Article

Population genomics of the Viking world

https://doi.org/10.1038/s41586-020-2688-8

Received: 12 July 2019

Accepted: 21 May 2020

Published online: 16 September 2020

Check for updates

Ashot Margaryan[1,2,3,71], Daniel J. Lawson[4,5,71], Martin Sikora[1,71], Fernando Racimo[1,71], Simon Rasmussen[6], Ida Moltke[7], Lara M. Cassidy[8], Emil Jørsboe[7,9], Andrés Ingason[1,10,11], Mikkel W. Pedersen[1], Thorfinn Korneliussen[1,12], Helene Wilhelmson[13,14], Magdalena M. Buś[15], Peter de Barros Damgaard[1], Rui Martiniano[16], Gabriel Renaud[1,17], Claude Bhérer[18], J. Víctor Moreno-Mayar[1,19], Anna K. Fotakis[3], Marie Allen[15], Raili Allmäe[20], Martyna Molak[21], Enrico Cappellini[3], Gabriele Scorrano[3], Hugh McColl[1], Alexandra Buzhilova[22], Allison Fox[23], Anders Albrechtsen[7], Berit Schütz[24], Birgitte Skar[25], Caroline Arcini[26], Ceri Falys[27], Charlotte Hedenstierna Jonson[28], Dariusz Błaszczyk[29], Denis Pezhemsky[22], Gordon Turner-Walker[30], Hildur Gestsdóttir[31], Inge Lundstrøm[3], Ingrid Gustin[13], Ingrid Mainland[32], Inna Potekhina[33], Italo M. Muntoni[34], Jade Cheng[1], Jesper Stenderup[1], Jilong Ma[1], Julie Gibson[32], Jüri Peets[20], Jörgen Gustafsson[35], Katrine H. Iversen[6,17], Linzi Simpson[36], Lisa Strand[25], Louise Loe[37], Maeve Sikora[38], Marek Florek[39], Maria Vretemark[40], Mark Redknap[41], Monika Bajka[42], Tamara Pushkina[43], Morten Søvsø[44], Natalia Grigoreva[45], Tom Christensen[46], Ole Kastholm[47], Otto Uldum[48], Pasquale Favia[49], Per Holck[50], Sabine Sten[51], Símun V. Arge[52], Sturla Ellingvåg[1], Vayacheslav Moiseyev[53], Wiesław Bogdanowicz[21], Yvonne Magnusson[54], Ludovic Orlando[55], Peter Pentz[46], Mads Dengsø Jessen[46], Anne Pedersen[46], Mark Collard[56], Daniel G. Bradley[8], Marie Louise Jørkov[57], Jette Arneborg[46,58], Niels Lynnerup[57], Neil Price[28], M. Thomas P. Gilbert[3,59], Morten E. Allentoft[1,60], Jan Bill[61], Søren M. Sindbæk[62], Lotte Hedeager[63], Kristian Kristiansen[64], Rasmus Nielsen[1,65,66] ✉, Thomas Werge[1,10,11,67] ✉ & Eske Willerslev[1,68,69,70] ✉

The maritime expansion of Scandinavian populations during the Viking Age (about AD 750–1050) was a far-flung transformation in world history[1,2]. Here we sequenced the genomes of 442 humans from archaeological sites across Europe and Greenland (to a median depth of about 1×) to understand the global influence of this expansion. We find the Viking period involved gene flow into Scandinavia from the south and east. We observe genetic structure within Scandinavia, with diversity hotspots in the south and restricted gene flow within Scandinavia. We find evidence for a major influx of Danish ancestry into England; a Swedish influx into the Baltic; and Norwegian influx into Ireland, Iceland and Greenland. Additionally, we see substantial ancestry from elsewhere in Europe entering Scandinavia during the Viking Age. Our ancient DNA analysis also revealed that a Viking expedition included close family members. By comparing with modern populations, we find that pigmentation-associated loci have undergone strong population differentiation during the past millennium, and trace positively selected loci—including the lactase-persistence allele of *LCT* and alleles of *ANKA* that are associated with the immune response—in detail. We conclude that the Viking diaspora was characterized by substantial transregional engagement: distinct populations influenced the genomic makeup of different regions of Europe, and Scandinavia experienced increased contact with the rest of the continent.

The events of the Viking Age altered the political, cultural and demographic map of Europe in ways that are evident to this day. Scandinavian diasporas established trade and settlements that stretched from the American continent to the Asian steppe[1]. They exported ideas, technologies, language, beliefs and practices to these lands, developed new socio-political structures and assimilated cultural influences[2].

To explore the genomic history of the Viking Age, we shotgun-sequenced DNA extracted from 442 human remains from archaeological sites dating from the Bronze Age (about 2400 BC) to the Early Modern period (about AD 1600) (Fig. 1, Extended Data Fig. 1). The data from these ancient individuals were analysed together with published data from 3,855 present-day individuals across two reference panels (Supplementary Note 6), and data from 1,118 ancient individuals (Supplementary Table 3).

Scandinavian ancestry and Viking Age origins

Although Viking Age Scandinavian populations shared a common cultural background, there was no common word for Scandinavian identity at this time[1]. Rather than there being a single 'Viking world', a series of interlinked Viking worlds emerged from rapidly growing maritime exploration, trade, war and settlement, following the adoption

A list of affiliations appears at the end of the paper.

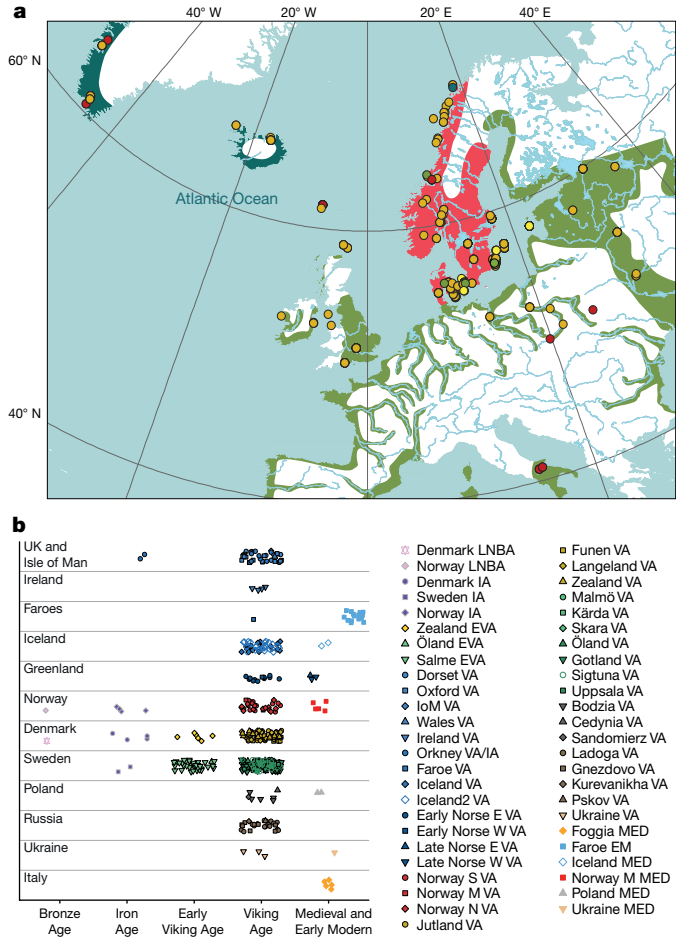


Fig. 1 | Overview of the Viking Age genomic dataset. a, Map of the Viking World from eighth to eleventh centuries AD, showing geographical location and broad age category of sites with ancient samples newly reported in this study. Age categories of sites (circles) are coloured-coded as: dark green, LNBA (2400–500 BC); light green, Iron Age (500 BC–AD 700); yellow, Early Viking Age (AD 700–800); Viking Age (AD 800–1100); Medieval and Early Modern (AD 1100–1600). Red region, area of Viking origins; green region, area of Viking raids, settlement and trade; dark blue region, area of pioneer Viking colonization. **b**, All of the ancient individuals from this study ($n = 442$), and previously published Viking Age samples from Sigtuna[10] and Iceland[18], categorized on the basis of their spatiotemporal origin. The ancient samples are divided into the following five broad categories: Bronze Age (BA), Iron Age (IA), Early Viking Age (EVA), Viking Age (VA), Medieval (MED) and Early Modern (EM). Random jitter has been added along the *x* axis in each category to aid visualization. LNBA, Late Neolithic and Bronze Age; Norse W, Norse western settlement; Norse E, Norse eastern settlement; Norway S, southern Norway; Norway N, northern Norway; Norway M, middle Norway.

of deep-sea navigation among coastal populations of Scandinavia and the area around the Baltic Sea[3,4]. Thus, it is unclear to what extent the Viking phenomenon refers to people with a recently shared genetic background or how far population changes accompanied the transition from the Iron Age (500 BC–AD 700) to the Viking Age in Scandinavia.

The Viking Age Scandinavian individuals of our study fall broadly within the diversity of ancient European individuals from the Bronze Age and later (Fig. 2, Extended Data Figs. 2, 3, Supplementary Note 8), but with subtle differences among the groups that indicate complex fine-scale structure. For example, many Viking Age individuals from the island of Gotland cluster with Bronze Age individuals from the Baltic region, which indicates mobility across the Baltic Sea (Fig. 2, Extended Data Fig. 3). Using f_4-statistics to contrast genetic affinities with steppe pastoralists and Neolithic farmers, we find that Viking Age individuals from Norway are distributed in a manner similar to that of earlier Iron Age individuals, whereas many Viking Age individuals from Sweden and Denmark show a greater affinity to Neolithic farmers from Anatolia (Extended Data Fig. 4a). Using the qpAdm program, we find that the majority of groups can be modelled as three-way mixtures of hunter-gatherer, farmer and steppe-related ancestry. The three-way model was rejected for some groups from Sweden, Norway and the Baltic region, which could be fit using four-way models that additionally included either Caucasus hunter-gatherer or East-Asian-related ancestry (Extended Data Figs. 4b, c)—the latter of which is consistent with previously documented gene flow from Siberia[5–7].

Investigating genetic continuity with Iron Age groups that are temporally more proximate to the Viking Age Scandinavian populations, we find that most Viking Age groups can be fit using a single Iron Age source and broadly fall into two categories: (i) English Iron Age sources (most of the Viking Age individuals from Denmark, as well as populations of the British Isles) and (ii) Scandinavian Iron Age sources (from Norway, Sweden and the Baltic region) (Extended Data Fig. 5a). Notable exceptions are individuals from Kärda in southern Sweden, for whom only early Medieval Longobard individuals from Hungary can be fit as a single source group ($P > 0.01$) (Extended Data Fig. 5a). Groups with poor one-way fits can be modelled by including either additional northeastern ancestry (for example, Viking Age individuals from Ladoga) or additional southeastern ancestry (for example, Viking Age individuals from Jutland) (Extended Data Fig. 5b). Overall, our analyses suggest that the genetic makeup of Viking Age Scandinavian populations largely derives from ancestry of the preceding Iron Age populations—but these analyses also reveal subtle differences in ancestry and gene flow from both the south and east. These observations are largely consistent with archaeological findings[8,9].

Viking Age genetic structure in Scandinavia

To elucidate the fine-scale population structure of Viking Age Scandinavia, we performed genotype imputation on a subset of 298 individuals with sufficient (>0.5×) coverage (289 from this study, along with 9 previously published individuals[10]) and inferred the genomic segments they shared via identity-by-descent with a reference panel of present-day European individuals ($n = 1{,}464$) (Supplementary Notes 6, 10, 11). Genetic clustering using multidimensional scaling and uniform manifold approximation and projection (UMAP) shows that Viking Age Scandinavian individuals cluster into three groups by geographical origin, with close affinities to their respective present-day counterparts (Fig. 3a, Supplementary Fig. 10.1). Some individuals—particularly those from the island of Gotland in eastern Sweden—have strong affinities with Eastern Europeans; this probably reflects individuals with Baltic ancestry, as clustering with Bronze Age individuals from the Baltic region is evident in the identity-by-state UMAP analysis (Fig. 2b) and through f_4-statistics (Supplementary Fig. 9.1).

We used ChromoPainter[11] and a reference panel enriched with Scandinavian individuals ($n = 1{,}464$) (Supplementary Notes 6, 11) to identify long, shared haplotypes and detect subtle population structure (Supplementary Figs. 11.1–11.10). We find ancestry components in Scandinavia with (inexact and indicative) affinities with present-day populations (Supplementary Fig. 11.11), which we refer to as 'Danish-like', 'Swedish-like', 'Norwegian-like' and 'North Atlantic-like' (that is, possible individuals from the British Isles entering Scandinavia). The sampling is heavily structured, so these complex results (Supplementary Fig. 11.12) are visualized over time and space (Fig. 4) using spatial interpolation[12] to account for sampling locations and report significant trends (Supplementary Table 11.2) using linear regression (Supplementary Notes 11, 12).

Norwegian-like and Swedish-like components cluster in Norway and Sweden, respectively, whereas Danish-like and North-Atlantic-like components are widespread (Fig. 4, Supplementary Fig. 11.12,

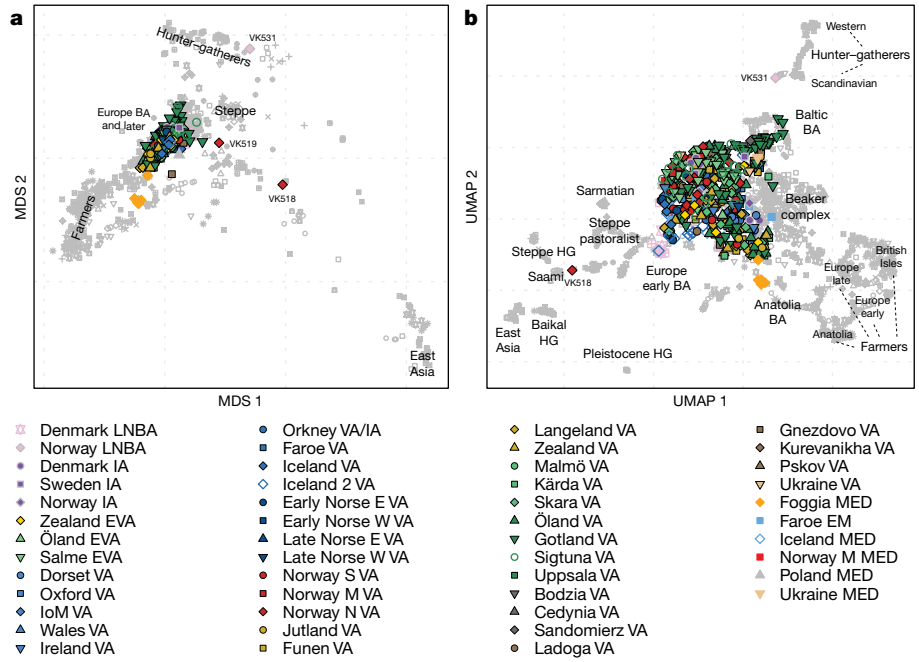


Fig. 2 | Genetic structure of Viking Age samples. a, Multidimensional scaling (MDS) of $n = 1{,}305$ ancient genomes, on the basis of a pairwise identity-by-state sharing matrix of the Viking Age and other ancient samples (Supplementary Table 3). Outlier individuals with hunter-gatherer (VK531) or Saami-related ancestry (VK518 and VK519) are highlighted. **b**, UMAP analysis of the same dataset as in **a**, with fine-scale ancestry groups highlighted. HG, hunter-gatherer.

Supplementary Table 6). Unexpectedly, Viking Age individuals from Jutland (Denmark) lack Swedish-like and Norwegian-like genetic components (Supplementary Fig. 11.12). We also find that gene flow within Scandinavia was broadly from south to north, dominated by movement from Denmark into Norway and Sweden (Supplementary Table 11.2).

We identified two ancient individuals from northern Norway (designated VK518 and VK519) with affinities to present-day Saami populations in Norway and Sweden. The VK519 individual probably also had Norwegian-like ancestors, which indicates genetic contacts between Saami groups and other Scandinavian populations.

The genetic data are structured by topographical boundaries rather than by the borders of present-day countries. Thus, the southwestern part of Sweden in the Viking Age is genetically more similar to Viking Age populations of Denmark than to those of central mainland Sweden, probably owing to geographical barriers that prevented gene flow.

We quantified genetic diversity using two measures: conditional nucleotide diversity (Supplementary Note 9) and variation in inferred ancestry on the basis of ChromoPainter results (Extended Data Fig. 6, Supplementary Note 11, Supplementary Fig. 11.13). We also visualized this diversity as the spread of individuals on a multidimensional scaling plot based on a pairwise identity-by-state sharing matrix (Fig. 3b).

Diversity varies markedly from the more-homogeneous inland and northern parts of Scandinavia to the diverse Kattegat (eastern Denmark and western Sweden) and Baltic Sea regions, which suggests an important role for these maritime regions in interaction and trade during the Viking Age. On Gotland, there are many more Danish-like and North-Atlantic-like genetic components (as well as an additional 'Finnish-like' ancestry component) than Swedish-like components, which indicates extensive maritime contacts for Gotland during the Viking Age.

Our results for Gotland and Öland agree with archaeological indications that these were important maritime communities from the Roman period (AD 1–400) onwards[13,14]. A similar pattern is observed on the central Danish islands (such as Langeland) but at a lower level. The data indicate that genetic diversity on the islands increased from the early (about eighth century AD) to the late Viking Age (about tenth to eleventh centuries AD), which suggests increasing interregional interaction over time. Evidence for genetic structure within Viking Age Scandinavia[2,4,15–17]—with diversity in cosmopolitan centres such as Skara, and trade-oriented islands such as Gotland—highlight the importance of sea routes during this period.

Viking migrations

Our fine-scale ancestry analyses of genomic data are consistent with patterns documented by historians and archaeologists (Figs. 3, 4, Supplementary Fig. 11.12): eastward movements mainly involved Swedish-like ancestry, whereas individuals with Norwegian-like ancestry travelled to Iceland, Greenland, Ireland and the Isle of Man. The first settlement in Iceland and Greenland also included individuals with North-Atlantic-like ancestry[18,19]. A Danish-like ancestry is seen in present-day England, in accordance with historical records[20], place names[21], surnames[22] and modern genetics[23,24], but Viking Age Danish-like ancestry in the British Isles cannot be distinguished from that of the Angles and Saxons, who migrated in the fifth to sixth centuries AD from Jutland and northern Germany.

Viking Age execution sites in Dorset and Oxford (England) contain North-Atlantic-like ancestry, as well as Danish-like and Norwegian-like ancestries. If these sites represent Viking raiding parties that were defeated and captured[25,26], then these raids were composed of individuals of different origins. This pattern is also suggested by isotopic data from a warrior cemetery in Trelleborg (Denmark)[27]. Similarly, the presence of Danish-like ancestry in an ancient sample from Gnezdovo in present-day Russia indicates that eastern migrations were not entirely composed of Viking individuals from Sweden.

Our results show that 'Viking' identity was not limited to individuals of Scandinavian genetic ancestry. Two individuals from Orkney who were buried in Scandinavian fashion are genetically similar to present-day Irish and Scottish populations, and are probably the first Pictish genomes published (see 'Evidence for Pictish genomes' in Supplementary Note 11, Supplementary Figs. 11.3, 11.12, 11.14, Supplementary Table 6). Two other individuals from Orkney had 50% Scandinavian

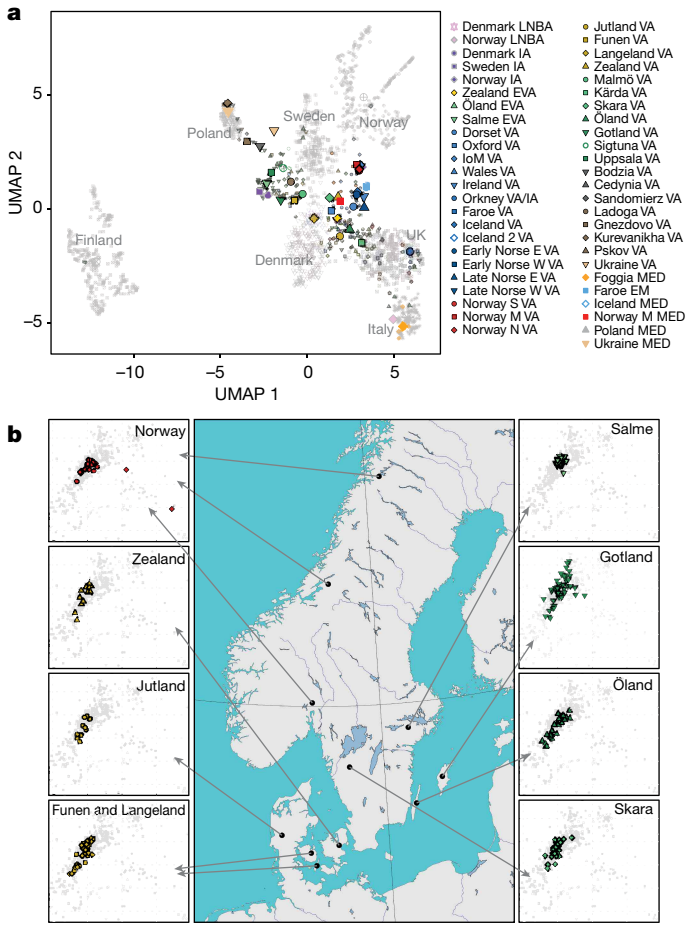


Fig. 3 | Genetic structure and diversity of ancient samples. a, UMAP analysis of $n = 1{,}624$ ancient and modern Scandinavian individuals, on the basis of the first 10 dimensions of MDS using identity-by-descent segments of imputed individuals. Large symbols indicate median coordinates for each group. **b**, Genetic diversity in major populations of the Scandinavian Viking Age. Plots next to the map show MDS analysis on the basis of a pairwise identity-by-state sharing matrix. Norway denotes all the sites from Norway. The scale is identical for all the plots.

ancestry, and five such individuals were found in Scandinavia. This suggests that Pictish populations may have been integrated into Scandinavian culture by the Viking Age.

Viking Age gene flow into Scandinavia

Non-Scandinavian ancestry in samples from Denmark, Norway and Sweden agrees with known trading routes (Supplementary Notes 11, 12): for example, Finnish and Baltic ancestry reached modern Sweden (including Gotland), but is absent in most individuals from Denmark and Norway. By contrast, western regions of Scandinavia received ancestry from the British Isles (Supplementary Notes 11, 12). The first evidence of South European ancestry (>50%) in Scandinavia is during the Viking Age in Denmark (for example, individuals VK365 and VK286 from Bogøvej) and southern Sweden (for example, VK442 and VK350 from Öland, and VK265 from Kärda) (Fig. 4, Supplementary Table 6).

Disappearance from Greenland

From around AD 980 to 1440, southwest Greenland was settled by people of Scandinavian ancestry (probably from Iceland)[28,29]. The fate of these populations in Greenland remains debated, but probable causes of their disappearance are social or economic processes in Europe (for example, political relations within Scandinavia and changed trading systems) and natural processes, including climatic change[29–31].

According to our data, the Greenland Norse populations were an admixture between Scandinavians (mostly from Norway) and individuals from the British Isles, similar to the first settlers of Iceland[18]. We see no evidence of long-term inbreeding in the genomes of Greenlandic Norse individuals, although we have only one high-coverage genome from the later period of occupation of the island (Supplementary Note 10, Supplementary Figs. 10.2, 10.3). This result could favour a relatively brief depopulation scenario, consistent with previous demographic models[32] and archaeological findings. We also find no evidence of ancestry from other populations (Palaeo-Eskimo, Inuit or Native American) in the Greenlandic Norse genomes (Supplementary Fig. 9.4), which accords with the skeletal remains[32]. This suggests that sexual interaction between the Greenland Norse populations and these other groups was absent, or occurred only on a very small scale.

Genetic composition of earliest Viking voyage

Although maritime raiding has been a constant of seafaring cultures for millennia, the Viking Age is partly defined by this activity[33]. However, the exact nature and composition of Viking war parties is unknown[5]. One raiding or diplomatic expedition has left direct archaeological traces: at Salme in Estonia, 41 men from Sweden who died violently were buried in two boats, accompanied by high-status weaponry[34,35]. Importantly, the Salme boat burial predates the first textually documented raid (on Lindisfarne (England) in 793) by nearly half a century.

Kinship analysis of the genomes of 34 individuals from the Salme burial reveals 4 brothers buried side by side and a third-degree relative of 1 of the 4 brothers (Supplementary Note 4). The ancestry profiles of the Salme individuals were similar to one another when compared to the profiles of other burials of the Viking Age (Supplementary Notes 10, 11), which suggests a relatively genetically homogeneous group of people of high status (including close kin).

The five Salme relatives are not the only kin in our dataset; we also identified two pairs of kin in which the related individuals were excavated hundreds of kilometres apart from each other, which markedly illustrates the mobility of individuals during the Viking Age.

Positive selection in northern Europe

We looked for single-nucleotide polymorphisms (SNPs) with allele frequencies that have changed significantly in the last 10,000 years[36,37], beyond what can be explained by temporal changes in ancestry alone (Supplementary Note 14). Extended Data Figure 8a shows the likelihood ratio scores in favour of selection in the entire 10,000-year period (the general scan), the period up to 4,000 years before present (the ancient scan) and the period from 4,000 years before present up to today (the recent scan).

As expected[38,39], the strongest candidates for selection are SNPs near *LCT*, the frequency of which increased after the Bronze Age[40,41]. Our dataset traces the frequency of the lactase-persistence allele (rs4988235) and its evolution since the Bronze Age. Extended Data Figure 8b shows that Viking Age groups had very similar allele frequencies at the *LCT* lactase-persistence SNP to those of present-day northern European populations. Conversely, Bronze Age Scandinavian individuals, as well as individuals from central Europe associated with Corded Ware and Bell Beaker assemblages, have a low frequency of this SNP despite evidence for milk consumption. Our Iron Age samples have intermediate frequencies, which suggests a rise in lactase persistence during this period. The frequency is higher in the Bronze Age of the Baltic Sea region than in Bronze Age Scandinavia, which is consistent with gene flow between the two regions explaining the increasing frequency of lactase persistence in Scandinavia.

Other candidates for selection include previously identified regions, including the one containing the *TLR1*, *TLR6* and *TLR10* genes, the HLA

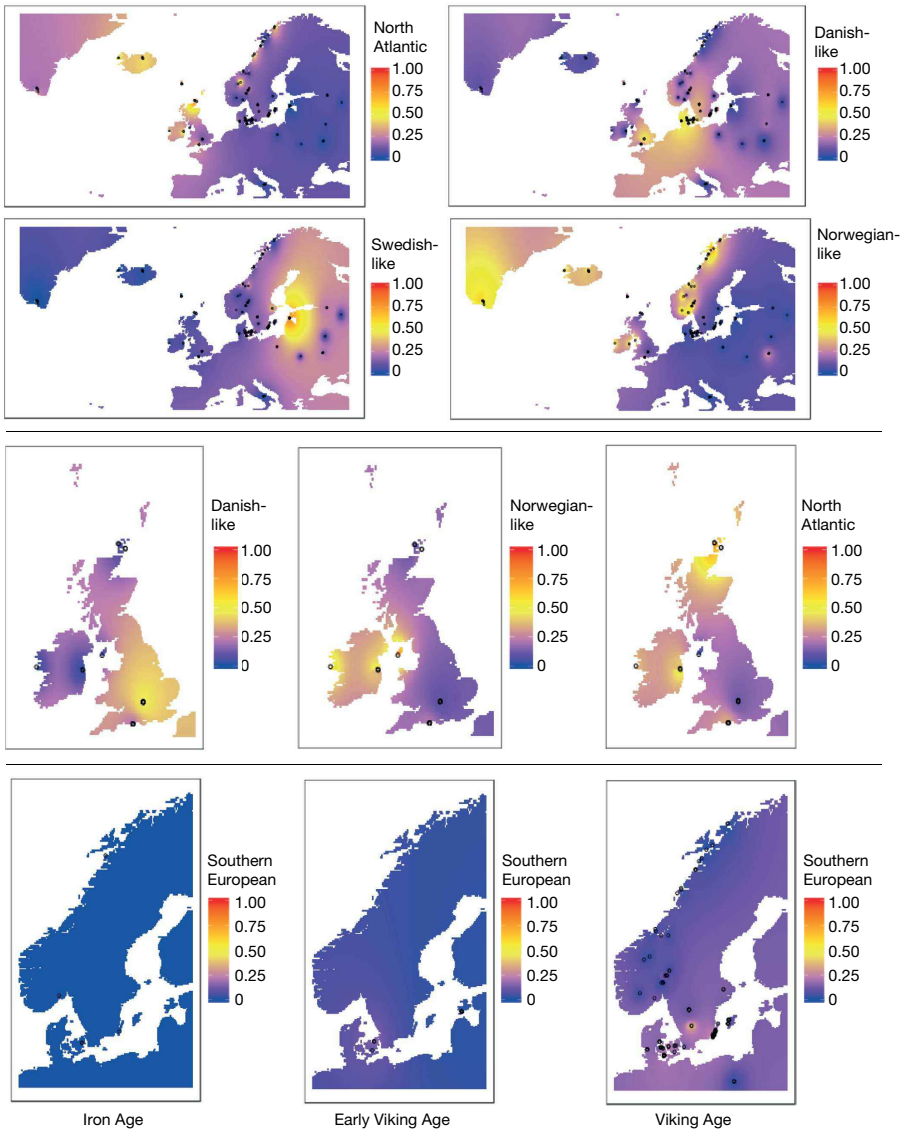


Fig. 4 | Spatiotemporal patterns of Viking and non-Viking ancestry in Europe during the Iron Age, Early Viking Age and Viking Age. We performed inverse distance-weighting interpolation of the ancestry painting proportions of each individual genome on a dense grid of points covering the European continent, to better visualize the distribution of ancestry paintings at different periods (Supplementary Note 12). Top, distinct spheres of influence in the Viking world. Middle, Danish Viking ancestry in southern Britain, Norwegian Viking ancestry in Ireland and Isle of Man and non-Scandinavian ('North Atlantic') ancestry in Orkney, Ireland and southern Britain. Bottom, Late southern European ancestry in southern Scandinavia. The Swedish-like ancestry is the highest in present-day Estonia owing to the ancient samples from the Salme ship burial, which originated from the Mälaren valley of Sweden (according to archaeological sources). $n = 289$ genomes used for interpolation.

region, and the genes *SLC45A2* and *SLC22A4*[41]. We also find additional candidate regions for selection that have associated trajectories that start before the Viking Age, which suggests shared phenotypes between ancient Viking and present-day Scandinavian populations (Supplementary Note 14). These regions include one that overlaps *DCC* and that is implicated in colorectal cancer[42], as well as one that overlaps *AKNA* and is involved in the secondary immune response[43].

Evolution of complex traits in Scandinavia

To search for signals of recent population differentiation at SNP markers associated with complex traits, we compared genotypes of Viking Age individuals with those of a panel of present-day Danish individuals[44]. We obtained summary statistics from 16 well-powered genome-wide association studies through the GWAS ATLAS[45] and tested for a difference in the distribution of polygenic scores between the two groups (Supplementary Note 15). The polygenic scores of Viking Age individuals and present-day Danish individuals differed for three traits: black hair colour ($P = 0.00089$), standing height ($P = 0.019$) and schizophrenia ($P = 0.0096$), although the latter two were not significant after accounting for the number of tests (Extended Data Fig. 7). Currently, we cannot conclude whether the observed differences in allele frequencies are due to selection acting on these alleles between the Viking Age and the present time or to some other factors (such as more ethnic diversity in the Viking Age sample). A binomial test of the number of black hair colour risk alleles at higher frequency in the Viking Age sample and the present-day sample was also significant (65/41; $P = 0.025$), which suggests that the signal is not entirely driven by a few large-effect loci.

Viking genetic legacy in populations today

To test whether present-day Scandinavian populations share increased ancestry with their respective counterparts in the Viking Age, we first

computed *D*-statistics of the form *D*(Yoruba (YRI), ancient; present-day population 1, present-day population 2), which measure whether an ancient test individual shares more alleles with present-day population 1 or with present-day population 2. Viking Age individuals shift subtly from Scandinavia towards their present-day counterparts in the distributions of these statistics (Extended Data Fig. 5c, Supplementary Figs. 9.2, 9.3).

We further examined ancient ancestry in present-day populations using fineSTRUCTURE (Supplementary Note 11, Supplementary Fig. 11.14). Within Scandinavia, most present-day populations resemble their Viking Age counterparts. The exception is Swedish-like ancestry, which is present at only 15–30% within Sweden today: one cluster from Sweden is closer to ancient Finnish populations, and a second is more closely related to Danish and Norwegian populations. Danish-like ancestry is now high across the whole region.

Outside of Scandinavia, the genetic legacy of Viking Age populations is consistent—although limited. A small Scandinavian ancestry component is present in Poland (up to 5%). Within the British Isles, it is difficult to assess how much of the Danish-like ancestry is due to pre-existing Anglo-Saxon ancestry, but the Viking Age contribution does not exceed 6% in England (Supplementary Note 11). The genetic effects are stronger in the other direction. Although some North-Atlantic-like individuals in Orkney became culturally Scandinavian, others found themselves in Iceland, Norway and beyond, leaving a genetic legacy that persists today. Present-day Norwegian individuals vary between 12 and 25% in North-Atlantic-like ancestry; this ancestry is more uniformly 10% in Sweden.

Discussion

Our genomic analyses shed light on long-standing questions raised by historical sources and archaeological evidence from the Viking Age. We largely confirm the long-argued movements of Vikings outside Scandinavia: Vikings from present-day Denmark, Norway, and Sweden going to Britain, the islands of the North Atlantic, and sailing east towards the Baltic region and beyond, respectively. However, we also see ancient Swedish-like and Finnish-like ancestry in the westernmost fringes of Europe, and Danish-like ancestry in the east, defying modern historical groupings. It is likely that many such individuals were from communities with mixed ancestries, thrown together by complex trading, raiding and settling networks that crossed cultures and the continent.

During the Viking Age, different parts of Scandinavia were not evenly connected, leading to clear genetic structure in the region. Scandinavia probably comprised a limited number of transport zones and maritime enclaves[46] with active external contacts, and limited external gene flow into the rest of the Scandinavian landmass. Some Viking Age Scandinavian locations are relatively homogeneous—particularly mid-Norway, Jutland and the Atlantic settlements. This contrasts with the strong genetic variation of populous coastal and southern trading communities such as in the islands of Gotland and Öland[47–49]. The high genetic heterogeneity in coastal communities implies increased population size, extending a previously proposed[10] urbanization model for the Late Viking Age city of Sigtuna (which suggested that more-cosmopolitan trading centres were already present at the end of the Viking Age in Northern Europe) both spatially and further back in time. The formation of large-scale trading and cultural networks that spread people, goods and warfare took time to affect the heartlands of Scandinavia, which retained pre-existing genetic differences into the Medieval period.

Finally, our findings show that Vikings were not simply a direct continuation of Scandinavian Iron Age groups. Instead, we observe gene flow from the south and east into Scandinavia, starting in the Iron Age and continuing throughout the duration of the Viking Age, from an increasing number of sources. Many Viking Age individuals—both within and outside Scandinavia—have high levels of non-Scandinavian ancestry, which suggests ongoing gene flow across Europe.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2688-8.

1. Brink, S. & Price, N. (eds) *The Viking World* (Routledge, 2008).
2. Jesch, J. *The Viking Diaspora* (Routledge, 2015).
3. Eriksen, M. H., Pedersen, U., Rundberget, B. & Axelsen, I. *Viking Worlds: Things, Spaces and Movement* (Oxbow Books, 2014).
4. Sindbæk, S. M. & Trakadas, A. *The World in the Viking Age* (Viking Ship Museum in Roskilde, 2014).
5. Sikora, M. et al. The population history of northeastern Siberia since the Pleistocene. *Nature* **570**, 182–188 (2019).
6. Lamnidis, T. C. et al. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nat. Commun.* **9**, 5018 (2018).
7. Saag, L. et al. The arrival of Siberian ancestry connecting the eastern Baltic to Uralic speakers further east. *Curr. Biol.* **29**, 1701–1711.e16 (2019).
8. Hedeager, L. in *The Viking World* (eds Brink, S. & Price, N.) 35–46 (Routledge, 2008).
9. Hedeager, L. *Iron Age Myth and Materiality: An Archaeology of Scandinavia AD 400–1000* (Routledge, 2011).
10. Krzewińska, M. et al. Genomic and strontium isotope variation reveal immigration patterns in a Viking Age town. *Curr. Biol.* **28**, 2730–2738.e10 (2018).
11. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).
12. Shepard, D. A Two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM National Conference* (eds Blue, R. B. & Rosenberg, A. M.) 517–524 (ACM, 1968).
13. Hansen, U. L. *Römischer Import im Norden: Warenaustausch zwischen dem Römischen Reich und dem freien Germanien während der Kaiserzeit unter besonderer Berücksichtigung Nordeuropas* (Det Kongelige Nordiske Oldskriftselskab, 1987).
14. Andersson, K. *I Skuggan av Rom: Romersk Kulturpåverkan i Norden* (Atlantis, (2013).
15. Bill, J. in *The Viking World* (eds Brink, S. & Price, N.) 170–180 (Routledge, 2008).
16. Sindbæk, S. M. The small world of the Vikings: networks in early Medieval communication and exchange. *Norw. Archaeol. Rev.* **40**, 59–74 (2007).
17. Hilberg, V. & Kalmring, S. in *Viking Archaeology in Iceland: Mosfell Archaeological Project* (eds Zori, D. & Byock, J.) 221–245 (Brepols, 2014).
18. Ebenesersdóttir, S. S. et al. Ancient genomes from Iceland reveal the making of a human population. *Science* **360**, 1028–1032 (2018).
19. Helgason, A. et al. mtDna and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am. J. Hum. Genet.* **68**, 723–737 (2001).
20. Downham, C. Viking ethnicities: a historiographic overview. *Hist. Compass* **10**, 1–12 (2012).
21. Fellows-Jensen, G. in *The Viking World* (eds Brink, S. & Price, N.) 391–400 (Routledge, 2008).
22. Bowden, G. R. et al. Excavating past population structures by surname-based sampling: the genetic legacy of the Vikings in northwest England. *Mol. Biol. Evol.* **25**, 301–309 (2008).
23. Leslie, S. et al. The fine-scale genetic structure of the British population. *Nature* **519**, 309–314 (2015).
24. Athanasiadis, G. et al. Nationwide genomic study in Denmark reveals remarkable population homogeneity. *Genetics* **204**, 711–722 (2016).
25. Loe, L., Boyle, A., Webb, H. & Score, D. *'Given to the Ground': A Viking Age Mass Grave on Ridgeway Hill, Weymouth* (Dorset Natural History and Archaeological Society, 2014).
26. Wallis, S. *The Oxford Henge and Late Saxon Massacre: With Medieval and Later Occupation at St John's College, Oxford* (Thames Valley Archaeological Services Limited, 2014).
27. Douglas Price, T., Frei, K. M., Dobat, A. S., Lynnerup, N. & Bennike, P. Who was in Harold Bluetooth's army? Strontium isotope investigation of the cemetery at the Viking Age fortress at Trelleborg, Denmark. *Antiquity* **85**, 476–489 (2011).
28. Price, T. D. & Arneborg, J. The peopling of the North Atlantic: isotopic results from Greenland. *Journal of the North Atlantic* **7**, 164–185 (2014).
29. Arneborg, J. in *The Viking World* (eds Brink, S. & Price, N.) 588–603 (Routledge, 2008).
30. Dugmore, A. J. et al. Cultural adaptation, compounding vulnerabilities and conjunctures in Norse Greenland. *Proc. Natl Acad. Sci. USA* **109**, 3658–3663 (2012).
31. Arneborg, J. in *Medieval Archaeology in Scandinavia and Beyond: History, Trends and Tomorrow* (eds Kristiansen, M. S. et al.) 247–271 (Aarhus Universitetsforlag, 2015).
32. Lynnerup, N. *The Greenland Norse: A Biological–Anthropological Study* (Museum Tusculanum, 1998).
33. Sindbæk, S. M. in *Routledge Handbook of Archaeology and Globalization* (ed. Hodos, T.) 553–565 (Routledge, 2016).
34. Peets, J. et al. Research results of the Salme ship burials in 2011–2012. *Archaeological Fieldwork in Estonia* **2012**, 43–60 (2012).
35. Douglas Price, T., Peets, J., Allmäe, R., Maldre, L. & Oras, E. Isotopic provenancing of the Salme ship burials in pre-Viking Age Estonia. *Antiquity* **90**, 1022–1037 (2016).
36. Cheng, J. Y., Racimo, F. & Nielsen, R. Ohana: detecting selection in multiple populations by modelling ancestral admixture components. Preprint at https://doi.org/10.1101/546408 (2019).
37. Alves, J. M. et al. Parallel adaptation of rabbit populations to myxoma virus. *Science* **363**, 1319–1326 (2019).
38. Enattah, N. S. et al. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30**, 233–237 (2002).

39. Bersaglieri, T. et al. Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74**, 1111–1120 (2004).
40. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
41. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
42. Fearon, E. R. et al. Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science* **247**, 49–56 (1990).
43. Siddiqa, A. et al. Regulation of CD40 and CD40 ligand by the AT-hook transcription factor AKNA. *Nature* **410**, 383–387 (2001).
44. Pedersen, C. B. et al. The iPSYCH2012 case–cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6–14 (2018).
45. Watanabe, K. et al. A global view of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
46. Westerdahl, C. The maritime cultural landscape. *Int. J. Naut. Archaeol.* **21**, 5–14 (1992).
47. Hyenstrand, Å. *Ancient Monuments and Prehistoric Society* (Central Board of National Antiquities, 1979).
48. Callmer, J. in *Regions and Reflections: In honour of Märta Strömberg* (eds Jennbert, K. et al.) 257–273 (1991).
49. Jakobsen, J. G. G. & Dam, P. *Atlas Over Danmark: Historisk-Geografisk Atlas* (Det Kongelige Danske Geografiske Selskab, 2008).

[1]Lundbeck Foundation GeoGenetics Centre, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark. [2]Institute of Molecular Biology, National Academy of Sciences, Yerevan, Armenia. [3]Section for Evolutionary Genomics, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark. [4]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [5]School of Statistical Sciences, University of Bristol, Bristol, UK. [6]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [7]The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [8]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland. [9]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [10]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [11]Institute of Biological Psychiatry, Mental Health Services Copenhagen, Copenhagen, Denmark. [12]HSE University, Russian Federation National Research University Higher School of Economics, Moscow, Russia. [13]Department of Archaeology and Ancient History, Lund University, Lund, Sweden. [14]Sydsvensk Arkeologi AB, Kristianstad, Sweden. [15]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden. [16]Department of Genetics, University of Cambridge, Cambridge, UK. [17]Department of Health Technology, Section for Bioinformatics, Technical University of Denmark (DTU), Copenhagen, Denmark. [18]Department of Human Genetics, McGill University, Montréal, Quebec, Canada. [19]National Institute of Genomic Medicine (INMEGEN), Mexico City, Mexico. [20]Archaeological Research Collection, Tallinn University, Tallinn, Estonia. [21]Museum and Institute of Zoology, Polish Academy of Sciences, Warsaw, Poland. [22]Anuchin Research Institute and Museum of Anthropology, Moscow State University, Moscow, Russia. [23]Manx National Heritage, Douglas, Isle of Man. [24]Upplandsmuseet, Uppsala, Sweden. [25]NTNU University Museum, Department of Archaeology and Cultural History, Trondheim, Norway. [26]The Archaeologists, National Historical Museums, Stockholm, Sweden. [27]Thames Valley Archaeological Services (TVAS), Reading, UK. [28]Department of Archaeology and Ancient History, Uppsala University, Uppsala, Sweden. [29]Institute of Archaeology, University of Warsaw, Warsaw, Poland. [30]Department of Cultural Heritage Conservation, National Yunlin University of Science and Technology, Douliou, Taiwan. [31]Institute of Archaeology, Reykjavík, Iceland. [32]UHI Archaeology Institute, University of the Highlands and Islands, Kirkwall, UK. [33]Department of Bioarchaeology, Institute of Archaeology of National Academy of Sciences of Ukraine, Kiev, Ukraine. [34]Soprintendenza Archeologia, Belle Arti e Paesaggio per le Province di Barletta, Andria, Trani e Foggia, Foggia, Italy. [35]Jönköping County Museum, Jönköping, Sweden. [36]Trinity College Dublin, Dublin, Ireland. [37]Heritage Burial Services, Oxford Archaeology, Oxford, UK. [38]National Museum of Ireland, Dublin, Ireland. [39]Institute of Archaeology, Maria Curie-Sklodowska University in Lublin, Lublin, Poland. [40]Västergötlands Museum, Skara, Sweden. [41]Department of History and Archaeology, Amgueddfa Cymru–National Museum Wales, Cardiff, UK. [42]Trzy Epoki Archaeological Service, Klimontów, Poland. [43]Historical Faculty, Moscow State University, Moscow, Russia. [44]Museum of Southwest Jutland, Ribe, Denmark. [45]Department of Slavic–Finnish Archaeology, Institute for the History of Material Culture, Russian Academy of Sciences, Saint Petersburg, Russia. [46]National Museum of Denmark, Copenhagen, Denmark. [47]Department of Research and Heritage, Roskilde Museum, Roskilde, Denmark. [48]Langelands Museum, Langeland, Denmark. [49]Department of Humanities, University of Foggia, Foggia, Italy. [50]Department of Molecular Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway. [51]Department of Archaeology and Ancient History, Uppsala University Campus Gotland, Visby, Sweden. [52]Tjóðsavnið – Faroe Islands National Museum, Tórshavn, Faroe Islands. [53]Peter the Great Museum of Anthropology and Ethnography (Kunstkamera), Russian Academy of Science, St Petersburg, Russia. [54]Malmö Museum, Malmö, Sweden. [55]Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse, CNRS UMR 5288, Université de Toulouse, Université Paul Sabatier, Toulouse, France. [56]Department of Archaeology, Simon Fraser University, Burnaby, British Colombia, Canada. [57]Department of Forensic Medicine, University of Copenhagen, Copenhagen, Denmark. [58]School of GeoSciences, University of Edinburgh, Edinburgh, UK. [59]Department of Natural History, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. [60]Trace and Environmental DNA (TrEnD) Laboratory, School of Molecular and Life Sciences, Curtin University, Perth, Western Australia, Australia. [61]Museum of Cultural History, University of Oslo, Oslo, Norway. [62]Centre for Urban Network Evolutions (UrbNet), School of Culture and Society, Aarhus University, Højbjerg, Denmark. [63]Institute of Archaeology, Conservation and History, Oslo, Norway. [64]Department of Historical Studies, University of Gothenburg, Gothenburg, Sweden. [65]Department of Integrative Biology, UC Berkeley, Berkeley, CA, USA. [66]Department of Statistics, UC Berkeley, Berkeley, CA, USA. [67]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Copenhagen, Denmark. [68]Department of Zoology, University of Cambridge, Cambridge, UK. [69]The Danish Institute for Advanced Study, University of Southern Denmark, Odense, Denmark. [70]The Wellcome Trust Sanger Institute, Cambridge, UK. [71]These authors contributed equally: Ashot Margaryan, Daniel J. Lawson, Martin Sikora, Fernando Racimo. ✉e-mail: rasmus_nielsen@berkeley.edu; Thomas.Werge@regionh.dk; ew482@cam.ac.uk

Methods

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

Laboratory work

Laboratory work was conducted in the dedicated ancient DNA clean-room facilities at the Globe Institute (University of Copenhagen), according to strict ancient DNA standards[50,51]. The overwhelming majority of ancient samples were petrous bones and teeth (Supplementary Table 1). The details of DNA extraction can be found in Supplementary Note 2. Double-stranded blunt-end DNA libraries were prepared using Illumina-specific adapters and NEBNext DNA Sample Pre Master Mix Set 2 (E6070) kit. We used an Agilent Bioanalyzer 2100 to quantify the amount of the purified DNA libraries. The libraries were sequenced 80-bp single-read chemistry on Illumina HiSeq 2500 machines at the Danish National High-throughput DNA Sequencing Centre.

Bioinformatics analysis and quality assessment

We used AdapterRemoval v.2.1.3[52] for removing Illumina adaptor sequences, keeping only sequences with a minimum length of 30 bp. Adaptor-free sequences were mapped against the human reference genome build 37 using BWA v.0.7.10 aligner[53] with the seed (-l parameter) disabled for higher sensitivity of ancient DNA reads[54]. DNA sequences were processed with samtools v.1.3.1[53], and only sequences with mapping quality ≥ 30 were kept. Picard v.1.127 (http://broadinstitute.github.io/picard) was used to sort the reads and remove duplicates. DNA libraries were combined at the sample level and realigned using GATK v.3.3.0[55] with Mills and 1000G gold-standard insertions and deletions (indels). At the end, realigned .bam files had the md-tag updated and extended base alignment qualities calculated using samtools calmd. Read depth and coverage were determined using pysam (http://code.google.com/p/pysam/) and BEDtools[56]. The mapping statistics for the ancient samples are summarized in Supplementary Table 2.

We used mapDamage v.2.0 to obtain approximate Bayesian estimates of damage parameters[57]. Data authenticity was assessed by estimating the rate of mismatches to the consensus mitochondrial sequence using contamMix[58] and Schmutzi[59], as well as the excess of heterozygous positions in male haploid X chromosomes using ANGSD[60]. The sex of ancient individuals was determined by calculating the Rγ parameter[61].

Uniparental haplogroup determination and kinship analysis

The mitochondrial haplogroups of the ancient individuals were assigned using haplogrep[62]. To get the mtDNA consensus sequences, we aligned the trimmed reads of ancient samples to the human mitochondrial reference genome: revised Cambridge Reference Genome (rCRS). Base quality ≥ 20 and mapping quality ≥ 30 filtering options were applied. Only SNPs at sites ≥3× coverage were considered for consensus calling using samtools mpileup/bcftools v.1.3.1[53].

We identified male Y chromosome lineages using the pathPhynder workflow (https://github.com/ruidlpm/pathPhynder) and Yleaf v.2[63]. For the latter, the analysis was restricted to 26,083 biallelic SNPs from the International Society of Genetic Genealogy (ISOGG) 2019 database (https://isogg.org/tree/ISOGG_YDNA_SNP_Index.html).

We used NgsRelate[64] to detect family relationships between all pairs of individuals. NgsRelate is a maximum-likelihood based program that—for a pair of individuals based on genotype likelihoods—estimates the three coefficients, $k0$, $k1$ and $k2$, which denote the proportions of the genome in which the pair of analysed individuals share 0, 1 and 2 alleles identical-by-descent, respectively. We only included the 376 samples with sequencing depth above 0.1× for the analysis. From these, we estimated genotype likelihoods and allele frequencies with ANGSD[60] using the SAMtools genotype likelihood model (-gl 1) including reads with mapping quality ≥ 30 and bases with base quality ≥ 20. We estimated genotype likelihoods and allele frequencies only for the autosomal transversion sites for which the 1000 Genomes CEU population (Utah residents with northern and western European ancestry) has a minor allele frequency of 0.05, resulting in 1,752,719 sites. READ[65] was used to confirm the degree of relatedness between pairs of individuals. The pedigree reconstructions on the basis of the kinship coefficients were conducted using Pedigree Reconstruction and Identification of a Maximum Unrelated Set (PRIMUS)[66].

Imputation

We imputed the genotypes of 298 ancient samples (289 from this study, and 9 from a previous study[10]) that had a sequencing depth greater than 0.5×. We used Beagle v.4.1[67] for imputations based on the genotype likelihood data, which was first estimated by GATK v.3.7.0 UnifiedGenotyper. To generate the genotype data, we called only biallelic sites present in the 1000 Genomes dataset, and only the observed alleles (--genotyping_mode GENOTYPE_GIVEN_ALLELES). The resulting .vcf files were filtered by setting genotype likelihoods to 0 for all three genotypes (for example, hom ref, het and hom alt) for sites with potential deamination (C>T and G>A), as described in a previous study[68]. Following this, the per-individual .vcf files were merged using bcftools v.1.3.1. The combined .vcf files were then split into 15,000 markers each and imputed separately using Beagle 4.0 using the 1000 Genomes phase-3 map included with Beagle (*.phase3.v5a.snps.vcf.gz and plink.chr*.GRCh37.map) with input through the genotype likelihood option. Run time for imputing using Beagle was approximately 280,000 core hours.

Merge with existing panels

Scandinavian panel. To assess the genetic relationships of various Viking Age groups with their present-day counterparts, we constructed a reference panel enriched with Scandinavian populations on the basis of published datasets: the EGAD00010000632 data set from a previous publicaton[23] (UK dataset) and the EGAD00000000120 dataset from The International Multiple Sclerosis Genetics Consortium and The Wellcome Trust Case Control Consortium 2 (ref. [69]) (EU dataset) (see Supplementary Note 6 for details). The seven most relevant populations from Denmark, Sweden, Norway, Finland, Poland, UK and Italy were considered (n = 1,464) with a total number of 414,264 SNPs. The Han Chinese (CHB) and Yoruba (YRI) populations from the 1000 Genomes project phase-3 database were merged to this panel as outgroups.

The 1000 Genomes panel. We used a set of 1,995 individuals from 20 populations (excluding individuals from the AMR super-population, as well as admixed ASW and ACB populations) of the 1000 Genomes project phase-3 release 5 (ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/). We restricted the dataset to a set of 12,731,663 biallelic transversion SNPs located within the strict mappability mask regions (ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/accessible_genome_masks/).

Analyses of phenotype associated SNPs were carried out using five European-ancestry populations: Spanish (IBS), Tuscan (TSI), CEU, British (GBR) and Finnish (FIN), along with CHB and YRI as outliers. These were used to assess genome-wide allele frequencies for various SNPs associated with pigmentation phenotypes and lactose intolerance.

Ancient panels. We constructed datasets for population genetic analyses by merging the newly sequenced Viking Age individuals as well as other previously published ancient individuals[40,41,68,70–96] with the two modern reference panels. Ancient individuals were represented with pseudohaploid genotypes, by using mpileup command of samtools and randomly sampling an allele passing filters (mapping quality ≥ 30 and base quality ≥ 30), further requiring that it matched one of the two alleles observed in the reference panel (Supplementary Table 3).

Clustering analyses

On the basis of the pseudohaploid individuals from the ancient panels, we ran ADMIXTURE[97] by thinning the dataset for linkage disequilibrium using plink with recommended settings (--indep-pairwise 50 10 0.1). This dataset contained 1,324 individuals for 151,235 markers for the autosomal chromosomes. Only samples with >20,000 SNPs overlapping with the Human Origins panel were kept in the analysis, resulting in 378 samples from this study. We did 50 replicates with different seeds for $k = 2$ to $k = 10$. We used pong[98] to identify the best run for each k and similar components between different k values.

The large number of ancient individuals included in the analysis panels facilitates genetic clustering using the ancient individuals themselves, rather than projecting them on axes of variation inferred from modern populations. We carried this out using MDS on a distance matrix obtained from pairwise identity-by-state sharing between individuals, using the cmdscale function in R. We performed the main genetic clustering on a set of 1,306 ancient Eurasian individuals with >50,000 SNPs with genotype data, restricting to the batch-corrected SNP set described in Supplementary Note 8. Results from the batch-corrected MDS were combined with further dimensionality reduction using UMAP, implemented in the uwot package in R.

Population genetics

We used f_4 statistics to investigate allele-sharing between sets of test individuals and different modern and ancient groups (Supplementary Note 9). To characterize the deep ancestry relationship of the study individuals we calculated f_4(YRI, test individual; Barcin_EN.SG, Yamnaya_EBA.SG) for all ancient Europeans from the Bronze Age onwards (1000 Genomes panel merge). This statistic contrasts genetic affinities of the test individuals with two major ancestry groups that contributed to the gene pool of ancient Europeans from the Bronze Age onwards: Anatolian farmers and Steppe pastoralists. Genetic continuity with Scandinavian Iron Age groups was investigated using f_4(YRI, test group; test individual, Scandinavia Iron Age group) (1000 Genomes panel merge). This statistic measures whether a test individual is consistent with forming a clade with Scandinavian Iron Age groups to the exclusion of a test group from outside of Scandinavia. Genetic affinities between ancient groups and present-day populations were investigated using f_4(YRI, test individual; present-day test population, present-day reference population) (Scandinavian panel).

Ancestry modelling using qpAdm

We estimated ancestry proportions of Viking Age groups using qpAdm[70], which is based on f_4-statistics of the from $f_4(X,O1;O2,O3)$, in which X is either the source or target population, and $O1$, $O2$ and $O3$ are triplets of outgroups to the source and target groups. To minimize batch effects and/or biases due to ancient DNA damage or SNP ascertainment, we used a set of 1,800,038 transversion-only sites that were found polymorphic with minor allele frequency ≥ 0.5% and missing genotype rate of ≤ 15% in the 1000 Genomes panel merge.

Genetic diversity

The genetic diversity of ancient groups was assessed using conditional nucleotide diversity, as previously described[73]. For this analysis, pairwise differences between individuals were calculated using SNPs polymorphic in an outgroup population (YRI) and with a minor allele count ≥ 5 in the 1000 Genomes merge.

Identity-by-descent analysis

The imputed genotypes of 298 individuals were used to infer genomic segments shared via identity-by-descent within the context of a reference panel of 1,464 present-day Europeans, using IBDseq[99] (version r1206) with default parameters. We conducted genetic clustering by MDS on a distance matrix obtained from pairwise identity-by-descent sharing and UMAP to reveal fine-scale population structure among Viking Age individuals.

Painting

To assess the fine-scale variation in genetic ancestry proportions of Viking Age individuals we used Chromosome Painting[11]. The following describes the general workflow of the Chromosome Painting analysis (see Supplementary Note 11 for details).

First, we created a modern reference panel using 1,675 modern individuals sampled from northern Europe, using the standard FineSTRUCTURE pipeline. We applied ChromoPainter to paint all modern individuals using the remaining individuals as donors using fs2.0.8. Related individuals were identified through increased haplotype similarity, and admixed individuals were identified by their FineSTRUCTURE clustering. These were removed, leading to 1,554 unrelated individuals who were re-painted. These individuals were then clustered using FineSTRUCTURE, resulting in 40 populations. After removal of small populations and merging of the CHB and YRI subpopulations, this resulted in 23 modern populations with geographical meaning. We named the resulting clustering the modern reference panel, which consists of 23 modern surrogate populations and 23 modern donor populations (Supplementary Fig. 11.2).

Second, we created an ancient reference panel using the modern reference panel, by applying ChromoPainter to paint all ancient individuals using the modern population palette (Supplementary Fig. 11.3). We then created a supervised ancient population palette consisting of 14 populations which either (a) represent a modern ancestry direction or (b) are best associated with a modern ancestry direction. The paintings consider the average per-individual donor rate to each of the seven modern populations, normalizing each donor label to have a mean of 1 (Supplementary Fig. 11.4). The individuals that contribute most to a population represent it (above a threshold amount chosen by identifying a change point). The remaining individuals are assigned to the population that they are best associated with. We create an ancient population surrogate for each modern population, consisting of the individuals that represent each modern population. For $k = 7$ modern populations, this results in a matrix of $k = 7$ rows (surrogate populations) and $2K = 14$ columns (donor palette populations), which captures the ancient population structure (Supplementary Fig. 11.6).

Third, we inferred ancestry by learning about population structure in modern individuals or ancient individuals, painting them with respect to the ancient population panel and fitting them as a mixture using the ancient population surrogates, using the non-negative least squares implemented in GLOBETROTTER[100] (Supplementary Information section 11) with uncertainty estimated using 100 bootstrap replicates. All samples were analysed by leaving out one individual per donor population so that modern and ancient individuals are exchangeable (as the ancient individual is itself excluded from its own ancient donor population). We report this in a number of ways. The inferred ancestry results (Supplementary Table 6) are summarized by taking the mean across inferred populations in Supplementary Fig. 11.11; Supplementary Fig. 11.12 shows the means over sample information labels. We performed a spatiotemporal regression (Supplementary Table 11.2) using the model $a_{ik} = \alpha_{jk}t_i + \beta_{jk}x_i + \gamma_{jk}y_i + \varepsilon_{ijk}$ in which a_{ik} is the amount of ancestry individual i possesses from population k in regional analysis j, t_i is the age category of the individual (1 = Iron Age, 2 = Early Viking Age, 3 = Viking Age, 4 = Medieval) and x_i and y_i are the longitude and latitude of the burial location of the individual, respectively. The modern ancestry results are estimated using the spatial median instead of the mean, to account for ancestry being constrained in a k-dimensional simplex (Supplementary Fig. 11.14), with uncertainty quantified by bootstrap resampling of individuals (Supplementary Fig. 11.15).

Fourthly, we performed sensitivity analyses to ensure that the inference procedure performed as expected. We checked that sequence depth was not associated with cluster membership (Supplementary

Fig. 11.7), and that sequence depth did not significantly affect inferred ancestry (Supplementary Fig. 11.8) by downsampling individuals with high-depth data available, rephasing, re-imputing and repainting them, and assigning ancestry using the above procedure. Results 2× and above were extremely similar, whereas at 1× there was some loss of precision but the broad structure remained clear.

Finally, we ran a principal components analysis of the ancient + modern populations painted against our donor populations (Supplementary Fig. 11.9) as well as an all-versus-all ChromoPainter analysis including modern and ancient individuals (Supplementary Fig. 11.10).

Ancestry diversity measure

We wish to quantify diversity in ancestry for a population of individuals, with diverse meaning a large deviation of individual ancestry estimates from the average ancestry in that population. We compute the average Kullback–Leibler divergence for each individual label from the average of that label:

$$D(A^{(l)}) = \frac{1}{n_l}\sum_{i=1}^{n_l} KL(A_i^{(l)} \,||\, p^{(l)})$$

in which $A^{(l)}$ is the n_l by K matrix of ancestry estimates in label l, $p(l)$ is the length **K** vector of average ancestries in that label, and $KL(Q \,||\, P) = \sum_{k=1}^{K} q_k \log_2\left(\frac{q_k}{p_k}\right)$. We performed a simulation study to validate this measure (Supplementary Information section 11, Supplementary Fig. 11.13), which allowed us to calibrate the expected diversity as a function of sample size.

Spatiotemporal patterns

To visualize the migration patterns of the Vikings, we used inverse distance weighting interpolation—implemented in the function idw of the R package gstat—to interpolate the proportion of each ancient genome that was attributed by our fineSTRUCTURE analysis (Supplementary Table 6) to one of the predefined ancestry groups: UK, Denmark, Norway, Sweden, Italy, Poland and Finland. We used the Shepard method of interpolation[12,101] with the weight for a given interpolation location x equal to $1/(d(x,v)^2)$, in which v is the location of an observed sample and $d(a,b)$ is the distance between two points a and b. For plotting maps, we used a Mercator projection and downloaded coastal contours at 1:50-m scale from Natural Earth (https://www.naturalearthdata.com/).

Lactase persistence and pigmentation SNPs

For ancient populations we estimated the derived A allele frequency of the SNP rs4988235, known to affect expression of the lactase (*LCT*) gene. The ancestral G allele is responsible for lactase intolerance in adult Europeans[39]. We used ANGSD[60] to estimate the allele frequencies of the ancient population on the basis of the genotype likelihood data. We used the five European populations (CEU, FIN, GBR, TSI and IBS) and two outgroups (YRI and CHB) from the 1000 Genomes Project as comparative groups. We also included the present-day Danish population from the IPSYCH case–cohort study[44] and geographically proximate Iron and Bronze Age populations to trace frequency shifts of SNP rs4988235 through time. We also used ANGSD[60] to estimate the frequencies of 22 SNPs (HIrisPlex[102]) with strongest influence on human pigmentation phenotypes in the Viking Age and Early Viking Age Scandinavian population.

Signatures of selection

We aimed to find SNPs with allele frequencies that changed significantly in the last 10,000 years, using our ancient human genomes to look at the frequencies of alleles in the past. We combined our Viking Age and Iron Age genomes with previously published present-day, Bronze Age, Neolithic and Mesolithic sequence data typed at the Human Origins array (Supplementary Note 6). We filtered for genomes that were younger than 8000 BC and that were located within a bounding box encompassing the European continent: 30°< latitude < 75° and −15°< longitude < 45°. We then used neoscan in Ohana[36,103] to scan for variants with allele frequencies that were strongly associated with time, after controlling for genome-wide changes in ancestry that might have also occurred over time. We analysed only sites with a minor allele frequency >1% (Supplementary Note 14).

Tracking the evolution of complex traits in Scandinavia

We wanted to examine whether we could identify signals of recent population differentiation of complex traits by comparing genotypes of Viking Age samples excavated in Scandinavia (that is, Denmark, Sweden and Norway) with those of a present-day Scandinavian population. For the latter, we used imputed genotypes from subjects born in Denmark between 1981 and 2011 from the IPSYCH case–cohort study[44]. We downloaded summary statistics from the genome-wide association study ATLAS webpage (https://atlas.ctglab.nl)[45], from studies of 16 disease- and anthropometric traits (excluding those related to cognition) published in 2017 or later with SNP heritability estimated at >0.1, sample size of >100,000 and >100 identified genome-wide significant loci. We calculated polygenic risk scores based on independent (R^2 < 0.1 within 10-Mb range) genome-wide significant allelic effects and standardized them to a unit representing the standard deviation of the mean of their distribution. We then removed outliers (anyone with a value for any of the 25 principal components falling more than 4 standard deviations away from the group mean) reiteratively from within each ancestry group (treating the Scandinavian Viking age samples as one ancestry group), and subsequently tested for difference in polygenic risk score distribution between Viking Age samples and Danish-ancestry IPSYCH random population samples using a linear regression model correcting for sex and the 25 principal components.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Sequence data are available at the European Nucleotide Archive under accession number PRJEB37976.

Code availability

Functions for calculating *f*-statistics are available as an R package at GitHub (https://github.com/martinsikora/admixr).

50. Willerslev, E. & Cooper, A. Ancient DNA. *Proc. R. Soc. Lond. B* **272**, 3–16 (2005).
51. Gilbert, M. T. P., Bandelt, H.-J., Hofreiter, M. & Barnes, I. Assessing ancient DNA studies. *Trends Ecol. Evol.* **20**, 541–544 (2005).
52. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
53. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
54. Schubert, M. et al. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012).
55. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
56. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
57. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
58. Fu, Q. et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
59. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
60. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
61. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013).
62. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).

63. Ralf, A., Montiel González, D., Zhong, K. & Kayser, M. Yleaf: software for human Y-chromosomal haplogroup inference from next-generation sequencing data. *Mol. Biol. Evol.* **35**, 1291–1294 (2018).
64. Korneliussen, T. S. & Moltke, I. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics* **31**, 4009–4011 (2015).
65. Monroy Kuhn, J. M., Jakobsson, M. & Günther, T. Estimating genetic kin relationships in prehistoric populations. *PLoS ONE* **13**, e0195491 (2018).
66. Staples, J., Nickerson, D. A. & Below, J. E. Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis. *Genet. Epidemiol.* **37**, 136–141 (2013).
67. 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).
68. Martiniano, R. et al. The population genomics of archaeological transition in west Iberia: investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genet.* **13**, e1006852 (2017).
69. International Multiple Sclerosis Genetics Consortium & The Wellcome Trust Case Control Consortium 2. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
70. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
71. Gamba, C. et al. Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
72. Jones, E. R. et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
73. Skoglund, P. et al. Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344**, 747–750 (2014).
74. Schiffels, S. et al. Iron Age and Anglo-Saxon genomes from east England reveal British migration history. *Nat. Commun.* **7**, 10408 (2016).
75. Olalde, I. et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).
76. Sikora, M. et al. Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* **358**, 659–662 (2017).
77. Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
78. Jones, E. R. et al. The Neolithic transition in the Baltic was not driven by admixture with early European farmers. *Curr. Biol.* **27**, 576–582 (2017).
79. Seguin-Orlando, A. et al. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
80. Raghavan, M. et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
81. Hofmanová, Z. et al. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl Acad. Sci. USA* **113**, 6886–6891 (2016).
82. Damgaard, P. B. et al. 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018).
83. Günther, T. et al. Population genomics of Mesolithic Scandinavia: investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol.* **16**, e2003703 (2018).
84. Mittnik, A. et al. The genetic prehistory of the Baltic Sea region. *Nat. Commun.* **9**, 442 (2018).
85. Kılınç, G. M. et al. The demographic development of the first farmers in Anatolia. *Curr. Biol.* **26**, 2659–2666 (2016).
86. Lazaridis, I. et al. Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218 (2017).
87. de Barros Damgaard, P. et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).
88. Valdiosera, C. et al. Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. *Proc. Natl Acad. Sci. USA* **115**, 3428–3433 (2018).
89. Martiniano, R. et al. Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nat. Commun.* **7**, 10326 (2016).
90. Mathieson, I. et al. The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018).
91. Gallego Llorente, M. et al. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**, 820–822 (2015).
92. Broushaki, F. et al. Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499–503 (2016).
93. Veeramah, K. R. et al. Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria. *Proc. Natl Acad. Sci. USA* **115**, 3494–3499 (2018).
94. Amorim, C. E. G. et al. Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nat. Commun.* **9**, 3547 (2018).
95. Olalde, I. et al. A Common genetic origin for early farmers from Mediterranean Cardial and Central European LBK cultures. *Mol. Biol. Evol.* **32**, 3132–3142 (2015).
96. Olalde, I. et al. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196 (2018).
97. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
98. Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P. & Ramachandran, S. pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* **32**, 2817–2823 (2016).
99. Browning, B. L. & Browning, S. R. Detecting identity by descent and estimating genotype error rates in sequence data. *Am. J. Hum. Genet.* **93**, 840–851 (2013).
100. Hellenthal, G. et al. A genetic atlas of human admixture history. *Science* **343**, 747–751 (2014).
101. Fortin, M.-J. & Dale, M. R. T. *Spatial Analysis: A Guide for Ecologists* (Cambridge Univ. Press, 2005).
102. Walsh, S. et al. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet.* **7**, 98–115 (2013).
103. Cheng, J. Y., Mailund, T. & Nielsen, R. Fast admixture analysis and population tree estimation for SNP and NGS data. *Bioinformatics* **33**, 2148–2155 (2017).

Acknowledgements This work was supported by the Mærsk Foundation, the Lundbeck Foundation, the Novo Nordisk Foundation, the Danish National Research Foundation, University of Copenhagen (KU2016) and the Wellcome Trust (grant no. WT104125MA). E.W. thanks St John's College, Cambridge for providing an excellent environment for scientific thoughts and collaborations. S.R. was supported by the Novo Nordisk Foundation (NNF14CC0001). F.R. was supported by a Villum Fonden Young Investigator Award (project no. 00025300). G.S. and E.C. were supported by a Marie Skłodowska-Curie Individual Fellowship 'PALAEO-ENEO', a project funded by the European Union EU Framework Programme for Research and Innovation Horizon 2020 (grant agreement number 751349). R.M. was supported by an EMBO Long-Term Fellowship (ALTF 133-2017). M.C. is supported by the Canada Research Chairs Program (231256), the Canada Foundation for Innovation (36801) and the British Columbia Knowledge Development Fund (962-805808). I. Moltke was supported by a YDUN grant from Independent Research Fund Denmark (DFF-4090-00244) and a Villum Fonden Young Investigator Award (project no. 19114). N.P. and C.H.J. are supported by the Swedish Research Council (2015-00466). N.G. was supported by the Program of Fundamental Scientific Research of the State Academies of Sciences, Russian Federation, state assignment no. 0184-2019-0006. D.G.B. and L.M.C. were supported by Science Foundation Ireland/Health Research Board/Wellcome Trust award no. 205072. We thank the iPSYCH Initiative, funded by the Lundbeck Foundation (grant numbers R102-A9118 and R155-2014-1724), for supplying SNP frequency estimates from the present-day Danish population for comparison with Viking Age samples; M. Jakobsson and A. Götherström for providing preliminary access to the sequencing data of 23 Viking Age samples from Sigtuna; M. Corrente for providing access to the skeletal remains from Cancarro; and N. M. Mangialardi and M. Maruotti for the useful suggestions; Greenland National Museum and Archives, as well as the Gotland Museum, for permission to sample their skeletons; J. Kavanagh for providing information on his excavation, and L. Buckley, D. Keating and B. Ó Donnabháin for analysing the remains; R. Breward and J. Murden from the Dorset County Museum for allowing access to their assemblage for DNA sampling; J. Hansen and M. B. Henriksen at Odense Bys Museer for allowing sampling of skeletal material from Hessum and Galjedil; Moesgaard Museum for allowing sampling of skeletal material from Hesselbjerg; C. Bertilsson, P. Lingström, B. Lundberg, K. Lidén and J. Andersson for their help in sampling the ancient human remains; L. Drenzel for permission to sample the human remains; C. Ödman for suggesting relevant material for this study; Ł. Stanaszek, M. Zaitz and the Regional Museum in Cedynia for providing samples; L. Vinner, A. Seguin-Orlando, K. Magnussen, L. Petersen, C. Mortensen and M. J. Jacobsen at the Danish National Sequencing Centre for producing the analysed sequences; P. S. Olsen and T. Brand for technical assistance in the laboratories; R. M. Durbin and J. H. Barrett for comments and suggestions; and J. Wilson, J. Jesch, E. Harlitz-Kern and F. Martín Racimo for their feedback.

Author contributions E.W. initiated and led the study. E.W., A.M., D.J.L., Martin Sikora, F.R., R.N., K.K., L.H., S.M.S., J.B., N.P., T.W., A.I., M.E.A., M.W.P., N.L., J.A., I. Moltke and A.A. designed the study. A.M., P.d.B.D., L.M.C., M.M.B., A.K.F., I.L. and J.S. produced the data. A.M., D.J.L., Martin Sikora, F.R., S.R., I. Moltke, R.N., T.W., L.M.C., E.J., A.I., M.W.P., T.K., R.M., G.R., C.B., J.V.M.-M., H.M., A.A., J.C., K.H.I. and M.E.A. analysed or assisted in analysis of data. E.W., A.M., D.J.L., Martin Sikora, F.R., S.M.S., K.K., L.H., R.N., M.C. and A.I. interpreted the results with considerable input from I. Moltke, M.E.A., M.W.P., T.K., H.W., R.M., G.R., T.W., C.H.J., J.A., N.L., N.P., J.B., A.A., M.T.P.G., L.O. and other authors. E.W., A.M., D.J.L., Martin Sikora, F.R., S.M.S., K.K. and L.H. wrote the manuscript with considerable input from M.C., J.B., N.P., I. Moltke, N.L., A.I., R.M., E.J., J.A., M.L.J., C.H.J., M.W.P., M.E.A., G.R. and M.M., with contributions from all authors. A.M., L.M.C., M.W.P., H.W., M.M.B., P.d.B.D., A.K.F., M.A., R.A., M.M., E.C., G.S., A.B., A.F., B. Schütz, B. Skar, C.A., C.F., D.B., D.P., G.T.-W., H.G., I.L., I.G., I. Mainland, I.P., I.M.M., J.M., J. Gibson, J.P., J. Gustafsson, L. Simpson, L. Strand, L.L., Maeve Sikora, M.F., M.V., M.R., M.B., T.P., M. Søvso, N.G., T.C., O.K., O.U., P.F., P.H., S.S., S.V.A., S.E., V.M., W.B., Y.M., P.P., M.D.J., A.P., D.G.B., M.L.J., J.A., N.L., N.P., M.T.P.G., M.E.A., J.B. and E.W. excavated, curated, sampled and/or described analysed skeletons.

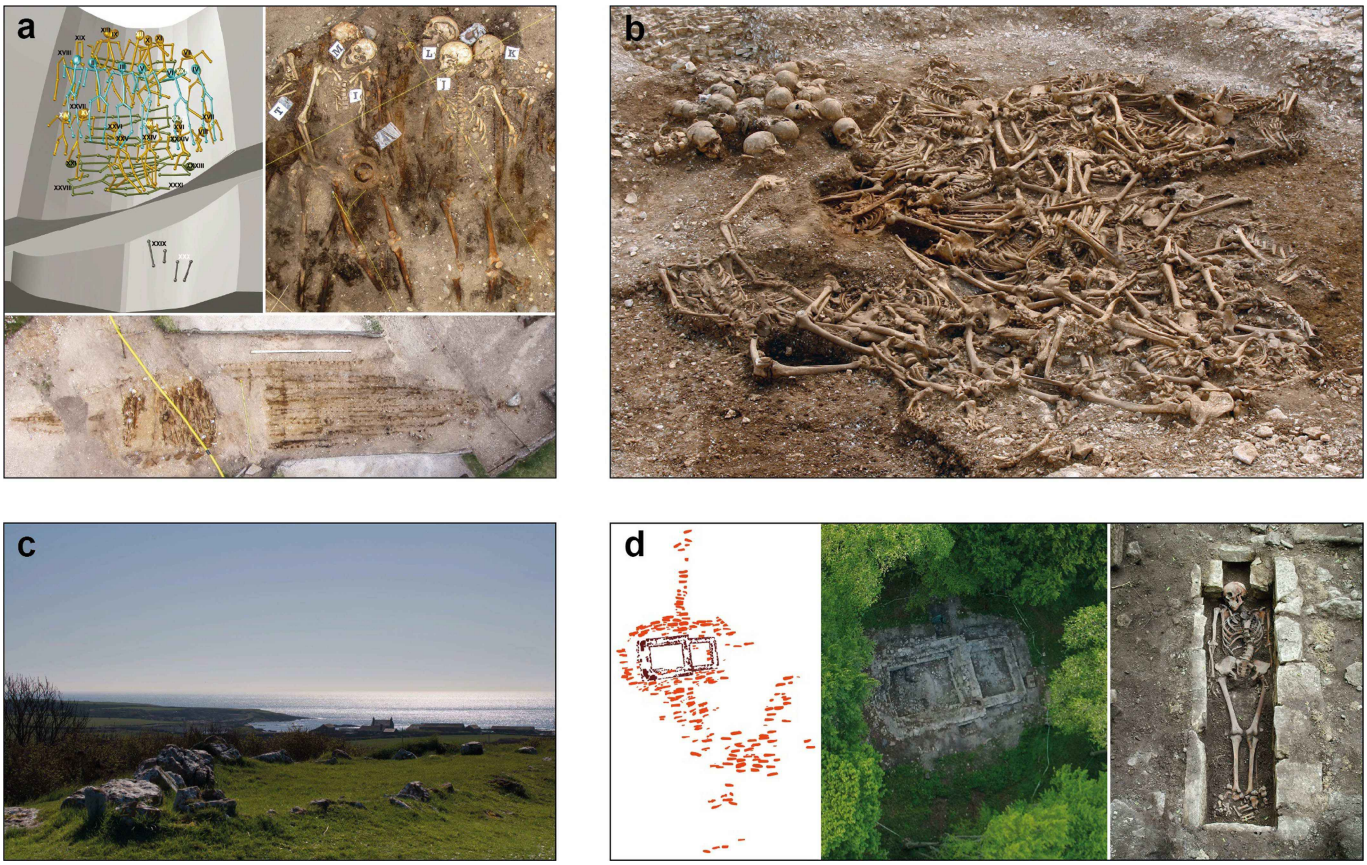
Competing interests The authors declare no competing interests.

Additional information

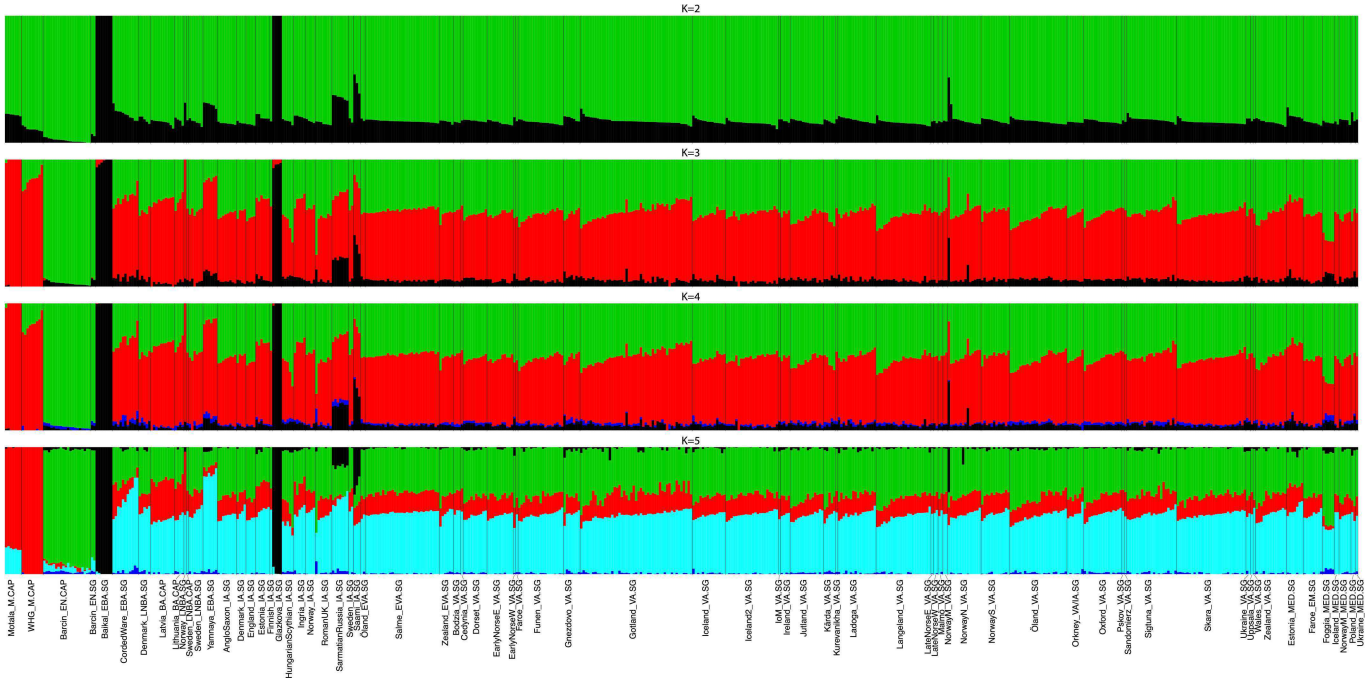
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2688-8.

Correspondence and requests for materials should be addressed to R.N., T.W. or E.W.

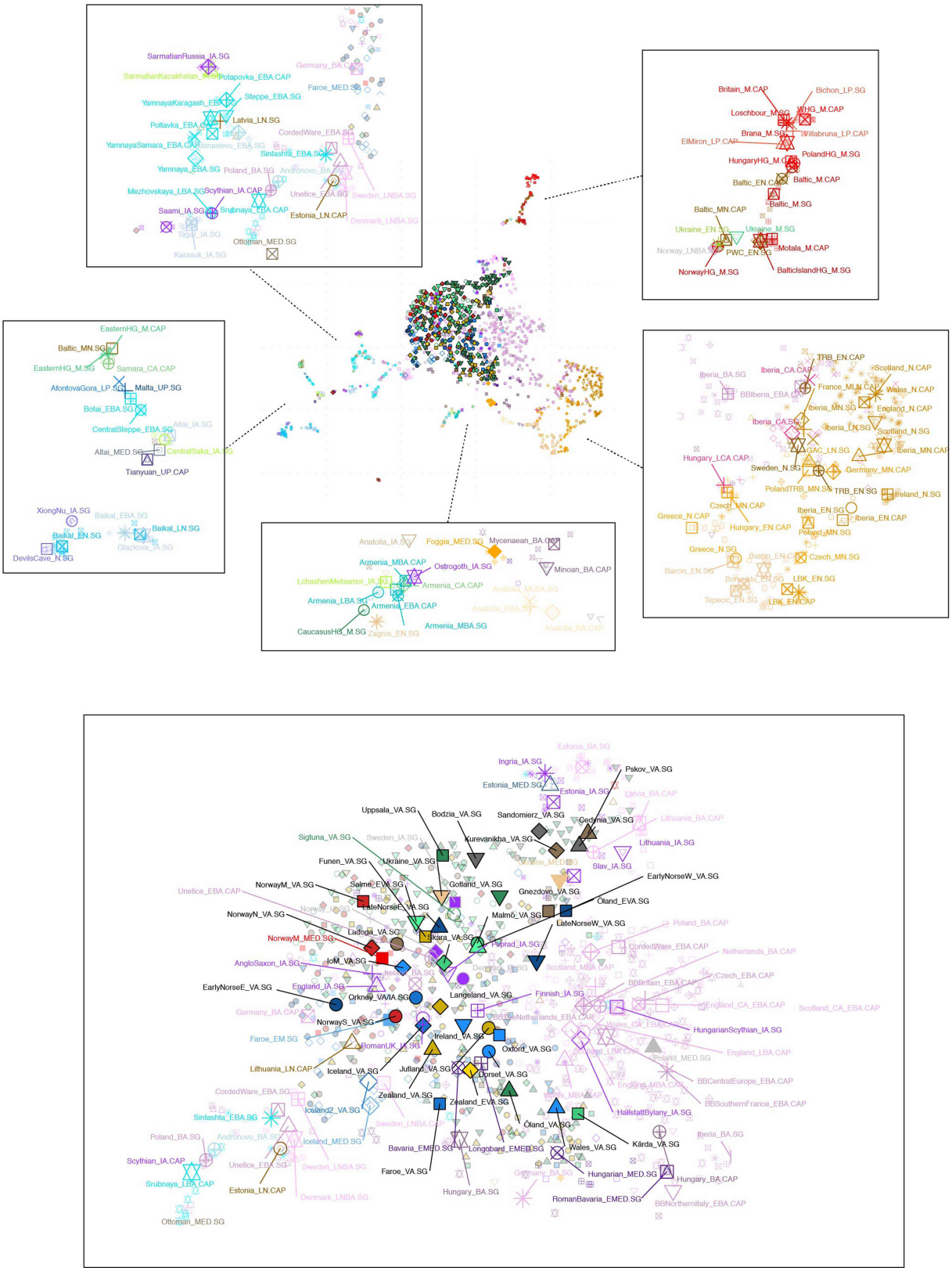
Peer review information *Nature* thanks James Barrett, Wolfgang Haak and Pontus Skoglund their contribution to the peer review of this work.



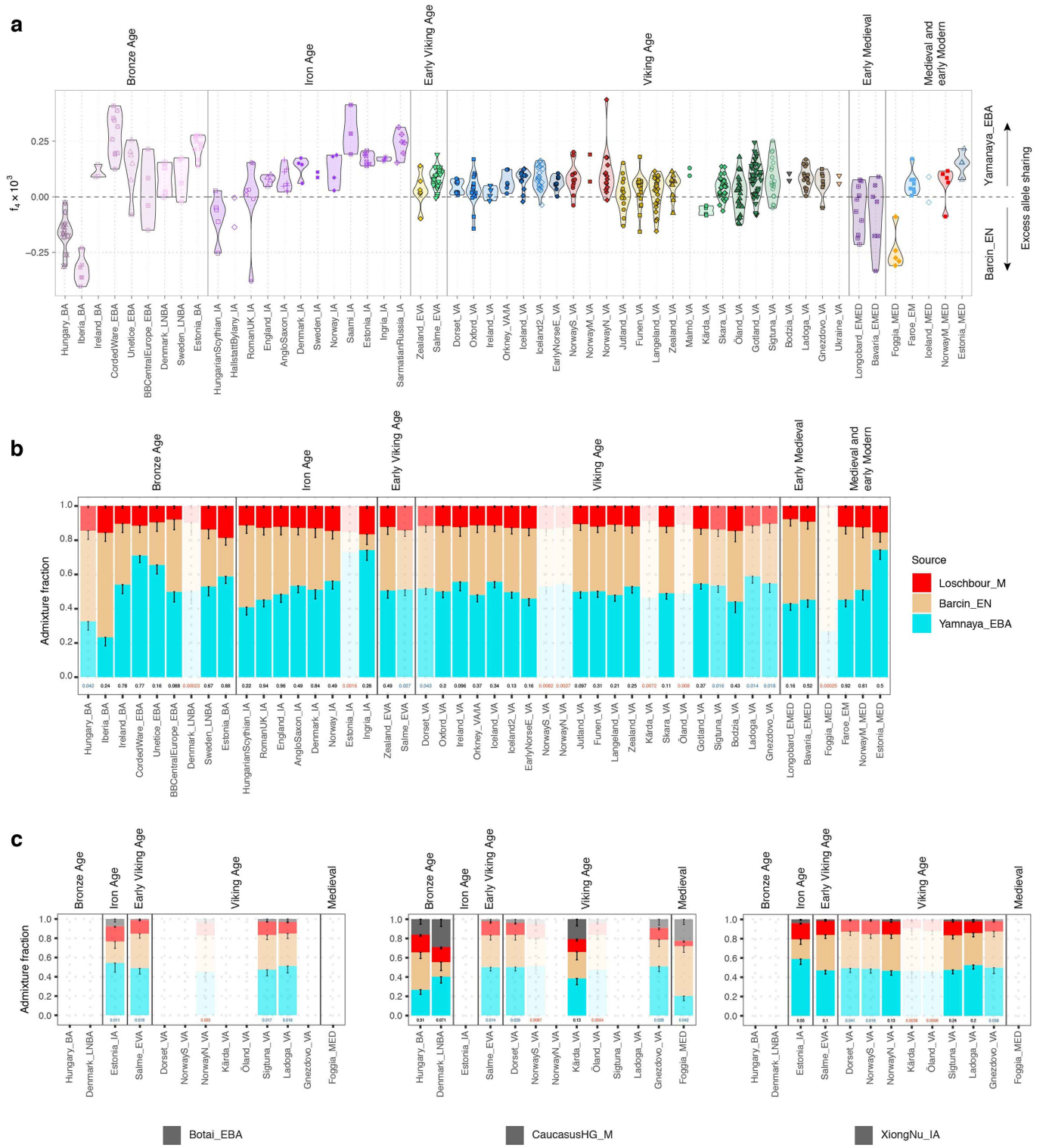
Extended Data Fig. 1 | Viking Age archaeological sites. Examples of a few archaeological Viking Age sites and samples used in this study. **a**, Salme II ship burial site of the Early Viking Age, excavated in present-day Estonia: schematic of skeletons (top left) and aerial images of skeletons (top right, and bottom). **b**, Ridgeway Hill mass grave dated to the tenth or eleventh century AD, located on the crest of Ridgeway Hill near Weymouth, on the south coast of England (reproduced with permission from Dorset County Council/Oxford Archaeology). Around 50 predominantly young adult male individuals were excavated. **c**, The site of Balladoole, around AD 900, a Viking was buried in an oak ship at Balladoole (Arbory) in the south east of the Isle of Man. **d**, Viking Age archaeological site in Varnhem, in Skara municipality (Sweden). Schematic map of the church foundation (left) and the excavated graves (red markings) at the early Christian cemetery in Varnhem; foundations of the Viking Age stone church in Varnhem (middle) and the remains of a 182-cm-long male individual (no. 17) buried in a lime stone coffin close to the church foundations (right).



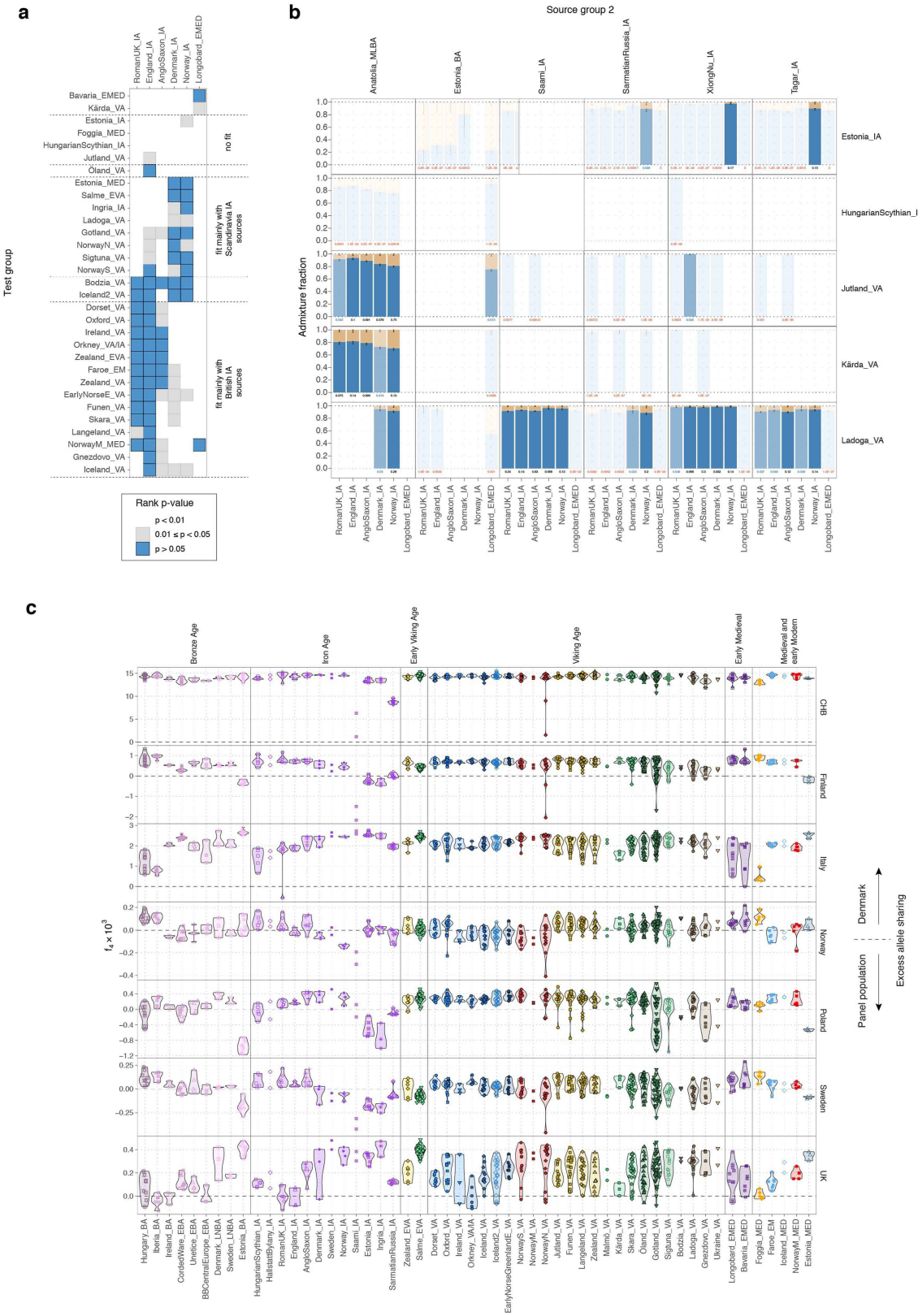
Extended Data Fig. 2 | Model-based clustering analysis. Admixture plot ($K = 2$ to $K = 5$) for 567 ancient individuals, spanning 71 populations. This figure is a subset of the most relevant individuals and populations from Supplementary Fig. 7.2; see Supplementary Note 7 for further details. This plot consists of 378 ancient samples from this study; Viking Age samples from Sigtuna (Sweden)[10] ($n = 21$), Iceland[18] ($n = 22$) and other ancient comparative groups ($n = 146$).



Extended Data Fig. 3 | Fine-scale population structure. The point cloud at the top centre shows an alternative view of the UMAP result from Fig. 2b, with all ancient individuals coloured on the basis of analysis group. The framed panels surrounding the point cloud highlight particular ancestry clusters (as indicated), with labels and larger symbols corresponding to the median coordinates for the respective group. Similarly, the larger bottom panel shows median group coordinates for the large central point cloud, which includes the vast majority of European individuals from the Bronze Age onwards.

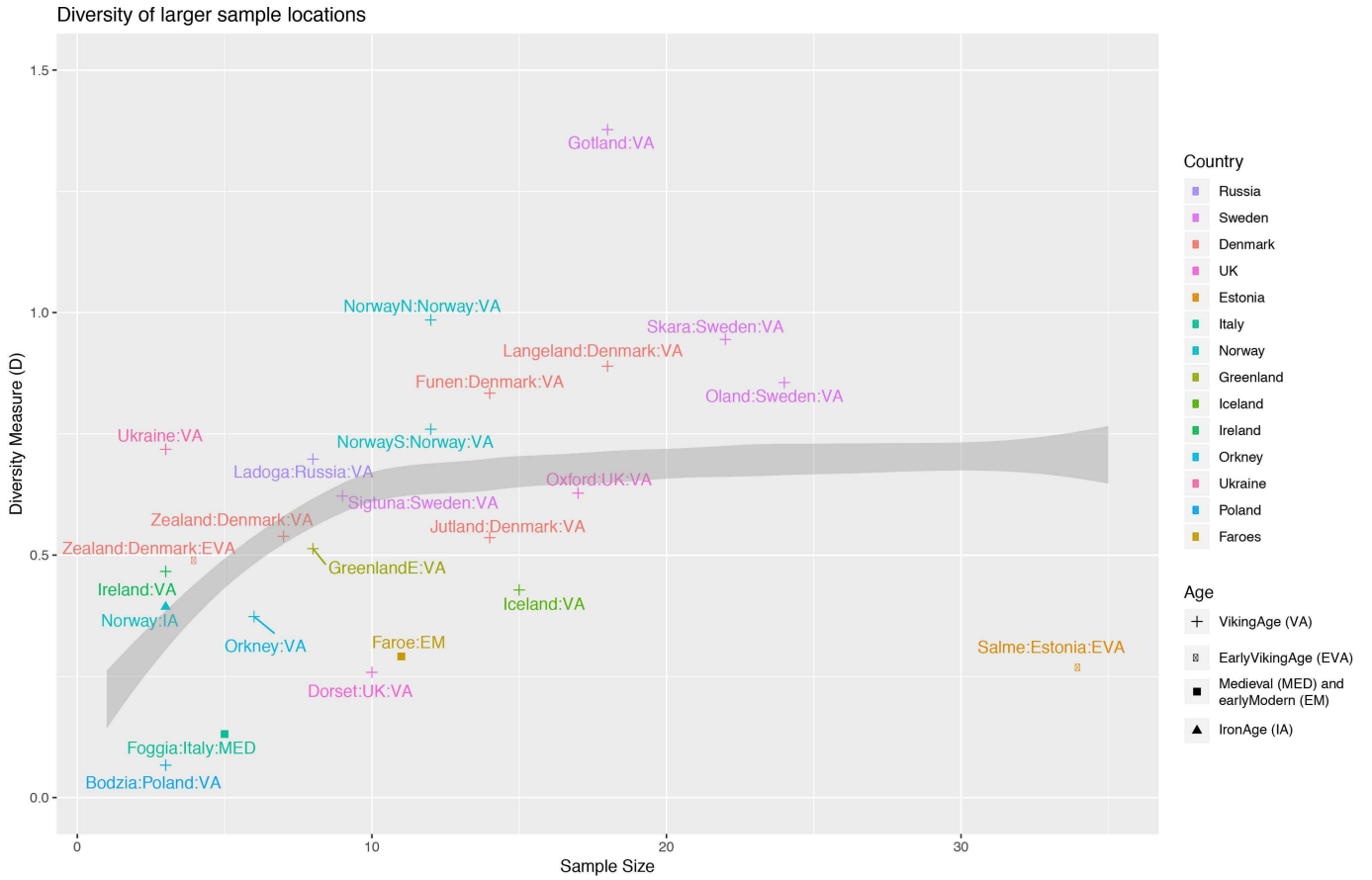


Extended Data Fig. 4 | Ancestry modelling for distal sources. a, Contrasting allele-sharing between Anatolian farmers (Barcin_EN) and Steppe pastoralists (Yamnaya_EBA) for European individuals from the Bronze Age and later. Violin plots showing distributions of statistics f_4(YRI,test individual;Barcin_EN,Yamnaya_EBA) for $n = 515$ individuals with a minimum of 1,000,000 SNPs with genotypes and groups with at least 2 such individuals. **b**, Ancestry proportions of analysis groups from the Bronze Age and later inferred using qpAdm. Target groups were modelled using three distal sources representing European hunter-gatherer (Loschbour_M), Anatolian farmer (Barcin_EN) and Steppe pastoralist (Yamnaya_EBA) ancestry. Sample sizes for target groups can be found in Supplementary Table 10. Error bars indicate standard error obtained from qpAdm. **c**, Ancestry proportions of analysis groups for which the three-source model was rejected using qpAdm ($P < 0.05$). Target groups were modelled including one additional distal source representing either Steppe hunter-gatherer (Botai_EBA), Caucasus hunter-gatherer (CaucasusHG_M) or East-Asian-related (XiongNu_IA) ancestry.

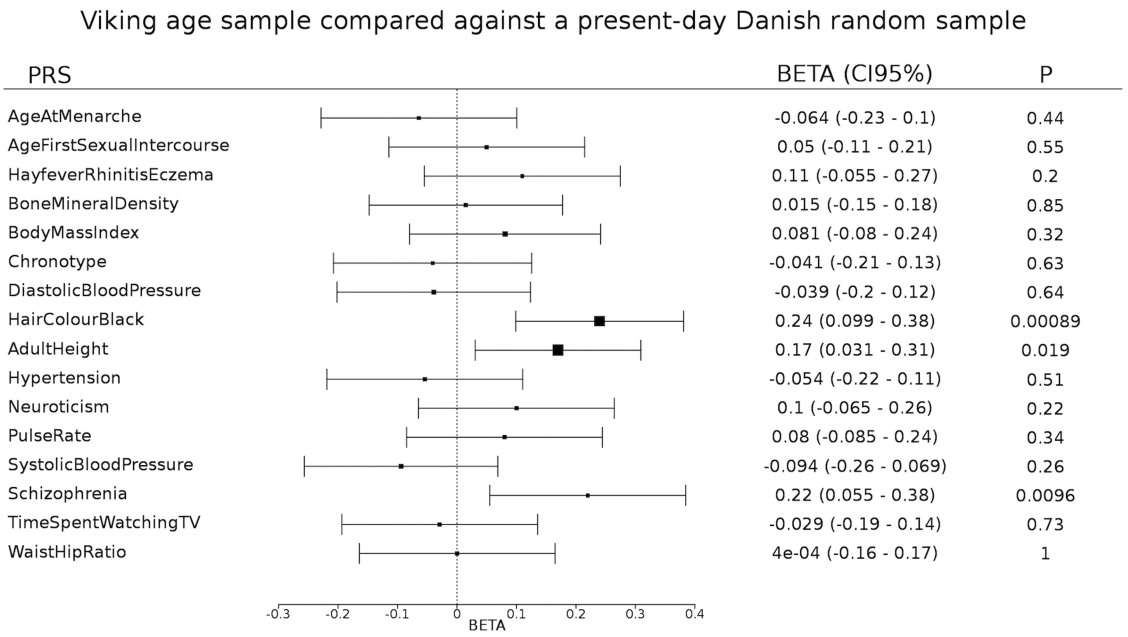


Extended Data Fig. 5 | See next page for caption.

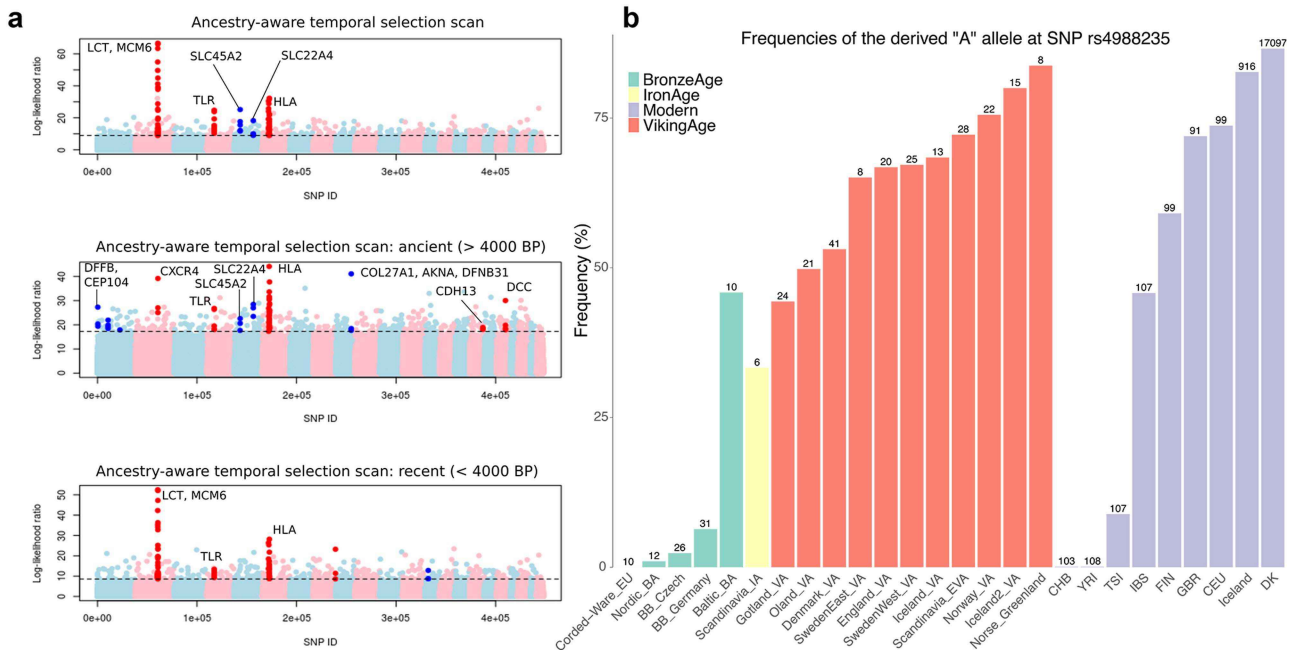
Extended Data Fig. 5 | Ancestry modelling for proximate sources. a, Testing for continuity between European Iron Age and later Viking Age and Medieval groups. Coloured squares depict whether a particular target group (row) can be modelled using a single source group (column). *P* values for f_4 rank of 0 (corresponding to a single source group) were obtained using qpAdm with a set of 15 outgroups, which included European Bronze Age groups that preceded the source groups. Sample sizes for target groups can be found in Supplementary Table 12. **b**, Two-way admixture ancestry proportions of target groups for which a single source was rejected ($P \leq 0.05$). Target groups were modelled using additional proximate Bronze and Iron Age sources. Sample sizes for target groups can be found in Supplementary Table 13. For both **a** and **b**, only ancient groups containing at least 3 individuals with a minimum of 1,000,000 SNPs with genotypes are plotted. **c**, Contrasting allele-sharing between populations of present-day Denmark and other populations. Violin plots showing distributions of statistics f_4(YRI,test individual;panel population,Denmark) for $n = 489$ individuals with a minimum of 50,000 SNPs with genotypes and groups with at least 2 such individuals. Median values for distributions are indicated with horizontal lines.



Extended Data Fig. 6 | Ancestry diversity of different population groups. Diversity of different labels (that is, sample locations combined with historical age) are shown as a function of their sample size. The diversity measure is the Kullback–Leibler divergence from the label means, capturing the diversity of a group with respect to the average of that group (see Supplementary Note 11 for details). Larger values are more diverse, although a dependence on sample size is expected. The simulation expectation for the best fit to the data ($\theta = 0.2$) is shown.



Extended Data Fig. 7 | Polygenic risk scores. Polygenic risk scores (PRS) for 16 complex human traits in 148 Viking Age samples from Denmark, Sweden and Norway, compared against a reference sample of 20,551 Danish-ancestry individuals randomly drawn from all individuals born in Denmark in 1981–2005. The PRS is in each case based on allelic effects for >100 independent genome-wide significant SNPs from recent genome-wide association studies of the respective traits and standardised to a mean of 0 and standard deviation of 1 in the entire sample. Difference in PRS was estimated in a linear regression correcting for sex and 25 principal components of overall genetic structure. The plotted BETA indicates the coefficient for the test-group (Viking Age sample) PRS compared to that of the Danish comparison sample, with error bars indicating the 95% confidence interval of BETA, and *P* indicating the two-tailed *P* value of the corresponding *t*-test (not corrected for number of tests). Only PRS for black hair colour is significantly different between the groups after taking account of multiple testing.



Extended Data Fig. 8 | Positive selection in Europe. a, Manhattan plots of the likelihood ratio scores in favour of selection looking at the entire 10,000-year period (top, general scan), the period up to 4,000 years before present (middle, ancient scan) and the period from 4,000 years before present up to the present day (bottom, recent scan). The highlighted SNPs have a score larger than the 99.9% quantile of the empirical distribution of log-likelihood ratios, and have at least two neighbouring SNPs (± 500 kb) with a score larger than the same quantile. $n = 1,185$ genomes are used in the selection scan. **b**, Frequencies of the derived A allele rs4988235 SNP responsible for lactase persistence in humans for different Viking Age groups, present-day populations from the 1000 Genomes Project as well as relevant Bronze Age population panels. The numbers at the top of the bars denote the sample size on which the allele frequency estimates are based.

nature research

Corresponding author(s): Eske Willerslev

Last updated by author(s): May 3, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No specific software was used for data collection. All software used in this study is listed below.
Data analysis	All software used in this work is publicly available. Corresponding publications are cited in the main text and supplementary material. List of software and respective versions: CASAVA v1.8.2 AdapterRemoval v2.1.3 bwa v0.7.10 bwa mem 0.7.10 picard tools v1.127 bamUtil v1.0.14 samtools v1.3.1 GATK v3.3.0 pysam 0.7.4 (python module) bedtools 2.27.1 mapDamage2.0 contamMix v1.0-5 SHRiMP 2.2.3 YFitter v0.2 Haplogrep 2.0 FineSTRUCTURE ANGSD v0.915 IBDseq v.r1206 PRANK v.150803 pathPhynder BEASTv1.8.2

schmutzi v.1.5.4
Yleaf v2
admixtools v4.1
NGSrelate v.1
GLOBETROTTER
READ
PRIMUS v1.9
plink and v1.9
ADMIXTURE v1.3
RAxML-8.1.15
SnpEff
SPAdes-3.9.0
R 3.2.3
Ohana
Beagle v4.1
python 2.7.12
perl v5.22.1
CALIB
FigTree v.1.4.4

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence data are available at the European Nucleotide Archive under accession number PRJEB37976.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We did not rely on statistical methods to predetermine sample sizes. Sample sizes in ancient population genetic studies are limited by the number of samples yielding endogenous DNA proportions amenable to whole genome sequencing.
Data exclusions	We selected 442 samples for whole-genome sequencing, out of all (n=528) screened samples, based on their endogenous content and low contamination estimates. These criteria are described in detail in Supplementary Notes 2 and 3. Furthermore, closely related individuals were excluded from analyses requiring population allele frequencies.
Replication	We did not attempt to specifically replicate experimental findings. But we note that samples from the same population carry similar genetic signatures. Moreover, genome-wide data allows for the analysis of multiple realisations of the sample history, by studying hundreds of thousands of SNP sites.
Randomization	We did not implement any randomization as no experimental groups or effect sizes were measured in this study.
Blinding	No blinding techniques were implemented, as experimental group assignment is not relevant for population genetic studies of this kind.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging