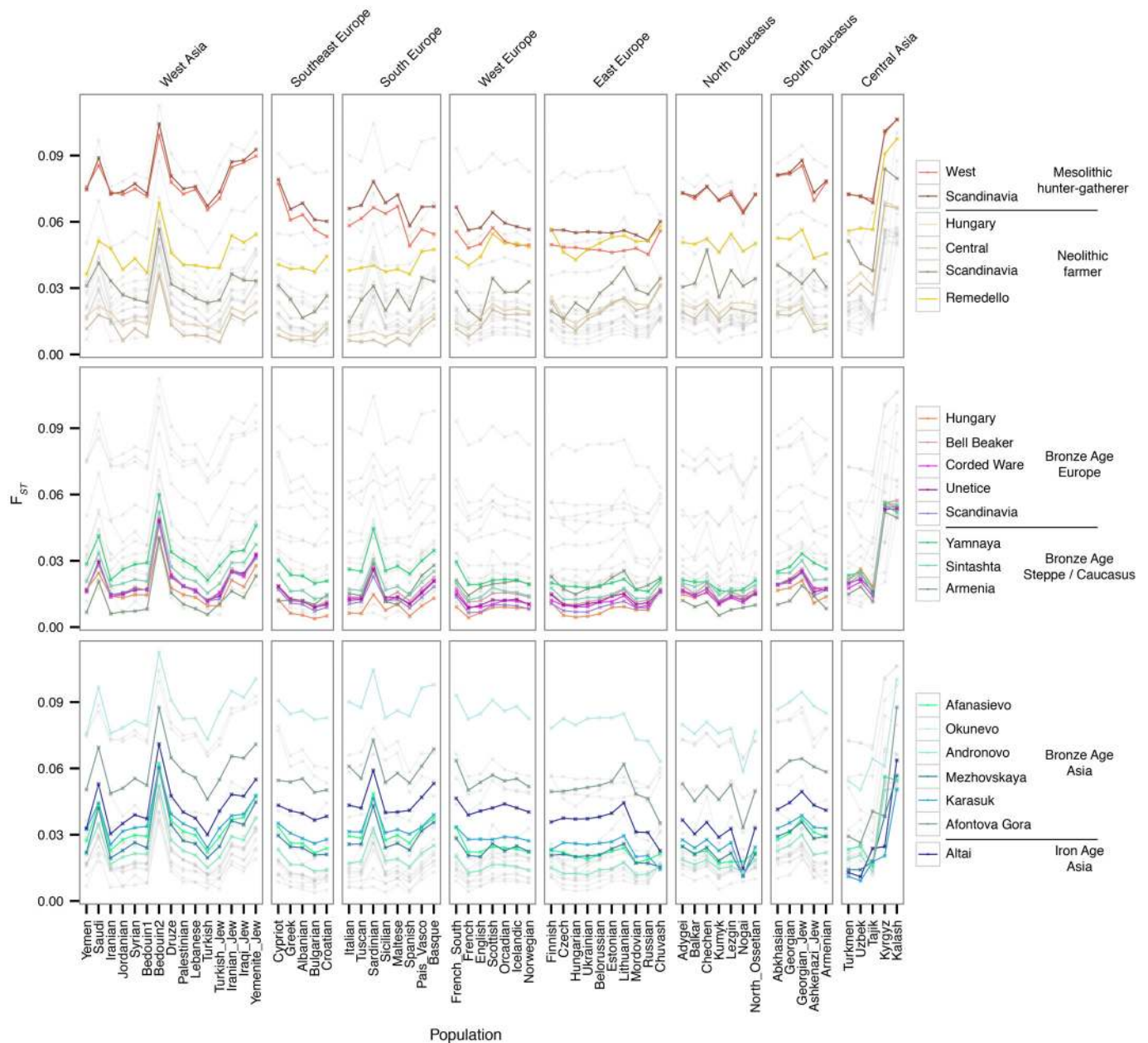
Extended Data Figure 4 | Genetic differentiation between ancient and modern groups in Human Origins data set. Panels show F_{ST} between pairs of modern and ancient groups (coloured lines) for subsets of ancient groups, with results for the remaining groups in the background (grey). Top, early Europeans. Middle, Bronze Age Europeans and steppe/Caucasus. Bottom, Bronze Age Asians. Results based on Human Origins data set (panel A).
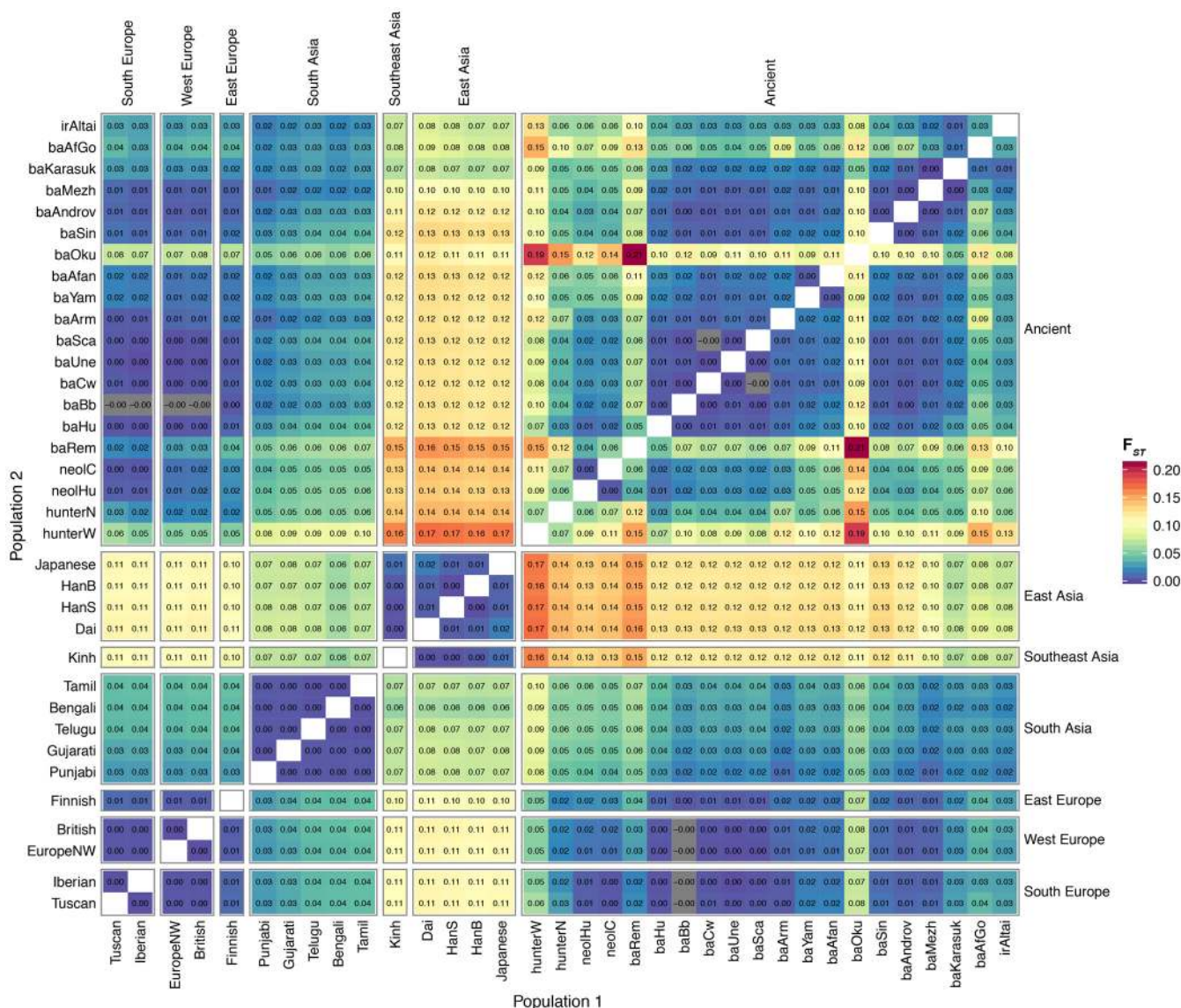
Extended Data Figure 5 | Genetic differentiation between ancient and modern groups in 1000 Genomes data set. Matrix of pairwise F_{ST} values between modern and ancient groups in the 1000 Genomes data set (panel B).
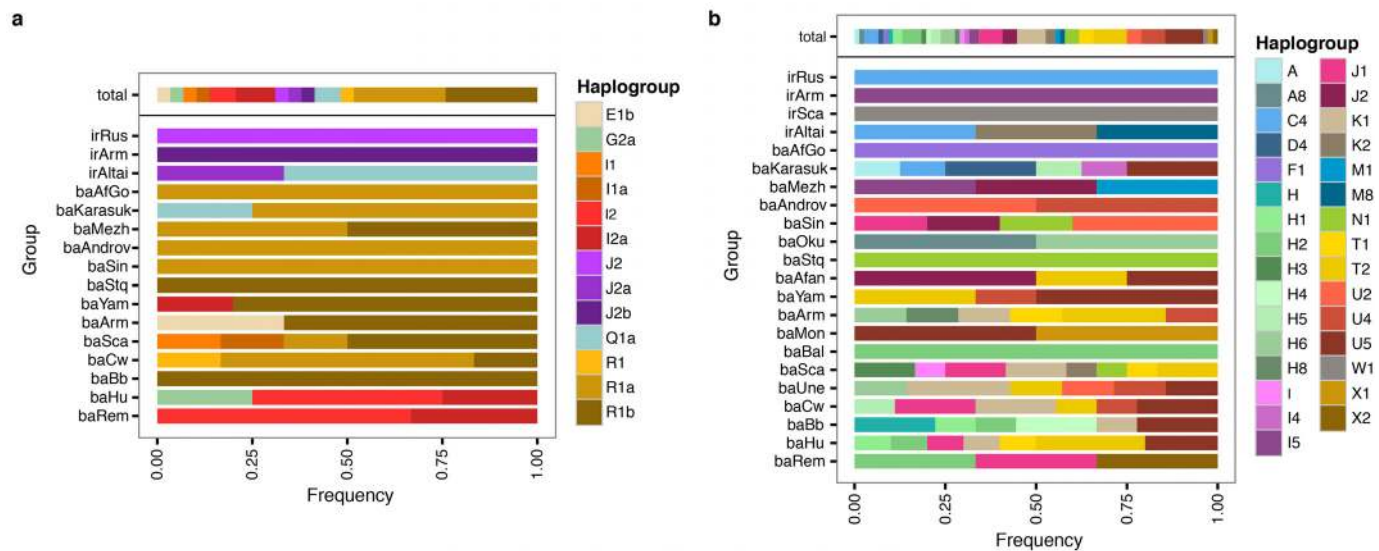
Extended Data Figure 6 | Distribution of uniparental lineages in Bronze Age Eurasians. a, b, Barplots showing the relative frequency of Y chromosome (**a**) and mitochondrial DNA lineages (**b**) in different Bronze Age groups. Top row shows overall frequencies for all individuals combined.
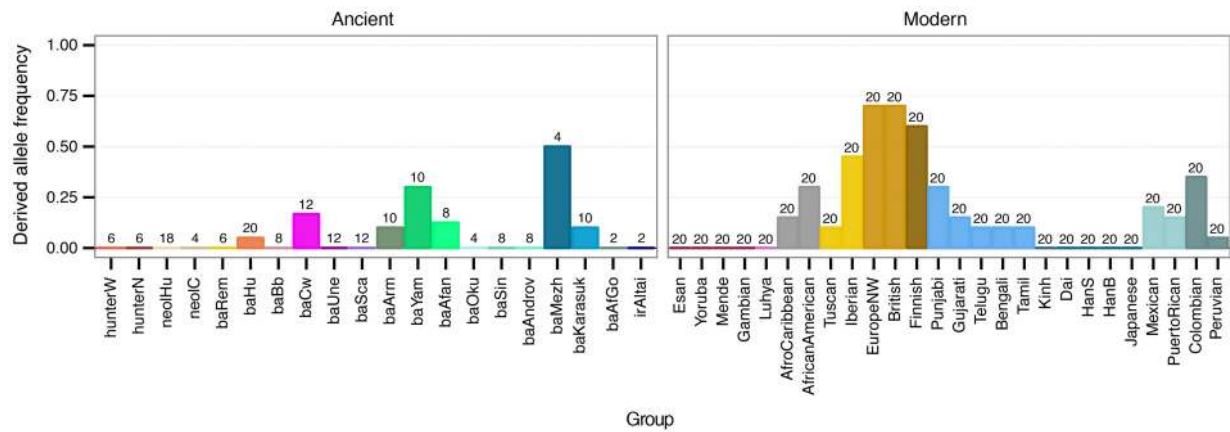
Extended Data Figure 7 | Derived allele frequencies for lactase persistence in modern and ancient groups. Derived allele frequency of rs4988235 in the *LCT* gene inferred from imputation of ancient individuals. Numbers indicate the total number of chromosomes for each group.

Extended Data Table 1 | Selected D-test results from 1000 Genomes data set (panel B)

Configuration	D	Z	Interpretation
D(Yoruba,Neolithic Central)(Remedello,Hungary)	-0.022	-6.2	Remedello is closer to Neolithic farmers than BA Hungarians
D(Yoruba,Neolithic Central)(Hungary,Corded Ware)	-0.014	-5.5	BA Hungarians are closer to Neolithic farmers than Corded Ware
D(Yoruba,Neolithic Central)(Bell Beaker,Corded Ware)	-0.009	-3.4	Bell Beaker is closer to Neolithic farmers than Corded Ware
D(Yoruba,Neolithic Central)(Corded Ware,Yamnaya)	-0.018	-7.0	Corded Ware is closer to Neolithic farmers than Yamnaya
D(Yoruba,Neolithic Central)(Sintashta,Yamnaya)	-0.014	-4.9	Sintashta is closer to Neolithic farmers than Yamnaya
D(Yoruba,Yamnaya)(Hungary,Bell Beaker)	0.011	4.5	Bell Beaker is closer to Yamnaya than BA Hungarians
D(Yoruba,Yamnaya)(Hungary,Corded Ware)	0.016	6.7	Corded Ware is closer to Yamnaya than BA Hungarians
D(Yoruba,Yamnaya)(Hungary,Sintashta)	0.011	4.8	Sintashta is closer to Yamnaya than BA Hungarians
D(Yoruba,Armenia)(Yamnaya,Corded Ware)	0.002	0.9	Corded Ware and Yamnaya form a clade to the exclusion of BA Armenians
D(Yoruba,Corded Ware)(Yamnaya,Armenia)	-0.015	-5.8	
D(Yoruba,Yamnaya)(Armenia,Corded Ware)	0.018	6.8	
D(Yoruba,Yamnaya)(Afanasievo,Karasuk)	-0.047	-17.7	Yamnaya and Afanasievo form a clade to the exclusion of other ancient groups*
D(Yoruba,Afanasievo)(Yamnaya,Karasuk)	-0.038	-16.0	
D(Yoruba,Karasuk)(Yamnaya,Afanasievo)	0.008	3.6	

*Results are shown for Karasuk as group X, which is the only ancient group with $Z > 3$ for D(Yoruba, X)(Yamnaya, Afanasievo)

Extended Data Table 2 | f_3 statistic results for ancient groups

Group	f_3 outgroup*			f_3 admixture†		
	Population with highest f_3 outgroup	f_3	SE	Example population pair with $f_3 < 0$	f_3	Z
Ust-Ishim	Surui	0.240	0.003			
Kostenki‡	Hunter-gatherer W	0.266	0.003			
Afontova Gora‡	Mal'ta	0.317	0.005			
Mal'ta†	Afontova Gora	0.317	0.005			
Hunter-gatherer W	Hunter-gatherer Scandinavia	0.313	0.002			
Hunter-gatherer Scandinavia	Hunter-gatherer W	0.313	0.002	(Hunter-gatherer W,Mal'ta)§	-0.005	-1.78
Neolithic Hungary	Neolithic Central	0.287	0.002	(Hunter-gatherer W,Neolithic Central)§	-0.002	-1.09
Neolithic Central	Neolithic Hungary	0.287	0.002			
Neolithic Scandinavia†	Neolithic Central	0.285	0.003			
Remedello	Neolithic Scandinavia	0.284	0.003			
BA Hungary	Hunter-gatherer W	0.284	0.002	(Hunter-gatherer W,BA Armenia)	-0.011	-8.69
Bell Beaker	Hunter-gatherer W	0.278	0.002	(Remedello,Yamnaya)	-0.007	-4.31
Corded Ware	Hunter-gatherer W	0.279	0.002	(Neolithic Hungary,Yamnaya)	-0.009	-6.16
Unetice	Hunter-gatherer W	0.278	0.002	(Neolithic Central,Afanasievo)	-0.008	-6.07
BA Scandinavia	Hunter-gatherer W	0.282	0.002	(Hunter-gatherer W,BA Armenia)	-0.013	-9.61
BA Baltic†	Hunter-gatherer Scandinavia	0.276	0.008			
BA Montenegro‖	Scottish	0.275	0.005			
BA Armenia	IA Armenia	0.270	0.005	(Neolithic Hungary,Okunevo)	-0.006	-2.66
Yamnaya	Afanasievo	0.284	0.002	(Neolithic Central,Mal'ta)§	-0.004	-2.32
Afanasievo	Yamnaya	0.284	0.002			
Stalingrad quarry†	BA Afontova Gora	0.287	0.007			
Okunevo	Karitiana	0.282	0.003			
Sintashta	Andronovo	0.278	0.002	(Neolithic Central,Okunevo)§	-0.004	-2.43
Andronovo	Hunter-gatherer Scandinavia	0.280	0.002	(Remedello,Mal'ta)	-0.005	-5.13
Mezhovskaya	Andronovo	0.277	0.002	(Sintashta,HanB)	-0.012	-5.62
Karasuk	BA Afontova Gora	0.276	0.003	(Andronovo,HanB)	-0.022	-27.49
BA Afontova Gora	Nganasan	0.282	0.003	(Remedello,Okunevo)§	-0.002	-0.15
IA Altai	BA Afontova Gora	0.274	0.004	(Afanasievo,HanB)	-0.017	-10.88
IA Hungary‡	Lithuanian	0.273	0.002			
IA Scandinavia‡	Hunter-gatherer Scandinavia	0.278	0.003			
IA Armenia‡	Bell Beaker	0.275	0.004			
IA Russia†	Nganasan	0.285	0.003			

*Human origins data set (panel A); †1000 Genomes data set (panel B); ‡group with single individual; §pair with lowest f_3 reported for groups with negative f_3 without significant Z-score after correcting for multiple hypothesis tests ($-4.1 < \min(Z) < 0$; 1,260 tests per group); ‖too few markers with data from more than one chromosome.