Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation

Torsten Günther   , Helena Malmström   , Emma M. Svensson   , Ayça Omrak   , Federico Sánchez-Quinto,
Gülşah M. Kılınç, Maja Krzewińska, Gunilla Eriksson, Magdalena Fraser, Hanna Edlund, Arielle R. Munters, Alexandra Coutinho,
Luciana G. Simões, Mário Vicente, Anders Sjölander, Berit Jansen Sellevold, Roger Jørgensen, Peter Claes, Mark D. Shriver,
Cristina Valdiosera, Mihai G. Netea, Jan Apel, Kerstin Lidén, Birgitte Skar, Jan Storå   , Anders Götherström   ,
Mattias Jakobsson

Abstract

Scandinavia was one of the last geographic areas in Europe to become habitable for humans after the Last Glacial Maximum (LGM). However, the routes and genetic composition of these postglacial migrants remain unclear. We sequenced the genomes, up to 57× coverage, of seven hunter-gatherers excavated across Scandinavia and dated from 9,500–6,000 years before present (BP). Surprisingly, among the Scandinavian Mesolithic individuals, the genetic data display an east–west genetic gradient that opposes the pattern seen in other parts of Mesolithic Europe. Our results suggest two different early postglacial migrations into Scandinavia: initially from the south, and later, from the northeast. The latter followed the ice-free Norwegian north Atlantic coast, along which novel and advanced pressure-blade stone-tool techniques may have spread. These two groups met and mixed in Scandinavia, creating a genetically diverse population, which shows patterns of genetic adaptation to high latitude environments. These potential adaptations include high frequencies of low pigmentation variants and a gene region associated with physical performance, which shows strong continuity into modern-day northern Europeans.

Author summary

The Scandinavian peninsula was the last part of Europe to be colonized after the Last Glacial Maximum. The migration routes, cultural networks, and the genetic makeup of the first Scandinavians remain elusive and several hypotheses exist based on archaeology, climate modeling, and genetics. By analyzing the genomes of early Scandinavian hunter-gatherers, we show that their migrations followed two routes: one from the south and another from the northeast along the ice-free Norwegian Atlantic coast.

These groups met and mixed in Scandinavia, creating a population more diverse than contemporaneous central and western European hunter-gatherers. As northern Europe is associated with cold and low light conditions, we investigated genomic patterns of adaptation to these conditions and genes known to be involved in skin pigmentation. We demonstrate that Mesolithic Scandinavians had higher levels of light pigmentation variants compared to the respective source populations of the migrations, suggesting adaptation to low light levels and a surprising signal of genetic continuity in *TMEM131*, a gene that may be involved in long-term adaptation to the cold.

Data Availability: All sequence data generated for this project are available from the European Nucleotide Archive under the study accession number PRJEB21940. All other data are within the paper and its Supporting Information files.

**Competing interests:** The authors have declared that no competing interests exist.

Abbreviations: aDNA, ancient DNA; BP, before present; cal, calibrated; CEU, Utah residents with Central European ancestry; CHG, Caucasus hunter-gatherer; Chimp, Chimpanzee; EHG, eastern hunter-gatherer; FIN, modern-day Finnish individual; kbp, kilo base pairs; LD, linkage disequilibrium; LGM, Last Glacial Maximum; MSMC, multiple sequentially

Introduction

As the ice sheet retracted from northern Europe after the Last Glacial Maximum (LGM), around 23,000 years ago, new habitable areas emerged [1], allowing plants [2,3] and animals [4,5] to recolonize the Scandinavian peninsula (hereafter referred to as Scandinavia). There is consistent evidence of human presence in the archaeological record from approximately 11,700 years before present (BP) both in southern and northern Scandinavia [6–9]. At this time, the ice sheet was still dominating the interior of Scandinavia [9,10] (Fig 1A, S1 Text), but recent climate modeling shows that the Arctic coast of (modern-day) northern Norway was ice free [10]. Similarities in late-glacial lithic technology (direct blade percussion technique) of Western Europe and the oldest counterparts of Scandinavia appearing around 11,000 calibrated (cal) BP [11] (S1 Text) have been used to argue for an early postglacial migration from southwestern Europe into Scandinavia, including areas of northern Norway. However, studies of another lithic technology, the "pressure blade" technique, which first occurred in the northern parts of Scandinavia around 10,200 cal BP, indicates contact with groups in the east and possibly an eastern origin of the early settlers [7,12–15] (S1 Text). The first genetic studies of Mesolithic human remains from central and eastern Scandinavian hunter-gatherers (SHGs) revealed similarities to two different Mesolithic European populations, the "western hunter-gatherers" (WHGs) from western, central, and southern Europe and the "eastern hunter-gatherers" (EHGs) from northeastern and eastern Europe [16–24]. Archaeology, climate modeling, and genetics suggest several possibilities for the early postglacial migrations into Scandinavia, including migrations from the south, southeast, northeast, and combinations of these; however, the early postglacial peopling of Scandinavia remains elusive [1,4,6–19,25,26]. In this study, we contrast genome sequence data and stable isotopes from Mesolithic human remains from western, northern, and eastern Scandinavia to infer the early postglacial migration routes into Scandinavia—from where people came, what routes they followed, how they were related to other Mesolithic Europeans [17–21,27]—and to investigate human adaptation to high-latitude environments.
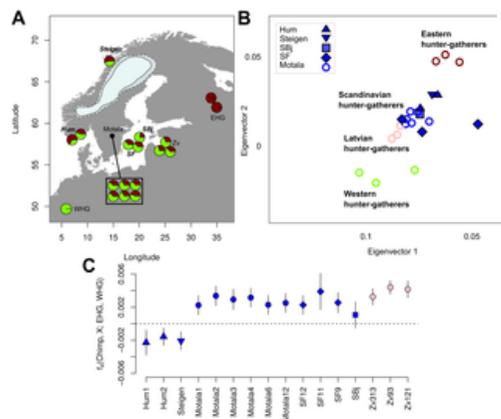
Fig 1. Mesolithic samples and their genetic affinities.
(A) Map of the Mesolithic European samples used in this study. The pie charts show the model-based [18,19] estimates of genetic ancestry for each SHG individual. The map also displays the ice sheet covering Scandinavia 10,000 cal BP (most credible [solid line] and maximum extend [dashed line] following [10]). Newly sequenced individuals are shown with bold and italic site names. SF11 is excluded from this map due to its low coverage (0.1×). Additional European EHG and WHG individuals used in this study derive from sites outside this map. The map was plotted using the R package rworldmap [28]. (B) Magnified section of genetic similarity among ancient and modern day individuals using PCA, featuring only the Mesolithic European samples (see S6 Text for the full plot). Symbols representing newly sequenced individuals have a black contour line. (C) Allele sharing between the SHGs, Latvian Mesolithic hunter-gatherers (Zv) [29], and EHGs versus WHGs measured by the statistic f_4(Chimp, SHG; EHG, WHG) calculated for the captured SNPs [20]. Error bars show two block-jackknife standard errors. Data shown in this figure can be found in S1 Data. BP, before present; cal, calibrated; Chimp, Chimpanzee; EHG, eastern hunter-gatherer; PCA, principal component analysis; SHG, Scandinavian hunter-gatherer; WHG, western hunter-gatherer; Zv, Latvian Mesolithic hunter-gatherer from Zvejnieki.
https://doi.org/10.1371/journal.pbio.2003703.g001

Results and discussion

We sequenced the genomes of seven hunter-gatherers from Scandinavia (Table 1; S1, S2 and S3 Text) ranging from 57.8× to 0.1× genome coverage, of which four individuals had a genome coverage above 1×. The remains were directly dated to between 9,500 cal BP and 6,000 cal BP, and were excavated in southwestern Norway (Hum1, Hum2), northern Norway (Steigen), and the Baltic islands of Stora Karlsö and Gotland (SF9, SF11, SF12, and SBj), and represent 18% (6 of 33) of all known human remains in Scandinavia older than 8,000 years [30]. All samples displayed fragmentation and cytosine deamination at fragment termini characteristic for ancient DNA (aDNA) (S3 Text). Mitochondrial (mt) DNA-based contamination estimates were <6% for all individuals (confidence intervals ranging from 0% to 9.5%) and autosomal contamination was <1% for all individuals except for SF11, which showed approximately 10% contamination (Table 1, S4 Text). Four of the seven individuals were inferred to be males and three were females. All the western and northern Scandinavian individuals and one eastern Scandinavian carried U5a1 mt haplotypes, whereas the remaining eastern Scandinavians carried U4a haplotypes (Table 1, S5 Text). These individuals represent the oldest U5a1 and U4 lineages detected so far. The Y chromosomal haplotype was determined for three of the four males, all carried I2 haplotypes, which were common in pre-Neolithic Europe (Table 1, S5 Text).

Table 1. Information on the seven SHGs investigated in this study, including cal BP (corrected for the marine reservoir effect, given as a range of two standard deviations), average genome coverage, average mt coverage, mt and Y chromosome haplogroups, and contamination estimates based on the mt, the X-chromosome for males and the autosomes.

The high coverage and Uracil-DNA-glycosylase (UDG)-treated genome (used in order to reduce the effects of postmortem DNA damage) [31] of SF12 allowed us to confidently discover new and hitherto unknown variants at sites with 55× or higher sequencing depth (S3 Text). Based on SF12's high-coverage and high-quality genome, we estimate the number of SNPs hitherto unknown (not recorded in dbSNP [v142]) to be approximately 10,600. This number is close to the median per European individual in the 1000 Genomes Project [32] (approximately 11,400, S3 Text), although a direct comparison is difficult due to the lower sequencing depth, different data processing, and larger sample sizes in the 1000 Genomes Project. At least 17% of these SNPs that are not found in modern-day individuals were in fact common among the Mesolithic Scandinavians (seen in the low coverage data conditional on the observation in SF12), and in total 24.2% were found in other prehistoric individuals (S3 Text), suggesting a substantial amount of hitherto unknown variation 9,000 years ago (S3 Text). Thus, many genetic variants found in Mesolithic individuals have not been carried over to modern-day groups. Among the novel variants in SF12, four (all heterozygous) are predicted to affect the function of protein coding genes [33] (S3 Text). The "heat shock protein" *HSPA2* in SF12 carries an unknown mutation that changes the amino acid histidine to tyrosine at a protein–protein interaction site, which likely disrupts the function of the protein (S3 Text). Defects in *HSPA2* are known to drastically reduce fertility in males [34]. It will be interesting to see how common such variants were among Mesolithic groups as more genome sequence data become available. The genomic data further allowed us to study the physical appearance of SHGs (S8 Text); for instance, they show a combination of eye color varying from blue to light brown and light skin pigmentation. This is strikingly different from the WHGs—who have been suggested to have the specific combination of blue eyes and dark skin [18,20,21,23] and EHGs—who have been suggested to be brown-eyed and light-skinned [19,20].

Demographic history of Mesolithic Scandinavians

In order to compare the genomic sequence data of the seven SHGs to genetic information from other ancient individuals and modern-day groups, data were merged with shotgun sequence data and SNP capture data from six published Mesolithic individuals from Motala in central Scandinavia, and 47 published Stone Age (Upper Paleolithic, Mesolithic, and Early Neolithic) individuals from other parts of Eurasia (S6 Text) [17–22,26,27,29,35–38], as well as with a world-wide set of 203 modern-day populations [18,32,39]. All 13 SHGs—regardless of geographic sampling location and age—display genetic affinities to both WHGs and EHGs (Fig 1A and 1B, S6 Text). One individual, SF11, seems to be a slight genetic outlier in the principal component analysis (PCA), which could be due to the lower coverage or driven by nuclear contamination (Table 1, S6 Text). Generally, the pattern of dual ancestry is consistent with a scenario in which SHGs represent a mixed group tracing parts of their ancestry to both the WHGs and the EHGs [17–19,22,24,40].

The SHGs from northern and western Scandinavia show a distinct and significantly stronger affinity to the EHGs compared to the central and eastern SHGs (Fig 1). Conversely, the SHGs from eastern and central Scandinavia were genetically more similar to WHGs compared to the northern and western SHGs (Fig 1). Using *qpAdm* [19], the EHG genetic component of northern and western SHGs was estimated to 48.9% (± 5%) and differs from the 37.8% (± 3.2%) observed in eastern and south-central SHGs. The latter estimate is similar to ancestry estimates obtained for eastern Baltic hunter-gatherers from Latvia [29] (33.7% ± 4.7%, Fig 1A). Although the difference in ancestry estimates between northern and western SHG, and eastern and south-central SHG is only marginally significant (Z = 1.87, *p* = 0.062), this pattern is in agreement with other analyses such as ADMIXTURE and TreeMix (S6 Text). Furthermore, the direct comparison using D statistics with Chimpanzee (Chimp) as an outgroup (D(Chimp, WHG; eastern or south-central SHG, northern or western SHG) < 0, Z = −5.14 and D(Chimp, EHG; eastern or south-central SHG, northern or

western SHG) > 0, Z = 1.72) show that WHG are genetically closer to eastern and south-central SHG, whereas EHG tend to share more alleles with northern and western SHGs (S2 Fig). These patterns of genetic affinity within SHGs are in direct contrast to the expectation based on geographic proximity with EHGs and WHGs.

From about 11,700 cal BP, consistent archaeological evidence of human presence exists in southern Scandinavia following the retreat of the ice sheet [6,41,42] (S1 Text). Artifacts and tools found at these sites show similarities with the Ahrensburgian tradition of northern central Europe [15,43], suggesting that these hunter-gatherers likely had a southern origin from a WHG-like gene pool as no EHG ancestry has been found in central and western Europe [18,21,24,27]. Although this genetic component would have entered from today's northern Germany and Denmark (Fig 2, Scenario a), it remains unclear how and where the EHG component entered Scandinavia (Fig 2, Scenarios b, c and/or d). The EHG-related migration likely took place after the migration of WHGs from the south as the earliest eastern-associated pressure blade finds postdate the southwestern-associated direct blade finds in Scandinavia (S1 Text). Two migrations with admixture at different time-periods would generate a genetic gradient with the highest contribution of a source close to its geographic region of entry. The observed genetic pattern is consistent with a migration of the EHGs from the northeast moving southwards along the ice-free Norwegian Atlantic coast where the two groups started mixing (Fig 2, Scenarios a and b), which would cause more EHG ancestry in western SHGs. If the EHG migration had crossed the Baltic Sea into Scandinavia, where it would meet and mix with a WHG-like population (Fig 2, combination of Scenarios a and c), a gradient with most EHG ancestry in eastern SHGs would have been created—exactly opposite to the observed pattern. A similar pattern would be expected if the EHG migration went around the Baltic Sea along current day's Finnish west coast and down via today's Swedish east coast (scenario not depicted in Fig 2). An EHG migration along the southern Baltic coast (Fig 2, Scenarios a and d) should cause a related pattern to a crossing of the Baltic Sea with more EHG ancestry in central and eastern SHGs. Furthermore, such a scenario would likely also make the Latvian Mesolithic hunter-gatherers the group with most EHG ancestry, which is in stark contrast to the empirical data in which the Latvian group shows the lowest proportion of EHG ancestry (33.7% ± 4.7%), and also not consistent with chronology, as the dated settlements east of the Baltic Sea are younger than the early settlements in Scandinavia (S1 Text). Thus, the only scenario consistent with both genetic and archaeological data is a migration of a WHG-related group migrating into Scandinavia from the south, followed by an EHG-related group migrating to Scandinavia from the northeast along the Norwegian Atlantic coast. Notably, such a migration along the Norwegian coast could have been facilitated by the use of the more specialized pressure blade technique (S1 Text) [12,14]. The individuals sequenced here postdate these migrations, but a genetic east-west gradient would be maintained over time in Scandinavia and only additional large-scale migrations from different sources would alter this pattern. This observation is important as the geographic pattern still holds without the chronologically much younger Steigen individual, which might represent local continuity or later migrations into north-western Scandinavia from the east.
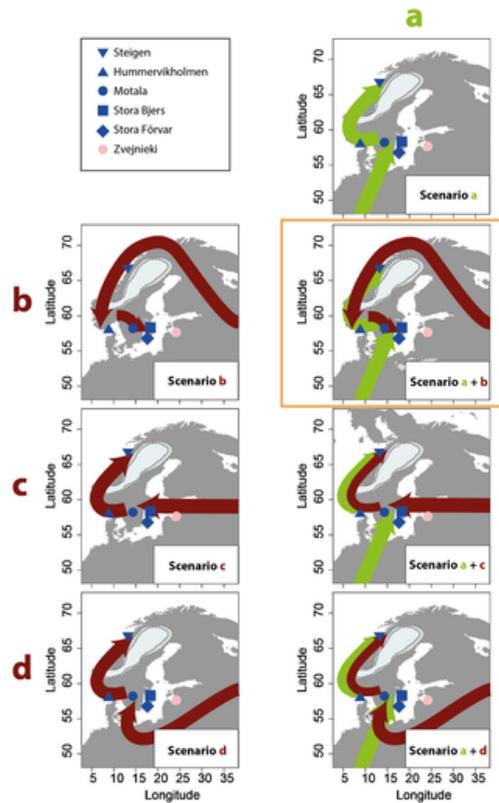
Fig 2. Migration scenarios into postglacial Scandinavia.
Maps showing potential migration routes into Scandinavia. Scenario (a) shows a migration related to the Ahrensburgian tradition from the south (S1 Text). Scenarios (b), (c), and (d) show different possible routes into Scandinavia for the EHG ancestry. The scenarios are discussed in the text and the scenario most consistent with genetic data and stone tools is a combination of routes (a) and (b). All maps were plotted using the R package rworldmap [28]. EHG, eastern hunter-gatherer.
https://doi.org/10.1371/journal.pbio.2003703.g002

Interestingly, stable nitrogen and carbon isotope analysis of northern and western SHGs revealed an extreme marine diet, suggesting a pronounced maritime subsistence, in contrast to the more mixed terrestrial and aquatic diet of eastern and central SHGs (S1 Text). Mobility is difficult to trace based solely on carbon and nitrogen isotope data; however, the patterns are consistent with a migration along the Norwegian Atlantic coast relying on local resources.

Genetic diversity in Mesolithic Scandinavia

By sequencing complete ancient genomes, we can compute unbiased estimates of genetic diversity, which are informative of past population sizes and population history. Here, we restrict the analysis to WHGs and SHGs because only SNP capture data are available for EHGs (S7 Text). In modern-day Europe, there is greater genetic diversity in the south compared to the north. During the Mesolithic period, by contrast, we find lower levels of runs of homozygosity (RoH) (Fig 3A) and linkage disequilibrium (LD) (Fig 3B) in SHGs compared to WHGs (represented by Loschbour and Bichon [18,35]). By using a multiple sequentially Markovian coalescent (MSMC) approach [44] for the high-coverage, high-quality genome of SF12, we find that right before the SF12 individual lived, the effective population size of SHGs was similar to that of WHGs (Fig 3C). At the time of the LGM and back to approximately 50,000 years ago, both the WHGs and SHGs go through a bottleneck, but the ancestors of SHGs retained a greater effective population size in contrast to the ancestors of WHGs who went through a more severe bottleneck (Fig 2C), which is consistent across 100 bootstrap replicates (S2 Fig). These differences in effective population size estimates may be attributed to the admixture in SHGs as migration events can have delayed effects on estimates of effective population size over time [45]. Around 50,000–70,000 years ago, the effective population sizes of the ancestors of SHGs, WHGs, Neolithic groups (represented by Stuttgart [18]), and Paleolithic Eurasians (represented by Ust-Ishim [38]) align, suggesting that these diverse groups all trace their ancestry back to a common ancestral group, which likely represents the early migrants out of Africa.
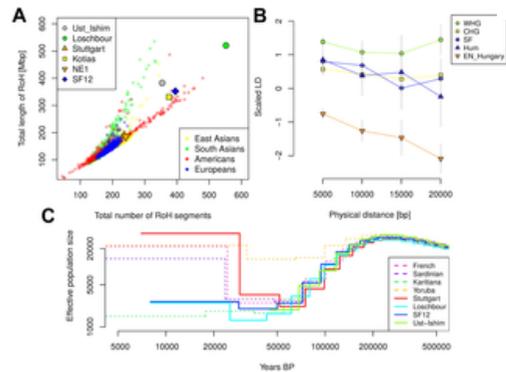
Fig 3. Genetic diversity in prehistoric Europe.
(A) RoH for the six prehistoric humans that have been sequenced to >15× genome coverage, (Kotias is a hunter-gatherer from the Caucasus region [35], NE1 is an early Neolithic individual from modern-day Hungary [27], the other individuals are described in the text), compared to all modern-day non African individuals from the 1000 Genomes Project [32]. (B) LD decay for five prehistoric populations each represented by two individuals (eastern SHGs: SF [SF9 and SF12], western SHGs: Hum [Hum1 and Hum2], CHGs [35]: [Kotias and Satsurblia], WHGs [18,35] [Loschbour and Bichon], and early Neolithic Hungarians [27]: EN_Hungary [NE1 and NE6]). LD was scaled in each distance bin by using the LD for two modern populations [32] as 0 (TSI) and as 1 (PEL). LD was calculated from the covariance of derived allele frequencies of two haploid individuals per population (S7 Text). Error bars show two standard errors estimated during 100 bootstraps across SNP pairs. (C) Effective population size over time as inferred by PSMC' [44] for four prehistoric humans with high genome coverage. The dashed lines show the effective population sizes for selected modern-day populations. All curves for prehistoric individuals were shifted along the x-axis according to their radiocarbon date. S4 Fig. shows 100 bootstrap replicates per individual. Data

Adaptation to high-latitude environments

With the aim of detecting signs of adaptation to high-latitude environments and selection during and after the Mesolithic period, we employed two different approaches that utilize the Mesolithic genomic data. In the first approach, we assumed that SHGs adapted to high-latitude environments of low temperatures and seasonally low levels of light, and searched for gene variants that carried over to modern-day people in northern Europe. Modern-day northern Europeans trace limited amounts of genetic material back to the SHGs (due to the many additional migrations during later periods), and any genomic region that displays extraordinary genetic continuity would be a strong candidate for adaptation in people living in northern Europe across time. We designed a statistic, D_{sel} (S9 Text), that captures this specific signal and scanned the whole genome for gene variants that show strong continuity (little differentiation) between SHGs and modern-day northern Europeans while exhibiting large differentiation to modern-day southern European populations [46] (Fig 4A; S9 Text). Six of the top 10 SNPs with greatest D_{sel} values were located in the *TMEM131* gene that has been found to be associated with physical performance [47], which could make it part of the physiological adaptation to cold [48]. This genomic region was more than 200 kbp (kilo base pairs) long and showed the strongest haplotypic differentiation between modern-day Tuscan individuals (TSIs) and modern-day Finnish individuals (FINs) across the genome (S9 Text). The particular haplotype was relatively common in SHGs, it is even more common among today's Finnish population (S9 Text) and showed a strong signal of local adaptation (S9 Text). Other top hits included genes associated with a wide range of metabolic, cardiovascular, and developmental and psychological traits (S9 Text) potentially linked to physiological adaptation to cold environments [48].
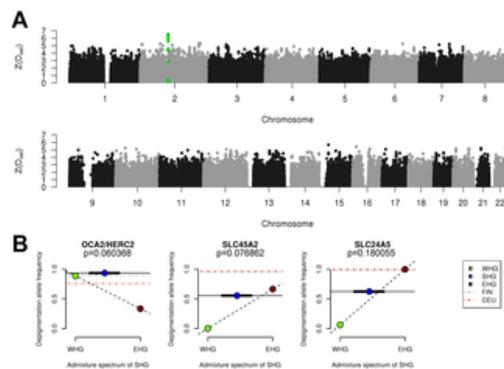
Fig 4. Adaptation to high-latitude environments.
(A) Plot of similarity between Mesolithic allele frequency and FIN allele frequency in contrast to difference to TSI allele frequency using the statistic D_{sel}. The figure shows all positive Z scores representing the number of standard deviations each SNP deviates from the mean. The green-highlighted SNPs are all located in the *TMEM131* gene. The plot was made with

qqman [49]. (B) Derived allele frequencies for three pigmentation-associated SNPs (*SLC24A5*, *SLC45A2*, which are associated with skin pigmentation, and *OCA2/HERC2*, which is associated with eye pigmentation). The dashed line connecting EHG and WHG represents potential allele frequencies if SHG were a linear combination of admixture between EHG and WHG. The solid horizontal line represents the derived allele frequency in SHG. The blue symbols representing SHGs were set on the average genome-wide WHG and EHG mixture proportion (on x-axis) across all SHGs, and the thick black line represents the minimum and maximum admixture proportions across all SHGs. Dashed horizontal lines represent modern European populations (CEU). The *p*-values were estimated from simulations of SHG allele frequencies based on their genome-wide ancestry proportions (S9 Text). Data shown in this figure can be found in S1 Data. CEU, Utah residents with Central European ancestry; EHG, eastern hunter-gatherer; FIN, modern-day Finnish individual; SHG, Scandinavian hunter-gatherer; TSI, modern-day Tuscan individual; WHG, western hunter-gatherer.
https://doi.org/10.1371/journal.pbio.2003703.g004

In addition to performing this genome-wide scan, we studied the allele frequencies in three pigmentation genes (*SLC24A5*, *SLC45A2*, which have a strong effect on skin pigmentation, and *OCA2/HERC2*, which has a strong effect on eye pigmentation) in which the derived alleles are virtually fixed in northern Europeans today. The differences in allele frequencies of those three loci are among the highest between human populations, suggesting that selection was driving the differences in eye color, skin, and hair pigmentation as part of the adaptation to different environments [50–53]. All of the depigmentation variants at these three genes are in high frequency in SHGs in contrast to both WHGs and EHGs (Fig 4B). We conduct neutral simulations of the allele frequencies in an admixed SHG population to estimate *p*-values for observing these allele frequencies without selection (S9 Text). The *p*-values for all three SNPs are lower than 0.2; the combined *p*-value [54] for all three pigmentation SNPs is 0.028. Therefore, the unique configuration of the SHGs is not fully explained by the fact that SHGs are a mixture of EHGs and WHGs, but could rather be explained by a continued increase of the allele frequencies after the admixture event, likely caused by adaptation to high-latitude environments [50,52].

Conclusion

By combining information from climate modeling, archaeology, and Mesolithic human genomes, we were able to reveal the complexity of the early migration patterns into Scandinavia and human adaptation to high-latitude environments. We disentangled two migration routes and linked them to particular archaeological patterns. We also demonstrated greater genetic diversity in Mesolithic northern Europe compared to southern and central Europe—in contrast to modern-day patterns—and showed that many genetic variants that were common during the Mesolithic period have been lost today. These findings reiterate the importance of human migration for dispersal of novel technology in human prehistory [14–20,27,40,55–58].

Materials and methods

Sample preparation

Genomic sequence data were generated from teeth and bone samples belonging to seven (eight, including SF13) Mesolithic SHGs (S1 Text). A detailed description on the archaeological background of the samples as well as post-LGM Scandinavia can be found in S1 Text. Additional libraries were sequenced for two previously published Neolithic hunter-gatherers, Ajvide58 and Ajvide70 [17] (S2 Text). All samples were prepared in the dedicated aDNA facilities at Uppsala University (SF9, SF11, SF12, SF13, SBj, Hum1, Hum2, Ajvide58, Ajvide70) and at Stockholm University (Steigen).

DNA extraction and library building

Bones and teeth were decontaminated prior to analysis by wiping them with a 1% Sodiumhypoclorite solution and DNA-free water. Furthermore, all surfaces were UV irradiated (6 J/cm^2 at 254 nm). After removing 1 millimeter of the surface, approximately 30–300 mg of bone was powderized and DNA was extracted following silica-based methods as in [59] with modifications as in [57,60] or as in [61] and eluted in 25–110 µl of EB buffer. Between one and 16 extractions were made from each sample and one extraction blank with water instead of bone powder was included per six to 10 extracts. Blanks were carried along the whole process until quantitative PCR (qPCR) and/or PCR and subsequent quantification.

DNA libraries were prepared using 20 µl of extract, with blunt-end ligation coupled with P5 and P7 adapters and indexes as described in [57,62]. From each extract one to five double stranded libraries were built. Because aDNA is already fragmented, the shearing step was omitted from the protocol. Library blank controls, including water as well as extraction blanks, were carried along during every step of library preparation. In order to determine the optimal number of PCR cycles for library amplification, qPCR was performed. Each reaction was prepared in a total volume of 25 µl, containing 1 µl of DNA library, 1X MaximaSYBRGreen mastermix, and 200 nM each of IS7 and IS8 [62] reactions were set up in duplicates. Each blunt-end library was amplified in four to 12 replicates with one negative PCR control per index-PCR. The amplification reactions had a total volume of 25 µl, with 3 µl DNA library, and the following in final concentrations: 1 X AmpliTaq Gold Buffer, 2.5 mM MgCl$_2$, 250 µM of each dNTP, 2.5 U AmpliTaq Gold (Thermo Fisher Scientific, Waltham, MA), and 200 nM each of the IS4 primer and index primer [62]. PCR was done with the following conditions: an activation step at 94°C for 10 min followed by 10–16 cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s, and a final elongation step of 72°C for 10 min. For each library, four amplifications with the same indexing primer were pooled and purified with AMPure XP beads (Agencourt; Beckman Coulter, Brea, CA). The quality and quantity of libraries was checked using Tapestation or BioAnalyzer using the High Sensitivity Kit (Agilent Technologies, Cary, NC). None of the blanks showed any presence of DNA comparable to that of a sample and were therefore not further analyzed. For initial screening, 10–20 libraries were pooled at equimolar concentrations for sequencing on an Illumina HiSeq 2500 using v.4 chemistry, and 125 bp paired-end reads or HiSeqX, 150 bp paired-end length using v2.5 chemistry at the SNP & SEQ Technology Platforms at Uppsala University and Stockholm University. After evaluation of factors such as clonality, proportion of human DNA, and genomic coverage samples were selected for resequencing, aiming to yield as high coverage as possible for each library.

Generation of a high-coverage UDG-treated genome

Based on the results of the non-damage-repair sequencing, the SF12 individual was selected for large-scale sequencing in order to generate a high-coverage genome of high quality where damages had been repaired using UDG. In addition to the 15 extracts previously prepared and used for non-damage repair libraries, another 111 extracts were made based on a variety of silica-based methods [27,57,59,60]. From these 126 extracts, a total of 258 damage-repaired double-stranded libraries were built for Illumina sequencing platforms. Libraries were built as above, except a DNA repair step in which UDG and endonuclease VIII or USER enzyme (NEB) treatment was included in order to remove deaminated cytosines [63]. qPCR was performed in order to quantify the number of molecules and the optimal number of PCR cycles prior to amplification for each DNA library. Furthermore, this step included extraction blanks, library blanks, and amplification blanks to monitor potential contamination. All of these negative controls showed an optimal cycle of amplification significantly higher to those of our aDNA libraries (>10 cycles) and they were thus deemed as negative. Our experimental results show minimal levels of contamination, which is in concordance with mt DNA and X chromosome estimates of contamination (see S4 Text and Table 1). Each reaction was done in a total volume of 25 µl, containing 1 µl of DNA library, 1 X MaximaSYBRGreen mastermix (Thermo Fisher Scientific), and 200 nM each of IS7 and IS8 [62], reactions were set up in duplicate. The PCRs were set up using a similar system as for the nondamage repair samples (in quadruplicates that were pooled prior to cleanup of the PCR products), except for using AccuPrime DNA polymerase (Thermo Fisher Scientific) instead

of AmpliTaqGold (Thermo Fisher Scientific) and the following PCR conditions: an activation step at 95°C for 2 min followed by 10–16 cycles of 95°C for 15 s, 60°C for 30 s, and 68°C for 1 min, and a final elongation step of 68°C for 5 min. Blank controls, including water as well as extraction blanks, were carried out during every step of library preparation. Amplified libraries were pooled, cleaned, quantified, and sequenced in the same manner as non-damage repaired libraries. A small proportion of the libraries (*n* = 14) were also subjected to whole genome capture (WGC) using European MYbaits from MYcroarray, and following the manufacturers protocol as done in [64].

In order to sequence libraries to depletion, two to eight libraries were pooled together and sequenced until reaching a clonality of >50%; if sequencing was halted before reaching that clonality level, it was either because the library was classified as unproductive based on the genome coverage generated, or that the sequencing goal (>55 × coverage) was already reached and further sequencing was deemed unnecessary. Sequencing was performed as above.

Bioinformatic data processing and authentication

Paired-end reads were merged using MergeReadsFastQ_cc.py [65]; if an overlap of at least 11 base pairs was found, the base qualities were added together and any remaining adapters were trimmed. Merged reads were then mapped single-ended with bwa aln 0.7.13 [66] to the human reference genome (build 36 and 37) using the following nondefault parameters: seeds disabled -l 16500 -n 0.01 -o 2 [17,18]. To remove PCR duplicates, reads with identical start and end positions were collapsed using a modified version, to ensure random choice of bases, of FilterUniqSAMCons_cc.py [65]. Reads with less than 10% mismatches to the human reference genome, reads longer than 35 base pairs, and reads with mapping quality higher than 30 were used to estimate contamination.

The genetic data obtained from the two bone elements SF9 and SF13 showed extremely high similarities, which suggested that the two individuals were related. Using READ [67], a tool to estimate kin-relationship from aDNA, SF9 and SF13 were classified as either identical twins or the same individual. Therefore, we merged the genetic data for both individuals and refer to the merged individual as SF9 throughout the genetic analysis.

All data show damage patterns indicative of authentic aDNA (S3 Text). Contamination was estimated using three different sources of data: (1) the mt genome [68], (2) the X chromosome if the individual was male [69,70], and (3) the autosomes [71]. The data mapping to the human genome can be considered largely endogenous, as the contamination estimates were low across all three methods (S4 Text).

Analysis of demographic history

Most population genomic analyses require a set of reference data for comparison. We compiled three different data sets from the literature and merged them with the data from ancient individuals (S6 Text). The three reference SNP panels were as follows:

> The Human Origins genotype data set of 594,924 SNPs genotyped in 2,404 modern individuals from 203 populations [18,39].

> A panel of 1,055,209 autosomal SNPs, which were captured in a set of ancient individuals by Mathieson et al. [20].

> To reduce the potential effect of ascertainment bias on SNP array data and of cytosine deamination on transition SNPs, we also ascertained 1,797,398 transversion SNPs with a minor allele frequency of at least 10% (to avoid the effect of Eurasian admixture into Yorubans) in Yorubans of the 1000 Genomes Project [32]. Those SNPs were extracted using vcftools [72].

These data sets were merged with ancient individuals of less than 15× genome coverage using the following approach: for each SNP site, a random read covering that site with minimum mapping quality 30 was drawn (using samtools 0.1.19 mpileup [73]) and its allele was assumed to be homozygous in the ancient individual. Transition sites were coded as missing data for individuals that were not UDG-treated, and SNPs showing additional alleles or indels in the ancient individuals were excluded from the data.

The six high-coverage ancient individuals (SF12, NE1 [27], Kotias [35], Loschbour [18], Stuttgart [18], and Ust-Ishim [38]) used in this study were treated differently, as we generated diploid genotype calls for them. First, the base qualities of all Ts in the first five base pairs of each read as well as all As in the last five base pairs were set to 2. We then used Picard [74] to add read groups to the files. Indel realignment was conducted with GATK 3.5.0 [66] using indels identified in phase 1 of the 1000 Genomes Project as reference [32]. Finally, GATK's UnifiedGenotyper was used to call diploid genotypes with the parameters -stand_call_conf 50.0, -stand_emit_conf 50.0, -mbq 30, -contamination 0.02, and—output_mode EMIT_ALL_SITES using dbSNP version 142 as known SNPs. SNP sites from the reference data sets were extracted from the VCF files using vcftools [72] if they were not marked as low-quality calls. Plink 1.9 [75,76] was used to merge the different data sets.

We performed PCA to characterize the genetic affinities of the ancient Scandinavian genomes to previously published ancient and modern genetic data. PCA was conducted on 42 present-day west Eurasian populations from the Human Origins data set [18,39] using *smartpca* [77] with numoutlieriter: 0 and lsqproject: YES options. A total of 59 ancient genomes (52 previously published and 7 reported here) (S6 Text) were projected into the reference PCA space and computed from the genotypes of modern individuals. For all individuals, a single allele was selected randomly—making the data set fully homozygous. The result was plotted using the *ploteig* program of EIGENSOFT [77] with the–x and–k options.

D and f statistics

popstats [78] was used to calculate D statistics to test deviations from a tree-like population topology of the shape $((A,B);(X,Y))$ [39]. Standard errors were calculated using a weighted block jackknife of 0.5 Mbp. The tree topologies are balanced at zero, indicating no recent interactions between the test populations. Significant deviations from zero indicate a deviation from the proposed tree topology depending on the value. Positive values indicate an excess of shared alleles between A and X or B and Y, whereas negative values indicate more shared alleles between B and X or A and Y. Using an outgroup as population A limits the test results to depend on the recent relationships between B and Y (if positive) or B and X (if negative). Here, we used high-coverage Mota [37], Yoruba [32], and Chimp genome as (A) outgroups. *popstats* [78] was used to calculate f_4 statistics in order to estimate shared drift between groups. Standard errors and Z scores for f_4 statistics were estimated using a weighted block jackknife (Fig 1C).

Model-based clustering

A model-based clustering algorithm, implemented in the *ADMIXTURE* software [79], was used to estimate ancestry components and to cluster individuals. *ADMIXTURE* was conducted on the Human Origins data set [18,39], which was merged with the ancient individuals as described above. Data was pseudo-haploidized by randomly selecting one allele at each heterozygous site of present-day individuals. Finally, the data set was filtered for LD using PLINK [75,76] with parameters (--indep-pairwise 200 25 0.4), this retained 289,504 SNPs. *ADMIXTURE* was run in 50 replicates with different random seeds for ancestral clusters from K = 2 to K = 20. Common signals between independent runs for each K were identified using the LargeKGreedy algorithm of *CLUMPP* [80]. Clustering was visualized using rworldmap, ggplot2, SDMTools, and RColorBrewer packages of GNU R version 3.3.0. Starting from K = 3, when the modern samples split up into an African and eastern and western Eurasian clusters, the Mesolithic Scandinavians

from Norway show slightly higher proportions of the Eastern cluster than Swedish Mesolithic individuals. This pattern continues to develop across higher values of K and it is consistent with the higher Eastern affinities of the Norwegian samples seen in the PCA and *D-* and f4 statistics. The results for all Ks are shown in S1 Fig.

In addition to *ADMIXTURE*, we assessed the admixture patterns in Mesolithic Scandinavians using a set of methods implemented in *ADMIXTOOLS* [39], *qpWave* [81], and *qpAdm* [19]. Both methods are based on f_4 statistics, which relate a set of test populations to a set of outgroups in different distances from the potential source populations. We used the following set of outgroup populations from the Human Origins data set: Ami_Coriell, Biaka, Bougainville, Chukchi, Eskimo_Naukan, Han, Karitiana, Kharia, and Onge. We first used *qpWave* to test the number of source populations for Mesolithic west Eurasians (WHG). *qpWave* calculates a set of statistics X(u,v) = f_4(u0, u; v0, v) where u0 and v0 are populations from the sets of test populations L and outgroups R, respectively. To avoid having more test populations than outgroups, we built four groups consisting of (1) genetically western and central hunter-gatherers (Bichon, Loschbour, KO1, LaBrana), (2) EHGs (UzOO74/I0061, SVP44/I0124, UzOO40/I0211), (3) Norwegian hunter-gathers (Hum1, Hum2, Steigen), and (4) Swedish hunter-gatherers (individuals from Motala and Mesolithic Gotland). qpWave tests the rank of the matrix of all X(u,v) statistics. If the matrix has rank m, the test populations can be assumed to be related to at least m + 1 "waves" of ancestry, which are differently related to the outgroups. A rank of 0 is rejected in our case (*p* = 3.13e-81), whereas a rank of 1 is consistent with the data (*p* = 0.699). Haak et al. [19] already showed, using the same approach, that WHG and EHG descend from at least two sources (confirmed with our data as rank 0 is rejected with *p* = 1.66e-86, whereas rank 1 is consistent with the data), and adding individuals from Motala does not change these observations. Therefore, we conclude that European Mesolithic populations, including Swedish and Norwegian Mesolithic individuals, have at least two source populations.

We then used *qpAdm* to model Mesolithic Scandinavian individuals as a 2-way admixture of WHG and EHG. *qpAdm* was run separately for each Scandinavian individual x, setting T = x as target and S = (EHG, WHG) as sources. The general approach of *qpAdm* is related to *qpWave*: target and source are used as L (with T being the base population), and f_4 statistics with outgroups from R (same as above) are calculated. The rank of the resulting matrix is then set to the number of sources minus one, which allows to estimate the admixture contributions from each population in S to T. The results are shown in Fig 1. We calculate a Z score for the difference between Norwegian and Swedish SHG as $Z = \frac{a_{nor} - a_{swe}}{\sqrt{SE_{nor}^2 + SE_{swe}^2}}$ where *a* are the ancestry estimates and *SE* are the respective block-jackknife estimates of the standard errors.

RoH

Heterozygosity is a measurement for general population diversity and its effective population size. Analyzing the extent of homozygous segments across the genome can also give us a temporal perspective on the effective population sizes. Many short segments of homozygous SNPs can be connected to historically small population sizes, whereas an excess of long RoH suggests recent inbreeding. We restricted this analysis to the six high-coverage individuals (SF12, NE1, Kotias, Loschbour, Stuttgart, Ust-Ishim) for which we obtained diploid genotype calls and we compared them to modern individuals from the 1000 Genomes Project. The length and number of RoH were estimated using Plink 1.9 [75,76] and the parameters--homozyg-density 50,--homozyg-gap 100,--homozyg-kb 500,--homozyg-snp 100,--homozyg-window-het 1,--homozyg-window-snp 100,--homozyg-window-threshold 0.05, and--homozyg-window-missing 20. The results are shown in Fig 3A.

LD

Similar to RoH, the decay of LD harbors information on the demographic history of a population. Long-distance LD can be caused by a low effective population size and past bottlenecks. Calculating LD for aDNA data is challenging, as the low amounts of authentic DNA usually just yields haploid allele calls with unknown phase. In order to estimate LD decay for ancient populations, we first combine two haploid ancient individuals to a pseudo-diploid individual (similar to the approach chosen for conditional nucleotide diversity, S7 Text). Next, we bin SNP pairs by distance (bin size 5 kb) and then calculated the covariance of derived allele frequencies (0, 0.5, or 1.0) for each bin. This way, we do not need phase information to calculate LD decay because we do not consider multilocus haplotypes, which is similar to the approach taken by ROLLOFF [39,82] and ALDER [83] to date admixture events based on admixture LD decay. For Fig 3B, we used two modern 1000 Genomes Project populations to scale the LD per bin. The LD between two randomly chosen PELs (modern-day Peruvian individuals) was set to 1 and the LD between two randomly chosen TSIs was set to 0. This approach is used to obtain a relative scale for the ancient populations, and we caution against a direct interpretation of the differences to modern populations because technical differences in the modern data (e.g., SNP calling or imputation) may have substantial effects.

Effective population size

We are using MSMC's implementation of PSMC' [44] to infer effective population sizes over time from single high-coverage genomes. We restrict this analysis to UDG-treated individuals (SF12, Loschbour, Stuttgart, Ust-Ishim) as postmortem damage would cause an excess of false heterozygous transition sites. Input files were prepared using scripts provided with the release of MSMC (https://github.com/stschiff/msmc-tools) and MSMC was run with the nondefault parameters--fixedRecombination and -r 0.88 in order to set the ratio of recombination to mutation rate to a realistic level for humans. We also estimate effective population size for six high-coverage modern genomes [84] (Fig 3C). We plot the effective population size assuming a mutation rate of 1.25×10^{-8} and a generation time of 30 years. The curves for ancient individuals were shifted based on their average C14 date. Additionally, we used multihetsep_bootstrap.py to generate 100 bootstraps per individual. The results are shown in S4 Fig.

Detecting adaptation to high-latitude environments

We scanned the genomes for SNPs with similar allele frequencies in Mesolithic and modern-day northern Europeans and contrasted it to a modern-day population from southern latitudes. Pooling all Mesolithic Scandinavians together, we obtain an allele frequency estimate for SHGs, which is compared to FINs and TSIs from the 1000 Genomes Project [32]. We use the Finnish population as representatives of modern-day northern Europeans (this sample contains the largest number of sequenced genomes from a northern European population). Tuscans are used as an alternative population, who also trace some ancestry to Mesolithic populations, but who do not trace their ancestry to groups that lived at northern latitudes in the last 7,000–9,000 years. Our approach is similar to PBS [85] and inspired by DAnc [46]. For each SNP, we calculated the statistic D_{sel}, comparing the allele frequencies between one ancestral and two modern populations:

$$D_{sel} = |DAF_{SHG} - DAF_{TSI}| - |DAF_{SHG} - DAF_{FIN}|$$

This scan was performed on all transversion SNPs extracted from the 1000 Genomes Project data. Only sites with a high-confidence ancestral allele in the human ancestor (as used by the 1000 Genomes Project [32]) and with coverage for at least six ancient Scandinavians were included in the computation. More information can be found in S9 Text.

Supporting information

S1 Table. Allelic states at phenotypically relevant SNPs (see also S8 Text).

https://doi.org/10.1371/journal.pbio.2003703.s001
(XLSX)

S1 Fig. Complete results of unsupervised ADMIXTURE K = 2 to K = 20.

https://doi.org/10.1371/journal.pbio.2003703.s002
(PDF)

S2 Fig. Distributions of all individual D statistics.

Positive D statistics for WHGs are all involving the low quality and slightly contaminated SF11 as Swedish SHGs. Data shown in this figure can be found in S1 Data. SHG, Scandinavian hunter-gatherer; WHG, western hunter-gatherer.
https://doi.org/10.1371/journal.pbio.2003703.s003
(PDF)

S3 Fig. f_4 statistics (see also Fig 1C) plotted against chronological age and longitude of the samples.

Pink symbols indicate Latvian samples. Data shown in this figure can be found in S1 Data.
https://doi.org/10.1371/journal.pbio.2003703.s004
(PDF)

S4 Fig. MSMC results with 100 bootstraps.

Data shown in this figure can be found in S1 Data. MSMC, multiple sequentially Markovian coalescent.
https://doi.org/10.1371/journal.pbio.2003703.s005
(PDF)

S1 Text. Archaeological background.

https://doi.org/10.1371/journal.pbio.2003703.s006
(PDF)

S2 Text. DNA sample preparation.

https://doi.org/10.1371/journal.pbio.2003703.s007
(PDF)

S3 Text. Processing of NGS data.

NGS, next-generation sequencing.
https://doi.org/10.1371/journal.pbio.2003703.s008

(PDF)

S4 Text. Estimates of contamination.

https://doi.org/10.1371/journal.pbio.2003703.s009
(PDF)

S5 Text. Uniparental markers.

https://doi.org/10.1371/journal.pbio.2003703.s010
(PDF)

S6 Text. Basic population genomic analysis.

https://doi.org/10.1371/journal.pbio.2003703.s011
(PDF)

S7 Text. Diversity estimates.

https://doi.org/10.1371/journal.pbio.2003703.s012
(PDF)

S8 Text. Functional variation in ancient samples.

https://doi.org/10.1371/journal.pbio.2003703.s013
(PDF)

S9 Text. Adaptation to high-latitude climates.

https://doi.org/10.1371/journal.pbio.2003703.s014
(PDF)

S1 Data. Data shown in figures.

https://doi.org/10.1371/journal.pbio.2003703.s015
(XLSX)

Acknowledgments

References

1. Andersen B, Borns HW. The ice age world: an introduction to quaternary history and research with emphasis on North America and Northern Europe during the last 2.5 million years. Oslo [u.a].: Scandinavian Univ. Press; 1994.

2. Parducci L, Jorgensen T, Tollefsrud MM, Elverland E, Alm T, Fontana SL, et al. Glacial Survival of Boreal Trees in Northern Scandinavia. Science. 2012;335: 1083–1086. pmid:22383845
   View Article • PubMed/NCBI • Google Scholar

3. François O, Blum MGB, Jakobsson M, Rosenberg NA. Demographic History of European Populations of Arabidopsis thaliana. PLoS Genet. 2008;4: e1000075. pmid:18483550
   View Article • PubMed/NCBI • Google Scholar

4. Hewitt G. The genetic legacy of the Quaternary ice ages. Nature. 2000;405: 907–913. pmid:10879524
   View Article • PubMed/NCBI • Google Scholar

5. Xenikoudakis G, Ersmark E, Tison J-L, Waits L, Kindberg J, Swenson JE, et al. Consequences of a demographic bottleneck on genetic structure and variation in the Scandinavian brown bear. Mol Ecol. 2015;24: 3441–3454. pmid:26042479
   View Article • PubMed/NCBI • Google Scholar

6. Riede F. The resettlement of Northern Europe. In: Cummings V, Jordan P, Zvelebil M, editors. The Oxford Handbook of the Archaeology and Anthropology of Hunter-Gatherers. Oxford University Press; 2014. pp.556–581.

7. Kankaanpää J, Rankama T. Fast or slow pioneers? A view from northern Lapland. In: Riede F, Tallavaara M, editors. Lateglacial and Postglacial Pioneers in Northern Europe, BAR International Series. Oxford: Oxford Archaeopress; 2014. pp. 147–154.

8. Breivik HM. Palaeo-oceanographic development and human adaptive strategies in the Pleistocene–Holocene transition: A study from the Norwegian coast. The Holocene. 2014;24: 1478–1490.
   View Article • PubMed/NCBI • Google Scholar

9. Stroeven AP, Hättestrand C, Kleman J, Heyman J, Fabel D, Fredin O, et al. Deglaciation of Fennoscandia. Quaternary Science Reviews. 2016;147: 91–121.
   View Article • PubMed/NCBI • Google Scholar

10. Hughes AL, Gyllencreutz R, Lohne ØS, Mangerud J, Svendsen JI. The last Eurasian ice sheets–a chronological database and time-slice reconstruction, DATED-1. Boreas. 2016;45: 1–45.
View Article • PubMed/NCBI • Google Scholar

11. Bjerck HB, Breivik HM, Piana EL, Zangrando AF. Exploring the role of pinnipeds in the human colonization of the seascapes of Patagonia and Scandinavia. In: Bjerck HB, Breivik HM, Fretheim SE, Piana EL, Skar B, Tivoli AM, et al., editors. Equinox Sheffield; 2016. pp. 53–74.

12. Sørensen M, Rankama T, Kankaanpää J, Knutsson K, Knutsson H, Melvold S, et al. The First Eastern Migrations of People and Knowledge into Scandinavia: Evidence from Studies of Mesolithic Technology, 9th-8th Millennium BC. Norwegian Archaeological Review. 2013;46: 19–56.
View Article • PubMed/NCBI • Google Scholar

13. Knutsson H, Knutsson K, Molin F, Zetterlund P. From flint to quartz: Organization of lithic technology in relation to raw material availability during the pioneer process of Scandinavia. Quaternary International. 2016;424: 32–57.
View Article • PubMed/NCBI • Google Scholar

14. Damlien H. Eastern pioneers in westernmost territories? Current perspectives on Mesolithic hunter–gatherer large-scale interaction and migration within Northern Eurasia. Quaternary International. 2014;419: 5–16.
View Article • PubMed/NCBI • Google Scholar

15. Knutsson K, Knutsson H. The postglacial colonization of humans, fauna and plants in northern Sweden. Arkeologi i norr. 2012;13: 1–28.
View Article • PubMed/NCBI • Google Scholar

16. Skoglund P, Malmström H, Raghavan M, Storå J, Hall P, Willerslev E, et al. Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. Science. 2012;336: 466–469. pmid:22539720
View Article • PubMed/NCBI • Google Scholar

17. Skoglund P, Malmström H, Omrak A, Raghavan M, Valdiosera C, Günther T, et al. Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. Science. 2014;344: 747–750. pmid:24762536
View Article • PubMed/NCBI • Google Scholar

18. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 2014;513: 409–413. pmid:25230663
View Article • PubMed/NCBI • Google Scholar

19. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015; 522:207–211. pmid:25731166
View Article • PubMed/NCBI • Google Scholar

20. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, et al. Genome-wide patterns of selection in 230 ancient Eurasians. Nature. 2015;528: 499–503. pmid:26595274
View Article • PubMed/NCBI • Google Scholar

21. Olalde I, Allentoft ME, Sánchez-Quinto F, Santpere G, Chiang CWK, DeGiorgio M, et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature. 2014;507: 225–228. pmid:24463515
View Article • PubMed/NCBI • Google Scholar

22. Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature. 2014;505: 87–91. pmid:24256729
View Article • PubMed/NCBI • Google Scholar

23. Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, et al. The genetic history of Ice Age Europe. Nature. 2016;534: 200–205. pmid:27135931
View Article • PubMed/NCBI • Google Scholar

24. Mittnik A, Wang C-C, Pfrengle S, Daubaras M, Zariņa G, Hallgren F, et al. The Genetic History of Northern Europe. bioRxiv. 2017;
View Article • PubMed/NCBI • Google Scholar

25. Cummings V, Jordan P, Zvelebil M, editors. The Oxford Handbook of the Archaeology and Anthropology of Hunter-Gatherers. Oxford University Press; 2014.

26. Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, et al. Population genomics of Bronze Age Eurasia. Nature. 2015;522: 167–172. pmid:26062507
View Article • PubMed/NCBI • Google Scholar

27. Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, et al. Genome flux and stasis in a five millennium transect of European prehistory. Nature Communications. 2014;5: 5257. pmid:25334030
View Article • PubMed/NCBI • Google Scholar

28. South A. rworldmap: A New R package for Mapping Global Data. The R Journal. 2011;3: 35–43.
View Article • PubMed/NCBI • Google Scholar

29. Jones ER, Zarina G, Moiseyev V, Lightfoot E, Nigst PR, Manica A, et al. The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. Curr Biol. 2017;27: 576–582. pmid:28162894
View Article • PubMed/NCBI • Google Scholar

30. Sjögren K-G, Ahlström T. Early Mesolithic burials from Bohuslän, western Sweden. Mesolithic burials—Rites, symbols and social organisation of early postglacial communities. Beier & Beran; 2017.

31. Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Pääbo S. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res. 2010;38: e87–e87. pmid:20028723
View Article • PubMed/NCBI • Google Scholar

32. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015;526: 68–74. pmid:26432245
View Article • PubMed/NCBI • Google Scholar

33. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w [1118]; iso-2; iso-3. Fly. 2012;6: 80–92. pmid:22728672
View Article • PubMed/NCBI • Google Scholar

34. Nixon B, Bromfield E, Dun M, Redgrove K, McLaughlin E, Aitken RJ. The role of the molecular chaperone heat shock protein A2 (HSPA2) in regulating human sperm-egg recognition. Asian Journal of Andrology. 2015;17: 568. pmid:25865850
View Article • PubMed/NCBI • Google Scholar

35. Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nature Communications. 2015;6: 8912. pmid:26567969
View Article • PubMed/NCBI • Google Scholar

36. Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspinas A-S, Manica A, Moltke I, et al. Genomic structure in Europeans dating back at least 36,200 years. Science. 2014;346: 1113–1118. pmid:25378462
View Article • PubMed/NCBI • Google Scholar

37. Llorente MG, Jones ER, Eriksson A, Siska V, Arthur KW, Arthur JW, et al. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. Science. 2015;350: 820–822. pmid:26449472
View Article • PubMed/NCBI • Google Scholar

38. Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature. 2014;514: 445–449. pmid:25341783
View Article • PubMed/NCBI • Google Scholar

39. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient Admixture in Human History. Genetics. 2012;192: 1065–1093. pmid:22960212
View Article • PubMed/NCBI • Google Scholar

40. Günther T, Jakobsson M. Genes mirror migrations and cultures in prehistoric Europe-a population genomic perspective. Curr Opin Genet Dev. 2016;41: 115–123. pmid:27685850

View Article • PubMed/NCBI • Google Scholar

41. Petersen EB. The human settlement of southern Scandinavia 12 500–8700 cal BC. In: Barton N, Street M, Terberger T, editors. Humans, Environment and Chronology of the Late Glacial of the North European Plain. Mainz: Römisch-Germanisches Zentral museum; 2009. pp. 89–129.

42. Schmitt L, Larsson S, Burdukiewicz J, Ziker J, Svedhage K, Zamon J, et al. Chronological insights, cultural change, and resource exploitation on the west coast of Sweden during the Late Palaeolithic/Early Mesolithic transition. Oxford Journal of Archaeology. 2009;28: 1–27.
View Article • PubMed/NCBI • Google Scholar

43. Fuglestvedt I. The Ahrensburgian Galta 3 site in SW Norway. Dating, technology and cultural affinity. Acta Archaeologica. 2007;78: 87–110.
View Article • PubMed/NCBI • Google Scholar

44. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. Nature Genetics. 2014;46: 919–925. pmid:24952747
View Article • PubMed/NCBI • Google Scholar

45. Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L. On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? Heredity (Edinb). 2016;116: 362–371. pmid:26647653
View Article • PubMed/NCBI • Google Scholar

46. Key FM, Fu Q, Romagné F, Lachmann M, Andrés AM. Human adaptation and population differentiation in the light of ancient genomes. Nature Communications. 2016;7: 10775. pmid:26988143
View Article • PubMed/NCBI • Google Scholar

47. Vasan RS, Larson MG, Aragam J, Wang TJ, Mitchell GF, Kathiresan S, et al. Genome-wide association of echocardiographic dimensions, brachial artery endothelial function and treadmill exercise responses in the Framingham Heart Study. BMC Medical Genetics. 2007;8: S2. pmid:17903301
View Article • PubMed/NCBI • Google Scholar

48. Makinen TM. Different types of cold adaptation in humans. Frontiers in bioscience. 2010;2: 1047–1067.
View Article • PubMed/NCBI • Google Scholar

49. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. 2014; 005165.
View Article • PubMed/NCBI • Google Scholar

50. Fan S, Hansen MEB, Lo Y, Tishkoff SA. Going global by adapting local: A review of recent human adaptation. Science. 2016;354: 54–59. pmid:27846491
View Article • PubMed/NCBI • Google Scholar

51. Donnelly MP, Paschou P, Grigorenko E, Gurwitz D, Barta C, Lu RB, et al. A global view of the OCA2-HERC2 region and pigmentation. Human Genetics. 2012;131: 683–696. pmid:22065085
View Article • PubMed/NCBI • Google Scholar

52. Jablonski NG, Chaplin G. The colours of humanity: the evolution of pigmentation in the human lineage. Philosophical Transactions of the Royal Society B: Biological Sciences. 2017;372: 20160349. pmid:28533464
View Article • PubMed/NCBI • Google Scholar

53. Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. Tracing the peopling of the world through genomics. Nature. 2017;541: 302–310. pmid:28102248
View Article • PubMed/NCBI • Google Scholar

54. Fisher RA. Statistical methods for research workers. Edinburgh: Oliver and Boyd; 1925.

55. Hofmanová Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Diez-del-Molino D, et al. Early farmers from across Europe directly descended from Neolithic Aegeans. Proceedings of the National Academy of Sciences. 2016; 133:6886–91. pmid:27274049
View Article • PubMed/NCBI • Google Scholar

56. Cassidy LM, Martiniano R, Murphy EM, Teasdale MD, Mallory J, Hartwell B, et al. Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. Proceedings of the National Academy of Sciences. 2016; 113:368–373. pmid:26712024
View Article • PubMed/NCBI • Google Scholar

57. Günther T, Valdiosera C, Malmström H, Ureña I, Rodriguez-Varela R, Sverrisdóttir ÓO, et al. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. Proceedings of the National Academy of Sciences. 2015;112: 11917–11922. pmid:26351665
View Article • PubMed/NCBI • Google Scholar

58. Olalde I, Schroeder H, Sandoval-Velasco M, Vinner L, Lobón I, Ramirez O, et al. A common genetic origin for early farmers from mediterranean cardial and central european LBK cultures. Molecular Biology and Evolution. 2015;32: 3132–3142. pmid:26337550
View Article • PubMed/NCBI • Google Scholar

59. Yang DY, Eng B, Waye JS, Dudar JC, Saunders SR. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. Am J Phys Anthropol. 1998;105: 539–543. pmid:9584894
View Article • PubMed/NCBI • Google Scholar

60. Malmström H, Svensson EM, Gilbert MTP, Willerslev E, Götherström A, Holmlund G. More on contamination: The use of asymmetric molecular behavior to identify authentic ancient human DNA. Molecular Biology and Evolution. 2007;24: 998–1004. pmid:17255122
View Article • PubMed/NCBI • Google Scholar

61. Dabney J, Knapp M, Glocke I, Gansauge M-T, Weihmann A, Nickel B, et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc Natl Acad Sci USA. 2013;110: 15758–15763. pmid:24019490
View Article • PubMed/NCBI • Google Scholar

62. Meyer M, Kircher M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harbor Protocols. 2010; pmid:20516186
View Article • PubMed/NCBI • Google Scholar

63. Briggs AW, Heyn P. Preparation of next-generation sequencing libraries from damaged DNA. Methods in Molecular Biology. 2012;840: 143–154. pmid:22237532
View Article • PubMed/NCBI • Google Scholar

64. Schlebusch CM, Malmström H, Günther T, Sjödin P, Coutinho A, Edlund H, et al. Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. Science. 2017;358:652–655. pmid:28971970
View Article • PubMed/NCBI • Google Scholar

65. Kircher M. Analysis of high-throughput ancient DNA sequencing data. Methods Mol Biol. 2012;840: 197–228. pmid:22237537
View Article • PubMed/NCBI • Google Scholar

66. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760. pmid:19451168
View Article • PubMed/NCBI • Google Scholar

67. Kuhn JMM, Jakobsson M, Günther T. Estimating genetic kin relationships in prehistoric populations. bioRxiv. 2017; 100297.
View Article • PubMed/NCBI • Google Scholar

68. Green RE, Malaspinas A-S, Krause J, Briggs AW, Johnson PLF, Uhler C, et al. A Complete Neandertal Mitochondrial Genome Sequence Determined by High-Throughput Sequencing. Cell. 2008;134: 416–426. pmid:18692465
View Article • PubMed/NCBI • Google Scholar

69. Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, Albrechtsen A, et al. An Aboriginal Australian genome reveals separate human dispersals into Asia. Science. 2011;334: 94–98. pmid:21940856
View Article • PubMed/NCBI • Google Scholar

70. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 2014;15: 356. pmid:25420514
View Article • PubMed/NCBI • Google Scholar

71. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. American Journal of Human Genetics. 2012;91: 839–848. pmid:23103226
View Article • PubMed/NCBI • Google Scholar

72. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27: 2156–2158. pmid:21653522
View Article • PubMed/NCBI • Google Scholar

73. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079. pmid:19505943
View Article • PubMed/NCBI • Google Scholar

74. Broad Institute. Picard tools. https://broadinstitute.github.io/picard/. 2016. 21 Feb 2016. http://broadinstitute.github.io/picard/

75. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81: 559–575. pmid:17701901
View Article • PubMed/NCBI • Google Scholar

76. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4: 7. pmid:25722852
View Article • PubMed/NCBI • Google Scholar

77. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. PLoS Genet. 2006;2: e190. pmid:17194218
View Article • PubMed/NCBI • Google Scholar

78. Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hünemeier T, Petzl-Erler ML, et al. Genetic evidence for two founding populations of the Americas. Nature. 2015;525: 104–108. pmid:26196601
View Article • PubMed/NCBI • Google Scholar

79. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19: 1655–1664. pmid:19648217
View Article • PubMed/NCBI • Google Scholar

80. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23: 1801–1806. pmid:17485429
View Article • PubMed/NCBI • Google Scholar

81. Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, et al. Reconstructing Native American population history. Nature. 2012;488: 370–374. pmid:22801491
View Article • PubMed/NCBI • Google Scholar

82. Moorjani P, Patterson N, Hirschhorn JN, Keinan A, Hao L, Atzmon G, et al. The history of african gene flow into Southern Europeans, Levantines, and Jews. PLoS Genet. 2011;7: e1001373. pmid:21533020
View Article • PubMed/NCBI • Google Scholar

83. Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, et al. Inferring admixture histories of human populations using linkage disequilibrium. Genetics. 2013;193: 1233–1254. pmid:23410830
View Article • PubMed/NCBI • Google Scholar

84. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014;505: 43–49. pmid:24352235
View Article • PubMed/NCBI • Google Scholar

85. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. Science. 2010;329: 75–78. pmid:20595611
View Article • PubMed/NCBI • Google Scholar