

Population Genomics of Bronze Age Eurasia

Morten E. Allentoft^{1*}, Martin Sikora^{1*}, Karl-Göran Sjögren², Simon Rasmussen³,
Morten Rasmussen¹, Jesper Stenderup¹, Peter B. Damgaard¹, Hannes Schroeder^{1,4},
Torbjörn Ahlström⁵, Lasse Vinner¹, Anna-Sapfo Malaspinas¹, Ashot Margaryan¹,
Tom Higham⁶, David Chivall⁶, Niels Lynnerup⁷, Lise Harvig⁷, Justyna Baron⁸,
Philippe Della Casa⁹, Paweł Dąbrowski¹⁰, Paul R Duffy¹¹, Alexander V. Ebel¹²,
Andrey Epimakhov¹³, Karin Frei¹⁴, Mirosław Furmanek⁸, Tomasz Gralak⁸, Andrey
Gromov¹⁵, Stanisław Gronkiewicz¹⁶, Gisela Grupe¹⁷, Tamás Hajdu^{18,19}, Radosław
Jarysz²⁰, Valeri Khartanovich¹⁵, Alexandr Khokhlov²¹, Viktória Kiss²², Jan Kolář^{23, 24},
Aivar Kriiska²⁵, Irena Lasak⁸, Cristina Longhi²⁶, George McGlynn¹⁷, Algimantas
Merkevicius²⁷, Inga Merkyte²⁸, Mait Metspalu²⁹, Ruzan Mkrtchyan³⁰, Vyacheslav
Moiseyev¹⁵, László Paja^{31,32}, György Pálfi³², Dalia Pokutta², Łukasz Pospieszny³³, T.
Douglas Price³⁴, Lehti Saag²⁹, Mikhail Sablin³⁵, Natalia Shishlina³⁶, Václav Smrčka³⁷,
Vasilii I. Soenov³⁸, Vajk Szeverényi²², Gusztáv Tóth³⁹, Synaru V. Trifanova³⁸, Liivi
Varul²⁵, Magdolna Vicze⁴⁰, Levon Yepiskoposyan⁴¹, Vladislav Zhitenev⁴², Ludovic
Orlando¹, Thomas Sicheritz-Pontén³, Søren Brunak^{3, 43}, Rasmus Nielsen⁴⁴, Kristian
Kristiansen² and Eske Willerslev^{1†}

*These authors contributed equally to the work.

†Corresponding author: ewillerslev@snm.ku.dk

1) Centre for GeoGenetics, Natural History Museum, University of Copenhagen,
Øster Voldgade 5-7, 1350 Copenhagen K, Denmark.

- 2) Department of Historical Studies, University of Gothenburg, 405 30 Gothenburg, Sweden.
- 3) Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Kgs Lyngby, Denmark.
- 4) Faculty of Archaeology, Leiden University, 2300 Leiden, The Netherlands.
- 5) Department of Archaeology and Ancient History, Lund University, 221 00 Lund, Sweden.
- 6) Oxford Radiocarbon Accelerator Unit, University of Oxford, OX1 3QY Oxford, UK.
- 7) Unit of Forensic Anthropology, Department of Forensic Medicine, University of Copenhagen, 2100 Copenhagen, Denmark.
- 8) Institute of Archaeology, University of Wrocław, 50-139 Wrocław, Poland.
- 9) Archaeological Institute, University of Zurich, CH-8006, Zurich, Switzerland.
- 10) Wroclaw Medical University, 50-367 Wrocław, Poland.
- 11) Department of Anthropology, University of Toronto, ONM5S 2S2, Canada.
- 12) Department of Archeology and General History, Gorno-Altai State University, 649000 Gorno-Altai, Russia.
- 13) Institute of History and Archaeology RAS (South Ural Department), South Ural State University, 454080 Chelyabinsk, Russia.
- 14) Environmental Research and Material Science and Centre for Textile Research, The National Museum of Denmark, 1471 Copenhagen K, Denmark.

- 15) Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS,
199034 St. Petersburg, Russia.
- 16) Department of Anthropology, Polish Academy of Sciences, 50–449 Wrocław,
Poland.
- 17) Biocentre of the Ludwig-Maximilian-University München, 82152 Munich,
Germany.
- 18) Department of Biological Anthropology, Institute of Biology, Eötvös Loránd
University, H-1117 Budapest, Hungary.
- 19) Department of Anthropology, Hungarian Natural History Museum, H-1083
Budapest, Hungary.
- 20) The Archaeological Museum of Wrocław, 50-077 Wrocław, Poland.
- 21) Samara State Academy of Social Science and Humanities, 443099 Samara,
Russia.
- 22) Institute of Archaeology of the Hungarian Academy of Sciences, Research Center
for the Humanities, H-1250 Budapest, Hungary.
- 23) Institute of Archaeology and Museology, Faculty of Arts, Masaryk University,
CZ-602 00 Brno, Czech Republic.
- 24) Department of Vegetation Ecology, Institute of Botany of the Czech Academy of
Sciences, CZ-602 00 Brno, Czech Republic.
- 25) Department of Archaeology, University of Tartu, 51003 Tartu, Estonia.
- 26) Archaeological Superintendence of Lombardy, 20123 Milano, Italy.
- 27) Department of Archaeology, University of Vilnius, LT-01513 Vilnius, Lithuania.

- 28) The SAXO Institute, University of Copenhagen, 2300 Copenhagen S, Denmark.
- 29) Department of Evolutionary Biology, Estonian Biocentre and University of Tartu, 51010 Tartu, Estonia.
- 30) Department of History, Yerevan State University, 0025 Yerevan, Armenia.
- 31) Hungarian National Museum, H-1083 Budapest, Hungary.
- 32) Department of Biological Anthropology, University of Szeged, H-6726 Szeged, Hungary.
- 33) Institute of Archaeology and Ethnology of the Polish Academy of Sciences, 61-612 Poznań, Poland.
- 34) Laboratory for Archaeological Chemistry, University of Wisconsin-Madison, WI 53706 Madison, USA.
- 35) Zoological Institute of the Russian Academy of Sciences, 199034 St. Petersburg, Russia.
- 36) Department of Archaeology, State Historical Museum, 109012 Moscow, Russia.
- 37) Institute for History of Medicine and Foreign Languages of the First Faculty of Medicine, Charles University, 121 08 Prague, Czech Republic.
- 38) Research Center for the History and Culture of the Turkic Peoples, Gorno-Altai State University, 649000 Gorno-Altai, Russia.
- 39) Department of Pre- and Early History, Institute of Archaeological Sciences, Faculty of Humanities, Eötvös Loránd University, H-1088 Budapest, Hungary.
- 40) Matrica Museum, 2440 Százhalombatta, Hungary.

41) Laboratory of Ethnogenomics, Institute of Molecular Biology, National Academy of Sciences, 0014 Yerevan, Armenia.

42) Department of Archaeology, Faculty of History, Moscow State University, 119991 Moscow, Russia.

43) Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, 2200 Copenhagen, Denmark.

44) Center for Theoretical Evolutionary Genetics, University of California, Berkeley, CA 94720-3140 Berkeley, USA.

Summary

The Bronze Age of Eurasia (c. 3,000-1,000 years BC) was a period of major cultural changes. However, it remains debated if these resulted from the circulation of ideas or from large-scale human migrations, potentially also facilitating the spread of Indo-European 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 hypothesised spread of Indo-European languages during 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 the latter trait than previously believed.

Introduction

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 ka 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 ka BP³. 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 Eurasians⁴ 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) c. 8-5 ka BP⁷⁻¹² and possibly also from Northern Asia¹⁰. However, what happened hereafter, during the Bronze Age (BA), is much less clear.

The archaeological record testifies to major cultural changes in Europe and Asia after the Neolithic period. By 3 ka before Christ (BC), the Neolithic farming cultures in temperate Eastern Europe appear to be largely replaced by the Early BA Yamnaya Culture, which is associated with a completely new perception of family, property and personhood^{13,14}, rapidly stretching from Hungary to the Urals¹⁵. By 2,8 ka BC a new social and economic formation, variously named Corded Ware, Single Grave or Battle Axe cultures develops in temperate Europe, possibly deriving from the Yamnaya

background, and culturally replacing the remaining Neolithic farmers (NF)^{16,17} (Fig. 1). In Western and Central Asia, hunter-gatherers (HG) still dominate in Early BA, except in the Altai Mountains and Minusinsk Basin where the Afanasievo Culture exists with a close cultural affinity to Yamnaya¹⁵ (Fig. 1). From the beginning of 2 ka BC a new class of master artisans known as the Sintashta Culture emerges in the Urals, building chariots, breeding and training horses (Fig. 1), and producing sophisticated new weapons¹⁸. These innovations quickly spread across Europe and into Asia where they appear to give rise to the Andronovo Culture^{19,20} (Fig. 1). In the Late BA around 1.5 ka BC, the Andronovo Culture is gradually replaced by the Mezhovskaya, Karasuk, and Koryakova cultures²¹. It remains debated if these major cultural shifts during the BA in Europe and Asia resulted from the migration of people or through cultural diffusion among settled groups¹⁵⁻¹⁷, and if the spread of the Indo-European languages was linked to these events or predates them¹⁵.

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 (aDNA) making sequencing costly and time consuming²³. To overcome this, we increased the average output of authentic endogenous DNA 4-fold by (i) targeting the outer cementum layer in teeth rather than the inner dentine layer^{24,25}, (ii) adding a 'pre-digestion' step to remove surface contaminants^{24,26}, and (iii) developing a new binding buffer for aDNA extraction (Supplementary Information Section 3). This allowed us to obtain low

coverage genome sequences (0.01-7.4X average depth, overall average = 0.7X) of 101 Eurasian individuals spanning the entire BA, including some late Neolithic and Iron Age individuals (Fig. 1, Supplementary Information Section 1-2). Our dataset includes 19 genomes, between 1.1-7.4X average depth, thereby doubling the number of existing Eurasian ancient genomes above 1X²⁷.

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 BA (Fig. 2; Extended Data Fig. 1; Supplementary Fig. 5, 6). Populations in Northern and Central Europe are composed of a mixture of the earlier HG and NF¹⁰ groups, but receive 'Caucasian' genetic input at the onset of the BA (Fig 2). This coincides with the archaeologically well-defined expansion of the Yamnaya culture from the Pontic-Caspian steppe into Europe (Fig. 1, 2). This admixture event results 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). While European BA cultures such 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, Extended Data Table 1). Using D-statistics, we find that Corded Ware and Yamnaya individuals form a clade to the exclusion of BA Armenians (Extended Data Table 1) showing that the Caucasus component present in BA Europe has a steppe origin rather than a

Southern Caucasus origin. Earlier studies have shown that Southern Europeans receive substantial gene flow from NF during the Neolithic⁹. 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 NF (Fig. 2; Extended Data Fig. 1, Supplementary Figure 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 BA 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 Sintashta in Asia or the Middle East²⁸. While 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 NF ancestry in both the Corded Ware and the Sintashta, combined with the NF absence in the earlier Yamnaya, would suggest the former being more likely (Fig. 2b, Extended Data Table 1).

Bronze Age Asia

We find that the BA in Asia is equally dynamic and characterized by large-scale migrations and population replacements. The early BA Afanasievo Culture in the Altai-Sayan region is genetically indistinguishable from Yamnaya, confirming an eastward expansion across the steppe (Fig. 1, 3b; Extended Data Fig. 2b; 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 BA (Fig. 1). The Andronovo Culture, which arose in Central Asia during the later Bronze Age (Fig. 1), is genetically closely related to Sintashta peoples (Extended Data Fig. 2c), and clearly distinct from both Yamnaya and Afanasievo (Fig. 3b; Extended Data Table 1). Therefore, Andronovo represents a temporal and geographical extension of the Sintashta gene pool. Towards the end of the BA in Asia, Andronovo is replaced by the Karasuk, Mezhovskaya, and Iron Age cultures which appear multi-ethnic and show gradual admixture with East Asians (Fig. 3b; Extended Data Table 2), corresponding with anthropological and biological research²⁹. 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 BA Okunevo Culture from the Sayano-Altai region (Fig 1) are closely related to present day Native Americans (Extended Data Fig. 2d), which confirms previous craniometric studies³⁰. This finding implies that Okunevo could represent a remnant population related to the Upper Palaeolithic Mal'ta HG population from Lake Baikal that contributed genetic material to Native Americans⁴.

Spread of the Indo-European languages

Historical linguists have contended 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 3rd 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 BA (Fig. 1)

corresponds well with the hypothesised expansion of the Indo-European languages. In contrast to recent genetic findings³², however, we only find weak evidence for admixture in Yamnaya, and only when using BA 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 HG as potential source population for admixture in our dataset. Modern Europeans show some genetic links to Mal'ta⁴ 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 likely derives from the spread of the Yamnaya Culture that distantly shares ancestry with Mal'ta (Fig. 2b, 3b; Extended Data Fig. 3).

Formation of Eurasian genetic structure

It is clear from our autosomal, mtDNA, and Y-chromosome data (Extended Data Fig. 6) that the European and Central Asian gene pools towards the end of the BA mirror present-day Eurasian genetic structure to an extent not seen in the previous periods (Fig. 2, 3, Extended Data Fig. 1, Supplementary Figure 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 BA groups than with earlier Mesolithic and Neolithic groups (Extended Data Fig. 4, 5). Notable exceptions are contemporary populations from Southern Europe such as Sardinians and Sicilians, which show the lowest F_{ST} with NF. 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 HG and BA 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 BA through population growth, combined with continuing gene flow between populations, to generate the low differentiation observed in contemporary West Eurasians.

Temporal dynamics of selected SNPs

The size of our dataset allows us to investigate the temporal dynamics of 104 genetic variants associated with important phenotypic traits or putatively undergoing positive selection³³ (Supplementary Table 13). Focusing on four well-studied polymorphisms, we find that two 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 c. 3,000 years between the Mesolithic and BA 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 HG as previously described³³. We find it at intermediate frequency in BA Europeans, but it is notably absent from the Pontic-Caspian steppe and Caucasus 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. While tolerance is high in present-day Northern Europeans, we find it at most at low

frequency in the BA (10% in BA Europeans; Fig. 4), indicating a more recent onset of positive selection than previously estimated³⁴. To further investigate its distribution, we imputed all SNPs in a 2 MB region around rs4988235 in all ancient individuals using the 1000 Genomes phase 3 dataset as a reference panel, as previously described¹². Our results confirm a low frequency of rs4988235 in Europeans, with a derived allele frequency of 5% in the combined BA Europeans (genotype probability > 0.85) (Fig. 4B). Among BA Europeans, the highest tolerance frequency was found in Corded Ware and the closely-related Scandinavian BA cultures (Extended Data Fig. 7). Interestingly, the BA 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 BA 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 is directional into Central Asia and the Altai-Sayan region and likely without much local infiltration, whereas the resulting Corded Ware Culture in Europe is the result of admixture with the local Neolithic people. The enigmatic Sintashta Culture near the Urals bears genetic resemblance to Corded Ware and is therefore likely to be an eastward migration into Asia. As this culture spreads towards Altai it evolves into the the Andronovo Culture (Fig 1.), which is 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 BA is the most likely scenario for the spread of Indo-European languages, in line with reconstructions based on archaeological and historical linguistic data^{15,31}. In the light of our results, the existence of the Afanasievo Culture near Altai around 3 ka BC could also provide an explanation for the mysterious presence of one of the oldest Indo-European languages Tocharian in the Tarim basin in China³⁷. It seems plausible that Afanasievo, with their genetic western (Yamnaya) origin, spoke an Indo-European language and could have introduced this southward to Xinjiang and Tarim³⁸. Importantly, however, although our results have confirmed a tight correspondence between cultural changes, migrations, and linguistic patterns, we caution that such relationships cannot always be expected but must be demonstrated case by case.

References

- 1 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449, doi:10.1038/nature13810 (2014).
- 2 Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118, doi:10.1126/science.aaa0114 (2014).
- 3 Rasmussen, M. *et al.* An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* **334**, 94-98, doi:10.1126/Science.1211177 (2011).
- 4 Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87-91, doi:10.1038/nature12736 (2014).

- 5 Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345**, doi:10.1126/science.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, doi:10.1038/nature13025 (2014).
- 7 Bramanti, B. *et al.* Genetic discontinuity between local hunter-gatherers and central europe's first farmers. *Science* **326**, 137-140, doi:10.1126/science.1176869 (2009).
- 8 Malmstrom, H. *et al.* Ancient DNA Reveals Lack of Continuity between Neolithic Hunter-Gatherers and Contemporary Scandinavians. *Current Biology* **19**, 1758-1762, doi:10.1016/j.cub.2009.09.017 (2009).
- 9 Skoglund, P. *et al.* Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science* **336**, 466-469, doi:10.1126/science.1216304 (2012).
- 10 Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, doi:10.1038/nature13673 (2014).
- 11 Haak, W. *et al.* Ancient DNA from European Early Neolithic Farmers Reveals Their Near Eastern Affinities. *PLoS Biology* **8**, doi:10.1371/journal.pbio.1000536 (2010).
- 12 Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun* **5**, doi:10.1038/ncomms6257 (2014).
- 13 Kristiansen, K. in *The World System and the Earth System. Global Socioenvironmental Change and Sustainability since the Neolithic* (eds B Hornborg & C Crumley) (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 University Press, 2007).
- 16 Harrison, R. & Heyd, V. in *Praehistorische Zeitschrift* Vol. 82 129 (2007).
- 17 Vandkilde, H. *Culture and Change in the Central European Prehistory, 6th to 1st millennium BC*. (Aarhus University Press, 2007).
- 18 Kristiansen, K. & Larsson, T. *The Rise of Bronze Age Society. Travels, Transmissions and Transformations*. (Cambridge University 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, doi:doi:10.1017/S0003598X00095235 (2007).
- 20 Kuznetsov, P. F. The emergence of Bronze Age chariots in eastern Europe. *Antiquity* **80**, 638-645, doi:doi:10.1017/S0003598X00094096 (2006).
- 21 Koryakova, L. & Epimakhov, A. V. *The Urals and Western Siberia in the Bronze and Iron Ages*. (Cambridge University Press, 2007).
- 22 Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757-762, doi:10.1038/nature08835 (2010).
- 23 Carpenter, Meredith L. *et al.* Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *American Journal of Human Genetics* **93**, 852-864, doi:10.1016/j.ajhg.2013.10.002 (2013).
- 24 Damgaard, P. d. B. *et al.* Improving access to endogenous DNA in ancient bones and teeth. *bioRxiv* (2015).

- 25 Adler, C. J., Haak, W., Donlon, D., Cooper, A. & Consortium, G. Survival and recovery of DNA from ancient teeth and bones. *Journal of Archaeological Science* **38**, 956-964, doi:Doi 10.1016/J.Jas.2010.11.010 (2011).
- 26 Orlando, L. *et al.* True single-molecule DNA sequencing of a pleistocene horse bone. *Genome research* **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**, doi:doi: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? *American Journal of Physical Anthropology* **108**, 193-204, doi:10.1002/(SICI)1096-8644(199902)108:2<193::AID-AJPA5>3.0.CO;2-V (1999).
- 31 Kristiansen, K. in *Becoming European. The transformation of third millennium Northern and Western Europe* (eds C Prescott & H Glørstad) (Oxbow Books, 2012).
- 32 Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **advance online publication**, doi:10.1038/nature14317 (2015).

- 33 Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225-228, doi:10.1038/nature12960 (2014).
- 34 Itan, Y., Powell, A., Beaumont, M. A., Burger, J. & Thomas, M. G. The Origins of Lactase Persistence in Europe. *PLoS Computational Biology* **5**, doi:e100049110.1371/journal.pcbi.1000491 (2009).
- 35 Mallory, J. *In Search of the Indo-Europeans. Language, Archaeology and Myth.* (Thames and Hudson, 1987).
- 36 Renfrew, A. C. *Archaeology and Language. The Puzzle of Indo-European Origins.* (Penguin Books, 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, doi:10.1007/s00439-009-0683-0 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements

We thank K. Magnussen, L.A. Petersen, C.D. Mortensen and A. Seguin-Orlando at the Danish National Sequencing Centre for help with the sequencing. We thank C. G. Zacho for technical assistance. The project was funded by The European Research Council (FP/2007-2013, grant no. 269442, The Rise), The University of Copenhagen

(KU2016 programme), Marie Curie Actions of the European Union (FP7/2007-2013, grant no. 300554), The Villum Foundation (Young Investigator Programme, grant no. 10120), Frederik Paulsen, The Miller Institute, UC Berkeley, The Lundbeck Foundation, and The Danish National Research Foundation.

Author Contributions

E.W. and K.K. initiated and lead the study. M.E.A., J.S., L.V., H.S., P.B.D., A.M., M.R., L.S. performed the DNA lab work. M.S., S.R., M.E.A., A.S.M., P.B.D., A.M. analysed the genetic data. K.G.S., T.A., N.L., L.H., J.B., P.D.C., P.D., P.R.D., A.E., A.V.E., K.F., M.F., G.G., T.G., A.G., S.G., T.H., R.J., J.K., V.K., A.K., V.K., A.K., I.L., C.L., A.M., G.M., I.M., M.M., R.M., V.M., D.P., G.P., L.P., D.P., L.P., M.S., N.S., V.S., V.S., V.I.S., G.T., S.V.T., L.V., M.V., L.Y., V.Z. collected the samples and/or provided input to the archaeological interpretations. T.H. and D.C. conducted the radiocarbon dating. T.S.B., L.O., S.B., R.N. provided input to the genetic analyses. E.W., K.K., M.E.A., M.S., K.G.S. wrote the paper with input from all co-authors.

Author Information

DNA sequence alignments are available from the European Nucleotide Archive (<http://www.ebi.ac.uk/ena>) under accession number PRJEB9021. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to ewillerslev@snm.ku.dk.

Figure legends

Figure 1. Distribution maps

Left: Localities, cultural associations, and approximate timeline of 101 sampled ancient individuals from Europe and Central Asia. **Upper right:** Distribution of Early Bronze Age cultures Yamnaya, Corded Ware, and Afanasievo with arrows showing the Yamnaya expansions. **Lower right:** Middle and Late Bronze Age cultures Sintashta, Andronovo, Okunevo, and Karasuk with the eastward migration indicated. Black markers represent chariot burials (2-1,8 ka 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,LN:Middle-Late Neolithic. EBA,MBA, LBA:Early-Middle-Late Bronze Age. IA:Iron Age. BAC, CWC: Battle Axe and Corded Ware cultures.

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.

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.

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* gene inferred from imputation of ancient individuals. Numbers indicate the total number of chromosomes for each group. BA Bronze Age; IA Iron Age.

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 (aDNA) clean lab facilities at 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 adapters³⁹ following established protocols³⁹⁻⁴¹. 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 aDNA^{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 experiment²³ 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. Adapter 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-0.6.2⁴⁴ with the seed disabled to allow for higher sensitivity⁴⁵. Mapped reads were filtered for mapping quality 30 and sorted using Picard (<http://picard.sourceforge.net>) and samtools⁴⁶. 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 calmd⁴⁶. Read depth and

coverage were determined using pysam (<http://code.google.com/p/pysam/>) and BEDtools⁴⁷. 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 2.0⁴⁸ 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) δ s, 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 summarised in Supplementary Table 7.

We also estimated the levels of mtDNA contamination. We used contamMix 1.0-10⁴⁹ 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-chromosome³ as implemented in ANGSD⁵⁰ (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 aDNA authentication criteria, our final sample consisted 101 individuals (Supplementary Information Section 1).

Datasets

We constructed two datasets for population genetic analysis by merging aDNA data generated in this as well as previous studies with two reference panels of modern individual genotype data (Supplementary Information Section 6). For both datasets, 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 datasets 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 with EIGENSOFT⁵¹, projecting ancient individuals onto the components inferred from sets of modern individuals by using the ‘lsqproject’ option of *smartpca*. The dataset was converted to all homozygous genotypes prior to 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 ADMIXTURE⁵². 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 dataset of all 2,345 modern individuals in the Human Origins SNP array dataset, 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 framework⁵³ to investigate patterns of admixture and shared ancestry in our dataset. All statistics were calculated from allele frequencies using the estimators described previously⁵³, with standard errors obtained from a block jackknife with 5Mb 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 Hill⁵⁴, 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 SNPs³³ in all ancient and modern groups in the 1000 Genomes dataset. Genotypes for the LCT region were imputed from genotype likelihoods with the 1000 Genomes Phase 3 reference panel⁵⁵ using BEAGLE⁵⁶. See Supplementary Information Section 6 for more details.

References

- 39 Meyer, M. & Kircher, M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols* **2010**, pdb.prot5448, doi:10.1101/pdb.prot5448 (2010).
- 40 Orlando, L. *et al.* Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74-78, doi:10.1038/nature12323 (2013).
- 41 Malaspinas, A.-S. *et al.* Two ancient human genomes reveal Polynesian ancestry among the indigenous Botocudos of Brazil. *Current Biology* **24**, R1035-R1037 (2014).
- 42 Willerslev, E. & Cooper, A. Ancient DNA. *P Roy Soc B-Biol Sci* **272**, 3-16, doi:10.1098/rspb.2004.2813 (2005).
- 43 Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 14616-14621, doi:10.1073/pnas.0704665104 (2007).

- 44 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (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, doi:10.1093/bioinformatics/btp352 (2009).
- 47 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
- 48 Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, doi:10.1093/bioinformatics/btt193 (2013).
- 49 Fu, Q. M. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2223-2227, doi:Doi 10.1073/Pnas.1221359110 (2013).
- 50 Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *Bmc Bioinformatics* **15**, Doi 10.1186/S12859-014-0356-4 (2014).
- 51 Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genet* **2**, e190, doi:10.1371/journal.pgen.0020190 (2006).
- 52 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655-1664, doi:10.1101/gr.094052.109 (2009).

- 53 Patterson, N. J. *et al.* Ancient Admixture in Human History. *Genetics*, doi:10.1534/genetics.112.145037 (2012).
- 54 Weir, B. S. & Hill, W. Estimating F-statistics. *Annual review of genetics* **36**, 721-750 (2002).
- 55 Nyström, V. *et al.* Microsatellite genotyping reveals end-Pleistocene decline in mammoth autosomal genetic variation. *Molecular ecology* **21**, 3391-3402, doi:10.1111/j.1365-294X.2012.05525.x (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. *The American Journal of Human Genetics* **81**, 1084-1097 (2007).

Extended data legends

Extended Data Figure 1. Principal component analysis of ancient genomes.

Principal component analysis of ancient individuals projected onto contemporary individuals from (a) non-African populations, (b) Europe, West Asia and the Caucasus. 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 $f_3(\text{Ju'hoan North}; \text{Population}_1, \text{Population}_2)$, 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 dataset (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 errors from 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 $D(\text{Outgroup, Ancient})(\text{Bedouin, Modern})$, contrasting (a) Mal'ta (MA1) and Hunter-gatherers, and (b) MA1 and Yamnaya. Coloured labels indicate modern populations, with lines corresponding to ± 1 standard error of the respective D-statistic from block jackknife. Text off the diagonal line indicates 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 dataset.

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) BA Europeans and steppe/Caucasus. (bottom) Bronze Age Asians. Results based on Human Origins dataset (panel A).

Extended Data Figure 5. Genetic differentiation between ancient and modern groups in 1000 genomes dataset

Matrix of pairwise F_{ST} values between modern and ancient groups in the 1000 Genomes dataset (panel B).

Extended Data Figure 6. Distribution of uniparental lineages in Bronze Age Eurasians

Barplots showing the relative frequency of (a) Y chromosome and (b) mtDNA lineages in different BA 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 dataset
(panel B)**

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

Extended Data Table 2. f_3 -statistic results for ancient groups

*Human origins dataset (panel A); [†]1000 Genomes dataset (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 >1 chromosomes.







