Genomics of Mesolithic Scandinavia reveal migration routes and high-latitude adaptation

Supplementary Information

Table of Contents

S1 Archaeological background

S1.1 Introduction

The late-glacial and post-glacial dispersal of humans to Scandinavia has been the focus of extensive archaeological debate. The complex history of climatic fluctuations during the Eurasian deglaciation (Figure S1.1) created a variety of habitats for the first humans colonizing the region [1–7]. How the dispersal took place varied over time and in the beginning it followed a wave and advance model with expansions and contractions, linked both to changes in climate and to the geological development of Scandinavia and the Baltic Sea Sea [8–14]. The fact that many Pleistocene coastal areas became submerged due to the post-glacial rise in sea level makes it particularly challenging to explore the earliest coastal activities. In Northern Scandinavia, however, where the Eurasian Ice sheet was at its thickest, isostatic land-uplift in many places more than compensated sea-level rise, thus permitting studies of early Holocene coastal settlement. The first human colonization of the region followed a wave of advance model with expansions and contractions linked to these climatic and landscape developments in Scandinavia and the Baltic Sea Area [8–14].
Our present knowledge of the various pulses of human colonization of the areas influenced by the Eurasian Ice Sheet rests on typological and technological analyses of lithic assemblages, information on subsistence and dietary pattern but also natural history proxy data. We here present novel genetic data of human remains dating between 9,300 and 6,000 cal BP that offer a unique possibility to examine the demographic history of the Early Holocene in northern Europe.

The earliest evidence for human presence in southern Scandinavia consists of lithic artifacts assigned to the Hamburgian culture. These date to c. 16,000 cal BP; thus roughly contemporaneous with the earliest radiocarbon dated finds of reindeers, the most important prey during this period. However, around 14,000 cal BP the evidence of human presence becomes scarce [10,11], only to be followed by a new advance of apparently episodic pioneer settlements from around 12,800 cal BP [15–18].

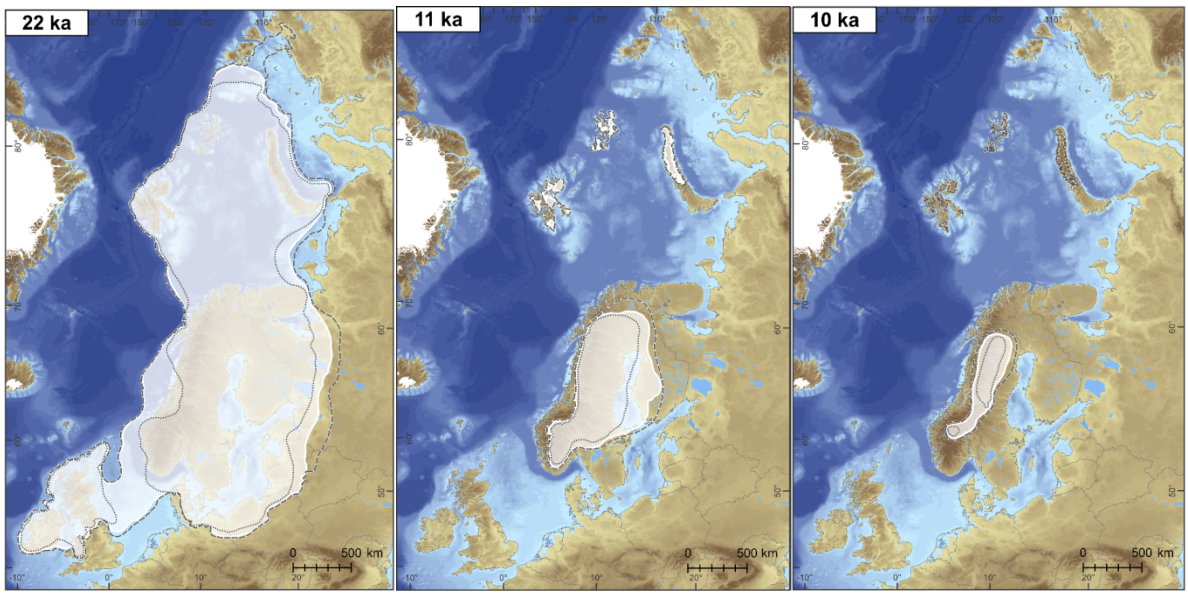


Figure S.1.1 Time-slice reconstruction of the extent of the Eurasian ice sheet, during the Last Glacial Maximum (LGM) and the relevant periods. Dates are expressed in calibrated years BP. From Hughes et al. [19], Figure 6. Reprinted with permission.

From around 11,700 cal BP the archaeological record indicates a more consistent northern expansion along the Norwegian coast, following the retreat of the ice sheet. By c. 11,500–11,300 cal BP there are approximately 800 settlements throughout the Norwegian coast-line connected to this pulse of migration [20–25]. Artefacts and tools associated with these pioneer settlements show links to mobile hunter-gatherer groups with a southern origin in the Ahrensburgian tradition [26]. The hunter-gatherers that had previously hunted reindeer, later on adopted a strictly marine economy. Seals and other marine mammals constituted an important pull factor for the terrestrial hunter-gatherers that became marine foragers in early Holocene Scandinavia [27–29]. The use of marine resources seems to have been the main mode of subsistence along the Norwegian coast, although terrestrial mammals like reindeer were also exploited during occasional mountain hunts [6,30–34]. Thus, a predominant hypothesis concerning the colonization of Scandinavia assumes that settlements rapidly advanced from the south to western Norway and finally to the Varanger area in the north [24,25,30,31,33]. Coastal areas of western Sweden and northern Norway were now free of ice and represented a partially protected archipelago rich in marine resources [14,35].

The oldest settlements on the eastern side of the Baltic Sea, from southern Finland, are dated to around 11,100 cal BP [36–38], and for Northern Finland around 10,300 cal BP [37–39]. Further south, in the Baltic countries dates fall around 10,800-10,200 cal BP [40]. The material culture found at these sites, especially the lithic technology, labelled as Post-Swiderian [41], has its closest counterparts within the Late Glacial microlithic Swiderian complexes of northeastern Europe, found to the east of the Ahrensburgian complexes [16,42]. Further east, to the north of the Ural Mountains in Russia, human settlements have been found on the northern coasts of the

continent dating to the Late Paleolithic, possibly as early as around 43,000-40,000 cal BP [43]. Recent data on the deglaciation [19,44], and archaeological evidence [45], such as the Pymva Shor and other sites (dated to c. 26,300 - 11,600 cal BP) indicate that a north-eastern route for human colonization of Scandinavia may also be considered. Several Early Mesolithic sites are known along the Kola Peninsula [46] and the White Sea coast, which are contemporaneous to the northern Scandinavian sites [35]. However, at present we do not have detailed knowledge of their chronology (Tikhonov, A. pers. comm.).

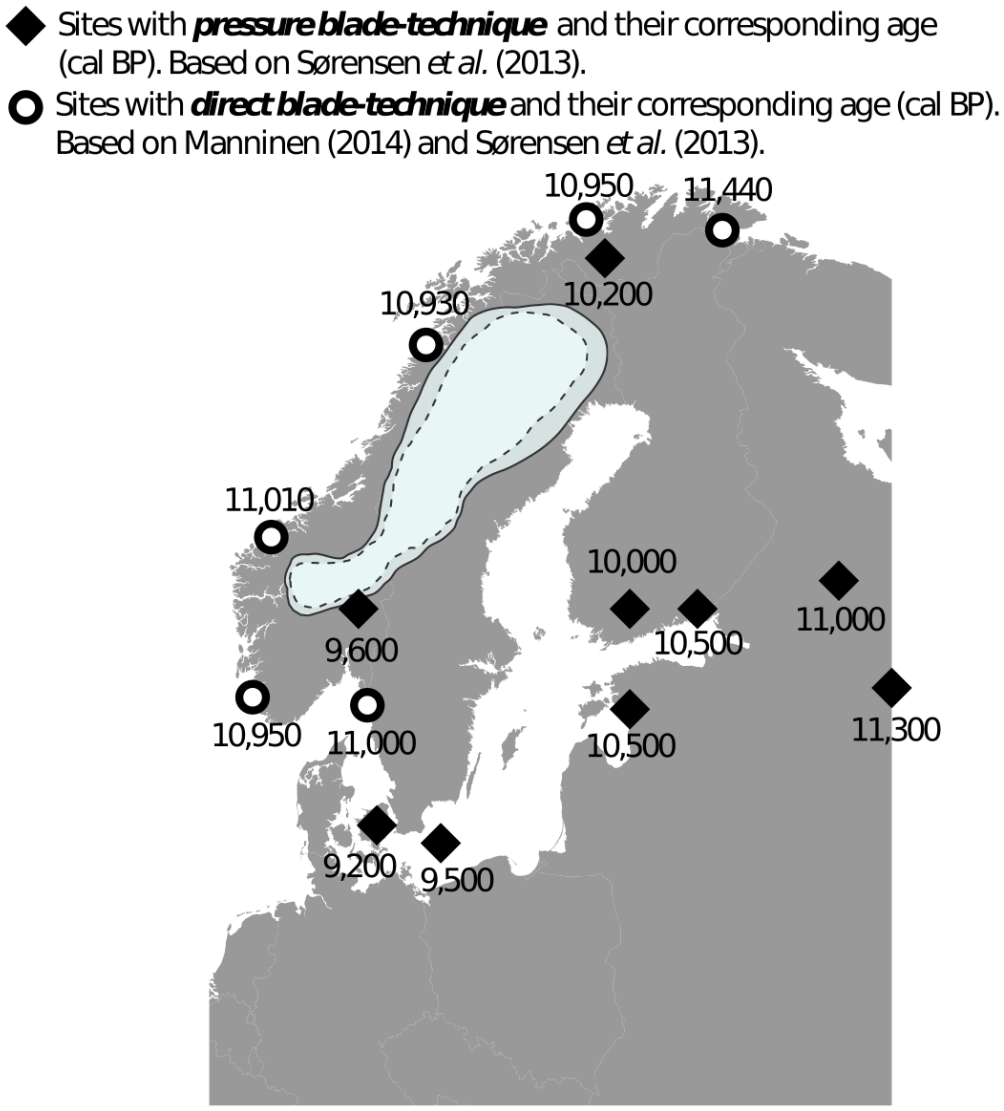


Figure S1.2 Distribution of 'direct blade' and 'pressure blade' lithic artifacts from radiocarbon dated archaeological sites. Locations and dates are based on [5,47].

S1.2 Lithic technology in Mesolithic Northern Scandinavia

The pioneer settlements of central and northern Scandinavia (c. 11,500-11,300 cal BP) shared a flint industry characterized by a macro-flake technology and a direct knapping blade technique that can also be found in southern Scandinavia. Certain artefact types such as tanged points and single-edged points, have their closest counterparts in the Ahrensburgian tradition in south Scandinavia and northern Europe [7,26,30,47–49]. This technology is widespread in western Sweden and Norway and evolved slowly for c. 1,200 years.

Around 10,300 cal BP archaeological finds of another lithic technology (pressure blade technique) appear in northern Finland and northernmost Norway, this time showing similarities with the post-Swiderian sites in north-western Russia dated to 11,300–7,700 cal BP [33,38,47–50]. This technology consists of slotted-bone tools with flint edges made of regular micro-blades that were pressed from conical blade cores. It has an eastern origin with its earliest appearance in northern East Asia between 30,000-20,000 cal BP [51–53] and is introduced to the Baltic area via the Upper Volga area in Russia [54]. From here, the technology spread, either by direct migration of people, or by knowledge transmission, via a northern or a southeastern route. Rankama & Kankaanpää[38] [38] conclude that the closest technological influences on the north Finnish site at Sujala and the Norwegian site Fállegoahtesajeguolbba provide strong evidence for people of the post-Swiderian groups coming from north-western Russia and establishing themselves in northern Lapland.

The pressure blade technology developed in areas where flint was an important raw material, but was adapted to local raw materials in northern Scandinavia and Finland [32,33] and eventually in middle Sweden [50]. Raw materials such as quartz, quartzite, sandstone, slate and flint have variable and different fracture patterns and they offer different possibilities for tool production. In northern Norway, Sweden and all of Finland [50,55] other techniques were also adopted and in e.g. Finland a flake based technology was much more common than a blade technology [56].

It has been suggested that the pressure blade technology dispersed to southern Scandinavia from the north [32,33,38,47,50](Figure S1.2). The technology is found in southern Norway as early as c. 10,300 cal BP [32,33,47,48,50,55] and in southern Scandinavia and western Europe around 9,000 cal BP to 7,800 cal BP [48,57]. It is interesting that the technology appears earlier in southern Norway than in both Denmark and south of the Baltic.

S1.3 Site description and analyzed individuals

We have studied the genetic variation in eight samples (representing 7 unique individuals) from four Mesolithic coastal sites: Hummervikholmen in Southern Norway, Steigen in Northern Norway, and Stora Förvar and Stora Bjers on islands off the Swedish coast in the Baltic proper. The samples and/or other skeletal elements from the same individuals have also been subjected

to stable carbon and nitrogen isotope analysis and radiocarbon dating (Table 1). The individuals from Hummervikholmen, Stora Förvar and Stora Bjers all date between c. 9,500 and 8,600 cal BP, i.e. the Boreal chronozone in the Early Holocene, while the individual from Steigen dates considerably later, c. 6,000–5,800 cal BP, i.e. the Atlantic chronozone in the Middle Holocene. Previously published genetic, stable isotope and radiocarbon data from two Mesolithic sites in Motala on the Swedish mainland, spanning the period 8,200–6,800 cal BP (the Atlantic chronozone in the Middle Holocene) were used for comparison.

Table S.1.1 Overview of all samples analyzed sorted by site: individuals, elements, analysis performed, and lab codes used. Detailed data in Tables S1.1-2.

Site	Short name	Individual	Skeletal element	DNA	IRMS $\delta^{13}C$/ $\delta^{15}N$	^{14}C date	AMS $\delta^{13}C$
Hummervikholmen, Norway	Hum1	Hummervik-holmen Ind. 1	Cranial fragment		x	x	x
Hummervikholmen, Norway	Hum1	Hummervik-holmen Ind. 1	Left 2nd maxillary incisor	x	x		
Hummervikholmen, Norway		Hummervik-holmen Ind. 2	Occipital bone (cranium)		x	x	x
Hummervikholmen, Norway	Hum2	Hummervik-holmen Ind. 2	Left 1st maxillary molar	x	x		
Hummervikholmen, Norway		Hummervik-holmen Ind. 3	Frontal bone (cranium)		x	x	x
Hummervikholmen, Norway			Tibia		x	x	x
Hummervikholmen, Norway			Femur			x	x
Steigen, Norway	Steigen	Steigen Ind. 1	Right 1st mandibular molar		x		
Steigen, Norway	Steigen	Steigen Ind. 1	Right 2nd mandibular molar	x	x		
Steigen, Norway	Steigen	Steigen Ind. 1	Left 3rd mandibular molar		x		
Steigen, Norway	Steigen	Steigen Ind. 1	Mandible		x	x	x
Stora Förvar, Sweden	SF9		Parietal bone (cranium)	x	x	x	
Stora Förvar, Sweden	SF11		Tibia	x	x	x	x
Stora Förvar, Sweden	SF12		Femur	x	x	x	x
Stora Förvar, Sweden	SF13		Os coxae	x	x	x	x
Stora Bjers, Sweden	SBj	Stora Bjers Ind. 1	Teeth, right M1 and M2 and M3 mandible, Tibia	x		x	x

S1.3.1 Hummervikholmen, Norway

In 1994, human skeletal remains were found at Hummervikholmen, a small island in the Søgne archipelago in Vest-Agder County in southernmost Norway [34,58,59]. These remains turned out to be the oldest dated skeletal remains from Norway[34,58,59]. The submerged site was investigated by the Norwegian Maritime Museum (NMM) in 1994-95 [58]. Recent dredging had damaged c. 60% of the site, and a stone foundation had disturbed the original beach-zone. The shallow sub-sea sediments had been partially removed down to the bedrock. During the sieving

of the re-deposited sediments several human bone fragments were recovered. Only a large, almost complete skull, a tooth, and a thighbone appeared to be *in situ [58]*. The two samples included in the present study were found *in situ*: the cranium of Individual 1 (Hum1) and a tooth from another individual (Hum 2).

Five distinct stratigraphic layers were identified, described from the top down: (1) marine silty sand, (2) compact oyster bank, (3) clay deposit with decomposed organic material, (4) intermittently occurring organic layer with skeletal remains, covering, (5) sterile moraine and/or bedrock [59]. The dating of the layer sequence is consistent and suggests that the oyster bank was deposited during the Tapes maximum (Atlantic period) in the relatively warm, calm waters, thus preserving the skeletal remains by sealing them in the thin organic bottom layer.

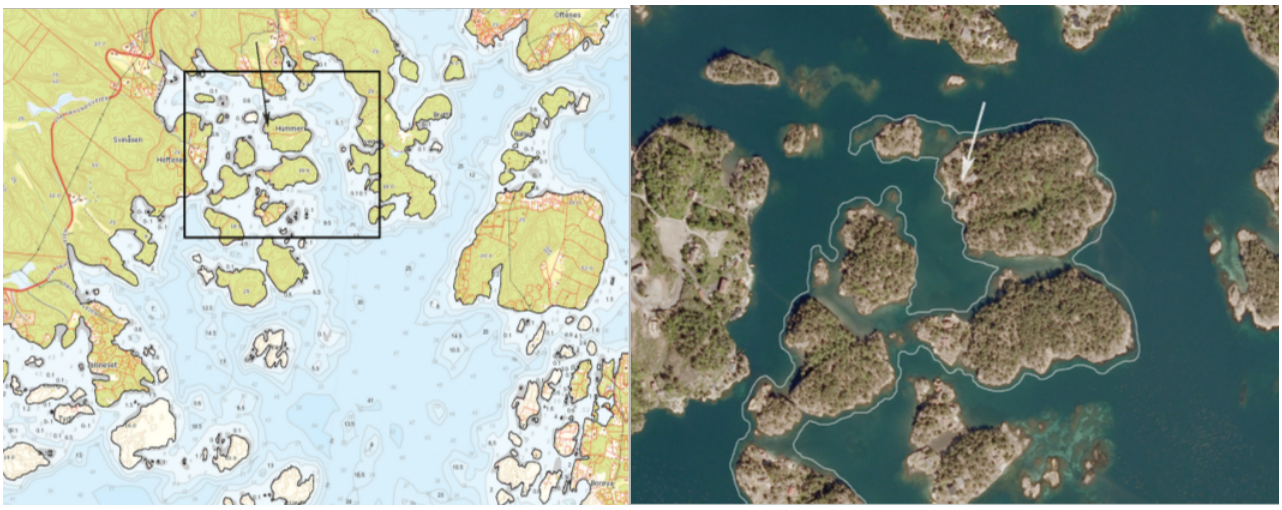


Figure S1.3 The Søgne archipelago indicating the location of Hummervikholmen (site marked with arrow) and the – 2 m contour line (marked in white). Drawing by Pål Nymoen. Source: www.kulturminnesok.no.

At least three to five individuals are represented in the material from Hummervikholmen, and the bone remains are well preserved, Figure S1.4. The assemblage consists of an almost complete skull (Individual1/Hum1), an occipital fragment (Individual2/Hum2), a frontal bone fragment (Individual 3), an almost intact left femur, and a damaged left tibia. The two postcranial bones are gracile; but it is not clear which of these adult bones belongs to which individual. The largest skull (Individual1/Hum1) is probably a female, *c*. 35–40 years of age at death. The skull is robust, resembling other Scandinavian Mesolithic female skulls [58,60]. The Hum1 and Hum 2 samples have been radiocarbon dated to 8,690±50 BP and 8,600±95 BP, respectively. Due to heavy δ13C and δN 15 values indicating an almost entirely marine diet a marine reservoir correction had to be applied rendering corrected dates (2 σ) 9,471 – 9,225 cal BP and 9,461 – 9,011 cal BP.

The findings indicate the presence of a burial tradition in Norway during the Middle Mesolithic. The remains represent unburnt burials, inhumations. The most complete skull was positioned

with the back of the head downwards—this is the only potential indication of a dorsal position, but it could have been caused by marine erosion during the Tapes transgression. A more detailed knowledge of the original position of the bodies cannot be established. There are neither signs that the dead were provided with any grave goods nor of rituals, e.g. ochre or physical cutting. The Hummervikholmen grave site thus adds to a diverse image of Middle Mesolithic burial traditions in Scandinavia that is clearly distinguished from the later, larger and more elaborate known Mesolithic grave sites from the Atlantic period like Skateholm and Vedbæk [61–63]. To date, Hummervikholmen is the most prominent submerged Mesolithic site in Norway.

Figure S1.4 Skeletal fragments from the Hummervikholmen site (Photo: Beate Kjørslevik, © NMM 2010)

S1.3.2. Steigen, Norway

The Steigen human remains were discovered in a cave on a small island, Måløya, in Steigen municipality, Nordland County, Norway. The island is situated in a small archipelago some 5 km west of Sørskott on the Norwegian mainland (Figure S1.5). The cave, situated in a crevice running in a north-south direction, was discovered in 1996, but was not properly analyzed and documented until 2011, when a human mandible was found inside the cave and brought to Tromsø Museum (Figure S1.6). Today the opening of the cave is situated c. 30 m above present sea level, whereas the actual find place for the mandible is situated some 25 m above the present-day sea level, which coincides with the sea level at c. 8,500–5,500 cal BP. It is a natural cave with an inner height exceeding 20 m in some places. An investigation in 2013, with the purpose of measuring the cave and assessing the security and potential for an archaeological excavation, indicated there was a great possibility to recover more finds. A large piece of wood, later to be determined as spruce (*Picea* sp.) or larch (*Larix* sp.) was also retrieved. The radiocarbon date of the mandible of the individual demonstrated that these were the oldest human skeletal remains from northern Norway, further stressing the importance of an excavation in the cave. Accordingly, in 2014 two archaeologists, Keth Lind and Roger Jørgensen from Tromsø Museum, performed an archaeological investigation. The access to the cave is somewhat limited, mainly

due to the fact that it has partly collapsed at the entrance, causing flooding of the cave, and it was consequently impossible to get to the inner part of the cave without an inflatable canoe. The site where the mandible was found is situated c. 100 m from the cave entrance some 10-15 m from the inner end of the cave. Here the cave was 2x3 m wide and the height was 2.5 m. The mandible was found on the gravel floor, next to a large boulder. The excavation of the surrounding area revealed no cultural layers and only a molar tooth belonging to the mandible, but no other skeletal remains. At some distance from the large boulder and the mandible, where the floor was more clayish, a few small fish bones were retrieved. It was decided that it was too risky to remove the large boulder where the mandible had been found, but also that the likelihood to find further human skeletal remains was small and no further excavations were carried out [64].

Figure S1.5. Måløya, in Steigen municipality, with the entrance to the cave marked. From [64].

Figure S1.6 Steigen mandible in situ (left, photo by Gunnar Svalbjørg) and close-up (right, photo by Kerstin Lidén)

S1.3.3 Stora Förvar, Sweden

The cave of Stora Förvar, situated on the small island of Stora Karlsö, off the west coast of Gotland in the Baltic Sea, was excavated between 1888 and 1893 when extensive cultural layers of up to 4.5 m thickness of the cave were removed [65] (Figure S1.7). Large amounts of finds dating from the Mesolithic and up to the medieval period were recovered. Radiocarbon dates show that the duration of the Mesolithic occupation lasted from around 9,300 cal BP and the ending around 7,000 cal BP. The zooarchaeological finds indicate that the site was mainly used for seal hunting, but the importance of fish and birds is difficult to evaluate because of the crude recovery techniques [66,67]. The layers were excavated by hand and finds were recovered in 0.3 m thick layers (one foot) in ten different sections/columns (Sw. *parcell*), labelled A to J. The Mesolithic layers (approximately Layers 13-9) produced more than 1,500 kg of bones. The assemblage included scattered human remains from up to ten individuals [68], among them two almost complete crania that unfortunately were lost during storage. The identified individuals are represented by a few skeletal elements each and in some cases, the bones may in fact originate from the same individual.

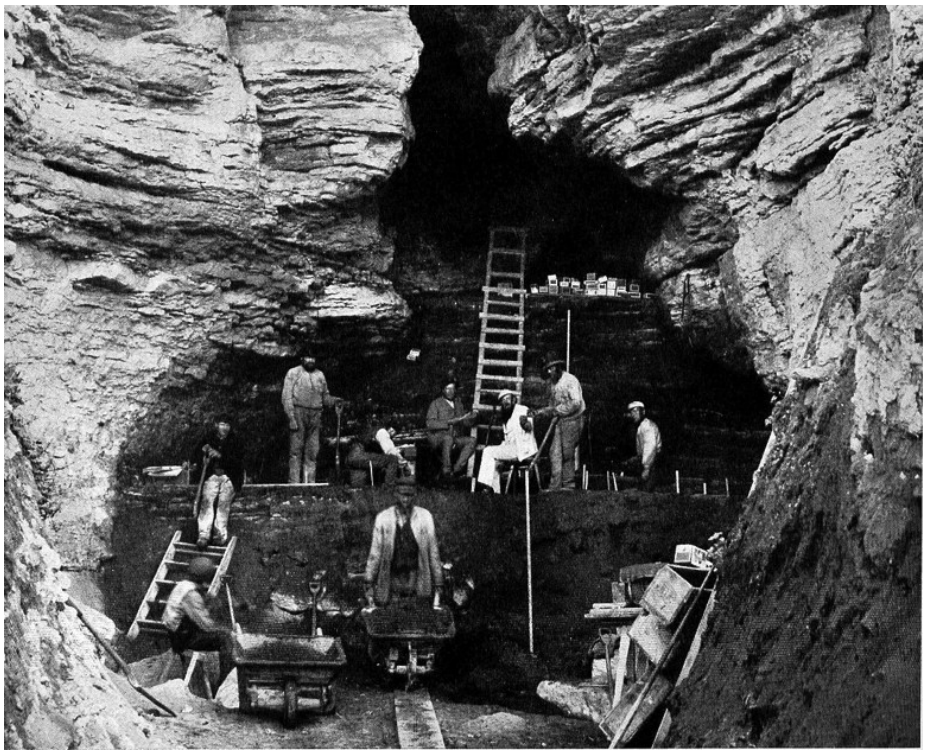


Figure S1.7 Photograph of the excavations of Stora Förvar on Stora Karlsö. Photo: Hjalmar Stolpe. Antiquarian Topographical Archives (ATA), Stockholm. Reproduced with permission.

The field documentation shows that a few human ribs and vertebrae were found in anatomical position indicating the presence of at least one – probably damaged – inhumation burial [68]. It is evident that the long-term use of the cave to some extent has affected the stratigraphic integrity and also the potential to estimate the number of individuals present in the assemblage. It is, however, clear that the human bones originate from at least one infant (maximum 6 months of age) and one or possibly two children at the age of 10-14 years. At least one juvenile male (10-21 years) is present and finally, bones from 3-5 adult individuals (this is a more conservative estimate than the one presented in Lindqvist and Possnert [68]. We took four samples from adult individuals: SF12 recovered in Section A and Layers 12-14, SF9 in Section G/Layer 9, SF11 from Section G/Layer 11 and SF13 found in Section F/Layer 13.

Lithic tools on the Mesolithic sites on Gotland were mostly manufactured from local Ordovician flint and procured with a simple direct technique in a manner similar to the tradition in the south (the Maglemose techno group I and II [66,69]). There are occasional examples of pressure technique and slotted bone point technology, but these are from a slightly younger period (9,000–8,500 cal BP) than those found in northern Scandinavia. At that time, the pressure technique was also known in Southern Scandinavia. The first pioneers on Gotland seem to have practiced a technology that had more in common with a south-west Scandinavian tradition than the south-east/northern tradition. Such a view is also supported by the find of a core axe of south

Scandinavian flint in the Mesolithic culture layers from the Stora Förvar cave [65,66] and of several finds on Gotland of polished Limhamn axes with a south Scandinavian origin.

S1.3.4. Stora Bjers, Sweden

The Stora Bjers burial, located in Stenkyrka parish, is one of the oldest burials on Gotland. We included the c. 45-year-old male that was found in a crouched position in a sandbank [70]. The grave was found in 1954 and the complete burial was brought to the Museum of Gotland where it is exhibited today. Interestingly, a fragmented slotted bone point with an associated regular microblade with a straight profile (Figure S1.8) were found, still attached to the pelvis bone. The regular shape and straight profile of the microblades are strong indicators of the use of pressure technique [57], they likely originate from a pressure blade core. The arrow had likely been shot into the body since only the tip of the point is preserved. The male from Stora Bjers also had severe crush wounds on the skull and left jaw, further indicating a violent death.

Figure S1.8 Fragmented slotted bone point and two blade fragments was found near the pelvis of the Stora Bjers male (left, photo by Jan Apel, right, photo by Johan Norderäng, Gotland Museum).

S1.3.5. Motala, Sweden

Motala is situated at the river Motala Ström, by the outlet of Lake Vättern in the province of Östergötland, in east central Sweden. Two sites on opposite sides of the river, Kanaljorden and Strandvägen, have been excavated during the past two decades, yielding a large amount of well-preserved organic remains [71]. Radiocarbon dates of faunal and human remains from the two sites span the period 8,200–6,800 cal BP. The two sites overlap chronologically but they differ contextually. While the Strandvägen site includes a cemetery, associated settlement remains and disarticulated human remains deposited along the shore of the river, the Kanaljorden site consists of a ceremonial wetland deposition of mainly human calvaria on a stone-packing. The lithic assemblages have common features with both the Lihult/Nøstvet Culture of south-west Sweden and south-east Norway [72–74] and the late Kongemose Culture of southern Scandinavia [75–79]. But it also resembles the stone technology found further to the north east [77–79], demonstrating that this is a border zone where different stone technologies met, a cultural mixture contact zone [80]. Previous genetic analyses on the skeletal material from Kanaljorden seem to confirm this

hypothesis, suggesting that the population has ancestry both from eastern and western hunter-gatherers [81,82].

Stable carbon and nitrogen isotope analysis demonstrate that all humans were high aquatic-protein consumers, where the aquatic resources ranged from freshwater to brackish/marine origin (Figure S1.9). The overall pattern in the dietary life histories for the Kanaljorden individuals is one of a stable diet throughout childhood, as reflected in $\delta^{13}C$ and $\delta^{15}N$ values [71]. Strontium isotope analysis ($^{87}Sr/^{86}Sr$) of human tooth enamel from both Kanaljorden and Strandvägen shows that there is a clear difference in strontium isotopic compositions between the two sites. While the Strandvägen humans generally exhibit strontium isotope values which fall within the local strontium bio-available range, by contrast, the Kanaljorden individuals generally have strontium isotopic values which fall outside the local range, indicative of a non-local population. The results of stable and radiogenic isotope analysis of human remains from Motala indicate a considerable variation in diet within the two sites as well as a substantial difference in origin between the sites.

S1.4 Stable isotope analysis and diet

Human bones and, where available, teeth from Hummervikholmen, Steigen and Stora Förvar were subjected to stable carbon and nitrogen isotope analysis to investigate dietary patterns. The $\delta^{13}C$ and $\delta^{15}N$ data are reported in Table S1.2 and Figure S1.9. The sampling, collagen extraction and EA-IRMS analysis were performed accordingly:

For the Hummervikholmen and Steigen samples, bone and dentine powder was obtained using a dentist's drill. Collagen was extracted following Brown et al. [83], which includes demineralization in a weak acid (0.25 M HCl) for two days, gelatinization with 0.01 M HCl in 58°C overnight, followed by ultrafiltration to remove remnants <30 kDa, and subsequent lyophilization. Approximately 0.5 mg of collagen was weighed into tin capsules, and combusted in a Carlo Erba NC2500 elemental analyzer connected to a Finnigan MAT Delta+ isotope ratio mass spectrometer run in continuous flow. All the sampling and extractions were performed at the Archaeological Research Laboratory, and the subsequent EA-IRMS analysis took place at the Stable Isotope Laboratory (SIL), Dept. of Geological Sciences, both at Stockholm University. The precision of the $\delta^{13}C$ and $\delta^{15}N$ measurements was ±0.15‰ or better. For the Stora Förvar samples, all sampling, extraction and EA-IRMS analysis were performed at the Beta analytic radiocarbon facility. No detailed data on extraction and measurement precision was reported, other than that collagen was extracted using HCl, followed by treatment with NaOH, and that no ultrafiltration took place.

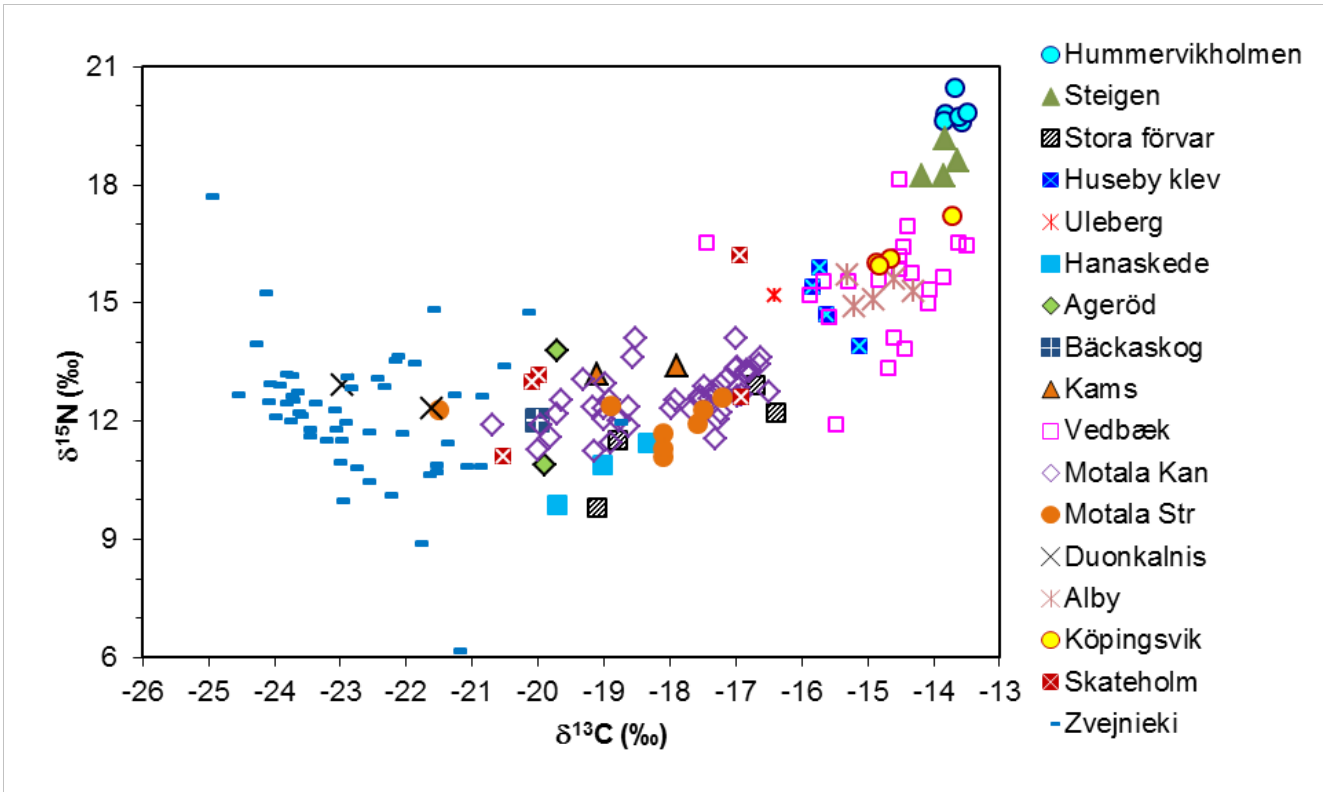


Figure S1.9 Stable carbon and nitrogen isotope data for Scandinavian and East Baltic individuals dated pre-6,000 cal BP. Data from Lidén 1996, Sten et al. 2000, Eriksson & Lidén 2003, Lidén et al. 2004, Eriksson 2007, Eriksson et al. 2009, Skar et al. 2016, Eriksson et al. 2016, Eriksson unpublished, and the present paper.

The Norwegian samples have values ranging between –14.2‰ and –13.5‰ for $\delta^{13}C$, and 18.2–20.5‰ for $\delta^{15}N$, indicative of a massive intake of marine mammal protein in both Hummervikholmen and Steigen (Figure S1.9). Although no faunal reference data from the sites are available to enhance interpretation, the human isotopic signatures are so elevated that they are only comparable to previously analyzed high marine-protein consumers [84,85]. For the two individuals with life history data, Hummervikholmen Individual 1 (Hum1) and Steigen, there is no suggestion of any significant changes in diet during their lifetimes.

The Stora Förvar samples have considerably lower values, between –18.8‰ and –16.4‰ for $\delta^{13}C$ and 9.8–12.9‰ for $\delta^{15}N$. The nitrogen isotope values suggest a diet predominated by protein from fish rather than seals. This is further corroborated by stable isotope data from a Stora Förvar grey seal (*Halichoerus grypus*) of approximately the same radiocarbon age (8100±30 BP, Beta-399028), and $\delta^{13}C$ and $\delta^{15}N$ values of –19.6‰ and 13.1‰, respectively. Because of the trophic-level effect, consumption of seal would result in human $\delta^{15}N$ values around 16‰, which is much higher than the measured human values. With regard to the $\delta^{13}C$ values, the present-day brackish Baltic Sea was at this time in the Ancylus Lake Phase, with freshwater conditions, which should typically render $\delta^{13}C$ values even lower than the ones for the Stora Förvar humans. However,

consumption of littoral fish, such as pike and perch, or possibly of migrating seals, may help to explain the range of human $\delta^{13}C$ values.

The genetic data indicate that the samples SF9 and SF13 derive from one single individual. The $\delta^{13}C$ values match (–18.8‰ and –19.1‰, respectively), while the $\delta^{15}N$ values differ by 1.7‰ (11.5‰ and 9.8‰, respectively). This is a larger difference than one would expect if the bones actually derive from the same person. However, since the analysis was performed on different parts of the skeleton, the cranium and the pelvis, respectively, it is possible that this difference merely reflects seasonal variation in protein intake, or a shift in diet [86]. Variation in collagen turnover rate between different bone elements could thus explain why the $\delta^{15}N$ values do not match, and accordingly still be compatible with the assumption that SF9 and SF13 represent one individual. The stable isotope data in this case are not conclusive.

For Stora Bjers there are no IRMS stable isotope data, only an AMS $\delta^{13}C$ measurement, for which the precision is poorer. Given that uncertainty, however, the AMS $\delta^{13}C$ value, –16.1‰, indicates that also this individual may have consumed substantial amounts of aquatic resources.

S1.5 Radiocarbon dating and reservoir effects

S1.5.1 Radiocarbon reservoir correction

As all the human individuals from the investigated sites included portions of aquatic protein in their diets, it is likely that the radiocarbon dates on the bones have been affected by radiocarbon reservoir effects. It is, consequently, crucial to correct for any reservoir age in order to get an accurate estimate of the true age of a sample. The radiocarbon dates and associated calibrated ranges, with and without reservoir correction for comparison, are presented in Table S1.3.

For the dates on human bone from Hummervikholmen and Steigen, we have applied a maximum reservoir age correction of 380±30 radiocarbon years, based on Mangerud et al. [87]. This correction is based on a 100% marine diet, which seems adequate in both cases, considering their stable isotope signatures. For Hummervikholmen, situated in southern Norway, this correction should be regarded as an absolute maximum, whereas for Steigen, located considerably further north, it may be closer to the truth. Applying an age offset correction, rather than a ΔR correction, results in differences in calibrated ranges of <30 years, which can be considered irrelevant given the time scale.

For the dates on human bone from Stora Förvar and Stora Bjers, we have applied a reservoir age correction of 70±40 radiocarbon years, following Eriksson [88]. Although this age offset was calculated for human seal hunters on Gotland during the Littorina stage of the Baltic Sea, with admittedly higher salinity, there is no better estimate for the prehistoric Baltic proper that we are aware of. Without more detailed dietary data it is currently not possible to make a more accurate estimate than this rather conservative approximation.

Table S1.2 Stable carbon and nitrogen isotope data (measured by IRMS) for the analyzed samples and individuals. Hummervikholmen data from Skar et al. [34]; Steigen data this study; Stora Förvar data obtained from Beta analytic, reported in conjunction with radiocarbon data. 2016. Nd = no data.

Short name	lab code	Skeletal element	$\delta^{13}C$ (‰)	$\delta^{15}N$ (‰)	collagen yield (%)	%C	%N	C/N
Hum1	SOG105	Cranial fragment	-13.6	19.7	11.2	40.2	14.6	3.2
Hum1	SOG106	Left 2nd maxillary incisor	-13.7	20.5	4.1	39.7	14.0	3.3
	SOG103	Occipital bone (cranium)	-13.8	19.6	7.9	41.0	14.8	3.2
Hum2	SOG107	Left 1st maxillary molar	-13.5	19.8	4.2	39.1	14.1	3.2
	SOG101	Frontal bone (cranium)	-13.8	19.8	5.2	40.5	14.5	3.2
	SOG104	Tibia	-13.6	19.6	6.4	41.0	14.8	3.2
Steigen	STE 01	Right 1st mandibular molar	-13.6	18.6	3.7	40.2	14.8	3.2
Steigen	STE 02	Right 2nd mandibular molar	-14.2	18.2	0.7	41.8	14.7	3.3
Steigen	STE 01	Right 1st mandibular molar	-13.6	18.6	3.7	40.2	14.8	3.2
Steigen	STE 03	Left 3rd mandibular molar	-13.8	18.2	4.0	40.0	14.7	3.2
Steigen	STE 04	Mandible	-13.8	19.2	0.6	37.2	13.4	3.2
SF9	Beta-399027	Parietal bone (cranium)	-18.8	11.5	nd	nd	nd	nd
SF11	Beta-448532	Tibia	-16.7	12.9	nd	43.5	15.8	3.2
SF12	Beta-448531	Femur	-16.4	12.2	nd	43.4	15.9	3.2
SF13	Beta-448533	Os coxae	-19.1	9.8	nd	43.1	15.4	3.3

S1.5.2 Calibrated ranges

There are nine radiocarbon dates from five human skeletal elements from Hummervikholmen (Table S1.3). The archaeological context strongly suggests that the human bones were deposited during one single event, so all the dates should match. Combining all the nine dates does not fulfill the chi-squared test for internal consistency (R_Combine [89]), but the exclusion of the two youngest dates (TUa-2105, TUa-2108) results in a combined date that does not violate the statistical test, giving a combined date of 8,703±27 BP. Alternatively, the selection of the four TRa dates only (which were produced on collagen extracted for stable isotope analysis, fulfilling the collagen quality criteria), results in a combined date of 8,730±34 BP. In both cases, the calibrated 2σ range (95.4% probability) is roughly 9,500–9,300 cal BP. For the Steigen individual, there was only one radiocarbon date, with a 2σ range of approximately 6,000–5,800 cal BP.

From the four Stora Förvar bones, seven radiocarbon dates were produced. The samples do not originate from a closed context, and therefore they should not be considered coeval. The three dates on SF9 and SF13, suspected to derive from one single individual, correspond well and thus support the suspicion, giving an approximate 2σ range of 9,300–9,000 cal BP. Of the four remaining dates, one (for SF11) is considerably younger (Ua-45742). Since there was a discrepancy between this date and the dates of other individuals, a second replicate dating was performed. This replicate date (Beta-448532) of the same sample element falls within the same range as the other two, and is also in accordance with stratigraphic observations. The resulting 2σ range for SF11 and SF12 is around 9,000–8,800 cal BP, or possibly slightly younger, taking the higher $\delta^{13}C$ values for these two individuals into account. The dates further suggest that the horizontal and vertical stratigraphic units (Sections and Layers) within the larger Mesolithic

stratum cannot be used for chronological differentiation. The bone from the Stora Bjers individual (SBj) generated a radiocarbon date with an approximate 2σ range of 9,000–8,600 cal BP, that is, roughly contemporaneous with SF11 and SF12.

Table S1.3 Radiocarbon dates sorted by site. Dates have been calibrated with Oxcal 3.10 [90], using the atmospheric curve Intcal13 [91]. The calibrated ranges (2σ, 95.4% probability) are given without and with correction for the marine reservoir effect, respectively. For the Norwegian samples, an offset of 380±30 radiocarbon years were applied, following Mangerud et al.[87], and for the Swedish samples from Gotland, an offset of 70±40 radiocarbon years was applied, following Eriksson [88].

Short name	AMS lab code	Skeletal element	AMS ^{14}C date (BP)	AMS δ^{13}C (‰)	Calibrated date (cal BP, 2σ, no reservoir correction)	Calibrated date (cal BP, 2σ, reservoir corrected)
Hum1	TRa-954	Cranial fragment	8690±50	-13.0	9865–9539	9471–9225
Hum1	TUa-1257	Cranial fragment	8600±95	-13.4	9890–9438	9461–9011
	TRa-952	Occipital bone (cranium)	8850±65	-13.4	10176–9700	9732–9368
	TUa-2106	Occipital bone (cranium)	8635±75	-13.3	9867–9488	9462–9102
	TRa-951	Frontal bone (cranium)	8665±100	-13.0	10125–9476	9555–9065
	TUa-2105	Frontal bone (cranium)	8095±55	-13.6	9249–8777	8789–8441
	TRa-953	Tibia	8680±85	-13.2	10115–9501	9534–9125
	TUa-2108	Tibia	8455±75	-12.9	9546–9297	9275–8895
	TUa-2107	Femur	8700±70	-12.6	9909–9534	9524–9191
Steigen	Beta-349961	Mandible	5450±30	-13.0	6300–6203	5950–5764
SF9	Beta-399027	Parietal bone (cranium)	8260 ±30	-18.8	9400–9128	9300–8988
SF11	Ua-45742	Tibia	6459±70	-19.1	7495–7255	7434–7147
SF11	Beta-448532	Tibia	8070±30	-16.7	9089–8786	9023–8760
SF12	Ua-45741	Femur	7952±53	-17.5	8992–8639	8955–8553
SF12	Beta-448531	Femur	8080±30	-16.4	9093–8798	9033–8757
SF13	Beta-386399	Os coxae	8330±40	-19.3	9467–9152	9421–9110
SF13	Beta-448533	Os coxae	8220±30	-19.1	9290–9033	9252–8978
SBj	Ua-46147	Tibia	7974±49	-16.1	8997–8649	8963–8579

S2 DNA Sample preparation

Genomic sequence data was generated from teeth and bone samples belonging to seven (eight, including SF13) Mesolithic Scandinavian hunter-gatherers (Table S2.1). Furthermore, additional libraries were sequenced for Ajvide58 and Ajvide70 [92] (Table S2.2). All samples were prepared in dedicated ancient DNA (aDNA) facilities at the Evolutionary Biology Centre in Uppsala (SF9, SF11, SF12, SF13, SBj, Hum1, Hum2) and at the Archaeological research laboratory, Stockholm University (Steigen).

S2.1 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. Further, all surfaces were UV irradiated (6 J/cm^2 at 254 nm). After removing one millimeter of the surface, approximately 30-100 mg of bone was powderized and DNA was extracted following silica-based methods as in [93] with modifications as in [94,95] or as in [96] 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 ten extracts. Blanks were carried along the whole process until qPCR and/or PCR and subsequent quantification.

DNA libraries were prepared using 20µl of extract, with blunt-end ligation coupled with P5 and P7 adapters and indexes as described in [95,97]. From each extract one to five double stranded libraries were built. Since 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 ul of DNA library, 1X MaximaSYBRGreen mastermix and 200 nM each of IS7 and IS8 [97] 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 ul DNA library, and the following in final concentrations; 1X AmpliTaq Gold Buffer, 2.5mM $MgCl_2$, 250uM of each dNTP, 2.5U AmpliTaq Gold (Thermofisher), and 200nM each of the IS4 primer and index primer [97]. 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 30s, 60°C for 30s and 72°C for 30s, and a final elongation step of 72°C for 10min. For each library four amplifications with the same indexing primer were pooled and purified with AMPure XP beads (Agencourt). The quality and quantity of libraries was checked using Tapestation or BioAnalyzer using the High Sensitivity Kit (Agilent Technologies). 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 paird-end reads or HiSeqX, 150bp

paired-end length using v2.5 chemistry at the SNP & SEQ Technology Platform at Uppsala University. After evaluation of factors such as clonality, proportion of human DNA and genomic coverage samples were selected for re-sequencing aiming to yield as high coverage as possible for each library.

S2.2 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 Uracil-DNA-glycosylase (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 [93–95,98]. 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 with (UDG and endonuclease VIII (endo VIII) or USER enzyme (NEB) treatment was included in order to remove deaminated cytosines [99]. Quantitative PCR (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 ancient DNA libraries (>10 cycles) and they were thus deemed as negative. Our experimental results show minimal levels of contamination, which is in concordance with mitochondrial DNA and X chromosome estimates of contamination (see section 4 and Table 1). Each reaction was done in a total volume of 25 µl, containing 1 ul of DNA library, 1X MaximaSYBRGreen mastermix and 200nM each of IS7 and IS8 [97] reactions were set up in duplicate.The PCRs were set up using a similar system as for the non-damage repair samples (in quadruplicates that were pooled when the PCR products were cleaned), with the difference of using AccuPrime DNA polymerase instead of AmpliTaqGold (Thermofisher) and the following PCR conditions; an activation step at 95°C for 2 min followed by 10-16 cycles of 95°C for 15s, 60°C for 30s and 68°C for 1min, and a final elongation step of 68°C for 5min. 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. In order to sequence libraries to depletion, two to eight libraries were pooled together and sequenced until reaching a clonality of >50%, if the clonality was lower, the library was either classified as unproductive or when the sequencing goal (>55X coverage) was reached and further sequencing was deemed unnecessary. Sequencing was performed as above.

Table S2.1 Summary the lab work and results from sequencing.

Sample name	Bone element	Number of Extracts	Number of libraries	Average read length	Average proportion human [%]
Hum1	Left 2nd maxillary incisor	2	10	60.639	3.4
Hum2	Left 1st maxillary molar	1	5	82.9063	5.3
Steigen	Mandibular molar	1	1	92.821	26.7
SF9	Parietal bone, Os coxae	8	29	74.447	2.9
SF11	Tibia	4	11	79.8543	2
SF12_damage repair	Femur	126*	258	74.9771	4.4
SF12	Femur	15	25	70.9709	6.8
SBj	Teeth, right M1 and M2 and M3 mandible, Tibia	5	15	78.5173	1.2

*This includes the 15 extracts below for which also non-damage repair libraries were built.

Table S2.2 Summary of additional data for two Ajvide samples previously published in Skoglund et al [92].

					Mitochondrial contamination estimate					
Sample	Average read length	Genome coverage	Mitochondrial coverage	Genetic sex	Point estimate	Informative sites	Consenus alleles	Total alleles	Lower C.I	Higher C.I
Ajv58	64.386	2.70	378.97	XY	2.90	5	1505	1550	2.07	3.74
Ajv70	61.365	1.34	108.56	XY	1.61	5	551	560	0.57	2.65

S3 Processing of NGS data

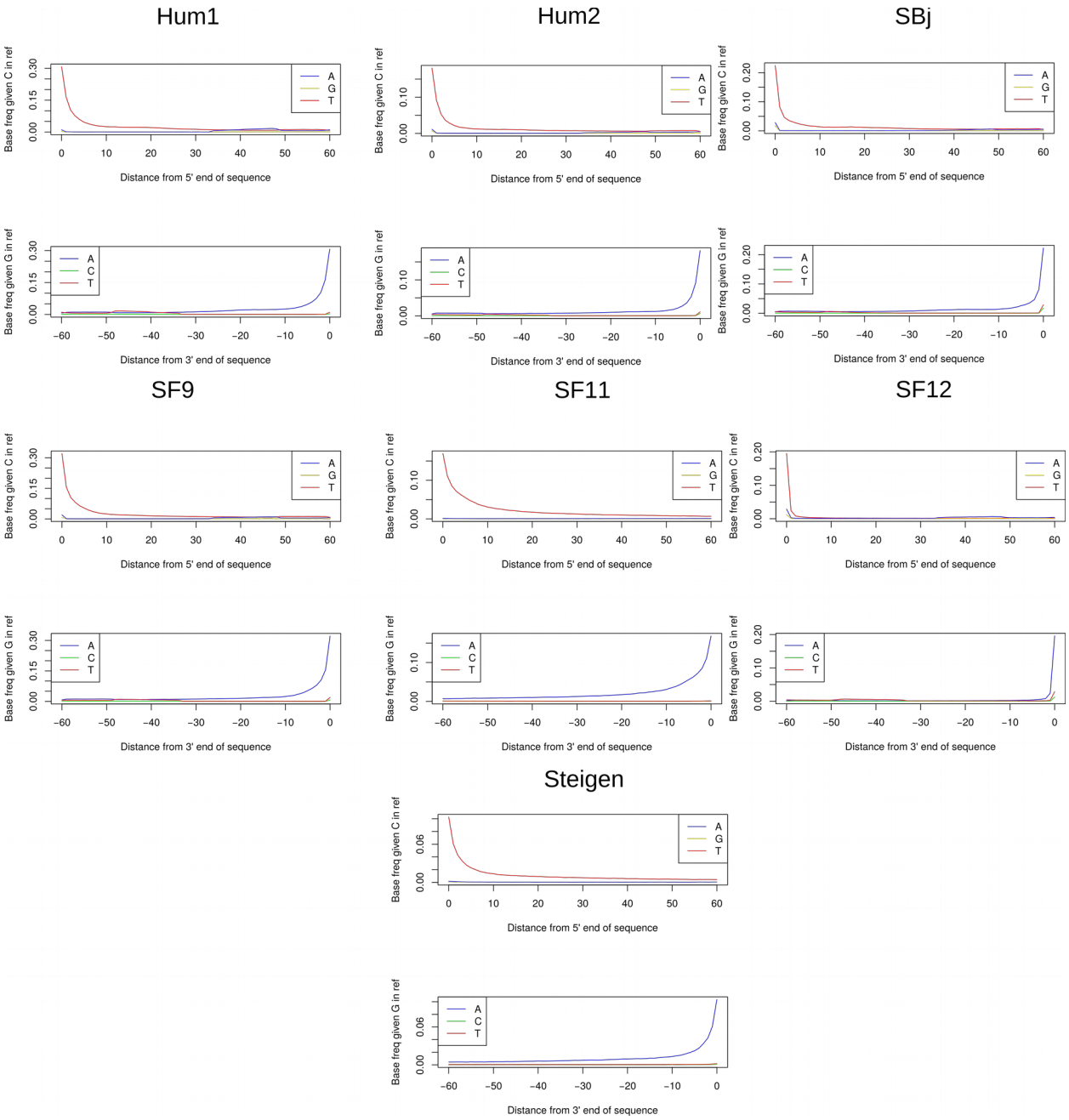
Paired-end reads were merged using MergeReadsFastQ_cc.py [100], if an overlap of at least 11 base pairs was found the base qualities were added together and any remaining adapters were trimmed. Merged reads were then mapped single ended with bwa aln 0.7.13 [101] to the human reference genome (build 36 and 37) using the following non-default parameters: seeds disabled -l 16500 -n 0.01 -o 2 [81,92]. 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 [100]. 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 [102], a tool to estimate kin-relationship from ancient DNA, 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.

S3.1 Authentication of aDNA

Each library, with an estimated contamination lower than 6% using the method developed by Green et al. [103] (see description below) (10 Hum1 libraries, 5 Hum2 libraries, 11 SBj libraries, 29 SF9 libraries, 11 SF11 libraries, 254 SF12 damage-repair libraries, and 1 Steigen library), was then merged (per sample) into a final bam-file using samtools merge [104]. The mtDNA contamination per library was estimated to between 0-4.5% (Table 1 and S4.1) after removal of potentially contaminated libraries (n=4) in SF12 (contamination higher than 5%). The data from all libraries also show the, for aDNA, characteristic deamination patterns towards the fragment-end [105] (Figure S3.1).

Figure S3.1 Damage patterns for all newly sequenced samples.



S3.2 Novel variants in SF12

The high sequencing coverage and the UDG treatment of the SF12 individual made it possible to call new variants in her genome. The number of unique variants per genome largely differed between populations among the individuals sequenced in the 1000 genomes project. The numbers ranged from an average of about 6,000 singletons per sequenced FIN individual to slightly more than 20,000 per sequenced individual from LWK [106]. SF12 represents a population that contributed to modern day European´s ancestry but with no direct continuity to any extant population [81,92,107]. Therefore, it is likely that some of the genetic variation present in SF12 has been lost since.

First, the base qualities of all Ts in the first five base pairs of each read together with all As in the last five base pairs were set to 2. This was done in order to avoid residual deamination among the last bases of each fragment. Further, we used Picard [108] to add read groups to the files. Indel realignment was conducted with GATK 3.5.0 [109] using indels identified in phase 1 of the 1000 genomes project as reference [106]. Finally, GATK's UnifiedGenotyper was applied 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.

GATK's VariantFiltration was used to filter variants applying the conservative filters QD < 3.0 || FS > 60.0 || MQ < 35.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0 || MQ0 >=5 and GQ < 50 || DP > 100. Call sets were created with different minimum coverages between 10 and 80. Last, we used bedtools [110] to restrict to regions uniquely mappable with 35 base pair reads [111] and evaluated the results using GATK's VariantEval. Using the transition-transversion ratio of comparable sites in dbSNP as a reference, we observe that the Ti/Tv ratio of novel SNPs in SF12 is too low for minimum coverages <45. This likely suggests an enrichment of false positives as the Ti/Tv ratio of random calls would be 1. We conclude that restricting the calls of new SNPs to sites with at least 55x coverage should provide high quality calls (Figure S3.2). This resulted in 5,502 autosomal SNP sites not reported in dbSNP. As this analysis excludes more than 40% of the human genome, we estimate that the total number of unknown SNP sites in SF12 would be approximately 10,600. This number is similar to the numbers of singletons found per European genome in the 1000 genomes project: 6,000 SNPs per Finnish genome, 9,500 SNPs per British genome, 12,000 SNPs per Spanish or CEU genome, and 14,500 per Tuscan genome [106].

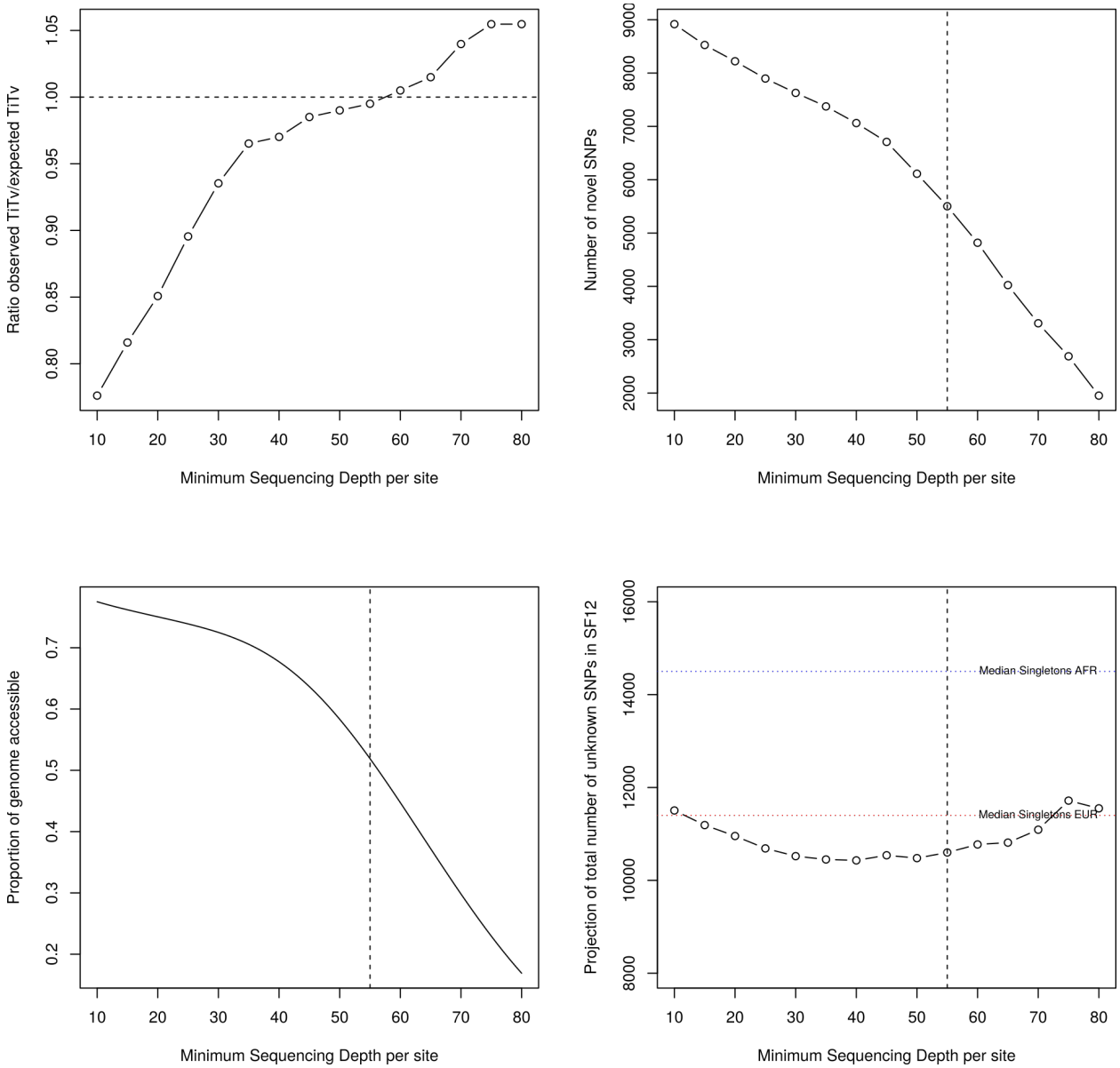


Figure S3.2 Quality control and number of novel variants in SF12. (a) Ratio of observed (called novel SNPs) and expected (for true SNPs) transition/transversion ratio as a function of the minimum sequencing depth per site considered in the analysis. (b) Number of novel variants called as a function of the minimum sequencing depth per site. (c) Proportion of the genome accessible when applying a minimum sequencing depth filter as a function of the minimum sequencing depth per site. (d) Projected number of novel variants (assuming the full human reference genome was accessible for SF12) as a function of the minimum sequencing depth per site. Dotted horizontal lines represent median numbers from the 1000 genomes project.

We also annotated the previously unobserved variants in SF12's genome using SNPeff 4.2 [112]. The novel SNPs are more common in genic regions than known SNPs also called in SF12 (Figure S3.3) which suggests that these novel variants could be younger and that they have not been subject to as much purifying selection. Only four of the novel SNPs in SF12 are annotated as "high impact", which includes such annotations as START_LOST, STOP_GAINED and mutations at splice sites [112]. One of those high impact SNPs falls on a splice acceptor site in *RP11-110I1.12*, the second SNP adds a stop codon to *REP15* and the third SNP affects a splice donor site in *PIGW*. Finally, a SNP affects a protein-protein binding site in *HSPA2*, a heat shock protein known to be involved in response to cold and heat. We did not find sequencing reads supporting these high impact variants in the other SHGs which suggests that they are either at low frequencies in the SHG population or some of them represent false positives. In order to obtain an upper bound on how many of the novel variants in SF12 are singletons, we checked all other SHGs at all 3,883 SNP sites that might not be due to deamination damage (reference allele C and alternative allele T or reference allele G and alternative allele A). 3,874 of these SNP sites were covered by reads in at least one of the other SHGs and at 668 sites at least one of the reads represented the alternative allele. This suggests that at least 17.2% of those novel variants were more frequent in Mesolithic Scandinavians. Extending this analysis to other prehistoric genomes studied in this paper increases this percentage to 24.2%.

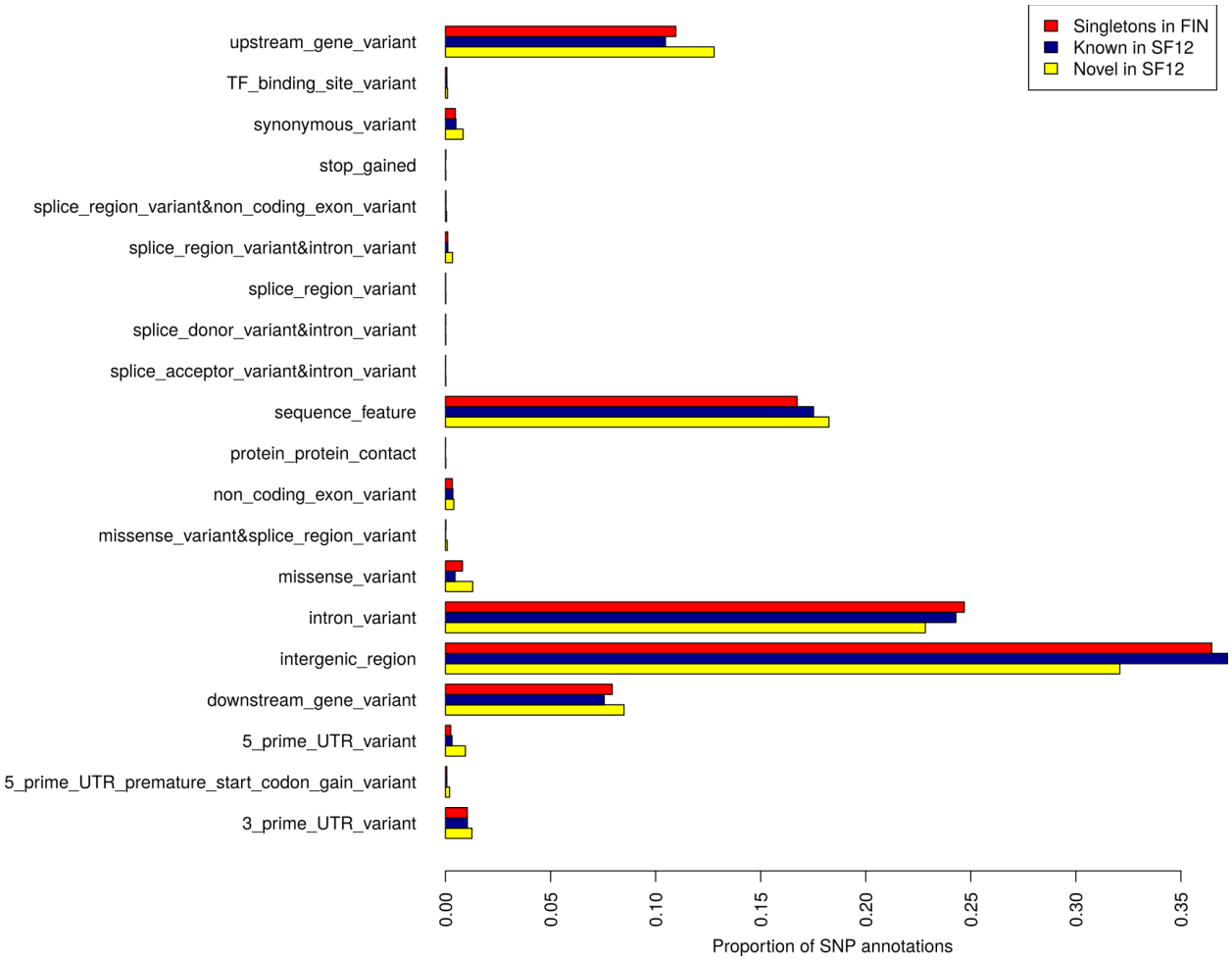


Figure S3.3 Annotation of novel and known SNPs called in SF12 compared to all singletons in the 1000 genomes FIN population.

S4 Estimates of contamination

Contamination was estimated using three different sources of data: (I) the mitochondrial genome, (II) the X chromosome if the individual was male and (III) the autosomes. Low contamination estimations over the three different approximations were interpreted as data mapping to the human genome being largely endogenous.

S4.1 Mitochondrial contamination

In order to estimate mtDNA contamination we used private or near-private consensus alleles [103] (<5% in 311 modern mtDNAs) with a base and mapping quality of 30 or higher as well as minimum 10x coverage for the ancient samples. To compensate for post-mortem damage, we filtered away positions with a consensus allele of either C or G, and where a transition substitution was detected. To obtain an estimate, the counts of consensus and alternative alleles were added together across all sites (Table S4.1).

Table S4.1 Mitochondrial contamination estimates

Sample	Point estimate	Informative sites	Consensus alleles	Total alleles	Lower C.I.	Higher C.I
Hum1	0.2898550725	2	1032	1035	0	0.6173812545
Hum2	0.1531393568	2	652	653	0	0.4506258182
SBj	3.719008264	6	466	484	2.033164827	5.404851702
SF9	5.357142857	4	106	112	1.869360277	9.527349687
SF11	3.418803419	2	113	117	0.1261463881	6.711460449
SF12	0.337938306	10	36864	36989	0.2787952963	0.3970813157
Steigen	0	2	506	506	0	0.5902928365

S4.2 X chromosome contamination

Contamination in male individuals (Hum2, Steigen, SF11 and SBj) was estimated using a method that examines heterozygous sites within the X chromosome. The method was first described in Rasmussen et al. [113] and is now implemented as part of ANGSD [114]. The X chromosome contamination module of ANGSD v.0.902 was run with two steps as described in the software manual. In the first step, a binary count file was built using the command “angsd -r X:5000000-154900000 -doCounts 1 -setMinDepth 3 -setMaxDepth 100 -iCounts 1 -minMapQ 30 -minQ 30”. In the second step the contamination estimate was obtained with the command “contamination -d 3 -e 100”. In this latter step, only transversion polymorphisms were screened

to avoid bias due to post-mortem DNA damage. We report contamination estimates and the confidence interval from method 1, which samples all reads from each site, thereby producing a more precise and sensitive estimate compared to method 2, which randomly samples one read per site (Table S4.2).

S4.3 Contamination in the nuclear genome

In addition to the estimates obtained from the two approaches described above, we used the *verifyBamId* tool [115], previously used by [116], to estimate nuclear contamination on ancient samples. This method estimates autosomal contamination using the 1000 Genomes reference panel, and thus provides a direct estimate of the nuclear contamination.
To estimate contamination at the sample level, we ran VerifyBamID v.1.1.2 [115] with the following command "verifyBamID -vcf <1000GenomesSitesFile> -bam <bamfile> -out <output> -verbose -ignoreRG". The 1000 genome vcf file (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ALL.wgs.phase3_shapeit2_mvncall_integrated_v5b.20130502.sites.vcf.gz) was further filtered to contain only transversion sites to avoid overestimation of contamination due to post-mortem damage. The program reports the contamination estimates in the FREEMIX column. It is important to note that while this method is a powerful direct tool to estimate nuclear contamination in modern samples, its accuracy to estimate contamination in ancient low coverage samples has not been formally tested. However, the consistency and low contamination estimates obtained using the three different methods for our low coverage samples such as SBj, indicate that our estimations are robust (Table S4.1 and S4.2).

Table S4.2 Two Nuclear contamination estimates.

	ANGSD Xchr		**VerifyBamID**			
Sample	**Contam [%]**	**# SNPs used**	**Contam [%]**	**Avgr Cov at SNPs**	**#Reads used**	**#SNPs used**
Steigen	0.4	2170	0	1.33	7042446	5286678
SF9	NA	NA	0	0.7	3682899	5286678
SF11	NA	NA	10.159	0.11	592735	5286678
SBj	1.4	133	0.063	0.49	2599852	5286678
Hum2	0.63	22209	0.73	4.42	23343202	5286678
Hum1	NA	NA	0	0.82	4323981	5286678
SF12	NA	NA	0.932	19.25	101772227	5286678

S5 Uniparental markers

S5.1 Y chromosome analyses

We used BAM files mapped to hg19 to call all single base substitutions from Phylotree [117] (version of 09/03/2016) using samtools mpileup. We extracted sites with mapping and base quality of at least 30. We excluded insertions and deletions and sites that displayed multiple alleles. We adopted a less stringent strategy, where we kept transition sites and A>T and G>C SNPs, to recover as many haplogroup defining substitutions as possible. Therefore, we report all derived states in the hierarchal phylogeny for each individual. We also report ancestral states downstream of the most derived allele to show the certainty of the haplogroup call within that branch. Additionally, we double checked that we had no ancestral alleles upstream of the defined haplogroup that would contradict the call. We used the nomenclature of the International Society of Genetic Geneaology (ISOGG) version 11.224 (http://isogg.org). For sites not present in ISOGG, we used the definitions in the minimal reference phylogeny of Phylotree (http://www.phylotree.org/Y/tree/). We caution that the allele state at a few sites may be erroneous due to post mortem deaminations, contamination or strand misidentifications. These were mainly seen as derived alleles sporadically scattered over the phylogeny and were usually contradicted by ancestral alleles at upstream positions.

Hum2
Individual Hum2 displayed the derived state at 20 sites leading in a hierarchal order to haplogroup I2-M438 (Table S5.1). We could not determine if Hum2 follows the I2c lineage as there were no sequence data for the defining markers L596 and L597. We could, however, exclude I2a and its subhaplogroups as Hum2 was ancestral for I2a-L460 as well as for an additional 20 downstream sites.

We noted a discrepancy in ancestral and derived allele state for CTS616 in the Y-SNP marker list of ISOGG and Phylotree, where the former describe the mutation as G>C and the latter as C>G. We assume that ISOGG is correct as it was most recently updated (9 December 2016) and as it makes phylogenetic sense with our data. Four additional sites displayed derived alleles for M118 (A>T), M236 (G>C), Y40 (C>T) and M417 (G>A). These are most likely erroneous because (I) they do not fit into the I2 phylogeny indicated by the majority of the sequence data (they define A1b1b2b1, B1, R1a and R1a1a1), (II) multiple upstream sites displayed ancestral alleles, and (III) they may have been caused by post mortem damage and/or possible strand misidentifications.

Steigen
Steigen belonged to I2a1b-M423 and displayed derived alleles at 14 sites leading to this haplogroup (Table S5.2). As this individual was ancestral for L161 and L621 downstream of I2a1b we conclude that he does not belong further down in this lineage. Ancestral states at 13

additional sites within the I2 phylogeny exclude the lineages I2a2 and I2c to further corroborate the I2a1b haplogroup call.

Steigen also displayed derived alleles for M281 (C>T) and Y40 (C>T) that defines E1b1b2 and R1a. As this individual displayed ancestral states for upstream sites (five and nine such sites respectively), these haplogroups do not make phylogenetic sense. Both sites were covered by a single read and they were C>T transitions that are likely caused by post mortem deamination.

SBj

SBj shows derived allele states at six sites leading to I2-L68 (xL597) (Table S5.3). We can exclude further downstream haplogroup calls within I2a2 (five sites with ancestral alleles, including I2a2-P217) and I2c (L597). As we had no data for the basal sites within I2a we could not determine if this individual belongs to I2 or to a subhaplogroup within I2a.

SBj displayed one additional site with a derived allele state at PF2259 (G>A) that defines G2a2b2b. This is likely erroneous and does not make phylogenetic sense as SBj also displayed ancestral alleles at three positions upstream of this haplogroup.

Haplogroup I

Three Mesolithic males in this study belong either to haplogroup I2 (Hum2 and SBj) or I2a1b Steigen). Unfortunately, there was not enough Y-chromosomal data from the SF11 male to define his haplogroup. Haplogroup I is one of the most frequently occurring lineages in pre-Neolithic samples from Europe from which both Y-chromosomal and genome-wide data have been generated. The earliest hg I individual has been found in a Gravettian context in the Paglicci cave in Italy and is dated to 34,580-31,210 BP [118]. This date is slightly older than a previous estimate of the age of the haplogroup (ca 20,000-25,000 years) [119], but more in line with newer estimates (ca 21,390-36,863 years) [120]. Haplogroups I and I2 have also been found in individuals dated to 16,000-12,830 BP from Magdalenian (Burkhardtshohle and HohleFels49), Azilian (Bichon) and Epipaleolithic (Rochedane) contexts in Germany, Switzerland and France respectively [118,121]. In line with our results, five other Mesolithic males from Sweden (Motala2, Motala3, Motala6, Motala9 and Motala12), belong to I2 lineages (I2a1, I2a1b and I2c) [81,82,118,122]. Three additional Mesolithic individuals from France (BerryAuBac and Chaudardes1) and Luxembourg (Loschbour) belong to I and I2a1b respectively [81,118]. Other, less prevalent, lineages found in pre-Neolithic European samples are C, F, J and R [82,118,121–126].

Haplogroup I is less common than other haplogroups in samples from the Early Neolithic and onwards. We note, however, that I lineages have been found in Early Neolithic males from Turkey (I0724 and I1096), Hungary (KO1, NE7) and Spain (Troc5) [82,98,122] as well as in Middle Neolithic to Bronze Age individuals from Germany (ESP30, ESP24, EUL57, ESP2, ESP4), Hungary (RISE254, RISE247), Italy (RISE487, RISE489, RISE486), Russia (RISE552)

and Spain (Mina4, MIR5/MIR6, MIR14, MIR19, MIR21, MIR25, MIR11, ATP12-1420) [82,95,122,127].

Four post-Mesolithic males from Scandinavia also belong within haplogroup I. These are one Scandinavian Middle Neolithic individual from a Pitted Ware hunter-gatherer context (Ajv58), who most likely belong to I2a1, and three Late Neolithic to Bronze Age individuals (RISE179, RISE210, RISE175) that were defined as haplogroup I [92,127]. The I and I2 lineages found in ancient Scandinavia are more common in the Balkans and Eastern Europe today [128]. Instead another I-lineage (I1-M253) is more prevalent in present-day Scandinavia [128–131].

Table S5.1 Y-chromosomal support for haplogroup I2 in individual Hum2. All markers are found in the ISOGG and Phylotree SNP databases (http://isogg.org/tree/ISOGG_YDNA_SNP_Index.html and http://www.phylotree.org/Y/marker_list.htm), except for markers with and asterisk (*) that are only found in Phylotree.

Haplogroup	SNP/ marker	Position (hg19)	Substitution	Observed state	No. reads	Inference
A0-T	L1085	2790726	T>C	C	3	derived
A0-T	L1130	16661010	T>G	G	3	derived
A1	V168	17947672	G>A	A	5	derived
A1	V171	4898665	C>G	G	3	derived
A1b	P108	15426248	C>T	T	3	derived
A1b	V221	7589303	G>T	T	3	derived
BT	M42	21866840	A>T	T	1	derived
BT	P97	14886273	G>T	T	2	derived
CT	M168	14813991	C>T	T	1	derived
CF	P143	14197867	G>A	A	3	derived
F	M213	15526751	T>C	C	1	derived
F	M89	21917313	C>T	T	4	derived
F	P14	17398598	C>T	T	4	derived
GHIJK	F1329	8589031	C>T	T	2	derived
IJK	M523	6753519	A>G	G	3	derived
I	M170	14847792	A>C	C	2	derived
I	M258	15023364	T>C	C	1	derived
I	U179	16354708	G>A	A	1	derived
I2	M438	16638804	A>G	G	3	derived
I2	L68	18700150	C>T	T	3	derived
I2a	L460	7879415	A>C	A	4	ancestral
I2a1a	CTS595	6874115	C>T	C	1	ancestral
I2a1a1	M26	21865821	G>A	G	3	ancestral
I2a1a2a1	L1287	21970862	G>T	G	3	ancestral
I2a1b	M423	19096091	G>A	G	1	ancestral
I2a1b2	L621	18760081	G>A	G	2	ancestral
I2a2	M436	18747493	G>C	G	4	ancestral
I2a2	P217	7628484	C>T	C	1	ancestral
I2a2a	M223	21717307	G>A	G	2	ancestral
I2a2a	U250	18888200	C>G	C	2	ancestral
I2a2a1	CTS616	6906332	G>C	G	6	ancestral
I2a2a1	CTS9183	18732197	A>G	A	2	ancestral
I-M284	CTS4544*	15707131	C>T	C	1	ancestral
I2a2a1a2a	L1229	14937828	C>A	C	5	ancestral
I2a2a1b	CTS10057	19232160	C>T	C	1	ancestral
I2a2a1b	CTS10100	19255890	G>A	G	5	ancestral
I2a2a1b1	L701	6753316	C>T	C	2	ancestral

I2a2a1b2a	L801	21763755	A>C	A	2	ancestral
I2a2a1b2a1	CTS1977	14140273	G>A	G	5	ancestral
I2a2a1b2a2	CTS6433	16889964	T>C	T	3	ancestral
I2a2b	L38	15668070	A>G	A	4	ancestral

Table S5.2 Y-chromosomal support for haplogroup I2a1b in individual Steigen. All markers are found in the ISOGG and Phylotree SNP databases (http://isogg.org/tree/ ISOGG_YDNA_SNP_Index.html and http://www.phylotree.org/Y/marker_list.htm), except for markers with and asterisk (*) that are only found in Phylotree.

Haplogroup	SNP/marker	Position (hg19)	Substitution	Observed state	No. reads	Inference
A0-T	L1155	22191266	G>C	C	1	derived
CT	M168	14813991	C>T	T	4	derived
F	M213	15526751	T>C	C	1	derived
F	M89	21917313	C>T	T	1	derived
GHIJK	F1329	8589031	C>T	T	1	derived
IJK	M522	7173143	G>A	A	1	derived
IJ	M429	14031334	T>A	A	1	derived
IJ	P126	21225770	C>G	G	1	derived
I	M170	14847792	A>C	C	2	derived
I	M258	15023364	T>C	C	1	derived
I	U179	16354708	G>A	A	1	derived
I2	M438	16638804	A>G	G	1	derived
I2a	L460	7879415	A>C	C	1	derived
I2a1b	M423	19096091	G>A	A	1	derived
I2a1b1	L161	22513718	C>T	C	3	ancestral
I2a1b2	L621	18760081	G>A	G	1	ancestral
I2a1a1	M26	21865821	G>A	G	2	ancestral
I2a1a2a1	L1287	21970862	G>T	G	1	ancestral
I2a2	M436	18747493	G>C	G	4	ancestral
I2a2a	U250	18888200	C>G	C	1	ancestral
I-M284	CTS4544*	15707131	C>T	C	1	ancestral
I2a2a1a2a	L1229	14937828	C>A	C	3	ancestral
I2a2a1b	CTS10057	19232160	C>T	C	1	ancestral
I2a2a1b	CTS10100	19255890	G>A	G	3	ancestral
I2a2a1b2	Z161	2696497	C>G	C	1	ancestral
I2a2a1b2a	L80	14640715	A>G	A	1	ancestral
I2a2a1b2a1	CTS1977	14140273	G>A	G	1	ancestral
I2c	L596	14197631	G>A	G	1	ancestral
I2c	L597	18887888	T>A	T	1	ancestral

Table S5.3 Y-chromosomal support for haplogroup I2 in individual SBj. All markers are found in the ISOGG and Phylotree SNP databases (http://isogg.org/tree/ ISOGG_YDNA_SNP_Index.html and http://www.phylotree.org/Y/marker_list.htm), except for markers with and asterisk (*) that are only found in Phylotree.

Haplogroup	SNP/ marker	Position (hg19)	Substitution	Observed state	No. reads	Inference
A0-T	L1130	16661010	T>G	G	3	derived
BT	P97	14886273	G>T	T	1	derived
GHIJK	F1329	8589031	C>T	T	1	derived
HIJK	M578	7202703	C>T	T	2	derived
I	U179	16354708	G>A	A	1	derived
I2	L68	18700150	C>T	C	1	derived
I2a1a2a1	L1287	21970862	G>T	G	1	ancestral
I2a1b2a1a	S17250	15531354	G>A	G	1	ancestral
I2a2	P217	7628484	C>T	C	1	ancestral
I2a2a	M223	21717307	G>A	G	1	ancestral
I2a2a1a2a	L1229	14937828	C>A	C	1	ancestral

I2a2a1b1	L701	6753316	C>T	C	2	ancestral
I2a2a1b2	Z161	2696497	C>G	C	1	ancestral
I2c	L597	18887888	T>A	T	4	ancestral

S5.2 Mitochondrial DNA results

Strict consensus sequences were generated using samtools' mpileup and vcfutils.pl [104,132]. A minimum base quality and mapping quality score of 30, and a coverage of at least three, were used to call the consensus sequences. Ambiguous SNPs in the consensus sequences were examined manually using samtools v1.3 and the mpileup command. The majority nucleotide (>75% of the reads) was called if the ambiguous position was covered by at least 10 reads with a minimum base quality and map quality score of 30. Haplogroups were assigned to the sequences using HaploFind [133] and PhyloTree mtDNA Build 17 [117]. The mutations are reported against the Reconstructed Sapiens Reference Sequence, RSRS [134]. The mitochondrial coverage, haplogroups, mutations supporting the called haplogroup and private mutations are reported in Table S5.4. There were a few regions where none of the consensus sequences (or only SF12) had any data after filtering. The majority of these positions are situated between the *ND1* and *CO3* genes and has previously been reported as regions that are difficult to map when working with short sequence reads [135].

SF9 belongs to U4a2 (Table S5.4). This mitochondrial genome has 46 haplogroup defining mutations (and no data for defining sites at np T4646C, T5999C, A6047G, A7521G and C8818T). SF9 also has two private mutations leading to U4a2f (A1978G and A12397G). As SF9 displays the ancestral state for the two additional defining positions for U4a2f (T1189C and G15172A), it is likely an ancestral lineage to U4a2f and we report it as U4a2.

SF11 belongs to U5a1, as previously reported [92]. This individual displayed 49 mutations and lacked data for one site leading to this haplogroup (A7521G, Table S5.4). SF11 had one private mutation, which is known to be recurrent (C16519T).

SF12 displays derived alleles at all 51 positions leading to U4a1 (Table S5.4). SF12 also has two private mutations, where one is a unique C6617T transition and the second is a recurrent G13708A mutation. Given these private mutations, it is likely that SF12 belongs to a previously unknown U4a1 sub-haplogroup.

SBj belongs to U4a1 (Table S5.4). This individual has 48 of the haplogroup defining mutations, and for the remaining three positions, there is no data available (np 4646, 5999 and 6047). Like SF12, SBj also has the private G to A mutation at np 13708A. However, SBj is ancestral at np 6617 where SF12 is derived.

Hum1 has all 50 SNPs leading to U5a1 (Table S5.4). This individual has two private mutations at positions that are known to be recurrent (G4769A and 16519T). Hum1 also has two chimeric sites with both C and T reads. These are most likely caused by deamination (C4799T and C7115T), but as the number of C reads were below the 75% SNP calling cut off (65% and 69% respectively), they were kept as Y in the consensus sequences.

Hum2 has 50 mutations leading to U5a1d (Table S5.4) but lacks the C152T transition. As this mutation defines an early branch in the phylogeny (L2'3'4'5'6), the most likely explanation for C at this site in Hum2 is that a back-mutation has occurred. Hum2 has a private mutation at a position that is known to be recurrently mutating (C16519T). Hum2 may share the G4769A transition that Hum1 had. However, Hum2 is chimeric and the amount of A reads were 62% (not above the 75% we use for SNP calling) which would mean that 38% of the reads could be due to G to A damage. Therefore Hum2 was kept as R at this particular site in the consensus sequence.

Steigen, like Hum2, has 50 mutations leading to U5a1d and lacks the C to T transition at np 152 (Table S5.4). Steigen has two private SNPs at positions that mutate recurrently (A189G and C16519T). Steigen is also chimeric for G4769A (as Hum2) and as 56% of the reads were As and 44% Gs, and was, like Hum2, kept as R in the consensus sequence.

The three newly reported Swedish Mesolithic individuals belong to U4 lineages. More specifically, SF12 from Stora Karlsö and SBj from Gotland belong to U4a1, and SF9 to U4a2 (Table S5.4, Figure S5.1). Only one other complete pre-Neolithic U4a genome, from an eastern hunter-gatherer from Karelia (I0211) contextually dated to 7,450-6,950 cal BP, has been reported [122]. This makes the Swedish Mesolithic U4 lineages the earliest thus far. However, the estimated coalescence time for the haplogroup is about 20,000 years BP [136,137]. The upstream haplogroup U and U2'3'4'7 are present in pre- and post-glacial Europe (Cioclovina1, Malta1, Paglicci108 and Rigney1) [118,138,139], making it plausible to discover more U4 haplotypes as more ancient samples are analyzed.

The three Mesolithic Norwegian samples from Hummerviksholmen and Steigen belong to U5a1 lineages. Hum1 belongs to U5a1, and Hum2 and Steigen belong to U5a1d (Table S5.4, Figure S5.1). Also the Swedish Stora Förvar sample, SF11 belongs to the same lineage as Hum1 (U5a1) [92]. In addition to our Mesolithic samples, two U5a1 haplotypes are found in Scandinavian hunter-gatherers from mainland Sweden dated to 7,900-7,500 cal BP (Motala1 and Motala3), one U5a1c has been found in a hunter-gatherer dated to 8,417-8,199 cal BP from Latvia (Latvia_HG1) [126], and one U5a1d in an eastern hunter-gatherer from Russia dated to 7,490-7,505 cal BP (Samara) [81,82,122] (Figure S5.1). The Norwegian Hum1 and Hum2 are the oldest representatives of U5a1 lineages found thus far. U5 is by far the most commonly found haplogroup among European Mesolithic individuals (24 of 36 individuals). U5a2 lineages are found in seven individuals from Sweden, France and Germany (Motala4, Motala6, Motala9, LesCloseaux13, MareuilLesMeaux1, Felsdach and Blätterhöhle20) and U5b lineages are found

in 11 individuals from Spain, France, Luxembourg, Germany, Italy and Ukraine (LaBrana1, Ranchot88, BerryAuBac, Chaudardes1, Loschbour, Bockstein, Ofnet, Falkenstein and HohlensteinStadel, Continenza, Ukraine_HG1) [81,82,118,122,125,126,138,140,141] (Figure S5.1). U5 is thought to have arisen in Upper Paleolithic Europe and molecular dating of modern mitochondrial genomes estimates the age of the haplogroup to ca. 25,000-36,000 years [136,142–144]. This is in line with recent findings of six pre-LGM individuals dated to between 31,000 and 26,000 BP carrying U5 haplotypes (KremsWA3, Vestonice15, Pavlov1, Vestonice16, Vestonice43 and Goyet2878-21) [118,138,145] as well as with U5b lineages found in individuals from Spain, Italy, France, Germany and Switzerland dated to 19,000-12,000 BP (ElMiron, Paglicci71, Villabruna, Iboussieres31-2, Iboussieres25-1, Iboussieres39, Rochedane, Oberkassel and Bichon) [118,121,138]. Except for the U4 and U5 lineages, there are also a few less common haplogroups (and subtypes thereof) in individuals from Mesolithic Europe. These are U2e lineages found in Sweden (Motala2 and Motala12) and Latvia (Latvia_HG2), C1 in Russia (Karelia), H13c in Georgia (Kotias) and R3 in Hungary (KO1, which is Neolithic but has been shown to be part of the Mesolithic gene pool) [81,82,98,122,126,146] (Figure S5.1).

Table S5.4 Mitochondrial coverage, haplogroup assignment, polymorphisms supporting the assigned haplogroup and private mutations in the Swedish and Norwegian Mesolithic consensus sequences.

Sample	MT coverage	MT hg	Polymorphisms for called hg (against RSRS)	Private mutations
SF9	93	U4a2	146T 152T 247G 310C 499A 769G 825T 1018G 1811G 2758G 2885T 3594C 4104A 4312C 7146A 7256C 7521G 8468C 8655C 8701A 9540T 10398A 10664C 10688G 10810T 10873T 10915T 11332T 11467G 11914G 12308G 12372A 12705C 13105A 13276A 13506C 13650C 14620T 15693C 16129G 16187C 16189T 16223C 16230A 16278C 16311T 16356C	1978G 12397G
SF11	45.	U5a1	146T 152T 195T 247G 769G 825T 1018G 2758G 2885T 3197C 3594C 4104A 4312C 7146A 7256C 8468C 8655C 8701A 9477A 9540T 10398A 10664C 10688G 10810T 10873T 10915T 11467G 11914G 12308G 12372A 12705C 13105A 13276A 13506C 13617C 13650C 14793G 15218G 16129G 16187C 16189T 16192T 16223C 16230A 16256T 16270T 16278C 16311T 16399G	16519T
SF12	9774	U4a1	146T 247G 499A 769G 825T 1018G 1811G 2758G 2885T 3594C 4104A 4312C 4646C 5999C 6047G 7146A 7256C 7521G 8468C 8655C 8701A 8818T 9540T 10398A 10664C 10688G 10810T 10873T 10915T 11332T 11467G 11914G 12308G 12372A 12705C 12937G 13105A 13276A 13506C 13650C 14620T 15693C 16129G 16134T 16187C 16189T 16223C 16230A 16278C 16311T 16356C	6617T 13708A
SBj	102	U4a1	146T 247G 499A 769G 825T 1018G 1811G 2758G 2885T 3594C 4104A 4312C 7146A 7256C 7521G 8468C 8655C 8701A 8818T 9540T 10398A 10664C 10688G 10810T 10873T 10915T 11332T 11467G 11914G 12308G 12372A 12705C 12937G 13105A 13276A 13506C 13650C 14620T 15693C 16129G 16134T 16187C 16189T 16223C 16230A 16278C 16311T 16356C	13708A
Hum1	597	U5a1	146T 152T 195T 247G 769G 825T 1018G 2758G 2885T 3197C 3594C 4104A 4312C 7146A 7256C 7521G 8468C 8655C 8701A 9477A 9540T 10398A 10664C 10688G 10810T 10873T 10915T 11467G 11914G 12308G 12372A 12705C 13105A 13276A 13506C 13617C 13650C 14793G 15218G 16129G 16187C 16189T 16192T 16223C 16230A 16256T 16270T 16278C 16311T 16399G	4769A 16519T
Hum2	432	U5a1d	146T 195T 247G 769G 825T 1018G 2758G 2885T 3027C 3197C 3594C	16519T

			4104A 4312C 7146A 7256C 7521G 8468C 8655C 8701A 9477A 9540T 10398A 10664C 10688G 10810T 10873T 10915T 11467G 11914G 12308G 12372A 12705C 13105A 13276A 13506C 13617C 13650C 14793G 15218G 16129G 16187C 16189T 16192T 16223C 16230A 16256T 16270T 16278C 16311T 16399G	
Steigen	278	U5a1d	146T 195T 247G 769G 825T 1018G 2758G 2885T 3027C 3197C 3594C 4104A 4312C 7146A 7256C 7521G 8468C 8655C 8701A 9477A 9540T 10398A 10664C 10688G 10810T 10873T 10915T 11467G 11914G 12308G 12372A 12705C 13105A 13276A 13506C 13617C 13650C 14793G 15218G 16129G 16187C 16189T 16192T 16223C 16230A 16256T 16270T 16278C 16311T 16399G	189G 16519T

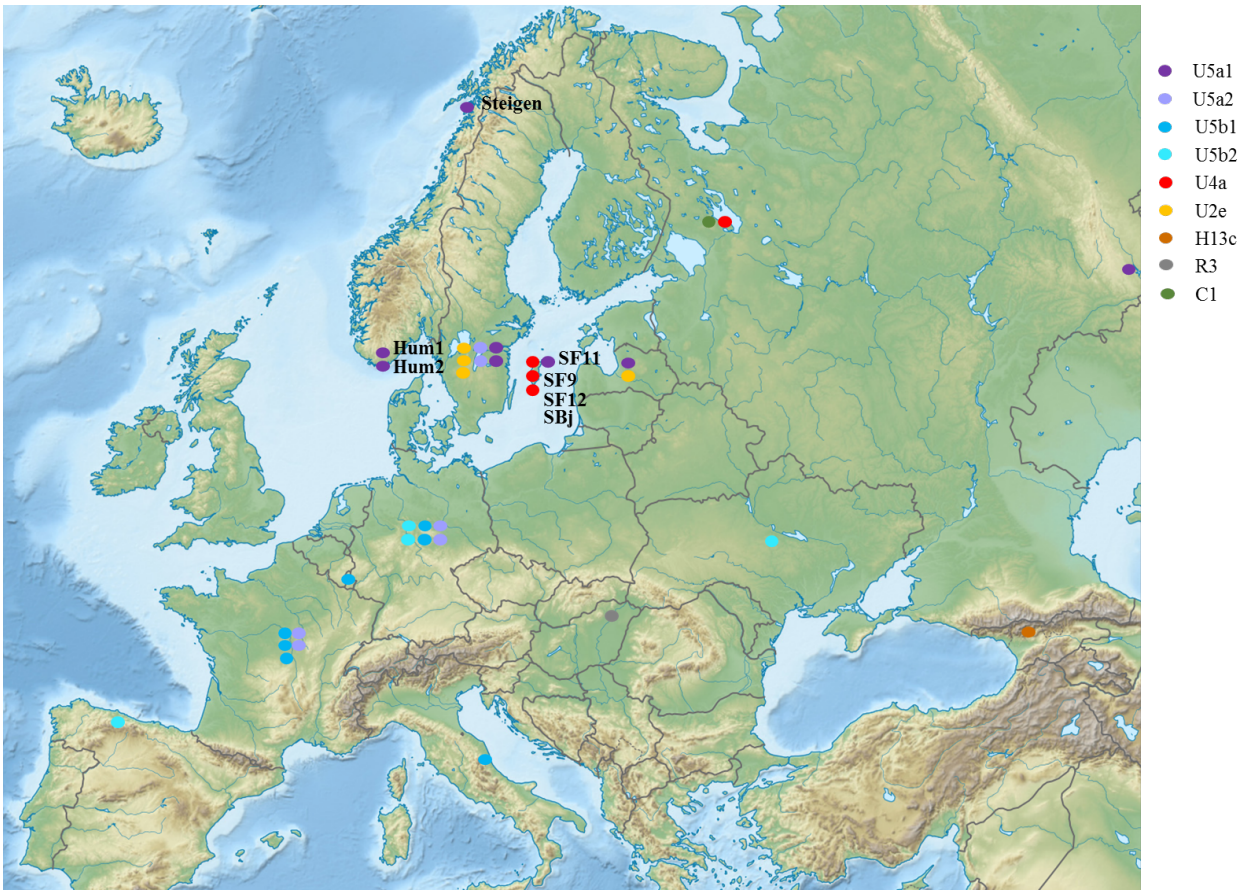


Figure S5.1 36 complete mitochondrial genomes from Mesolithic Europe including the three Norwegian samples Steigen, Hum1, Hum2, and the four Swedish samples SF9, SF11, SF12 and SBj. Seven previously published samples from Sweden [81,82,122], two from Latvia [126], one from Spain [125,141], five from France [118,138], one from Luxembourg [81], six from Germany [118,138,140], one from Italy [118], one from Hungary [98], one from Georgia [121], one from Ukraine [126] and three from Russia [81,82,122,146].

S6 Basic Population genomic analysis

S6.1 Reference data for comparison

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. The three reference SNP panels were (Table S6.1):

- The Human Origins genotype data set of 594,924 SNPs genotyped in 2,404 modern individuals from 203 populations [81,147].
- A panel of 1,055,209 autosomal SNPs which were captured in a set of ancient individuals by Mathieson et al [122].
- 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 [106]. Those SNPs were extracted using vcftools [132].

These data sets were merged with ancient individuals of less than 15x 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) 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.

Six high coverage ancient individuals (SF12, NE1 [98], Kotias [121], Loschbour [81], Stuttgart [81], Ust-Ishim [148]) 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 [108] to add read groups to the files. Indel realignment was conducted with GATK 3.5.0 [109] using indels identified in phase 1 of the 1000 genomes project as reference [106]. 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 [132] if they were not marked as low quality calls. Plink 1.9 [149,150] was used to merge the different data sets.

Table S6.1 Merged SNPs with genotype datasets

Individual	Number of SNPs overlapping with Human Origins	Number of SNPs overlapping with Mathieson2015 capture SNPs	Number of SNPs overlapping with 1000 genomes transversions	Reference
Bichon	102540	299715	1694024	Jones et al. (2015)
Kostenki14	65064	185278	954417	Seguin-Orlando et al. (2014)
LaBrana1	544119	958583	1435555	Olalde et al. (2014)
KO1	414976	745223	1039570	Gamba et al. (2014)
MA1	35313	99763	504287	Raghavan et al. (2014)
Mota	102604	300381	1699427	Gallego-Llorente2015
NE5	74916	203483	858144	Gamba et al. (2014)
NE6	73159	198106	974979	Gamba et al. (2014)
NE7	74626	202923	949847	Gamba et al. (2014)
Satsurblia	75549	221996	1158984	Jones et al. (2015)
Motala1/I0011	347021	587972	160770	Lazaridis et al. (2014), Mathieson et al. (2015)
Motala2/I0012	434301	731445	234654	Lazaridis et al. (2014), Mathieson et al. (2015)
Motala3/I0013	258476	298714	-*	Lazaridis et al. (2014), Mathieson et al. (2015)
Motala4/I0014	434714	698031	188548	Lazaridis et al. (2014), Mathieson et al. (2015)
Motala6/I0015	370945	556799	127750	Lazaridis et al. (2014), Mathieson et al. (2015)
Motala12/I0017	503322	857690	1457992	Lazaridis et al. (2014), Mathieson et al. (2015)
UzOO74/I0061	536949	957887	390487	Mathieson et al. (2015)
SVP44/I0124	282751	446671	99887	Mathieson et al. (2015)
UzOO40/I0211	73963	135875	25445	Mathieson et al. (2015)
BAR2/I0707	536086	941835	442887	Mathieson et al. (2015)
BAR6/I0708	525535	913818	389237	Mathieson et al. (2015)
BAR20/I0709	531783	920496	436761	Mathieson et al. (2015)
L11-216/I0736	462639	755679	218371	Mathieson et al. (2015)
M10-275/I0744	477950	822903	260522	Mathieson et al. (2015)
M11-363/I0745	533156	927837	464096	Mathieson et al. (2015)
L11-322/I0746	535594	935324	446015	Mathieson et al. (2015)
BAR26/I1096	390684	705558	147125	Mathieson et al. (2015)
BAR271/I1097	388345	702390	145044	Mathieson et al. (2015)
BAR99/I1098	409836	740392	159847	Mathieson et al. (2015)
M11-352a/I1101	339277	616629	122322	Mathieson et al. (2015)
M11-S-350/I1103	309136	561561	103912	Mathieson et al. (2015)
M13-72/I1579	423543	772168	200584	Mathieson et al. (2015)
L12-393/I1580	471099	855638	311850	Mathieson et al. (2015)
L12-502/I1581	425080	774697	213779	Mathieson et al. (2015)

L14-200/I1583	505043	914076	389573	Mathieson et al. (2015)
M11-59/I1585	427645	778039	232951	Mathieson et al. (2015)
LBK1992/I0025	476557	827905	244202	Mathieson et al. (2015)
LBK2155/I0026	485096	850232	234029	Mathieson et al. (2015)
HAL5/I0046	464381	829856	351506	Mathieson et al. (2015)
UWS4/I0054	547797	960468	626530	Mathieson et al. (2015)
HAL4/I0100	531901	922902	410387	Mathieson et al. (2015)
CB13	63867	183119	970351	Olalde et al. (2016)
Gok2	70093	187298	888578	Skoglund et al. (2014)
Hum1	36545	107443	562138	this study
Hum2	97637	286491	1603656	this study
SBj	30512	87942	574454	this study
SF11	11886	33728	171620	Skoglund et al. (2014) and this study
SF9	51038	146258	731468	this study
Steigen	76113	224241	1217162	this study
Ajv58	96273	281130	1531643	Skoglund et al. (2014) and this study
Ajv70	79856	228312	1171958	Skoglund et al. (2014) and this study
Kotias	97602	289363	1521960	Jones et al. (2015)
NE1	97147	1095486	1562267	Gamba et al. (2014)
LBK/Stuttgart	554264	290309	1611413	Lazaridis et al. (2014)
Loschbour	558885	1058564	1571764	Lazaridis et al. (2014)
Ust_Ishim	573301	1069826	1661490	Fu et al. (2014)
SF12	576797	1101606	1749144	this study
Zv313# (Latvia_HG1)	67736	241991	1017742	Jones et al. (2017)
Zv93# (Latvia_HG2)	98071	349433	1535533	Jones et al. (2017)
Zv121# (Latvia_HG3)	56402	200712	835147	Jones et al. (2017)

*Motala3 was excluded from the analysis of the shotgun BAM files as the BAM file seems identical to Motala12.
#The individuals from Jones et al. (2017) were named based on the burial number at the site. The original names used by Jones et al. (2017) are given in parentheses.

S6.2 Principal component analysis

We performed principal component analysis (PCA) to characterize the genetic affinities of the ancient Scandinavian genomes to previously published ancient and modern genetic variation. PCA was conducted on 42 present-day west Eurasian populations from the Human Origins dataset [81,147], using *smartpca [151]* with numoutlieriter: 0 and lsqproject: YES options. A total of 59 ancient genomes (52 previously published and 7 reported here) (Table S6.1) were projected into the reference PCA space, computed from the genotype 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 the EIGENSOFT [151] using with the –x and –k options.

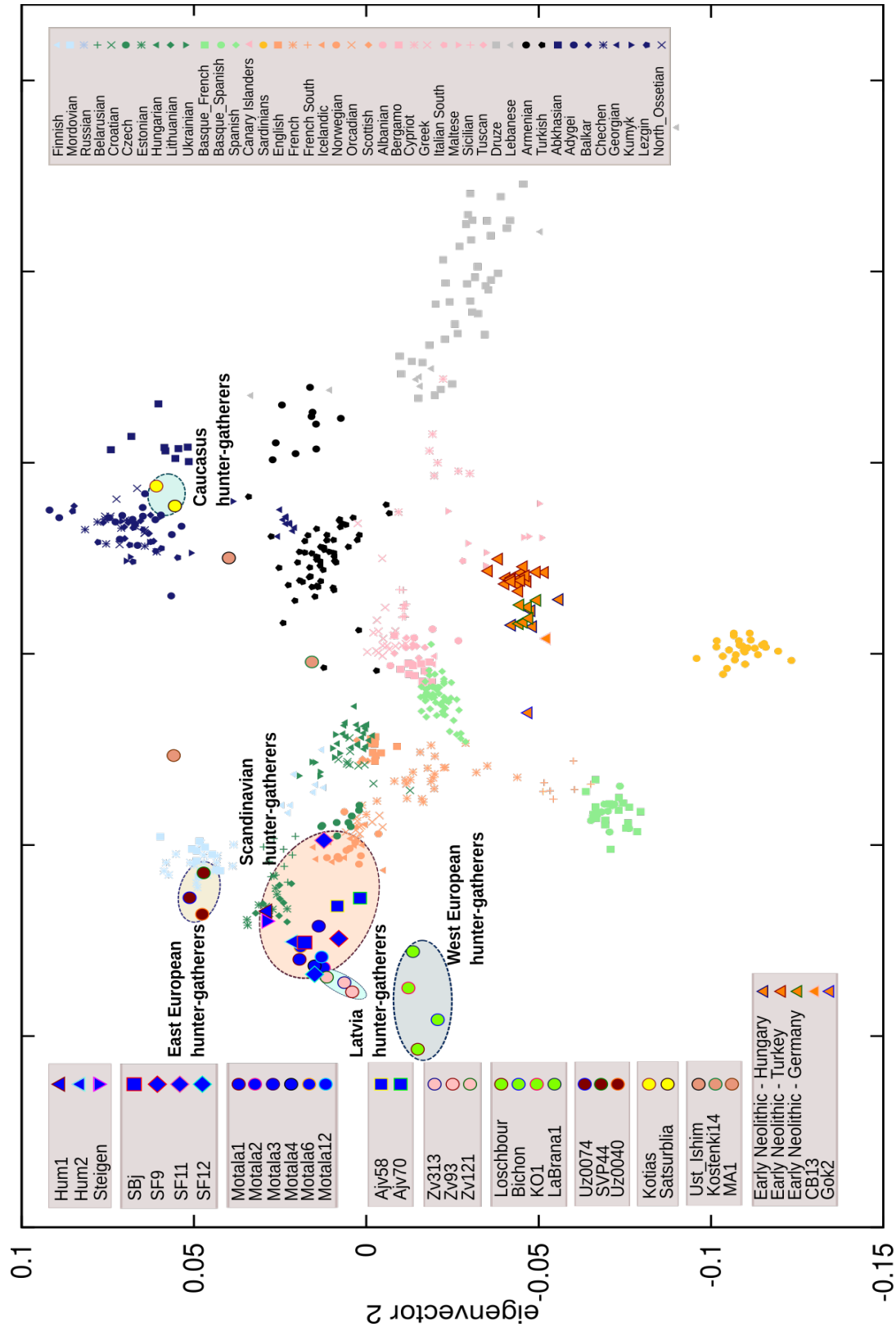


Figure S6.1 PCA of ancient individuals used in this study projected on a PCA of modern-day western Eurasians.

All Mesolithic individuals cluster tightly with individuals from a similar geographic region and cultural context (Figure 1 and Figure S6.1). The only exception appears to be SF11, instead of clustering with other Mesolithic individuals from Gotland, it clusters with Neolithic Pitted Ware culture (PWC) samples (also from Gotland). This grouping is consistent with a previous publication involving these individuals [92]. SF11 is also the lowest coverage individual with the highest estimate of nuclear contamination (Table 1, Table S4.2). To further investigate the clustering of SF11, we used SF9 as a representative of Mesolithic Gotland and down-sampled it to 34000, 19000, 15000, 8000 and 4000 SNPs. Additionally we produced ten different subsamples of SF9 with 10000 SNPs each to represent the noise caused by this low number of SNPs. While this subsampling added noise to the position of SF9 in the PCA, it was not possible to reproduce the position of SF11 (Figure S6.2). To assess whether modern contamination could have caused the clustering of SF11, we restricted the analysis to damaged reads using PMDTools [152] and PMD score cutoffs of 2 and 3. This filtering step changed the position of SF11 in the PCA but did not lead to a clustering with other Mesolithic samples from Gotland (here represented by SF9). Finally, we also tried to combine contamination and low coverage by artificially adding 10 and 20% nuclear contamination from a randomly chosen GBR sample from the 1000 genomes data [106] to the subsampled (8000 SNPs) SF9 sample. Adding contamination to SF9 results in a position closer to SF11's location in the PCA, but even 20% contamination is not sufficient to reproduce the clustering of SF11. We note that contamination from a source other than modern North-Western Europeans might cause other patterns. This led us to the conclusion that the position of SF11 in the PCA might be a combination of noise due to lower coverage, modern contamination, and possibly a slightly different genetic makeup of SF11.

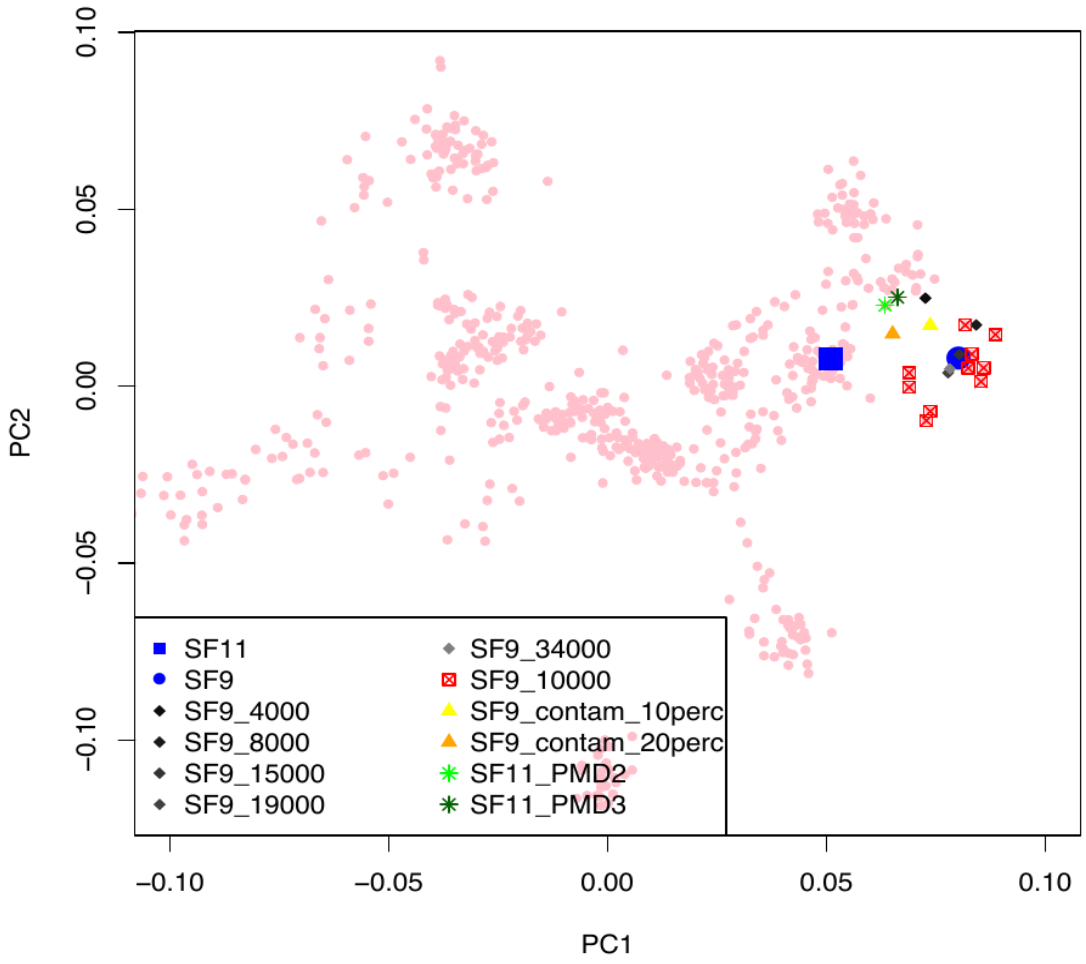


Figure S6.2 PCA for SF11 as well as different subsets and simulated contamination of SF9.

S6.3 D- and f-tests

The qpDstat program of ADMIXTOOLS was used to calculate *D*-statistics to test deviations from a tree-like population topology of the shape ((*A*,*B*);(*X*,*Y*)) [147]. Standard errors were calculated using a 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 while 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 [153], Yoruba [106] and Chimp genome as (A) outgroups. The software popstats [154] was used to calculate f4 statistics, in order to estimate shared drift between groups. Standard errors and Z scores for f4 statistics were estimated using a weighted block jackknife (Figure 1C).

S6.4 Model-based clustering

A model-based clustering algorithm, implemented in the ADMIXTURE software [155], was used to estimate ancestry components and to cluster individuals. ADMIXTURE was conducted on the Human Origins data set [81,147], 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 dataset was filtered for linkage disequilibrium using PLINK [149,150] 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 [156]. 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/f4 statistics. The results for all Ks are shown in Supplementary Figure 1.

In addition to ADMIXTURE, we assessed the admixture patterns in Mesolithic Scandinavians using a set of methods implemented in ADMIXTOOLS [147], qpWave [157] and qpAdm [81,82]. Both methods are based on f4 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, 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) = f4(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 (a) genetically western and central hunter-gatherers (Bichon, Loschbour, KO1, LaBrana), (b) Eastern hunter-gatherers (UzOO74/I0061, SVP44/I0124, UzOO40/I0211) (c) Norwegian hunter-gathers (Hum1, Hum2, Steigen) and (d) 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) while a rank of 1 is consistent with the data (p=0.699). Haak et al 2015 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, while 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 f4 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 populations in S to T. The results are shown in Figure 1.

S6.5 Admixture graphs

Relationships amongst ancient individuals were examined using a statistical framework as implemented in TreeMix [158]. Treemix builds a maximum likelihood tree of populations using the covariance matrix of allele frequencies and fits admixture edges to given populations. TreeMix was applied to SF12, Motala12 [81] (as representatives of Scandinavian hunter-gatherers from Sweden), Hum1, Hum2 (as representatives of Scandinavian hunter-gatherers from Norway), Loschbour [81] (as representative of west European hunter-gatherers), and UzOO74 [122] (I0061, as representative of east European hunter-gatherers) and the tree was rooted with the ancient Ethiopian Mota individual [153]. Since single individuals were used separately, TreeMix was run with "-noss" option to turn off the correction for low sample size and standard errors were estimated using blocks of 500 SNPs with (-k 500 -se) parameters.

Analysis was restricted with a total of 73,960 overlapping transversion SNPs that were ascertained in African Yoruba individuals from the 1000 Genomes project [106] and were genotyped in seven ancient individuals. TreeMix was run by modelling "known" gene flow from east European hunter-gatherers into all four Scandinavian hunter-gatherers using the (–cor-mig and –climb) parameters with 50 different random seeds. Starting proportion of admixture was set as 0.0 to allow 0% possibility of gene flow. The majority of runs (78%) supported the gene flow from EHGs to all SHGs. Hum1 and Hum2 received more EHG ancestry with admixture proportions of 48% and 45%, respectively while SF12 and Motala12 received approximately 12% EHG ancestry ($p < 0.05$). We report the graph with the highest likelihood and most common topology in Figure S6.3.

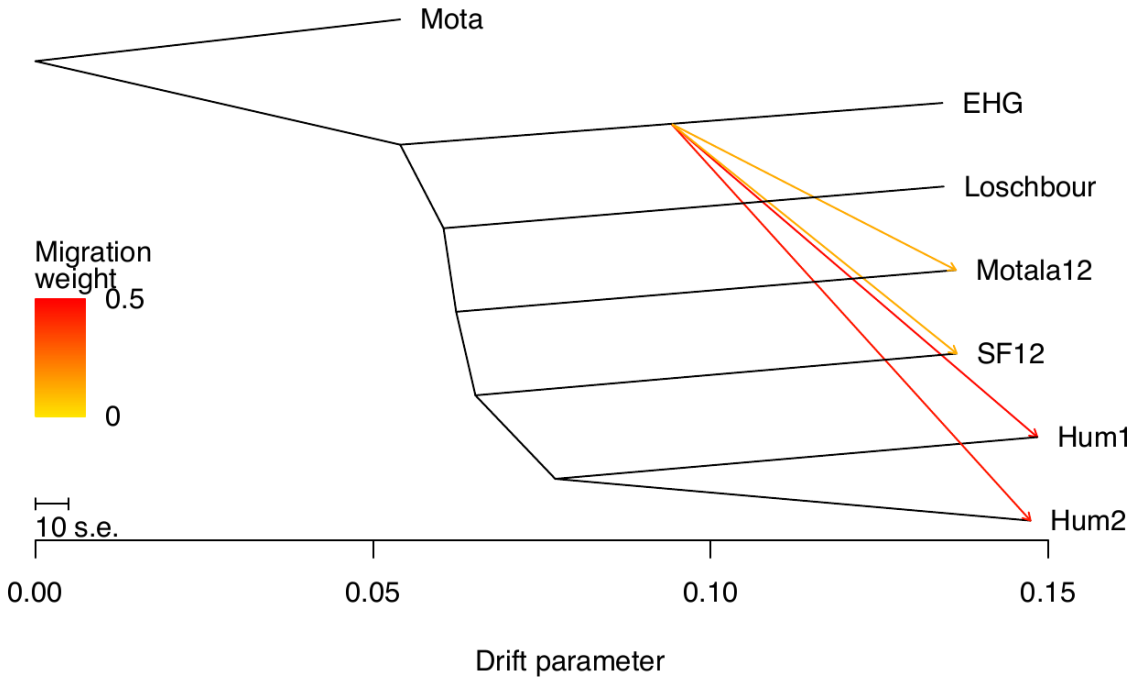


Figure S6.3 Treemix model of the data including gene flow from EHG into all SHGs.

S7 Diversity estimates

In order to assess genetic diversity and effective population size in different ancient populations, we used a number of methods described below.

S7.1 Pairwise mismatch

Conditional nucleotide diversity [92] is a simple diversity measure which should be approximately equivalent to the heterozygosity within a population. It is based on pairwise mismatches between two haploid ancient individuals combined into one pseudo-diploid individual. To avoid effects of post-mortem damage and ascertainment bias, we only calculated this estimate for 1,797,398 transversion SNPs ascertained in Yorubans [106]. Standard errors were estimated using a block jackknife procedure with a blocksize of 2000 SNPs.

First, diversities were calculated for a mixed data set of shotgun sequence data and SNP capture data. For the individuals from Motala, both shotgun sequence [81] and SNP capture data [82,122] was available. We noticed, however, that diversities for SNP capture from Motala (e.g. Motala12-Motala1: 0.22) were substantially higher than diversities estimated from shotgun sequence data of the same individuals (0.1997). The values obtained from SNP capture data would reach similar levels as shotgun-sequenced Neolithic individuals (e.g. NE5-NE6: 0.22). As it seems well established that Neolithic populations had a higher diversity than Mesolithic Europeans [81,92,95,159], we conclude that some technical bias in the comparison of capture data and shotgun-sequence data likely causes these inflated diversity estimates, which is why we restrict all diversity estimates in this paper to shotgun-data. A plausible explanation could be that the targeted capture of the alternative allele could cause less reference bias than shotgun sequencing. However, the differences between shotgun and SNP capture data are not expected to affect our results for PCA, Admixture, Treemix or f-tests in any way. Genotyping error or reference bias should just be observed as an excess of sample-specific drift without specific direction in those analyses. We note however, that diversity in SNP capture EHGs appeared to be higher (~0.24) than in SNP capture Motala data (~0.22).

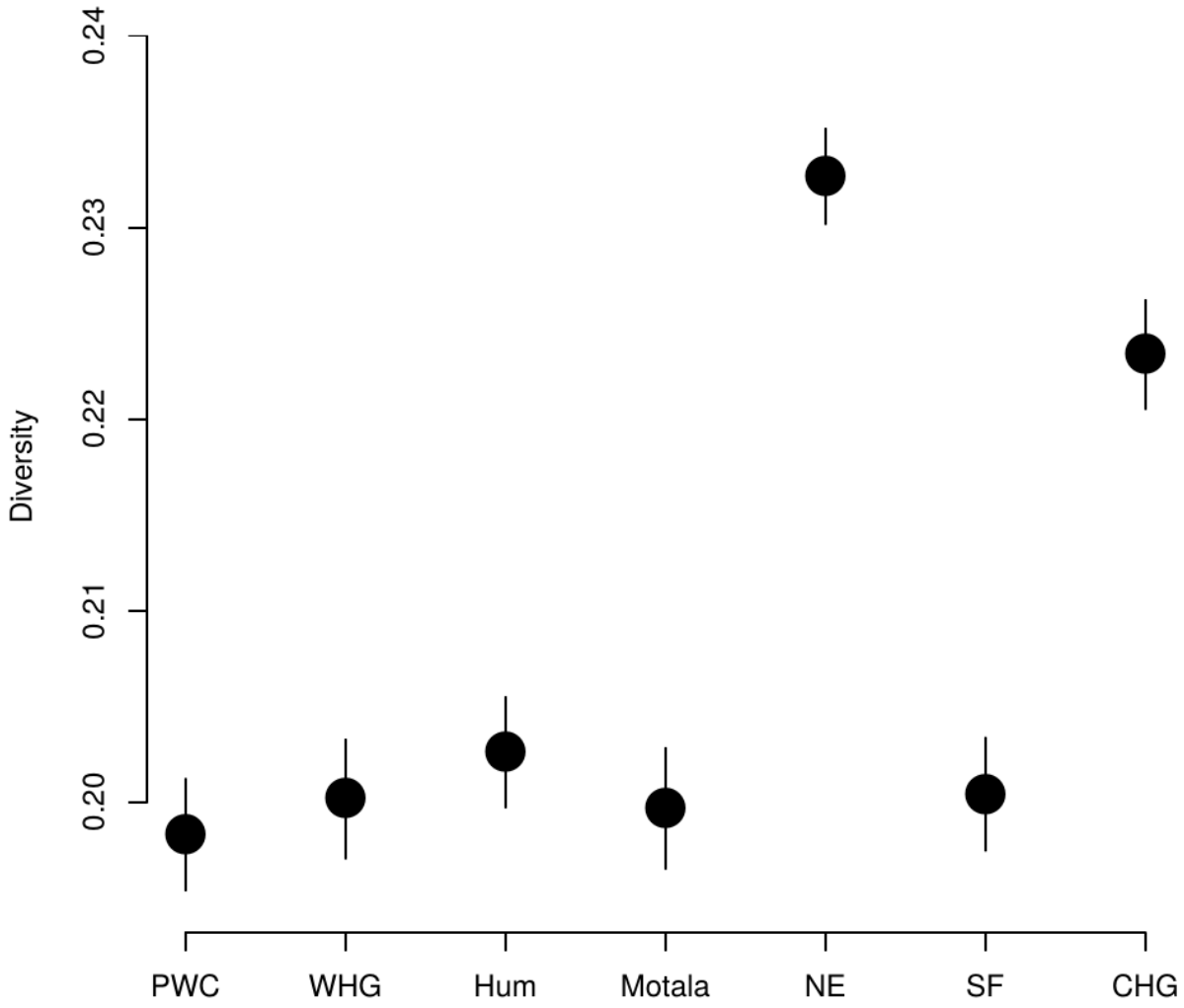


Figure S7.1 Conditional nucleotide diversity for seven prehistoric groups (using the two individuals with the highest shotgun sequencing coverage per group). Error bars show two standard errors.

Furthermore, we noticed that the differences between shotgun sequenced Motala12 and Motala3 were reduced to one third when compared to other pairs of individuals from the Motala site. This is even lower than expected for first degree relatives. This reduced diversity was not observed when we used capture data for both individuals. The mitochondrial haplogroup of shotgun sequenced Motala3 (U2e1) differs from the haplogroup reported in Lazaridis et al [81] and the haplogroup of SNP capture data for the same individual (U5a1 in both cases). As the haplogroup for Motala12 is also U2e1, we suspect that the BAM file for Motala3 which was uploaded to the European Nucleotide Archive is likely from Motala12 while all SNP capture files have the correct assignments. Motala3 was excluded from all analyses based on sequence data only.

The general results of the diversity analysis is consistent with expectations [92], showing a low diversity in hunter-gatherer groups and a higher diversity in early farmers (Figure S7.1). This

analysis did not pick up substantial differences between the hunter gatherer groups, so we continued with more fine-scale methods to measure genetic diversity (see below).

S7.2 Runs of homozygosity

Heterozygosity as investigated above 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 while an excess of long runs of homozygosity 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 runs of homozygosity were estimated using Plink 1.9 [149,150] 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 Figure 2a.

S7.3 Linkage disequilibrium

Similar to runs of homozygosity, the decay of linkage disequilibrium (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 ancient DNA 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). Next, we bin SNP pairs by distance (bin size 5kb) and then calculate 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 as we do not consider multilocus haplotypes, which is similar to the approach taken by ROLLOFF[147,160] and ALDER [161] to date admixture events based on admixture LD decay. For Figure 2b, we used two modern 1000 genomes populations to scale the LD per bin. The LD between two randomly chosen PEL (Peruvian) individuals was set to 1 and the LD between two randomly chosen TSI (Tuscan) individuals 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 as technical differences in the modern data (e.g. SNP calling or imputation) may have substantial effects.

S7.4 MSMC

We are using MSMC's implementation of PSMC' [162] 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 post-mortem 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 non-default 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 [111] (Figure 2c). We plot the effective population size assuming a mutation rate of 1.25x10e-8 and a generation time of 30 years. The curves for ancient individuals were shifted based on their average C14 date.

S7.5 Haplotype sharing among ancient individuals

Chromosomal segments that are identical by descent (IBD) between two individuals are due to shared ancestors. The number and length of these shared segments can be used to infer recent inbreeding as well as gene flow between populations [163]. IBD segments between the six high coverage ancient individuals (SF12, NE1, Kotias, Loschbour, Stuttgart, Ust-Ishim) were detected with BEAGLE 4.1 [164] (v03May16.862) and the genetic map files provided with BEAGLE (based on the HapMap genetic map) using SNPs ascertained in Yorubans. BEAGLE was run with default settings except [81] for the non-default parameters impute=false and ibdtrim=70.

The numbers and total length of IBD segments between the ancient individuals are shown in Figure S7.2. The highest proportion of the genome is shared between the two European Mesolithic hunter-gatherers SF12 and Loschbour which is consistent with their shared ancestry and the low effective population size in hunter-gatherers. The second highest amount of sharing is between the two Neolithic Europeans NE1 and Stuttgart, which is also consistent with the higher effective population size in farmers. Other pairs of individuals share lower proportions of the genome which is not surprising since they are all from different groups.

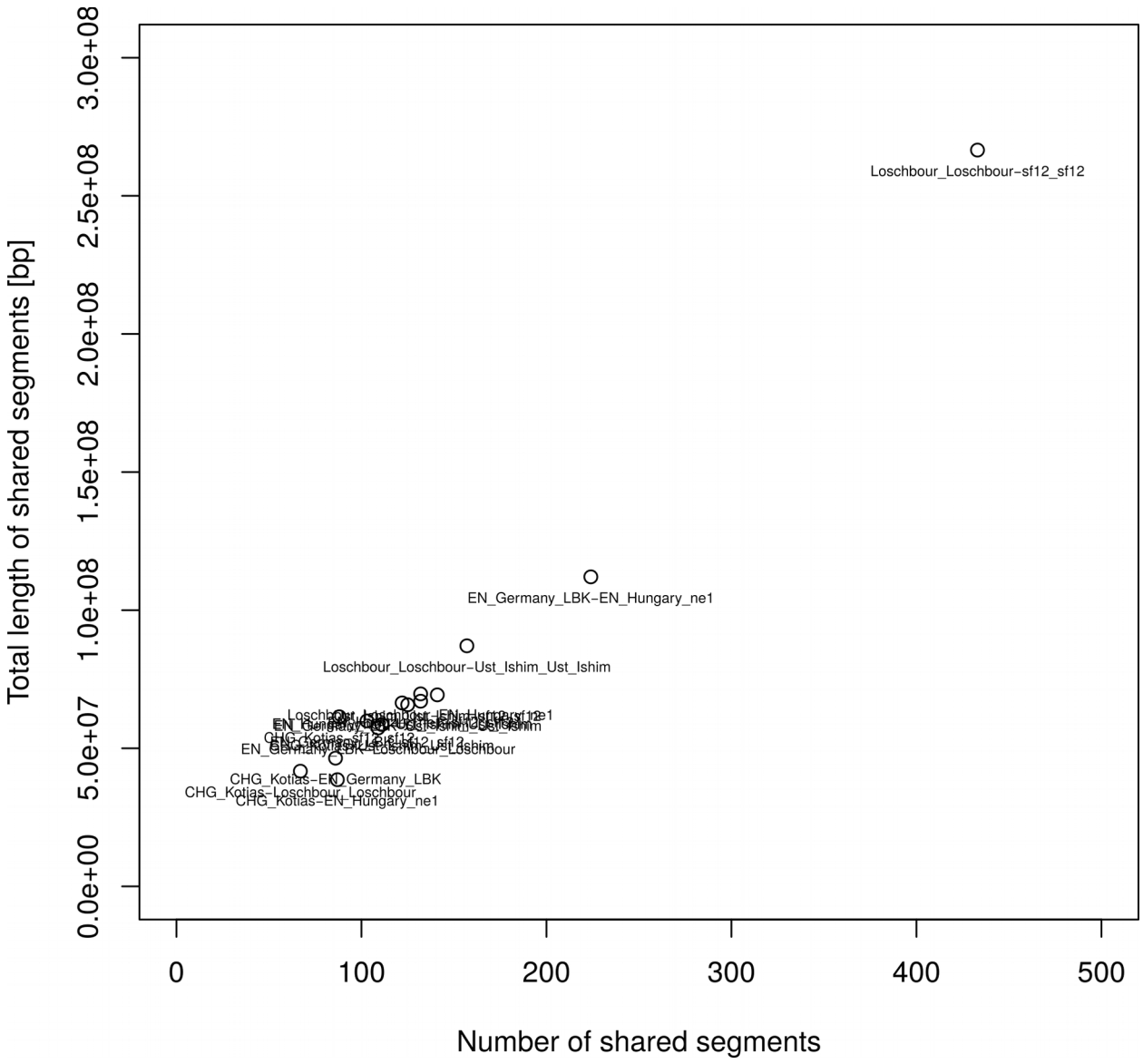


Figure S7.2 Results of the IBD-analysis plotting the number of shared IBD segments versus their cumulative length for all pairwise comparisons between ancient individuals.

S8 Functional variation in ancient samples

Methods

In order to generate a detailed portrait of the putative physical and metabolic phenotypes of Mesolithic hunter-gatherers, we screened all individuals presented in this study for variants associated with different traits of interest. We attempted to predict their physical appearance (pigmentation, scalp and facial hair phenotype), resistance to infectious diseases and other immune-related conditions, ability to digest certain animal or plant products, predisposition to genetic disorders, and other relevant conditions. The allelic states from these newly sequenced individuals were compared to those from the literature, including Motala, Western European HGs (WHG), Eastern European HGs (EHG), Caucasus hunter-gatherers (CHGs), as well as Eurasian Upper Paleolithic and Neolithic individuals to better understand the evolution and spread of these traits. Furthermore, only individuals with an average of >1x coverage for the captured SNPs were considered from [118] (ElMiron, Villabruna, GoyetQ116-1, Vestonice16 and Kostenki14). Genome sequence data and target-SNP capture data have been produced for some Motala individuals [81,82]. To jointly analyze those two types of data, we merged data per individual, after manually checking that no contradicting information was observed between the two BAM files at the investigated positions.

Both transversions and transitions were investigated for this analysis, as roughly two thirds of variants of interest are transitions. MapDamage V2.0.2-15 [165,166], was used with default parameters to rescale the base qualities of potentially deaminated sites in all BAM files. Data from individuals whose libraries had been damage-repaired (UDG treated) were not included in this base quality rescaling step. This approach may not completely remove all aDNA damages at the sequence fragment termini, but it controls for aDNA damage under a probabilistic framework by utilizing all information from the ancient genome sequence data. Samtools v1.3 [104] mpileup was used to restrict our analysis to variants with base qualities >= 30 and reads with MAPQ >= 30. The classification of published ancient samples can be found in table S8.1 and all calls at functional sites can be found in Supplementary Table 1.

Table S8.1 Samples used in the functional variants section.

Sample	Group	Group assignment for functional variant section	Country	Reference
Ajv58	NHG	Neolithic Hunter-gatherer Scandinavia	Sweden	Skoglund et al., 2014 and this study
Ajv70	NHG	Neolithic Hunter-gatherer Scandinavia	Sweden	Skoglund et al., 2014 and this study
Motala1/I0011	SHG	Mesoltihic Hunter-gatherer Scandinavia	Sweden	Haak et al 2015
Motala2/I0012	SHG	Mesoltihic Hunter-gatherer Scandinavia	Sweden	Lazaridis et al 2014, Haak et al 2015
Motala3/I0013	SHG	Mesoltihic Hunter-gatherer Scandinavia	Sweden	Lazaridis et al 2014, Haak et al 2015
Motala4/I0014	SHG	Mesoltihic Hunter-gatherer Scandinavia	Sweden	Lazaridis et al 2014, Haak et al 2015
Motala6/I0015	SHG	Mesoltihic Hunter-gatherer Scandinavia	Sweden	Lazaridis et al 2014, Haak et al 2015
Motala12/I0017	SHG	Mesoltihic Hunter-gatherer Scandinavia	Sweden	Lazaridis et al 2014, Haak et al 2015
Steigen	SHG	Mesoltihic Hunter-gatherer Scandinavia	Norway	This study
SF9	SHG	Mesoltihic Hunter-gatherer Scandinavia	Sweden	This study
SF11	SHG	Mesoltihic Hunter-gatherer Scandinavia	Sweden	This study and Skoglund et al 2014
SBj	SHG	Mesoltihic Hunter-gatherer Scandinavia	Sweden	This study
Hum2	SHG	Mesoltihic Hunter-gatherer Scandinavia	Norway	This study
Hum1	SHG	Mesoltihic Hunter-gatherer Scandinavia	Norway	This study
SF12	SHG	Mesoltihic Hunter-gatherer Scandinavia	Sweden	This study
LaBrana1	WHG	Mesoltihic Hunter-gatherer West Europe	Spain	Olalde et al 2014
Loschbour	WHG	Mesoltihic Hunter-gatherer West Europe	Luxembourg	Lazaridis et al 2014
Bichon	WHG	Mesoltihic Hunter-gatherer West Europe	Switzerland	Jones et al. 2015
KO1	WHG	Mesoltihic Hunter-gatherer West Europe	Hungary	Gambaet al. 2014
UzOO74/I0061	EHG	Mesoltihic Hunter-gatherer East Europe	Russia	Mathieson et al. 2015
SVP44/I0124	EHG	Mesoltihic Hunter-gatherer East Europe	Russia	Mathieson et al. 2015
Kotias	CHG	Mesoltihic Hunter-gatherer Caucasus	Georgia	Jones et al. 2015
Satsurblia	CHG	Mesoltihic Hunter-gatherer Caucasus	Georgia	Jones et al. 2015
ElMiron	PEHG	Paleolithic Hunter-gatherer Europe	Spain	Fu et al. 2016
Villabruna	PEHG	Paleolithic Hunter-gatherer Europe	Italy	Fu et al. 2016
GoyetQ116-1	PEHG	Paleolithic Hunter-gatherer Europe	Belgium	Fu et al. 2016
Vestonice16	PEHG	Paleolithic Hunter-gatherer Europe	Czech Republic	Fu et al. 2016
Kostenki14	PEHG	Paleolithic Hunter-gatherer Europe	Russia	Fu et al. 2016
Klei10	EEF	Early Neolithic Balkans	Greece	Hofmanova et al. 2016
Rev5	EEF	Early Neolithic Balkans	Greece	Hofmanova et al. 2016
Pal7	EEF	Early Neolithic Balkans	Greece	Hofmanova et al. 2016
Bar8	EEF	Early Neolithic Anatolian	Turkey	Hofmanova et al. 2016
Bar31	EEF	Early Neolithic Anatolian	Turkey	Hofmanova et al. 2016
Tep002	EEF	Early Neolithic Anatolian	Turkey	Kılınç et al. 2016
Tep003	EEF	Early Neolithic Anatolian	Turkey	Kılınç et al. 2016
Tep004	EEF	Early Neolithic Anatolian	Turkey	Kılınç et al. 2016
Tep006	EEF	Early Neolithic Anatolian	Turkey	Kılınç et al. 2016
Bon001	EEF	Early Neolithic Anatolian	Turkey	Kılınç et al. 2016
Bon002	EEF	Early Neolithic Anatolian	Turkey	Kılınç et al. 2016
Bon004	EEF	Early Neolithic Anatolian	Turkey	Kılınç et al. 2016
BAR2/I0707	EEF	Early Neolithic Anatolian	Turkey	Mathieson et al. 2015
BAR6/I0708	EEF	Early Neolithic Anatolian	Turkey	Mathieson et al. 2015
BAR20/I0709	EEF	Early Neolithic Anatolian	Turkey	Mathieson et al. 2015
M11-363/I0745	EEF	Early Neolithic Anatolian	Turkey	Mathieson et al. 2015
L11-322/I0746	EEF	Early Neolithic Anatolian	Turkey	Mathieson et al. 2015
L14-200/I1583	EEF	Early Neolithic Anatolian	Turkey	Mathieson et al. 2015
Stuttgart	EEF	Early Neolithic Central European	Germany	Lazaridis et al. 2014
HAL2/I0659	EEF	Early Neolithic Central European	Germany	Haak et al. 2015, Mathieson et al. 2015
HAL24/I0821	EEF	Early Neolithic Central European	Germany	Haak et al. 2015, Mathieson et

				al. 2015
HAL25/I0048	EEF	Early Neolithic Central European	Germany	Haak et al. 2015, Mathieson et al. 2015
HAL34/I0057	EEF	Early Neolithic Central European	Germany	Haak et al. 2015, Mathieson et al. 2015
HAL4/I0100	EEF	Early Neolithic Central European	Germany	Haak et al. 2015, Mathieson et al. 2015
HAL5/I0046	EEF	Early Neolithic Central European	Germany	Haak et al. 2015, Mathieson et al. 2015
KAR16A/I0797	EEF	Early Neolithic Central European	Germany	Haak et al. 2015, Mathieson et al. 2015
KAR6	EEF	Early Neolithic Central European	Germany	Haak et al. 2015, Mathieson et al. 2015
LBK1976/I0022	EEF	Early Neolithic Central European	Germany	Haak et al. 2015, Mathieson et al. 2015
LBK1992/I0025	EEF	Early Neolithic Central European	Germany	Haak et al. 2015, Mathieson et al. 2015
LBK2155/I0026	EEF	Early Neolithic Central European	Germany	Haak et al. 2015, Mathieson et al. 2015
SZEH4/I0176	EEF	Early Neolithic Central European	Hungary	Haak et al. 2015
Starcevo	EEF	Early Neolithic Central European	Hungary	Haak et al. 2015
Troc1/I0409	EEF	Early Neolithic Iberian	Spain	Haak et al. 2015
Troc3/I0410	EEF	Early Neolithic Iberian	Spain	Haak et al. 2015
Troc5/I0412	EEF	Early Neolithic Iberian	Spain	Haak et al. 2015
Troc7/I0413	EEF	Early Neolithic Iberian	Spain	Haak et al. 2015

Finally, it is noteworthy that the variants used in our subsequent phenotypic predictions have been validated in present-day human populations, thus it is possible that undetected functional variation within the analyzed genes (or even other regions) contributed to phenotypic variation in ancient human populations. Such variation would not be captured by the following analyses [81].

S8.1 Pigmentation

Predicting human eye, hair, and skin pigmentation scales and color has recently been reported by analyzing genetic variation at several genes and SNPs [167–172]. Here we reconstructed the phenotypic appearance of the SHGs, and screened these functional variants in other hunter-gatherer individuals as well as Early Neolithic farmers across Europe, and thus improved the understanding of the history and evolution of these variants.

Eye, hair, and skin color/pigmentation was assessed by screening two sets of pigmentation-associated markers recently validated in present-day human populations [170,172–174]. The probability for certain eye and hair color for each ancient individual was computed using the enhanced version 1.0 Hirisplex Microsoft Excel macro [174], a SNP system originally designed for forensic purposes. We attempted to predict skin pigmentation using the 8plex system described by Hart et al [172], but as previously reported [81,121], the results were inconclusive for ancient samples using this panel. Similar to previous investigations of ancient humans [81,121,125,175], we provide skin-color predictions based on the genetic variants at the rs16891982 (in the *SLC45A2* gene) and rs1426654 (*SLC24A5*) SNPs. The derived allele at rs1426654 (*SLC24A5*) has the greatest effect on skin pigmentation among Europeans [176–179].

Moreover, this variant is part of a light skin haplotype termed "*C11*", which is defined by a set of 16 SNPs across 78kb within the *SLC24A5* gene. In addition to rs1426654, we investigated the non-synonymous SNP rs16891982 in the *SLC45A2* gene, which is a second major-effect pigmentation locus in Europeans [180]. The derived alleles at both rs1426654 and rs16891982 positions are found at extremely high frequencies in present-day European populations (where the former allele is virtually fixed and the latter is ~90%), an observation that has been interpreted as the signature of recent positive selection in Europeans [181].

In studies of eye pigmentation, the derived allele at SNP rs12913832 (within the *HERC2* gene) has been associated with iris depigmentation. Individuals carrying the homozygous derived variant exhibit a blue eye-color, whereas carriers of the heterozygote genotype or the homozygote ancestral allele display an intermediate or a brown eye-color phenotype [182]. Across present-day Europeans populations from the 1000 genomes project, the blue eye color associated variant is found at an average frequency of 60%, and it reaches its highest frequency in the Finnish population (90%) and the British population (81%) [183]. The SNP rs12913832 is part of the *h-1* haplotype, which is carried by 97% of blue-eyed individuals in a present-day study population from Turkey, Jordan, and Denmark [184]. The presence of *h-1* can be assessed through screening the allelic states at a set of 13 SNPs in the *OCA2/HERC* region.

Allelic states of the investigated SNPs as well as the *C11* and *h-1* haplotypes for each ancient group is shown in Supplementary Table 1. A brief summary of the pigmentation phenotypic reconstruction can be found below for each group and individual, as well as interpretations of the history of these variants.

Western hunter-gatherers (WHGs)

The Mesolithic La Braña, Loschbour WHGs, the Upper Palaeolithic Villabruna individual, and the Neolithic KO1 individual, which has a genetic makeup of a WHGs [98], had dark skin and blue eyes based on their allelic states at the rs1426654, rs16891982 and rs12913832 SNPs [81,98,118,125]; an observation supported by their high probabilities of blue eye color inferred by Hirisplex in this study. The Upper Paleolithic Bichon from central Europe [121] had likely dark-skin and brown eyes (Supplementary Table 1). The dark skin and blue eyes phenotype combination was likely very common among WHGs and it has been inferred for four out of five Mesolithic and Upper Paleolithic individuals in western and central Europe. Moreover, the allelic states at the *h-1* defining positions suggest that La Braña, Loschbour and Villabruna carried this haplotype. The haplotype determination was inconclusive for the remaining individuals due to alternative alleles and missing data at informative positions. All individuals from this group were predicted to have a dark hair color, where Loschbour, Villabruna and Bichon probably had black hair, whereas La Braña and KO1 present similar probabilities of having brown or black hair (Supplementary Table 1).

Eastern hunter-gatherers (EHGs)

The Karelian and Samaran Russian Mesolithic hunter gatherers [82] are currently the best representatives of a group of EHGs that migrated and admixed with WHGs to form SHGs. Mathieson et al [122] recently reported data from another Karelian Mesolithic hunter gatherer (sample I0211), with low coverage (0.136X across all captured SNPs), preventing pigmentation characterization. Hirisplex skin and hair color predictions suggest some opposing pigmentation patterns for the EHG individuals. The Karelian individual presents high probabilities of being brown-eyed (0.99), and having a dark hair (0.96). Without speculating about the genetic architecture of skin pigmentation, we suggest an intermediate skin-pigmentation phenotype for the Karelia individual, as it carried the ancestral allele at rs16891982 and the derived allele at rs1426654 (Supplementary Table 1). The presence of the rs1426654 light-skin allele, in addition to five additional *C11*-associated alleles at haplotype defining SNPs (Supplementary Table 1) suggests that the Karelian individual carried the *C11* light-skin haplotype. The Samaran individual exhibits high probabilities of being blue-eyed (0.88), light hair shade (0.99); most likely being blond (0.75). The two skin pigmentation SNPs suggest that the Samaran individual was light-skinned. In summary, the EHGs had high frequencies of the light-skin variants and intermediate frequencies of the blue-eye variants.

Scandinavian hunter-gatherers (SHGs)

The SHG group includes the seven individuals from this study, the six Motala hunter-gatherers [81], and the Neolithic Ajvide 58 hunter-gatherer [92]. From SHGs sequenced in this study, we obtained an eye and hair pigmentation portrait for SF9, Hum2, SBj and SF12. No pigmentation predictions were generated for the SF11, Hum1 and Steigen samples, due to missing data at the relevant positions. Out of the four individuals with a Hirisplex prediction, both light and dark pigmentation phenotypes were observed. SBj and SF12 exhibited high probabilities of being blue-eyed (0.91 and 0.88, respectively), while SF9 and Hum2 were predicted to have been brown eyed. The high-coverage and high quality genome of SF12 carried two copies of the *h-1* haplotype (defined by the homozygous state of all *h-1* haplotype SNPs, Supplementary Table 1). SF12 and Hum2 likely had dark hair (p=0.75 and p=0.99), while SF9 and SBj had light hair (p=0.58 and p=0.90). SBj presented slightly higher probabilities of being blond (0.52) than having black, brown or red hair.

The common feature of the skin-pigmentation SNPs among these Scandinavian hunter-gatherers was a mix of light and dark skin alleles at either rs16891982 or rs1426654 (Supplementary Table 1). The Hum2 individual exhibited only derived haplotype-associated alleles at the 16 *C11*-defining positions, suggesting that Hum2 had at least one copy of this haplotype. SF9 and Steigen present derived variants at 8 and 13 out of the 16 screened positions, respectively. Even though both Steigen and SF9 lacked information for the core SNP rs1426654, Steigen only carried C11-associated alleles at rs1834640, rs2675345, rs938505 positions, and SF9 at rs1834640 and rs938505, which distinguish C11 from other *SLC24A5* haplotypes. SF12 exhibits

a different haplotype, C9, compared to the other SHGs, and carries the ancestral allele at the core SNP rs1426654, indicating dark skin pigmentation.

A Hirisplex eye and hair color prediction was obtained for the six hunter-gatherers from Motala [81,82]. Interestingly, all individuals exhibited high probabilities of being blue-eyed (0.71-0.92). The Motala2, Motala3, Motala4 and Motala12 individuals most likely had a dark hair color (0.70-0.99), while Motala1 and Motala6 had a light shaded hair (~0.91); they may have been blond (~0.60). Similar to SF9, SF11, SF12, SBj, Hum1, Hum2 and Steigen, the Motala hunter-gatherers presented a combination of light and dark skin pigmentation alleles. Only Motala2 presented exclusively light-skin variants at both rs16891982 and rs1426654. Motala12 exhibited only haplotype-defining derived alleles at both the *h-1* and the *C11* associated positions. The combination of C11-associated alleles at the rs1834640 and rs2675345 SNPs with a light-pigmentation allele at the core SNP rs1426654, suggest that Motala2 and Motala4 individuals carried the *C11* haplotype as well.

The Neolithic Ajvide 58 and 70 individuals from a Pitted Ware Culture context shares part of its phenotypic variation with SHGs. While they presented high probabilities of being blue-eyed (0.79 and 0.86) and having a dark hair color (0.94 and 0.97), Ajv58 exhibits the dark skin pigmentation alleles at rs16891982 and rs1426654 positions, while Ajv70 is heterozygous at those positions..

Interestingly, the eye and light skin pigmentation phenotypes observed in all SHGs could potentially be explained by admixture between WHG and EHG groups. The high relative-frequency of the blue-eye color allele in SHGs, resembles WHG, while the intermediate frequencies of the skin color determining SNPs in SHGs seem more likely to have come from EHG, since both light-pigmented alleles are virtually absent from WHG. However, for all three well-characterized skin and eye-color associated SNPs, the SHGs display a frequency that is greater for the light-skin variants and the blue-eye variant than can be expected from a mixture of WHGs and EHGs. This observation indicates that the frequencies may have increased due to continued adaptation to a low light conditions.

Caucasus hunter-gatherers (CHGs)

Both Kotias and Satsurblia CHGs [121] were predicted by Hirisplex to have brown eyes (>0.96) and a dark hair shade (>0.92). Looking at skin-pigmentation sites, both individuals carried the dark-skin allele at rs16891982 and the light-pigmentation allele at rs1426654. Similar to the Hum2 and Motala12 individuals, Kotias showed exclusively haplotype-associated alleles at the *C11*-defining positions (with at least 5 reads of such alleles per site), suggesting it carried the *C11* haplotype.

Palaeolithic European hunter-gatherers (PEHG)

The Hirisplex eye and hair color prediction for the Paleoltihic European hunter gatherers ElMiron, GoyetQ116-1, Vestonice16, Kostenki14 (dated from 18,000 to 36,000 ya), displayed high probabilities of being brown-eye color (> 0.99), and high probabilities of exhibiting a dark hair pigmentation (0.60-0.99). All four individuals presented only dark-skin alleles at rs16891982. The GoyetQ116-1 and Vestonice16 individuals missed information at the rs1426654, while El Miron carried only dark-skin alleles at that position, and Kostenki14 had four dark-skin alleles and one light-pigmentation allele.

Early European farmers (EEF)

The Hirisplex eye and hair color prediction of 37 Early Neolithic farmers sequenced across Europe and Anatolia [81,82,175,185] revealed that four individuals presented high probabilities of being blue-eyed (p=0.55-0.91), 23 were predicted to be brown-eyed (p=0.76-0.99), and 10 individuals did not have enough data to make a prediction. In total 23 individuals exhibit a high probability for dark hair pigmentation (p=0.57-0.99), five had most likely a light hair shade; the remaining nine farmers lacked data for meaningful hair pigmentation prediction (Supplementary Table 1). The Anatolian Barcin I1583 individual [122] exhibited blue-eye color variants at the core *h-1* SNP rs12913832 and its linked rs1129038 variant. Interestingly, it also presented *h-1*-haplotype-defining alleles at rs7170852, rs2240203, rs916977, suggesting that this individual might have harbored the blue-eye color founder haplotype as well. No other early farmer presented direct evidence for the *h-1* haplotype's presence.

While both pigmentation alleles were observed at rs16891982 (although the derived allele in much higher proportion, see below), virtually only the light-skin allele was observed at rs1426654. Unequivocal evidence for the presence of the *C11*-haplotype was observed for the central European Stuttgart sample, while (although at low coverage) only haplotype-associated alleles were observed in the Anatolian Klei10 early farmer individual.

Summary

After investigating the presence of light-pigmentation alleles and haplotypes in different hunter-gatherers and early farmers across Europe and Anatolia, it seems as if eye and light skin pigmentation alleles entered Europe several times during different migration events. In particular, the light skin pigmentation variant arrived to Scandinavia already in the Mesolithic with migrants from the northeast, whereas the blue-eye variants probably arrived in Scandinavia with migrants from the south.

Light eye pigmentation variants were present at high frequencies in WHG, SHG, EHG and EEF (not present in PEHG), while the blue-eye color founder haplotype *h-1* was found in the La Brana, Loschbour, Villabruna WHGs, SF12, Motala1 and Motala12 SHGs and at least one early farmer. Such results suggest that the blue eye-color allele is rather old. Using an ABC modeling

approach Nakagome et al. [186], predicted that the light-pigmentation allele at rs12913832 emerged around 42,000 years ago or earlier; a date close in time to the initial peopling of Europe. A plausible scenario of the origin of the blue-eye mutation that reconciles our results with findings from other studies is one where this variant appeared in an ancestral population before the ancestors of the WHG migrated from Near East into West and Central Europe [118].

The large effect light-skin alleles at rs16891982 and rs1426654 were present in SHG, EHG, CHG and EEF but absent in WHG and PEHG. Similarly, the *C11* haplotype is present in hunter-gatherers (SHG, EHG and CHG but not WHG and PEHG) throughout Europe, as well as in at least two early farmers. This pattern is consistent with reports that the rs1426654 derived allele arose ~22,000-28,000 years ago [186,187], and that the light-pigmentation allele at rs16891982 arose only once in Eurasians [186,188]. A possible geographical origin for these two major light-skin alleles is West Asia or the Near East [189]. Later migrations across the Caucasus (CHG) and Eastern Europe would have brought it to Scandinavia, while EEF migrations introduced both alleles into central Europe.

Moreover, [179] suggested that selective sweeps on rs1426654 and rs16891982 (light-pigmentation alleles), started at between ~15-19 kya (under a dominant model) and ~11-13 kya (under an additive model), respectively. In a comparison of SNP capture data of hunter-gatherers of mixed geographic origin, early farmers and modern-day Europeans, Mathieson et al [122] suggested that while light-pigmented alleles frequencies in extant Europeans at rs1426654 could be explained by demography, variation at the rs16891982 locus produced the second highest genome-wide selection signal observed in their study. These results, of a high frequency of rs1426654 light-skin allele in Mesolithic Scandinavia and Eastern Europe, at a time when it is not seen in central Europe, supports a scenario of environmental adaptation to northern latitudes.

To summarize the results of the skin and eye color section, we display the allele frequencies of light-pigmentation alleles of the three SNPs rs12913832, rs16891982 and rs1426654 for each group in Figure 3. Allele frequencies are estimated as the proportion of chromosomes carrying the light pigmentation allele. A distinction was made between low coverage (haploid calls) sites (SNPs covered by less than five reads in a particular individual), versus higher coverage (diploid calls) sites (SNPs covered by five or more reads in a particular individual). The maximum count of light-pigmentation alleles at haploid sites was 1, and 2 at diploid sites. Positions where both alleles were observed contributed either 0.5 (haploid call) or 1 (diploid call) to light-color allele frequencies in order to be consistent with only calling 2 alleles if we have 5 or more reads covering a site. Data from all individuals was used to calculate standard errors for each estimate.

S8.2 Scalp and facial hair associated SNPs

Genetic studies in modern-day humans have found certain SNPs to be associated with scalp hair (shape, greying, balding) and facial hair (beard thickness, monobrow, eyebrow thickness) [190].

Here, we investigate these SNPs in ancient Scandinavian individuals. Table S8.2 summarizes the results.

Regarding hair shape, we screened SNPs rs3827760 and rs11803731, where variants have been associated with straight hair in Europeans and Asians. Furthermore, it has been suggested that the associated variants have been under recent selection in both populations [191,192]. At rs11803731, we find both the associated and non-associated variant among SHGs (SF12 is homozygote for the straight-hair associated variant), which is similar to other hunter-gatherers throughout Europe. For rs3827760, within the EDAR gene, the derived G allele is associated with shovel-shaped teeth and hair thickness phenotype in East Asians. In the novel SHGs in this study, only the ancestral A allele is present (SF12 is homozygote AA). The derived variant was reported in three of the six Motala SHGs which are younger than most other SHGs in this study [122]. It is clear that the variant was present among SHGs [122], and it is possible that it has a continuous (but varying) distribution from Scandinavia to East Asia during the Mesolithic, and that the very low sample size of EHGs has failed to pick up the variant. It is also possible that the derived rs3827760 variant was brought to Scandinavia by migration in the Late Mesolithic, perhaps related to the specific Motala group.

In total 13 SNPs have been associated to facial hair presence and shape [190]. Among those positions, rs365060, rs4864809, rs6901317 and rs117717824 have been linked to beard thickness, rs112458845 to eyebrow thickness and rs2218065 to monobrow. These SNPs were ascertained in admixed Latin American populations and it is still unclear how well these variants influence facial hair presence and shape in non-Latin Americans. These are also highly polygenic traits, so the markers investigated here do not allow for reliable conclusions regarding the ancient individuals. Nonetheless, in order to get an indication for these traits among the ancient individuals, we investigated these SNPs.

For beard thickness, 2 SNPs (rs365060 and rs117717824) showed opposing patterns of only the thick beard variant for the former SNP, and only the thin beard variant for the latter SNP. For the 2 remaining beard thickness SNPs, we observe both variants (2 of 5 vs 3 of 5 for rs4864809) and 5.5 of 8 vs 2.5 of 8 in rs6901317). Hence, the SHGs would, on average, have intermediate and varied beard thickness. While SF12 is a female (she is heterozygote for 2 of the SNPs, and homozygote thick beard for one site and homozygote thin beard for one site), Steigen is a male and has the thick beard variant at 2 sites, possibly heterozygote at one site and carries one thin beard variant at the 4th site. Interestingly, the thick beard variant at rs365060 has been shown to be under selection in modern-day western Europeans [190]. A comparison of the allelic state of the individuals from this study to other HGs, showed that the thick beard variant at rs365060 was widely spread within WHGs, EHGs and CHGs. Thus, a selective event probably predated the split of hunter-gatherers in Europe.

The eyebrow thickness associated SNP rs112458845 showed variation among SHGs (3As and 2Gs called), whereas the monobrow variant G (rs2218065) appears to be in high frequency (6 of 6) among SHGs. SF12 was homozygote for both the thick eyebrow and monobrow variants. The gray hair associated variant was also quite common among SHGs (4 of 5).The SNPs rs2814331 and rs4258142 have been associated to balding [190,193]. Among all 7 individuals, we only find balding associated variants at these two SNPs (3 individuals had no data on these SNPs).

Table S8.2 Scalp and facial associated SNPs

	SNPID	gene	mutation	strnd	Steigen	SF9	SF11	SBj	Hum2	Hum1	SF12
Hair shape	rs11803731	*TCHH*	A > T	+	NA	1A	NA	NA	2T	NA	77T
Hair shape	rs3827760	*EDAR*	A > G	-	1A	1A	NA	NA	8A	NA	88A
Beard thickness	rs365060	*EDAR*	G > C	+	2C	NA	NA	NA	4C	NA	85C
Beard thickness	rs4864809	*LNX1*	A > G	+	2G	1G	NA	NA	3A	NA	29G,23A
Beard thickness	rs6901317	*PREP*	G > T	+	1T,1G	1G	1G	NA	2G	1G	35T,35G
Beard thickness	rs117717824	*FOXP2*	G > T	+	2G	2G	NA	NA	1G	1G	54G
Eyebrow thickness	rs112458845	*FOXL2*	G > A	+	NA	NA	NA	1G	2G	NA	31A
Monobrow	rs2218065	*PAX3*	A > G	+	2G	1G	NA	NA	1G	NA	41G
Hair greying	rs12203592	*IRF4*	C > T	+	NA	NA	NA	NA	4T	2T	40C,34T
Balding	rs2814331	*GRID1*	C > T	+	1T	1T	NA	NA	NA	NA	45T
Balding	rs4258142	*AR/EDA2R*	C > T	+	3T	2T	NA	NA	2T	NA	24T

S8.3 Acetylator phenotype

The genetic basis of drug response is a major focus of pharmacogenetics. Drugs are metabolized using metabolic pathways, which evolved to degrade harmful metabolic by-products, as well as dietary and environmental toxins. Therefore, individual drug response constitutes a complex genetic trait shaped by both genetic and environmental pressures. N-acetylotransferase 2 gene (*NAT2*) is involved in the metabolism of a wide variety of foreign chemical substances (xeniobiotics) [195]. The metabolism can be either fast or slow depending on the polymorphisms within NAT2 gene. This differential drug response has also been linked to response to a number of carcinogens, especially those causing bladder and colon cancers [196]. The global distribution of the two phenotypes suggests the incidence of the fast-acetylator phenotype is higher in modern hunter-gatherers, Native Americans and East Asians than in Africans and West Eurasians. The prevalence of each of the phenotypes has been reported to correlate with the type of subsistence practiced by various populations, ranging from high frequency of fast phenotype among hunter-gatherers (>77%) to medium levels among pastoralists and agriculturalists (51-54.9%) [197]. There is evidence from present-day populations for selection favoring the slow-

acetylator phenotype, which was hypothesised to be related to dietary changes accompanying the transition to agriculture [198,199]. Traditionally, *NAT2* phenotype is predicted based on phased *NAT2* haplotypes. However, in the absence of phased data, genome-wide shotgun sequence data from 7 informative SNPs (rs1801279, rs1801280, rs1799930, rs1799931, rs1041983, rs1799929, rs1208) can be used in phenotype prediction. The methodology was adopted in an online *NAT2* prediction tool - NAT2PRED (http://nat2pred.rit.albany.edu/) [200]. Alternatively to the 7-SNP phenotype prediction panel it is possible to use a single tag SNP (rs1495741) and/or a 2-SNP panel (rs1041983 and rs1801280) [201]. In general, presence of ancestral alleles is associated with the presence of rapid metabolizer phenotypes. It has however been suggested that the 2-SNP panel outperforms the prediction power of the single tag SNP, which may be Caucasian specific [202]. We have used the combination of all three methods to predict the acetylation state in 6 out of 7 investigated individuals. The 7-SNP panel was only useful for two individuals: SF12 and Hum2. While SF12 was estimated to be a slow acetylator, this type was previously described in the Stuttgart farmer [81], Hum2 was probably an intermediate or rapid acetylator (similarly to Motala12 and Loschbour foragers). Remaining individuals did not have enough genome coverage for reliable prediction based on the 7-SNP system. However, using the tag single- and two-SNP panels, it was possible to predict the following phenotypes: Steigen (slow), SF9 (slow), SF11 (intermediate). Thus, the Scandinavian hunter-gatherers from this study exhibit the whole range of acetylator phenotypic variation with excess of slow acetylator phenotype (N=4) over the rapid and intermediate phenotypes (N=3). While those phenotype frequencies are similar to those observed among present day Europeans, they deviate from previous observations of modern-day and two ancient hunter-gatherers [81,197]. In order to further investigate this scenario, we screened the aforementioned markers in 17 other European hunter-gatherers. We were able to assign the 7-SNP acetylator phenotypes in five of the tested individuals. Two of the individuals were slow (Bichon, Kotias), two were intermediate (LaBrana, Loshbour) and the Motala12 individual was a rapid metabolizer (based on NATPRED and 2-SNP prediction). We then applied the tag-SNP (rs1495741) in order to increase the number of individuals and managed to assign the metabolizer state in 17 hunter-gatherers (Supplementary Table 1). According to tag-SNP analyses only one individual was arapid metabolizer (KO1), five were intermediate metabolizers (Motala4, Motala12, Loshbour, Vestonice16 and Kostenki14), while 9 foragers were slow metabolizers (Ajvide58, Motala6, LaBrana, Bichon, Kotias, Karelia, Samara, ElMiron and Villabruna). In Motala12 we see an inconsistency between the scoring systems, where both 7-SNP and the 2-SNP give the same type, while the phenotype based on the 1 tag SNP system differs.

Identified mixed phenotypes with a slight excess of slow metabolizers in populations pre-dating introduction of farming, contradicts the widely held view that the slow-acetylator phenotype arose as a result of dietary changes accompanying the transition to agriculture [198,199], however, it might have increased its frequency as a consequence of adaptation to new dietary habits.

Table S8.3 Allelic state of *NAT2* associated SNPs in all seven new SHGs.

SNPID	gene	mutation	strand	Steigen	SF9	SF11	SBj	Hum2	Hum1	SF12
rs1801279	*NAT2*	G > A	+	3G	2G	NA	NA	1G	NA	79G
rs1801280	*NAT2*	T > C	+	NA	1T	NA	NA	1T,1C	NA	57T
rs1799930	*NAT2*	G > A	+	NA	1G	1A*	NA	4G	NA	46A
rs1799931	*NAT2*	G > A	+	2G,1A	NA	NA	1G	8G	1A	68G
rs1495741#	*NAT2*	G > A	+	1A	2A	1G,1A	NA	1G	NA	62A
rs1041983	*NAT2*	C > T	+	NA	1C	1C	NA	5C	NA	49T
rs1799929	*NAT2*	C -> T	+	NA	1C	NA	NA	NA	NA	55C
rs1208	*NAT2*	A > G	+	1G	NA	NA	1A	5G,5A*	1A	41A

Table S8.4 Predicted acetylation status in seven new SHGs (NAT2PRED: http://nat2pred.rit.albany.edu/) (Kuznetsov *et al.* 2009). Here (+) is indicative of presence of a fast metabolizer *NAT2*4* allele, (-) indicates the presence of a slow metabolizer *NAT2*4* allele, (+/-) denotes the intermediate allele and # is rs1495741 status prediction tag where genotypes are interpreted as metabolizers: (A;A) slow, (A;G) intermediate, (G;G) rapid.

SNPID	NAT2 allele	Steigen	SF9	SF11	SBj	Hum2	Hum1	SF12
rs1801279	G191A	(+)	(+)	nd	nd	(+)	nd	(+)
rs1801280	T341C	nd	(+)	nd	nd	(+/-)	nd	(+)
rs1799930	G590A	nd	(+)	(-)	nd	(+)	nd	(-)
rs1799931	G857A	(+/-)	nd	nd	(+)	(+)	(-)	(+)
rs1041983	C282T	nd	(+)	(+/-)	nd	(+)	nd	(-)
rs1799929	C481T	nd	(+)	nd	nd	(-)	nd	(+)
rs1208	A803G	(-)	nd	nd	(+)	(+/-)	(+)	(+)
NAT2PRD prediction		***nd***	***nd***	***nd***	***nd***	***interm.***	***nd***	***slow***
rs1495741	# tag SNP	(-)	(-)	(+/-)	nd	(+)	nd	(-)
rs1495741 prediction		***slow***	***slow***	***interm.***	***nd***	***rapid***	***nd***	***slow***
rs1801280	T341C	nd	(+)	nd	nd	(+/-)	nd	(+)
rs1041983	C282T	nd	(+)	(+/-)	nd	(+)	nd	(-)
2-SNP prediction		***nd***	***nd***	***nd***	***nd***	***rapid***	***nd***	***interm.***

(+) indicative of presence of a fast metabolizer *NAT2*4* allele
(-) indicative of presence of a slow metabolizer *NAT2*4* allele
(+/-) intermediate allele
rs1495741 is a prediction tag where genotypes are interpreted as metabolizers: (A;A) slow, (A;G) intermediate, (G;G) rapid

S8.4 Diet and taste related SNPs

We have typed all 7 SHG individuals from this study at a number of SNPs of known dietary function, closely linked to Vitamin D levels and rickets in modern populations. All those sites have been suggested to have undergone recent positive selection [122,203], and in the case of the SHGs of this study they may have conferred local adaption to high latitude climates.

Blood levels of circulating vitamin D are the result of diet, exposure to sunlight and genetic predisposition [204]. Some of those genetic components include the genes *DHCR7* and *NADSYN1* which are associated with circulating vitamin D levels. All sites associated with circulating vitamin D levels show high levels of heterozygosity among SHGs, clearly observed in SF12 and Hum2. Furthermore, high diversity was found among other tested ancient hunter-gatherer groups including WHGs, EHGs and CHGs (see Supplementary Table 1).

FADS1 and *FADS2* are involved in fatty acid metabolism and the derived rs174546 SNP has been linked to lower triglyceride levels in Europeans [205]. The high coverage individual (SF12) was a heterozygote at the site, while two other individuals were carriers of a single C allele associated with increased LC-PUFAS levels and one non-associated allele T respectively (SBj and SF9) [206]. Fumagalli et al [207] found selection on the *FADS* region in Greenlandic Inuit (GI) populations, which is thought to be an environmental adaptation. The presence of the derived T allele at position rs7115739 has been associated with reduced height and Body-Mass-Index (BMI) in GI. The authors also found similar albeit weaker associations in Europeans between the derived alleles (T) on rs7115739 and rs174570 and reduced height [207]. Neither Hum2 nor SF12 were carriers of derived allele at rs7115739 (T), however, almost all ancient individuals were homozygous for the associated alleles at rs174602 and rs174570 (*FADS1-3*) with only SF12 being heterozygous at rs174570. The latter is associated with low fasting serum insulin, LDL and cholesterol levels in Europeans [207]. Variation in *FADS1-3* probably reflects interactions between *FADS1* and the dietary intake of omega-3 and omega-6 fatty acids [208].

The ability to digest sugars found in milk during adulthood is widely perceived as an important cultural adaptation in humans. Initially thought to have been spread with early farmers, the ability to digest milk has now been shown to be uncommon among early farmers and likely only increased in frequency in Europe during the Iron Age [127]. We tested all our ancient individuals at both *LCTa* (rs4988235) and *LCTb* (rs182549) loci and found that the three individuals for whom there was overlapping data, were all carriers of non-lactase persistence alleles, indicating inability to digest milk in adulthood, thus in line with previous observations.

Additionally, we screened polymorphisms at two SNPs loci linked to celiac disease which have been identified as having undergone recent selection within the *SLC22A4* and *ATXN2* genes [122]. The two SNPs in *SLC22A4* show a consistent pattern in which rs1050152 is a C variant

not associated with increased risk of Crohn's disease (also in 12 other HGs) while rs272872 is heterogeneous among all investigated HGs. Furthermore, *ATXN2* associated with celiac disease displays a derived non associated T in four SHG individuals and 16 other hunter gatherers with available data at this position. The identification of only non-associated alleles at position rs1050152 in our dataset as well as in 12 other ancient hunter gatherers is concordant with Mathieson et al. [122] who suggested that the disease causing variant (L503F) did not reach significant frequency until recent times.

Finally, according to Fumagalli et al. [207] *WARS2/TBX15* might convey adaptation to cold in Greenlandic Inuits by decreasing the waist/hip ratio and possibly influencing the regulation of brown and beige adipose tissue. Subsequently, the type of tissue can influence lipid oxidation and increase of heat generation in response to cold. We found ancestral rs2298080 alleles in Steigen and Hum1, while both Hum2 and SF12 were heterozygotes.

Table S8.5 Diet related SNPs.

SNPID	gene	mutation	strand	Steigen	SF9	SF11	SBj	Hum2	Hum1	SF12
rs182549	*LCTb*	C > T	+	NA	NA	NA	NA	7C	NA	90C
rs4988235	*MCM6/LCTa*	G > A	-	2G	3G	1G	NA	2G	2G	66G
rs174546	*FADS1/FADS2*	T > C	+	NA	1C	NA	1T	NA	NA	44C,36T
rs7940244	*DHCR7&NADSYN1*	C > T	+	NA	4T,2C	1T	NA	4C,1T	NA	52C,38T
rs7944926	*DHCR7&NADSYN1*	G > A	+	NA	1A	NA	NA	2A	NA	41A,20G
rs12785878	*NADSYN1*	G > T	+	1T	NA	NA	NA	4G,1T	2T	48T,43G
rs1050152	*SLC22A4*	C > T	+	3C	NA	NA	NA	1C	1C	88C
rs272872	*SLC22A4*	G > A	-	1A	2G	NA	NA	6A	NA	68A
rs653178	*ATXN2*	C > T	-	1T	1T	NA	NA	3T	NA	52T
rs74771917	*FADS1-3*	C > T	+	1C	2C	NA	1C	NA	NA	91C
rs3168072	*FADS1-3*	A > T	+	NA	1A	NA	NA	1A	NA	62A
rs12577276	*FADS1-3*	A > G	+	3A	NA	NA	NA	4A	2A	87A
rs7115739	*FADS1-3*	G > T	+	NA	NA	NA	NA	3G	NA	72G
rs174602	*FADS1-3*	C > T	-	1T	NA	NA	2T	4T	NA	99T
rs174570	*FADS1-3*	C > T	+	NA	NA	NA	1T	7T	NA	31T,27C
rs2298080	*WARS2/TBX15*	G > A	+	2G	NA	NA	NA	6G,2A	2G	49A,29G

S8.5 Taste-perception related SNPs

The ability to taste phenylthiocarbamide (PTC) is closely linked with the ability to taste other compounds containing thiocyanate. This ability was linked to a gene called taste receptor 2 member 38 (*TAS2R38*) and is present among humans on two common haplotypes: taster (PAV haplotype) and non-taster (AVI haplotype). The latter is inherited as an autosomal recessive trait (OMIM 171200) [194]. In our study, the two individuals with sufficient genome coverage at the sites of interest (Hum2 and SF12) were carriers and were homozygous for the PAV haplotype (Table S8.6); thus, both had the ability to taste PTC [209]. Interestingly, the taster haplotype frequency in present-day Norway and Finland is in between PAV and AVI, and it has been suggested that the observed frequency arose as a result of drift rather than selection favoring specific haplotypes [210]. The distribution of PAV ad AVI haplotypes among WHG, EHG, CHG and PHG revealed that most ancient individuals were carriers of the PAV haplotype (see Table S8.6, Supplementary Table 1). Only two individuals were identified as carriers of AVI phenotype (Samara, Vestonice16), while further two were carriers of mixed haplotypes (Ajvide58, Kostenki14). Due to the limited number of available hunter-gatherer genomes, this result should be taken with caution, however based on available data it seems all Western and Northern hunter-gatherers were carriers of PAV phenotype (with the exception of heterozygous Ajvide58), while AVI were only identified in individuals from central Europe.

Table S8.6 Bitter taste receptor TAS2R38 alleles

SNPID	gene	mutation	strand	Steigen	SF9	SF11	SBj	Hum2	Hum1	SF12
rs713598	*TAS2R38*	G > C	-	NA	NA	NA	1G	4C	NA	96C
rs1726866	*TAS2R38*	G > A	-	3G	NA	NA	NA	4A	NA	85A
rs10246939	*TAS2R38*	C > T	+	1T	3T	NA	NA	7T	NA	68T
rs4726481	*TAS2R38*	T > G	+	2T,1G	NA	NA	NA	9G	NA	52G
rs17162635	*TAS2R38*	T > A	+	2T	NA	NA	NA	4T	NA	63T
Haplotype	*TAS2R38*	-	-	NA	NA	NA	NA	PAV	NA	PAV

S8.6 Copy number variation at the AMY1 region

Copy number variation (CNV) in the human loci encoding for salivary amylase (*AMY1*), an enzyme involved in starch degradation, has been reported as an example of recent positive selection associated to diet changes [211]. Different from other great apes and archaic hominids [111,211] – which have 2 *AMY1* copies – the reference human genome has 3 copies of the *AMY1* gene: *AMY1A*, *AMY1B* and *AMY1C* (a total of six diploid copies). Each of these loci has been observed to expand into a higher copy number in present-day human populations, which might

be explained by a starch-consumption adaptation upon the advent of agriculture [211]. The authors reported that modern human agriculturalist groups – known to have a higher starch consumption - present a higher *AMY1* gene copy number, in contrast to contemporary hunter gatherers – assumed to have a low-starch consumption - and exhibiting a lower copy number variation. Most present-day Europeans seem to have around 8-10 copies, with as low as two and as high as sixteen diploid copies. Other present-day African and non-African populations have similar distributions, however Asians present the widest range of copies in modern humans [111]. Recently it has been reported that the La Braña and Loschbour WHGs, the Motala12 SHG and the Stuttgart early farmer, had on average ~5, 13, 6 and 16 *AMY1* diploid gene copies, respectively [81,125]. While the *AMY1* diploid gene copy numbers in La Braña and the Motala12 individual agree with a low-starch hunter-gatherer diet, the prediction for the Loschbour WHG, falls well within the range of current European populations and farmers [81,111,211].

In order to continue to shed light on the distribution of *AMY1* gene copies in ancient hunter-gatherer groups, we aim to characterize the CNV in the SHGs from this study by predicting CNVs from the distribution of *de novo* mapped reads [125,212]. Illumina reads without adapter sequences were split into 36 bp reads, using the fastqutils module of the ngstils 0.5.9 suite [213]. Split reads were mapped against the human genome reference using mrFAST version 2.6.1.0 [212]. We used the hg19 version of the human reference masked for repeats detected with RepeatMasker (www.repeatmasker.org) and Tandem Repeat Finder [214]. Following mapping, mrCaNaVaR V 0.51 [212] was employed to calculate the mean read depth per base pair in 1 kb non-overlapping windows of non-repetitive sequence.

Using coordinates obtained from the UCSC browser for each of the *AMY1* genes (*AMY1A*, *AMY1B* and *AMY1C*), as well as the most upstream coordinate and most downstream positions, for *AMY1A* and *AMY1C* respectively, the average number of diploid *AMY1* gene copies was estimated for the SF12, Hum2 and Steigen individuals. CNV detection in the other individuals from this study was hindered by their low nuclear genome coverage. Given SF12's high nuclear genome coverage, and that the described methodology involves remapping raw sequencing reads, the following time-saving strategy was envisioned to analyze that individual's data. In brief, SF12 libraries were ranked by a high nuclear coverage and low mitochondrial contamination estimates. Following that order, libraries were grouped and merged together to generate three separate "test genomes" of an average coverage of 5x each. The protocol described above to call CNVs was performed for all three subsets separately as for the other individuals.

Table S8.7 Estimated copy numbers for the *AMY1* region.

Sample	*AMY1A*	*AMY1B*	*AMY1C*	Full region
SF12_1st_5x	10.3065	10.3256	9.53159	8.53405
SF12_2nd_5x	10.4606	10.5248	9.83696	8.38953

SF12_3rd_5x	8.08692	8.11287	7.59612	7.0647
Steigen	14.7399	14.8016	14.9535	12.2159
Hum2	12.8483	12.8709	12.8003	11.6797

The predicted copy number for each *AMY1* gene loci varied from 7.5 to ~15 for SF12, Hum2 and Steigen individuals. Given the high sequence similarity between the *AMY1* A, B and C genes and as mrFAST allows for each read to map against multiple sites, we expect most reads to map against all three genes and not only one of them. Therefore, we also use the depth estimate encompassing the whole region to report the number of *AMY1* copies per individual, which results in slightly smaller estimates. Our results, together with those from the LaBraña1, Loschbour, Motala12 and Stuttgart individuals, suggest that the amylase copy number expansion in modern-humans had occurred prior to the advent of agriculture in Europe. An observation also consistent with other recent studies on the worldwide distribution of the CNV [215].

S8.7 Metabolic syndrome related SNPs

Type 2 diabetes (T2D; OMIM 125853) is a complex disease with a variable world-wide distribution. Its development is preconditioned by a combination of genetic and environmental factors. A number of genetic risk scoring systems have been developed aming to facilitate T2D risk prediction. Following the approach of Lazaridis et al. [81] we use two scoring systems [216,217] to facilitate comparison of our ancient metabolic genotypes to modern non-diabetic genotype.

We were able to calculate phenotype prediction scores for two out of seven investigated ancient SHG individuals. Note that results obtained for Hum2 should be treated with caution due to low coverage at a number of polymorphic sites. According to both systems, SF12's count based risk scores seem slightly lower than those of Hum2 with weighted genetic risk scores of 6.26 and 80.51 for SF12 and 8.16 and 79.14 for Hum2 according to the two different models respectively [216,217]. Both individuals showed genetic risk scores (GRS) at the lower end of the score spectrum detected among modern-day Europeans. Even though the GRS prediction power is limited due to a number of other non-genetic factors influencing T2D, our results show that both SF12 and Hum2 had reduced MetS risk scores compared to both modern day Europeans [216,217], as well as the ancient forager and farmer Loschbour and Stuttgart individuals from continental Europe. We analyzed the genetic risk scores in another seven hunter-gatherers with >1x genome coverage. In all ancient individuals the genetic risk scores were either at the low (Motala12, Kotias, Kostenki14) or intermediate (Ajv58, Ajv70, LaBrana1, Bichon) levels compared to modern populations. The two ancient individuals with highest risk score, but still

within the median range compared to modern populations, were Bichon and Loschbour or Ajv58, depending on the SNP scoring panel used.

Table S8.8 Method 2 results from Cornelis et al [217](217)

SNPID	gene	mutation	risk allele	strand	Hum2	SF12
rs564398	*CDKN2A/B*	T > C	T	-	3C	38C,40T
rs10010131	*WFS1*	G > A	G	+	2A,1G	42G,39A
rs4402960	*IGF2BP2*	G > T	T	+	3T,3G	45G
rs1801282	*PPARG*	C > G	C	+	2C,1G	28G,26C
rs5219	*KCNJ11*	C > T	T	+	2C	58C
rs1111875	*HHEX*	T > C	C	-	3C	34T,37C
rs13266634	*SLC30A8*	C > T	C	+	3C	71C
rs10811661	*CDKN2A/B*	C > T	T	+	8T	47T,46C
rs7756992	*CDKAL1*	G > A	G	+	8A	53A
rs12255372	*TCF7L2*	G > T	T	+	6G	95G
The Genetic Risk Score (GRS)					**9**	**7**
Weighted Genetic Risk Score					**8.16**	**6.26**

Table S8.9 Method 1 results from Meigs et al. [216](216)

SNPID	gene	mutation	risk allele	strand	Hum2	SF12
rs7901695	*TCF7L2*	C > T	C	+	2C,1T	55T
rs7903146	*TCF7L2*	T > C	T	+	6C	54C
rs1470579	*IGF2BP2*	C > A	C	+	2A	66A
rs10811661	*CDKN2A/B*	C > T	T	+	8T	46C,47T
rs864745	*JAZF1*	T > C	T	-	1T,1C	35C
rs5219	*KCNJ11*	C > T	T	+	2C	58C
rs5215	*KCNJ11*	T > C	C	+	1T	75T
rs12779790	*CDC123/CAMK1D*	A > G	G	+	8A	57A
rs7578597	*THADA*	T > C	T	+	4T	63T
rs7754840	*CDKAL1*	G > C	C	+	2G	41G
rs7961581	*TSPAN8/LGR5*	C > T	C	+	3T	47T

rs4607103	*ADAMTS9*	C > T	C	+	5C	69C
rs1111875	*HHEX*	T > C	C	-	3C	37C,34T
rs10923931	*NOTCH2*	G > T	T	+	6G	85G
rs13266634	*SLC30A8*	C > T	C	+	3C	71C
rs1153188	*DCD*	A > T	A	-	4A	34T,27A
rs1801282	*PPARG*	C > G	C	+	2C,1G	28G,29C
rs9472138	*VEGFA*	C > T	T	+	6C	34C,32T
rs10490072	*BCL11A*	T > C	T	+	4C*	57C
rs689	*INS*	A > T	A	-	2A	35A,32T
The Genetic Risk Score (GRS)					**14**	**14**
Weighted Genetic Risk Score					**79.14**	**80.51**

S8.8 Blood type

AB0 blood group polymorphisms often referred to as classical genetic markers have been utilized in population studies ever since their discovery in the early 20th century. AB0 blood groups consist of three major alleles (A, B and 0) located on chromosome 9. Alleles A and B are codominant, while 0 is recessive. Allele A has two major subgroups: A_1 and A_2. Therefore, there are six major possible AB0 phenotypes encoded by various genotypes (listed in parenthesis): A_1 (A_1A_1, A_1A_2, A_10), A_2 (A_2A_2, A_20), A_1B (A_1B), A_2B (A_2B), B (B0 or BB) and 0(00). The Rh (Rhesus) system is encoded by the RHD gene encoding for the D polypeptide. Rh+ denotes presence of the D antigen, while Rh- indicates absence of D [218].

By screening a number of SNPs associated with ABO polymorphisms we aimed at describing blood group variation in the investigated ancient individuals. There are numerous different SNPs in the AB0 gene encoding amino acid changes defining various AB0 alleles. We have checked our individuals at a number of SNPs in the region, of which the following are the best known blood group defining variants: rs8176719 (G/-) defines blood group 0 in case of G deletion; rs7853989 (C;C or C;G) defines likely blood group B while (G;G) excludes B, rs8176746 (C;C) defines blood group A and (A;A) is characteristic of blood group B, rs8176747 (G;G) defines either allele A or O while rs590787 is a SNP defining Rhesus (Rh) with (C;C) defining Rh- and (C;T) and (T;T) defining Rh+ [219]. It was possible to make broad blood group inferences in Hum2 who was carrier of at least one 0 allele. However lack of coverage at rs590787 did not allow for evaluation of presence or absence of antigen D (Rh+/Rh-). The high coverage individual (SF12) was identified as a carrier of A0+.

Table S8.10 Blood defining SNPs

SNPID	gene	mutation	strand	Hum2	SF12
rs8176719	*ABO*	A > C	-	2-	25G,25-
rs1053878	*ABO*	G > A	-	NA	44G,36A
rs7853989	*ABO*	G > C	+	4G	102G
rs8176740	*ABO*	A > T	-	3A	83A
rs8176743	*ABO*	C > T	-	5C	99C
rs8176746	*ABO*	G > T	-	1G	84G
rs41302905	*ABO*	C > T	+	1C	80C
rs8176747	*ABO*	C > G	-	1C	79C
rs8176749	*ABO*	C > T	-	2C	81C
rs8176720	*ABO*	C > T	-	NA	50T
rs8176741	*ABO*	G > A	-	3G	89G
rs8176750	*ABO*	G > -	-	NA	55G
rs590787	*ABO*	A > C/G/T	-	NA	1A
Blood type ***(genotype)***				**NA** ***(0?)***	**A+** ***($A_1$0)***

(http://www.ncbi.nlm.nih.gov/projects/gv/mhc/xslcgi.cgi?cmd=bgmut/systems_info&system=abo)

S8.9 Other phenotypic traits

We have tested all individuals for a number of SNPs associated with various additional phenotypic traits. SF12, SF9 and Hum1 had normal risk of hypertension carrying the low-risk allele A at the *AGT* gene (similar to 12 other hunter-gatherers) [220]. This was not the case for Steigen and Hum2 who were carriers of risk associated allele G and had increased risk of hypertension (the allele was also found in Motala1, Loschbour, Bichon, Karelia and Kotias individuals) (see Supplementary Table 1 and Table S8.11).

KITLG conveys increased risk of testicular cancer, but also provides protection from sun damage causing UV protective tanning response in lighter skinned individuals [221]. The variant associated with UV protection was found in SF12 but not in SF11 or Hum2. The derived variant seems to be widespread among Scandinavian hunter-gatherers (identified in all Motala individuals), moderate in EHG and CHG individuals, but not particularly common within WHGs (See table S8.11).

Furthermore, SF9, SBj, Hum2 and SF12 as well as all other HG tested (with the exception of KO1), carried allelic variants at *ABCC1* gene associated with wet earwax, normal body odor and normal colostrum [222].

The lack of variant A alleles at the *ALDH2* (rs671) suggests absence of the Asian flush and reduced hangovers in SF12, SF9, SF11, SBj as well as all tested Hgs [223,224]. None of the tested individuals with sequence coverage at position rs1229984 were carriers of the protective A allele (SF12, SF9, Hum2, Hum1). This variant, in its heterozygous state, has only been found in four non Scandinavian hunter-gatherers (Loschbour, Villabruna, Karealia and Kostenki14). If present, *ADH1B**47His conveys reduced risk for oral cancer and is associated with alcohol metabolism in East Asians. The mutation is closely linked to Asian flush, which is believed to protect from alcoholism [199,225]. It has been suggested that prevalence of derived *ADH1B**47His alleles could be explained by the increased alcohol consumption in connection to Neolithic rice domestication [199].

Hereditary haemochromatosis is a genetic disease caused by mutations in the *HFE* gene leading to increased accumulation of iron in body tissues, especially the liver, often resulting in cirrhosis [226]. The mutation causing haemochromatosis was hypothesized to have been disseminated with either Viking or Celtic expansions [227,228]. Steigen and Hum2 carried at least one copy of the risk-allele so they could be affected by a mild form of haemochromatosis. The frequency was very similar in all other tested hunter-gatherers with not a single identified carrier of the risk allele A at position rs1800562. Thus, the oldest so far identified European carrier of the C282Y mutation was a Bronze Age individual from Ireland [159].

Table S8.11 Allelic state of selected metabolic and physical conditions in all seven new SHGs.

SNPID	gene	mutation	strand	Steigen	SF9	SF11	SBj	Hum2	Hum1	SF12
rs699	*AGT*	G > A	-	1G	2A	NA	NA	2G,2A	2A	98A
rs4590952	*KITLG*	G > A	+	NA	NA	1A	NA	5A,1G	NA	50G
rs3827760	*EDAR*	A > G	-	1A	3A	NA	NA	8A	NA	88A
rs17822931	*ABCC1*	C > T	+	NA	4C	NA	1C	1C	NA	79C
rs671	*ALDH2*	G > A	+	NA	3G	1G	2G	NA	NA	100G
rs3811801	*ADH1Ba*	G > A	-	2G	NA	NA	1G	1G	3G	73G
rs1229984	*ADH1Bb*	C > T	-	NA	1C	NA	NA	3C	NA	57C
rs1800562	*HFE*	G > A	+	NA	1G	NA	NA	NA	NA	87G
rs1799945	*HFE*	C > G	+	2G	2C	NA	NA	2C,1G	NA	82C

S8.10 Immunological variants

A host-pathogen “arms race” has shaped the human genome as a result of selective pressures acting upon genes involved in host resistance and immune response. The Neolithic revolution has been regarded as one of the most important events during that process, due to the

unprecedented exposure to zoonotic infections within relatively large new farmer settlements [229]. While several studies have suggested that a large number of the common modern infections in humans could be due to zoonotic post-Neolithic events [229,230], more recent findings suggest a more complex scenario.

It is reported that a Mesolithic hunter-gatherer from the Iberian Peninsula presented both derived and ancestral variants at SNPs influencing the susceptibility to infections, and within immune-related genes under positive selection in Europeans [125]. It was thus concluded that both pre- and post-Neolithic events contributed to shaping the immune system of present-day Europeans. However, if post-Neolithic events contributed or not to such adaptation remains unknown. The genomes of the seven SHG in this study, together with other Mesolithic and Paleolithic hunter-gatherers across Europe, provide an unprecedented sample size of 27 pre-Neolithic hunter-gatherers (Table S8.1 and S8.12) to investigate this topic.

We built an empirical distribution of derived allele frequency differences between present-day Europeans and HGs for all genic SNPs throughout the genome. We then compared this genome-wide distribution to the allele frequency differences for 67 SNPs from genes which show patterns of positive selection in Europeans and which include polymorphisms that influence susceptibility to infections in Europeans (Table S8.12) [125]. If these immune response sites are among the regions of the genome that have changed the most, then exposure to previously unknown environments and pathogens would have shaped the immune system since Mesolithic times.

In order to build the empirical distribution, derived allele frequencies were estimated for all 16,668,219 SNPs (as annotated in Ensemble 75) in all present-day Europeans of the 1000 genomes project (1KGP) data [106]. Next, we also estimated the derived allele frequency at each of these positions for the 27 screened European HGs as described above (S8.1). We restricted our analysis to sites were the derived allele was variable in both populations. Singletons in present-day Europeans, and singletons in the ancient population with data for less than four HGs were excluded as well. This filter avoids potential sequencing errors or unfiltered DNA damage (see beginning of Section S8) to be considered as low frequency variants. Our empirical distribution of allele frequency differences between the two groups was derived from 5,488,574 sites after filtering. However, since HG sequencing mainly came from low-coverage shotgun or capture data, large proportions of missing data adds noise to the derived allele frequency estimations for ancient populations. In order to obtain somewhat reliable allele frequency estimates, we restricted our analysis to SNPs that had sequence data from at least 25% (7/27) of the HGs.

Table S8.12 shows the results for all SNPs in this analysis. There appears to be an enrichment of the immunological SNPs among the sites that have changed the most after the Mesolithic period. Nine out of 67 immune SNPs (13.4%) were in the top 5% of the empirical distribution when restricting to sites which were covered in at least 25% of the HGs. Seven out of 42 sites (16.7%) were among the top 5% when restricting to sites covered in at least 50% of the HGs. Figure S8.1 was produced with sites that had information from at least 25% of all HGs.

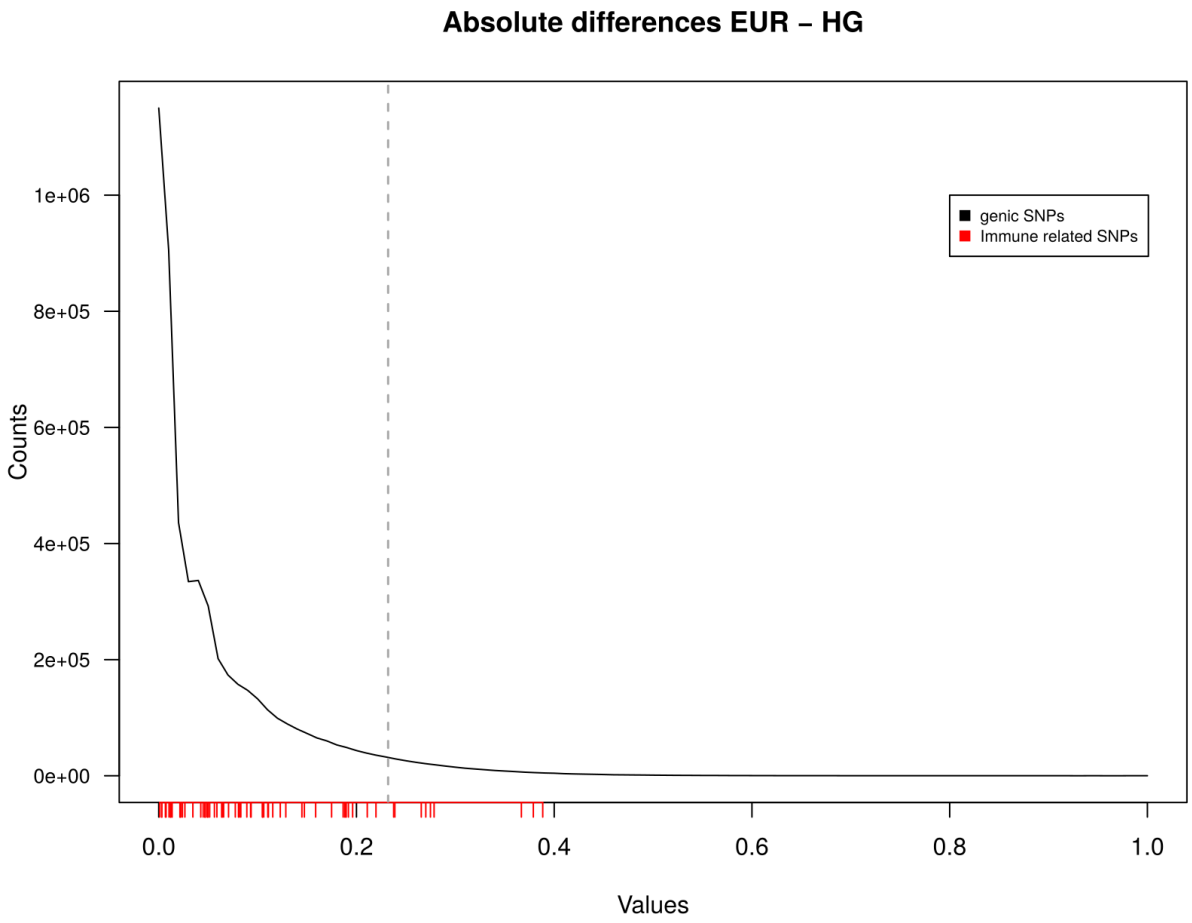


Figure S8.1 Distribution of absolute allele frequency differences between modern Europeans and Mesolithic Europeans for all genic sites covered in at least 25% of the Mesolithic individuals. The dashed vertical line represents the top 5% of all SNPs, the allele frequency differences of immune SNPs are marked in red.

Table S8.12 Comparison of the allele frequency differences between immunological SNPs and a genome-wide distribution for all genic SNPs.

				Derived allele frequencies			Percentile of the genome-wide distribution conditional on the proportion of HGs covered		
SNP ID	**Gene**	**Derived allele**	**% of covered HGs**	**EUR**	**HG**	**Absolute derived allele frequency difference**	**> 25 % HG w/ data**	**> 50% HG w/ data**	**> 75% HG w/ data**
rs497116	*CASP12*	A	44.4444	0.999	1.000	0.001	3.064	NA	NA
rs2039381	*IFNE*	A	44.4444	0.003	0.000	0.003	12.836	NA	NA
rs9302752	*NOD2*	T	85.1852	0.255	0.258	0.003	17.541	17.926	10.625
rs4986790	*TLR4*	G	81.4815	0.057	0.050	0.007	26.389	26.951	16.591
rs10982385	*TNFSF15*	G	66.6667	0.573	0.580	0.007	28.252	28.84	NA
rs1800872	*IL-10*	T	88.8889	0.240	0.250	0.010	32.885	33.554	21.468
rs1024611	*CCL2/MCP-1*	G	62.963	0.316	0.304	0.012	34.328	35.039	NA
rs5743315	*TLR3*	A	37.037	0.012	0.000	0.012	34.381	NA	NA
rs2015070	*CCL18*	T	66.6667	0.096	0.083	0.013	36.369	37.111	NA
rs2066842	*NOD2*	T	74.0741	0.247	0.260	0.014	36.514	37.272	NA
rs1800451	*MBL2*	T	77.7778	0.012	0.033	0.021	43.669	44.667	31.852
rs2015086	*CCL18*	A	85.1852	0.860	0.838	0.022	43.729	44.726	32.015
rs61752945	*RIG-1*	T	40.7407	0.009	0.031	0.022	44.421	NA	NA
rs5743708	*TLR2*	A	88.8889	0.024	0.000	0.024	45.441	46.519	33.989
rs4251545	*IRAK4*	A	74.0741	0.087	0.113	0.026	47.339	48.444	NA
rs2569190	*CD14*	A	77.7778	0.486	0.452	0.034	52.3	53.591	43.228
rs17217280	*RIG-1*	T	37.037	0.136	0.094	0.042	57.226	NA	NA
rs12212067	*FOXO3A*	T	48.1481	0.872	0.917	0.045	58.924	NA	NA
rs30461	*IL-29*	A	88.8889	0.888	0.934	0.047	59.871	61.951	50.358
rs10930046	*IFIH1/MDA5*	T	44.4444	0.988	0.941	0.047	60.179	NA	NA
rs3802814	*Mal/TIRAP*	A	55.5556	0.172	0.125	0.047	60.213	62.277	NA
rs2069727	*IFNG*	C	85.1852	0.463	0.414	0.049	61.437	63.584	51.74
rs4986791	*TLR4*	T	74.0741	0.058	0.107	0.049	61.6	63.755	NA
rs2779249	*NOS2A*	A	51.8519	0.325	0.275	0.050	61.977	64.098	NA
rs14304	*CCL18*	T	62.963	0.318	0.370	0.051	62.751	64.871	NA
rs5743836	*TLR9*	A	44.4444	0.869	0.813	0.056	65.004	NA	NA
rs8078340	*NOS2A*	A	40.7407	0.130	0.071	0.059	66.014	NA	NA
rs361525	*TNF*	A	81.4815	0.064	0.000	0.064	67.709	69.389	59.441
rs179008	*TLR7*	T	29.6296	0.231	0.167	0.064	67.969	NA	NA
rs10114470	*TNFSF15*	T	81.4815	0.325	0.391	0.066	68.398	70.065	60.443
rs16910526	*Dectin-1/CLEC7A*	C	40.7407	0.071	0.000	0.071	70.047	NA	NA
rs4574921	*TNFSF15*	C	44.4444	0.266	0.344	0.077	72.168	NA	NA

rs2274894	*NOS2A*	T	55.5556	0.396	0.476	0.081	73.137	74.867	NA
rs3753344	*TNFRSF18*	A	37.037	0.052	0.133	0.082	73.435	NA	NA
rs1873613	*LRRK2*	C	44.4444	0.695	0.778	0.083	73.81	NA	NA
rs8177374	*Mal/TIRAP*	T	55.5556	0.186	0.275	0.089	75.571	77.395	NA
rs7137054	*SOCS2*	G	48.1481	0.157	0.250	0.093	76.61	NA	NA
rs1891467	*TGFB2*	A	85.1852	0.761	0.855	0.093	76.757	78.636	71.363
rs1800629	*TNF*	A	85.1852	0.134	0.029	0.105	79.595	81.444	74.994
rs10745657	*SOCS2*	G	48.1481	0.506	0.611	0.105	79.639	NA	NA
rs2305619	*PTX3*	G	62.963	0.544	0.438	0.106	79.916	81.748	NA
rs333	*CCR5*	-	40.7407	0.110	0.000	0.110	80.798	NA	NA
rs42490	*RIPK2*	A	81.4815	0.611	0.722	0.111	80.929	82.703	76.786
rs7215373	*NOS2A*	T	44.4444	0.532	0.647	0.115	81.798	NA	NA
rs2228428	*CCR4*	T	81.4815	0.314	0.191	0.123	83.216	84.899	79.875
rs8057341	*NOD2*	A	85.1852	0.295	0.167	0.129	84.203	85.857	81.199
rs3750920	*TOLLIP*	T	44.4444	0.457	0.313	0.145	86.742	NA	NA
rs11754268	*IFNGR1*	T	37.037	0.231	0.083	0.147	87.098	NA	NA
rs3775291	*TLR3*	T	81.4815	0.324	0.483	0.159	88.605	90.109	87.081
rs5030737	*MBL2*	A	70.3704	0.060	0.234	0.175	90.444	91.843	NA
rs900	*TGFB2*	T	44.4444	0.280	0.094	0.187	91.601	NA	NA
rs3802813	*Mal/TIRAP*	A	77.7778	0.042	0.230	0.188	91.739	93.053	91.301
rs3764880	*TLR8*	A	70.3704	0.731	0.920	0.189	91.833	93.14	NA
rs3135499	*NOD2*	A	85.1852	0.565	0.375	0.190	91.887	93.19	91.497
rs2043055	*IL-18BP*	G	70.3704	0.372	0.180	0.192	92.085	93.378	NA
rs4804803	*DC-SIGN*	A	44.4444	0.784	0.588	0.196	92.464	NA	NA
rs6478108	*TNFSF15*	C	88.8889	0.336	0.547	0.211	93.619	94.788	93.638
rs4833095	*TLR1*	T	51.8519	0.720	0.500	0.220	94.225	95.339	NA
rs2333227	*MPO*	T	44.4444	0.238	0.000	0.238	95.317	NA	NA
rs2074158	*LGP2*	C	70.3704	0.178	0.417	0.239	95.386	96.384	NA
rs5743899	*TOLLIP*	T	62.963	0.789	0.524	0.265	96.648	97.488	NA
rs4077515	*CARD9*	T	66.6667	0.397	0.667	0.270	96.825	97.64	NA
rs11096957	*TLR10*	G	74.0741	0.398	0.672	0.275	97.01	97.798	NA
rs4073	*IL-8*	T	40.7407	0.579	0.300	0.279	97.143	NA	NA
rs5743810	*TLR6*	A	66.6667	0.409	0.042	0.367	99.088	99.474	NA
rs2430561	*IFNG*	A	29.6296	0.462	0.083	0.379	99.225	NA	NA
rs3184504	*SH2B3*	T	81.4815	0.464	0.076	0.388	99.316	99.636	99.693

S9 Genetic reconstruction of SF12's face

A facial prediction for SF12 was obtained using a two-step procedure following Claes et al. [231] and based on Claes et al. [232,233], but using larger datasets. In a first phase, a base-face was created accounting for the effects of sex, genomic ancestry, age, height and weight, using diverse panel of 4,110 participants with 3D facial images and SNP-chip genotyping data calibrated against 11 Hapmap project populations [234] and using a partial least squares regression model. Average values for age (25 years), height (165 cm) and weight (55 kg) were used. In a second phase, the effects of 364 individual SNPs, identified and modelled with BRIM [233] on 1,984 participants of European-descent, were overlaid on the base-face forming the predicted-face in a process akin to a photomontage or image blending [231,231,232].

Sample and Genotypes

We performed predictive facial modeling employing two reference datasets. The first of these two datasets (FaceDS1) was used to establish the ancestry and sex base-face and includes a large number of participants (N=4,110) from many populations. In addition to large scale local sampling in State College, PA, sampling efforts in Europe (Paris, Rome, Warsaw, Dublin, Porto, and London), Africa (Cape Verde), South America (Brasilia and Baependi, Brazil), Central America (Chiapas, Mexico), and the U.S. (State College, PA, Twinsburg, OH, Williamsport, PA, Austin and San Antonio, TX, Champaign-Urbana, IL, Orlando, FL, and the Bronx and Manhattan, NY) have been performed. Human subject's oversight was provided by the Penn State University Institutional Review Board and local ethics panels. All participants provided informed consent. The individuals were genotyped on three different genotyping platforms (23andMe v3 and v4 arrays (23andMe, Mountainview, CA).and a custom Affymetrix Axiom array (Santa Clara, CA). We selected a core set of 118,420 LD-pruned autosomal SNPs, which represent the intersection across all platforms, to characterize the genetic structure of the sample calibrated against the HapMap project. The second dataset (FaceDS2) was composed of persons of primarily European ancestry (N=1984) and was used to investigate and model the facial effects of ~12,000 SNPs, which were selected based on candidate genes.

Facial phenotyping: Imaging, remapping, processing, and deconstructing

Facial surfaces were collected using photogrammetry (3dMDFace System, 3dMD, Atlanta, GA), which produces three dimensional point clouds defining the shape of the face and a texture map that contains the surface appearance information. An anthropometric mask (AM) [235] is non-rigidly mapped onto 3D images [236]. The AM is essentially a template covering the facial area of interest and the mapping thereof onto the facial images is a process equivalent to the indication of traditional landmarks. This establishes homologous spatially dense (~10,000) quasi-landmark (QL) configurations for all 3D images. Note that, by homologous, we mean that each quasi-landmark occupies the same position on the face relative to all other quasi-landmarks for all individuals. In other words, image data from different individuals are standardized and can be

analyzed in a spatially dense manner. The remapped QL meshes are next trimmed (to 7,250 QLs) to help eliminate edge effects, a common artifact for the QLs on the outer edge. Once aligned by superimposition methods, a set of remapped faces comprise a QL face-space, a rectangular matrix where the rows are different faces and the columns the X, Y, Z, coordinates for the QL positions. In other words, we have a vector of length 21,450 (= 3 x 7,150) defining the position of each face in a QL face-space. Subsequently, principal component analysis (PCA) is employed to reduce the number of dimensions from 21,450 by selecting the first set of principal components (PCs), which together capture 98.5% of the total variance observed in all 7,150 3D points over all facial images in FaceDS1 (#PCs=86) and FaceDS2 (#PCs= 78), separately. Such a dimension reduction is not required but lowers the computational burden and allows for the dimension expansion to reconstruct and display 7,150 3D points and therefore 3D facial images from 86/78 dimensional single points in the face-space. Although these reconstructions are not exact compared to the face-space prior to PCA reduction, they approximate 3D facial images with sub-millimeter (<1mm) accuracy for 3D facial images in FaceDS1 and FaceDS2, separately.

Composing the base face

The ancestry space was defined by a set of reference populations from the publicly available HapMap 3 dataset (YRI: Yorubans from Ibadan; MKK: Masai from Kenya; LWK: Luhya from Kenya; CEU: Utah residents of Northern and Western European ancestry; TSI: Italians from Tuscany; CHB: Han Chinese from Beijing; JPT: Japanese from Tokyo; CHD: Han Chinese living in Denver; GIH: Gujarati Indians from Houston; MEX: Mexicans from the Southwest; ASW: African Americans from the Southwest). The ancestry space was constructed by carrying out PCoA on the allele sharing distance matrix genotypes of 988 unrelated HapMap individuals for the same 118,420 SNPs of FaceDS1. Into this space, we projected all 4,110 individuals from FaceDS1 using PLINK. This ensures that the shape of the ancestry space is only defined by individuals from the HapMap dataset.

We tested each of the HapMap genomic PC axes for significant effects on facial variation in FaceDS1 using partial least squares regression (PLSR) [237]. A total of 23 genomic PC axes were found to have significant ($p<=0.05$) effects and were retained to model the base face. These 23 PCs were combined with self-reported sex and age, and either self-reported or measured height and weight in a PLSR prediction model. To estimate the genomic PCs for SF12, or another person not in the model, we superimpose the persons into the genomic PCoA space. We next use the 23 genomic PCs, sex deduced from zygosity at X-linked SNPs, and given age, weight, and height to predict the base-face. In the case of SF12 we stipulated an age of 25, weight of 55 kg, and height of 165 cm for females.

Identifying SNPs directly affecting facial variation and modeling their effects onto other faces and the base-face

We screened ~12,000 SNPs in FaceDS2 for genetic association with facial effects using a BRIM-based four-fold cross validation framework [232]. In brief, the full sample (N=1984) is divided

into four non-overlapping subsamples. Each fold in turn, acts like a test panel, for the other three folds combined as training panel. A potential facial effect is modeled in the training panel using PLSR under and additive SNP model, and continuous RIP values are imputed for the test panel. Subsequently these replication test values are investigated using ANOVA with SNP genotype coding for group membership. A single p-value is reported using Fisher's method of combining p-values across the four folds. 364 significant SNPs (P<=0.05) were identified as such, and for these SNPs a single PLSR model with associated RIP distributions per genotype group was created using the full sample (N=1984).

A RIP value can be obtained for any given facial phenotype and can be evaluated within the RIP distributions per genotype, therefore it forms a bridge from the facial phenotype to an underlying SNP, e.g., it allows to evaluate whether a given face matches a given genotype. This aspect was used to overlay individual SNP effects onto the base-face. Following the sequential optimization [232], and starting with the average face from the FaceDS1 face-space as an initial and current solution, the solution is evaluated against a single SNP and updated to improve its match to the given genotype. The updated solution then becomes the current solution. Updating the solution is done by moving the face along the PLSR direction associated with the SNP variant until it matches the given genotype (a single line-search optimization). Subsequently, the current solution is evaluated against another SNP and is updated again. The updating of the current solution is performed sequentially for all 364 SNPs and this is done 10 times with randomized SNP order. Finally, the facial difference on the level of quasi-landmarks between the initial solution (average of FaceDS1 face-space) and the final solution is added to the predicted base-face, to create the final predicted-face (Figure S9.1).

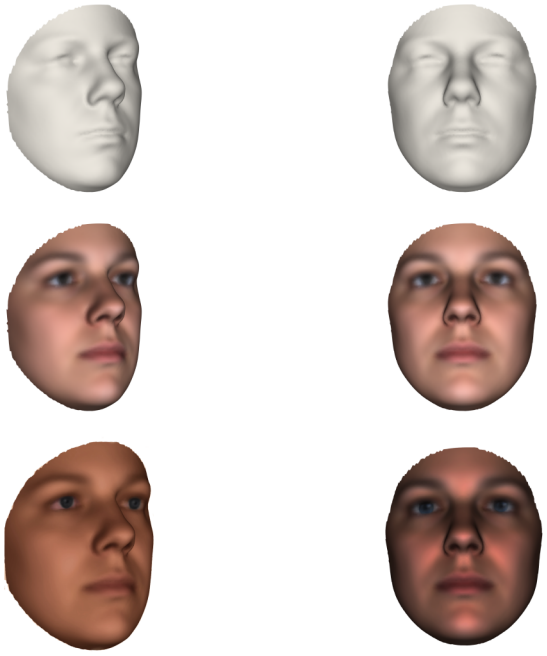


Figure S9.1 Facial reconstruction of SF12 based on diploid genotype calls. The first row shows the predicted facial shape, the second row adds European pigmentation to the shape, and the third row shows a modified pigmentation based on major pigmentation loci (Section S8).

S10 Adaptation to high-latitude climates

S10.1 Genome-wide scan for patterns of adaptation

This study presents the largest number of Mesolithic Scandinavians (to date) that have had their genomes sequenced. These individuals were among the first pioneering inhabitants of Scandinavia and northern Europe. While genetic variation of the Mesolithic populations falls outside the modern-European genetic variation, it is known that modern-day Europeans trace some ancestry to these groups [81,92]. Assuming that Mesolithic as well as modern-day northern Europeans were adapted to similar climatic conditions, possibly by sharing some genetic material (i.e., continuity) as previously demonstrated [92,107], we investigated if certain alleles or gene-regions show a long term continuity in the region. Signals of such allele/gene-region continuity will be informative of local adaptation, possibly linked to the environment at northern latitudes. A strong selective pressure in high-latitude regions is cold temperature. The response to cold stress is cardiovascular, metabolic and endocrinological while physiological adaptation to cold climates is mainly insulative or metabolic [238]. A recently detected example of adaptation to arctic climates is the gene cluster for fatty acid desaturase enzymes (FADS), in the Greenlandic Inuit population, which modulate fatty acid composition [207].

We scanned the genomes for SNPs with similar allele frequencies in Mesolithic and modern-day northern Europeans, and contrast it to a modern-day population from southern latitudes. Pooling all Mesolithic Scandinavians together, we obtain an allele frequency estimate for Scandinavian hunter-gatherers (SHG) which is compared to modern-day Finnish individuals (FIN) and Tuscan individuals (TSI) from the 1000 genomes project [106]. 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 traces some ancestry to Mesolithic populations, but who do not trace their ancestry to groups that lived at northern latitudes in the last 7-9,000 years. Our approach is similar to PBS [239] and inspired by DAnc [240], for each SNP, we calculate the statistic D_{sel} comparing the allele frequencies between an 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 data. Only sites with a high confidence ancestral allele in the human ancestor (as used by the 1000 genomes project [106]) and with coverage for at least six ancient Scandinavians were included in the computation. Outliers detected using this approach appear to have functional relevance as the upper end of the distribution of D_{sel} values is enriched with SNPs at conserved sites (measured by GERP score > 3; [241]) (Figure S10.1).

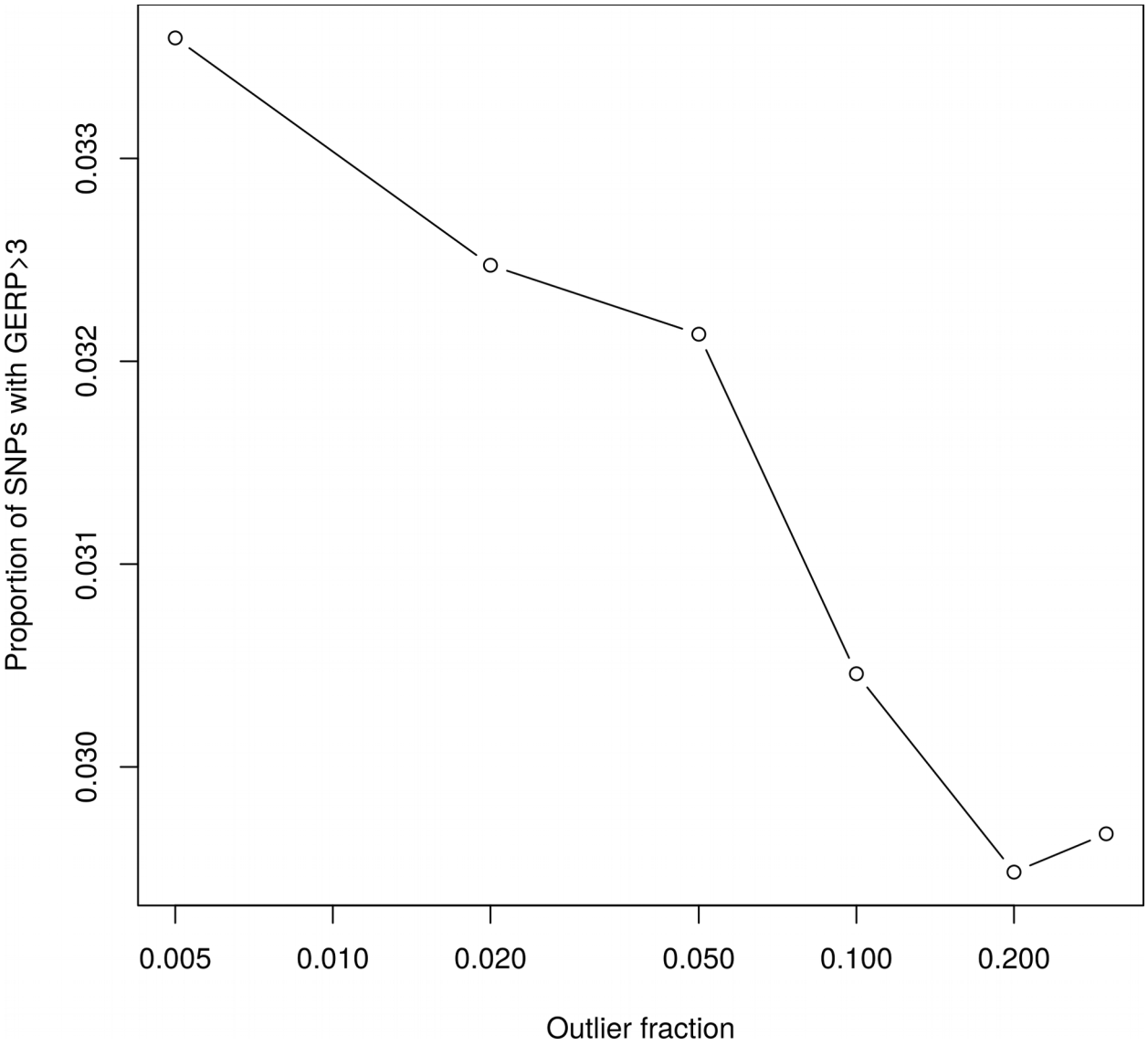


Figure S10.1 Enrichment of SNPs with high conservation among the outliers of the selection scan

We explored the most extreme and positive values of our statistic since those represent SNPs similar in modern-day and Mesolithic northern Europeans, but different to southern Europeans. We used SNPnexus [242] to obtain annotation for the top ranking SNPs in our genome-wide scan. Notably, six of the ten SNPs with the highest D_{sel} values are located in the transmembrane gene *TMEM131* (https://www.ncbi.nlm.nih.gov/gap/phegeni?tab=1&gene=23505). GWAS have associated SNPs in this gene with performance in exercise tests (rs10520549, $p<10^{-5}$ [243]) and heart rate (rs1026015, $p<10^{-5}$ [244]). The heritability of the 17 tested physical exercise phenotypes are in the range 0.30-0.52 [243]. These cardiovascular traits are likely connected to the climatic conditions in northern Europe [238]. Four of the top 100 ranking SNPs (Table S10.1) are located in *FHIT*, which has been associated with a wide range of phenotypes (Table S10.2). These

include psychological traits (sleep [245], attention-deficit hyperactivity disorder [246], major depressive disorder [247], Tobacco Use Disorder [248], Asperger Syndrome [249], metabolic traits (body mass index [250], type 2 diabetes [251]), cardiovascular traits (blood pressure [252]), and developmental traits (Cleft Lip [252], menopause [253]). Due to this large range of different phenotypes it is difficult to find a clear link to adaptation to high-latitude climates, although several of the traits involved have been linked to cold adaptation [238]. *GPC5* harbors three of the top 100 ranking SNPs. This gene has also been associated with a wide range of phenotypes,including metabolic traits (serum metabolites [254], Cholesterol and HDL [255]), immune phenotypes (Monocyte Chemoattractant Protein-1 [256], multiple sclerosis [257–259], Crohn's disease [260]) and developmental traits (Mental Competency [253], height [261], hair thickness [262], kidney aging [263]. For both *GPC5* and *FHIT*, a majority of the phenotypes are possibly involved in local adaptation, e.g. handling changes in light exposure (psychological and developmental traits) and the increased energy demand during cold seasons (metabolic and cardiovascular traits), or general physiological changes to adapt to the environment (developmental traits). The genes *PLD1* and *GABPB1* also harbor three or more SNPs out of the first 100 SNPs of D_{sel}. Unfortunately, we do not find any GWAS results for these genes. *PLD1* is involved in Ras protein signal transduction [264], so it could be connected to the response to external signals. *GABPB1* might be involved in physical performance in competitions [265], which is similar to the associations of *TMEM131* and may also be well connected to the climatic conditions in northern Europe. Other genes among the top 100 SNPs are associated with a wide range of metabolic, cardiovascular and psychological traits (Table S10.2).

All six of the highest scoring SNPs that fell within *TMEM131* show similar allele frequency differences between FIN and TSI which suggests that two different haplotypes are present in high frequencies in these two modern populations. In total, the region comprises at least 264 kilobases with allele frequency differences of up to 40%. This region is a genome-wide outlier in its allele frequency difference between FIN and TSI compared to other regions of similar length (Figure S10.3d). To produce Figure S10.3d we defined blocks as follows: For each SNP, we scanned the next 50 kbp for other SNPs with maximally 5% difference between the allele frequency differences between the two populations. A block is a sequence of such SNPs with less than 50 kbp between neighboring SNPs (note that SNPs with highly different allele frequencies were allowed if another SNP within 50 kbp had a similar allele frequency difference). Figure S10.3d shows the block with the maximum allele frequency per 1 Mbp window of the genome. In order to investigate the haplotype structure in the TMEM131 region, we phased chromosome 2 of all FIN and TSI individuals plus SF12 and Hum2 using FastPHASE 1.4 [266] (with parameters -T25 -C25 -w -Pm -Pp -H100 -K25 -Kp.1). The GNU R package pegas [267] was used to draw a haplotype network for the region (Figure S10.3e). All major haplotype configurations seem to be segregating in both modern populations but there appears to be a clear gradient between the two populations. Both of the most extreme haplotype configurations are predominantly found in either TSI or FIN, and the haplotype found in SF12 and Hum2 is more common in FIN than in TSI. This pattern of haplotype differentiation is also visible in a

haplotype bifurcation diagram around the highest scoring SNP in *TMEM131* (Figure S10.3c, drawn using the R package rehh2 [268]). The *TMEM131* region is the second strongest signal of haplotype differentiation on chromosome 2 when using a haplotype based selection scan (rsb [269], calculated with rehh2 [268]). Only the region around the Lactase gene shows a higher haplotype differentiation between TSI and FIN (Figure S10.3a,b). Both alleles at the highly differentiated SNPs are also found in other prehistoric Europeans included in this study. The SNPs are polymorphic in both hunter-gatherers and early farmers, but the haplotype found in sf12 is found in slightly higher frequencies in other hunter-gatherers (5 out of 7 with sufficient data) than in early farmers (3 out of 7 with sufficient data).

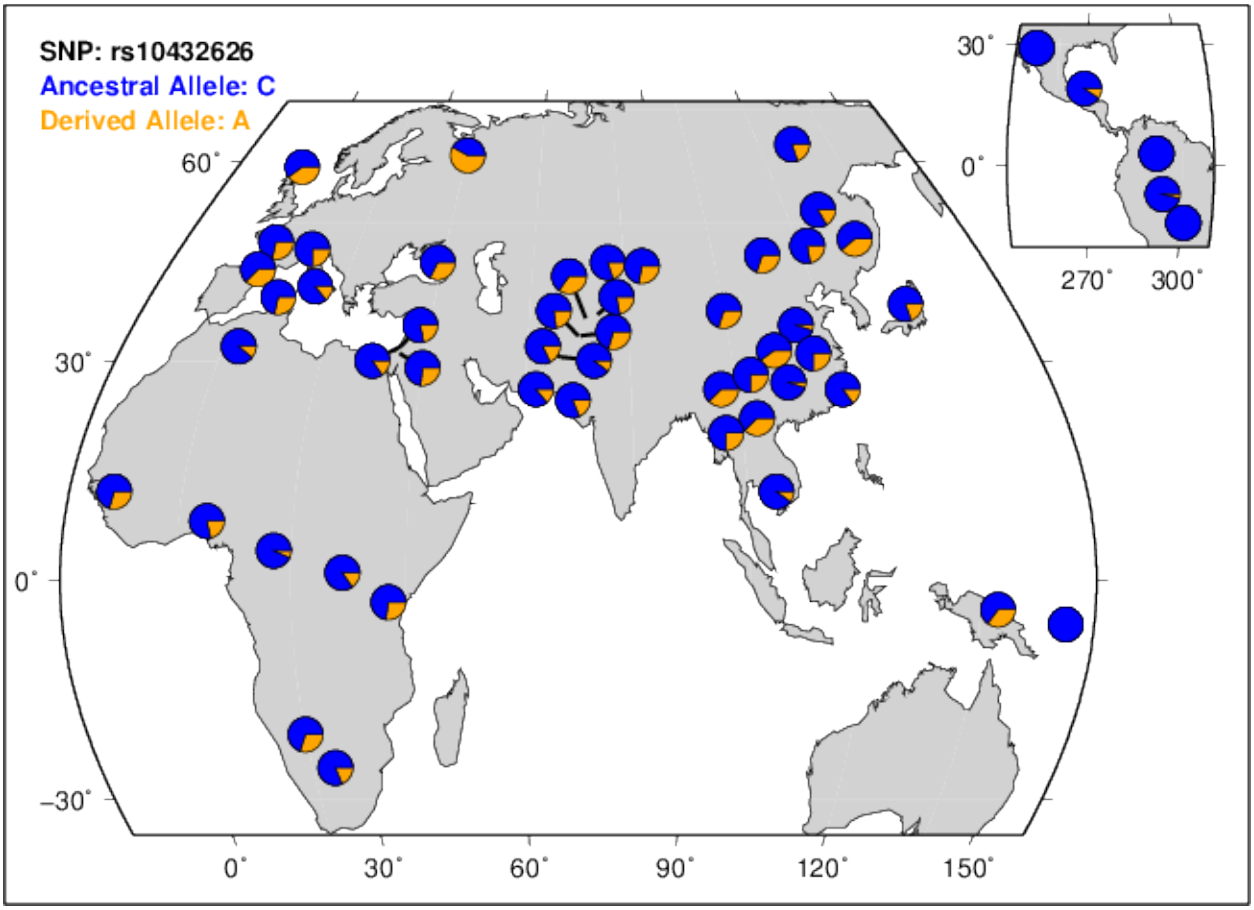


Figure S10.2 Allele frequencies of rs10432626, one of the SNPs in TMEM131. The plot was obtained from the HGDP selection browser (http://hgdp.uchicago.edu/).

In addition to investigating the genes among the top 100 SNPs, we also looked at biological process GO terms among the top 0.5% of D_{sel} scores (1298 SNPs) compared to all SNPs tested. For GO term enrichment we used Gowinda [270]. Gowinda employs a permutation approach to detect GO terms overrepresented among a subset of all SNPs analyzed. This accounts for the different lengths of genes and the number of SNPs expected for each gene. Gowinda was run in gene mode while counting all SNPs 20kbp up- or downstream to that gene. We only used categories with at least 10 genes and conducted 1,000,000 permutations. Only ten GO terms have

a FDR of less than 20%, we show the top 20 GO terms in Table S10.3. In contrast to the genes among the top 100 SNPs, these GO terms mainly include developmental processes but also some involved in signaling processes. These could be involved in polygenic adaptation to high-latitide climates and may have changed morphology in northern Europeans as represented by individuals from Pitted War Culture individuals (PWC) in osteological analyses. Comparing individuals from Funnel beaker (TRB) and PWC contexts in Sweden osteologically, a certain degree of morphological differences between the skeletons has been found. It has been shown that PWC individuals exhibit skeletal traits characteristic of cold-adaptation, such as certain facial features and limb proportions (crural index) which are absent in TRB individuals [271,272].

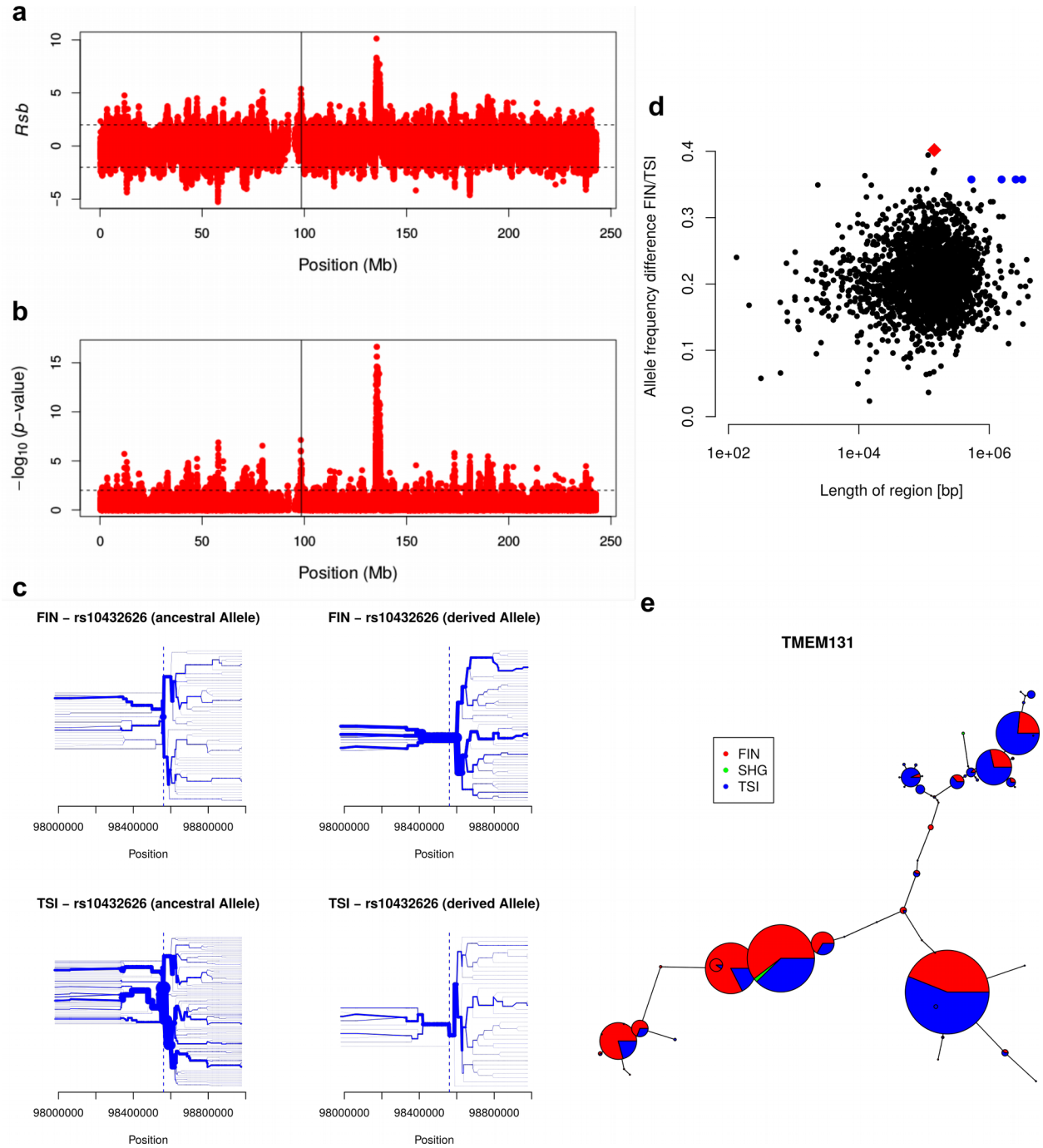


Figure S10.3 Candidate gene TMEM131 for adaptation to high-latitude climates. (a) Haplotype differentiation measured using rsb between TSI and FIN populations on chromosome 2 (calculated with rehh2 , Supplementary Information 10), and (b) -log 10 (p-value) for rsb. (c) Bifurcation of the different haplotypes around the highest scoring SNP (drawn with GNU R and rehh2). (d) Block length versus allele frequency differences between southern Europeans, TSI, and northern Europeans, FIN. Blocks are defined as the maximum physical distance (in base-pairs) between two SNPs of similarly high allele frequency difference between TSI and FIN, requiring that the block contains other SNPs with similarly high allele frequency difference, but with a maximum distance of 50 kbp between neighboring SNPs. The red diamond represents the

TMEM131 gene region, blue dots represent the OCA2/HERC region. (e) Haplotype network of the TMEM131 region (drawn with the GNU R package pegas).

Table S10.1 Top 100 SNPs of the D_{sel} analysis

SNP-ID	D_{sel}	DAF_{SHG}	DAF_{FIN}	DAF_{TSI}	Consequence type	Gene(s)
rs10432626	0.397	0	0.37	0.77	intronic	*TMEM131*
rs13020776	0.393	0.17	0.37	0.77	intronic	*TMEM131*
rs1838797	0.393	0.83	0.63	0.23	intronic	*TMEM131*
rs10210880	0.382	0	0.15	0.53	intronic	*TMEM131*
rs11692671	0.38	0.83	0.64	0.26	intronic	*ZAP70*
rs7402734	0.368	0.83	0.74	0.37	intergenic	
rs11894541	0.341	0.17	0.15	0.53	intronic	*TMEM131*
rs35940587	0.341	0.17	0.15	0.53	intronic	*TMEM131*
rs168714	0.332	0.83	0.82	0.49	intergenic	
rs6726062	0.325	0.25	0.4	0.73	intronic	*PARD3B*
rs6546608	0.321	0.83	0.74	0.42	intergenic	
rs10276954	0.319	0.61	0.58	0.26	intergenic	
rs537672	0.318	0.43	0.44	0.76	intronic, non-coding intronic	*SLC9A8*
rs6012753	0.317	0.67	0.56	0.24	intronic, non-coding intronic	*SLC9A8*
rs6118443	0.317	0.29	0.32	0.64	intergenic	
rs452203	0.315	0	0.22	0.54	intronic	*FHIT*
rs2734389	0.314	0.08	0.23	0.55	intronic	*FHIT*
rs306169	0.314	0.93	0.7	0.39	intergenic	
rs3104821	0.314	0	0.42	0.74	intergenic	
rs4882475	0.314	0.17	0.24	0.56	intergenic	
rs2218657	0.312	0.83	0.55	0.23	non-coding intronic	*MIR4435_1HG*
rs3112591	0.31	0	0.23	0.54	intergenic	
rs369278	0.304	0.08	0.24	0.55	intronic	*FHIT*
rs4845824	0.304	0.58	0.61	0.26	intergenic	
rs7963463	0.302	0	0.34	0.64	intergenic	
rs11251448	0.3	0.31	0.36	0.66	intergenic	

rs7698798	0.3	0	0.18	0.48	intronic	*MTTP*
rs7173127	0.3	0.17	0.37	0.67	intronic	*GABPB1*
rs11638564	0.3	0.83	0.63	0.33	intronic	*GABPB1*
rs1972701	0.3	1	0.63	0.33	intronic	*GABPB1*
rs1972700	0.3	0	0.37	0.67	intronic	*GABPB1*
rs28372114	0.3	0.14	0.37	0.67	intronic	*GABPB1*
rs11853236	0.3	0.17	0.37	0.67	intronic	*GABPB1*
rs11070768	0.3	0.83	0.63	0.33	intronic	*GABPB1*
rs2033115	0.3	0.14	0.37	0.67	intronic	*GABPB1*
rs11069744	0.299	0.08	0.31	0.61	intergenic	
rs12465488	0.299	0.08	0.43	0.73	intronic	*PARD3B*
rs556682	0.298	0.93	0.8	0.5	intergenic	
rs9571939	0.298	0.94	0.8	0.5	intergenic	
rs3784296	0.298	0.71	0.55	0.25	3’ downstream	*GABPB1*
rs12811599	0.297	0.17	0.28	0.58	intronic	*ANO4*
rs11184569	0.295	0.81	0.76	0.46	intergenic	
rs12034143	0.294	0.83	0.69	0.39	intergenic	
rs6958292	0.294	0.64	0.56	0.26	intergenic	
rs11688847	0.293	0.17	0.2	0.5	intergenic	
rs62256379	0.292	0.75	0.65	0.36	intronic	*SUCLG2*
rs7129877	0.292	0.42	0.47	0.77	intergenic	
rs57090061	0.291	0.25	0.61	0.9	intronic	*PLD1*
rs6806989	0.291	0.93	0.39	0.1	intronic	*PLD1*
rs7616441	0.291	0.44	0.61	0.9	intronic	*PLD1*
rs6773632	0.291	0.42	0.61	0.9	intronic	*PLD1*
rs9839305	0.29	0.79	0.51	0.21	3’ utr	*GPD1L*
rs378022	0.29	0.08	0.25	0.54	intronic	*FHIT*
rs8007792	0.29	0.14	0.19	0.48	intronic	*TTLL5*
rs6434424	0.288	0	0.15	0.44	intergenic	
rs2908871	0.287	0.29	0.41	0.7	intergenic	

rs7145573	0.287	0.75	0.53	0.24	intronic	*ACTN1*
rs6883098	0.285	0.94	0.75	0.47	non-coding intronic	*LOC102467224*
rs7987488	0.285	0	0.25	0.53	intronic	*GPC5*
rs1002420	0.285	0.14	0.25	0.54	intronic	*PCSK5*
rs6598159	0.285	0.71	0.62	0.34	intergenic	
rs2005127	0.284	0.83	0.8	0.51	intergenic	
rs4805487	0.284	0.75	0.55	0.26	intergenic	
rs28679562	0.283	0.57	0.57	0.86	intergenic	
rs79176913	0.283	0.08	0.16	0.44	intergenic	
rs7581814	0.283	0.06	0.16	0.44	intergenic	
rs12230024	0.283	0.69	0.53	0.25	intergenic	
rs2704516	0.282	0.92	0.71	0.43	intergenic	
rs41377545	0.282	1	0.96	0.68	intergenic	
rs7332756	0.282	0.83	0.77	0.49	intronic	*GPC5*
rs10804805	0.281	0.67	0.57	0.29	intergenic	
rs494428	0.281	0	0.25	0.53	intergenic	
rs6492597	0.281	0.17	0.25	0.53	intronic	*GPC5*
rs59740759	0.28	0.17	0.26	0.54	intergenic	
rs41204	0.28	0.1	0.51	0.79	intergenic	
rs62109766	0.279	0.86	0.55	0.27	5’ upstream, intronic	*ARHGEF18*
rs6859099	0.279	0.92	0.73	0.45	non-coding intronic	*LOC102467224*
rs10819439	0.279	0.86	0.79	0.51	intronic	*ZER1*
rs9541386	0.279	0.08	0.15	0.43	intergenic	
rs4869761	0.279	0	0.21	0.49	intronic	*SYNE1*
rs10203341	0.278	0.83	0.72	0.44	intronic	*THSD7B*
rs9521695	0.278	0.25	0.4	0.68	intronic	*COL4A2*
rs2819419	0.278	0.75	0.6	0.32	coding	*AHNAK2*
rs7162536	0.278	0.83	0.65	0.37	intergenic	
rs80353268	0.278	0.75	0.71	0.43	intronic, non-coding intronic	*CCNT2*
rs1319222	0.277	1	0.83	0.56	intronic, 5’ upstream	*SEMA5A, SNHG18*

rs793084	0.277	0.5	0.45	0.18	intergenic	
rs4748302	0.277	0.92	0.64	0.36	3’ downstream, 3’ utr	*PTER, C1QL3*
rs10258475	0.276	0.14	0.24	0.52	intergenic	
rs7296207	0.276	0.21	0.37	0.64	intergenic	
rs1300237	0.276	0	0.37	0.64	intronic	*SLC46A3*
rs11221793	0.275	0	0.13	0.41	intergenic	
rs9884570	0.275	0.25	0.38	0.66	intronic	*DCHS2*
rs6067275	0.275	0.17	0.38	0.66	intergenic	
rs7958156	0.275	0	0.32	0.6	intergenic	
rs7631636	0.275	0.42	0.45	0.72	intronic	*SUCLG2*
rs9419673	0.274	0.88	0.73	0.46	intergenic	
rs7213892	0.274	0.93	0.48	0.21	intronic	*ALOXE3*
rs17050803	0.274	0.33	0.33	0.61	intergenic	
rs28647713	0.274	1	0.67	0.39	intergenic	

Table S10.2 GWAS results for the genes found among the top 100 SNPs of the D_{sel} analysis.

Gene	GWAS associated phenotype
PCSK5	Dehydroepiandrosterone, Body Height
PTPRN2	C-Reactive Protein
THSD7B	Brain, Cholesterol, HDL, Cholesterol, LDL
TPO	Respiratory Function Tests
AFF3	Cholesterol, Cholesterol, HDL
TMEM131	Exercise Test, Heart Rate
MLL3	Schizophrenia
CNTNAP2	Heart Failure
IRG1	Waist Circumference
SPEN	Heart Failure
EPHB2	Insulin, Insulin Resistance
SEPT10	Blood Pressure
PDE4DIP	Respiratory Function Tests
AGBL3	Attention Deficit Disorder with Hyperactivity
MTF1	Hypothyroidism
NAV2	Arteries, Asthma, Cell Adhesion Molecules, Lipoproteins, Myocardial Infarction, Stroke, Attention Deficit Disorder with Hyperactivity, HIV-1
FAM23A	Blood Coagulation Factors, Body Weight
MRC1L1	Aspartate Aminotransferases
MRC1	Aspartate Aminotransferases
TPH2	Waist Circumference
PUS7	Erythrocyte Indices
PARD3B	Knee osteoarthritis, C-Reactive Protein, Platelet Count, Cholesterol, HDL, Body Height, Osteoarthritis, Knee, E-Selectin, Tuberculosis, Acquired Immunodeficiency Syndrome
SLC26A5	Triglycerides
SEMA5A	Autism, Parkinson's disease, Blood Pressure Determination, Breath Tests, Glucose, Myocardial Infarction, Tunica Media, Parkinson Disease, Alkaline Phosphatase, Peroxidase, Mortality, Hip, Hemoglobin A, Glycosylated, Cholesterol, Cholesterol, LDL, Body Weight, Blood Pressure, Carotid Artery Diseases
HP	Apolipoproteins B, Cholesterol, LDL

HPR	Apolipoproteins B, Cholesterol, LDL
GC	Erythrocytes, Vitamin D
ANK3	Arteries, Creatinine, Glomerular Filtration Rate, Cholesterol, LDL, Triglycerides, Bipolar Disorder, Schizophrenia
ZNF32	Body Mass Index
RET	Hirschsprung Disease
FHIT	lung cancer and preneoplastic bronchial lesions, tumour kinetics and chromosomal instability, transcriptional inactivation of the FHIT gene, smoking, cervical cancer, prostate cancer, ADHD | attention-deficit hyperactivity disorder, major depressive disorder , Albumins, Body Composition, Coronary Artery Disease, Erythrocyte Count, Lipids, Lipoproteins, Myocardial Infarction, Schizophrenia, Stroke, Waist Circumference, Creatinine, Glomerular Filtration Rate, Fibrinogen, Body Mass Index, Body Weight, Blood Pressure, Sleep, Asperger Syndrome, Aorta, Anticonvulsants, Cleft Lip
LRRN1	Blood Pressure, Menopause, Cholesterol, HDL, Triglycerides, Body Weight, Echocardiography
TKT	Waist Circumference, Heart Function Tests
ZNF717	Hippocampus
IGHG1	Sjogren's syndrome, atopy
ANG	Stroke
ABCB1	Phospholipids
TNP2	Diabetes Mellitus, Type 1
C7orf10	Precursor Cell Lymphoblastic Leukemia-Lymphoma
ZNF107	Calcium
UNC13A	Hemoglobins, Amyotrophic Lateral Sclerosis
ZNF92	Smoking
ZNF138	Smoking
BMP8A	Atrial Natriuretic Factor
BMP8B	Atrial Natriuretic Factor
IL12RB2	Liver Cirrhosis, Biliary
SYNE1	Ovarian cancer , tonometry, Body Height, Erythrocyte Count, Forced Vital Capacity, Diabetes Mellitus, Type 2, Triglycerides, Echocardiography
PIGF	Body Height
STK4	Neuroblastoma
LRRN4	Menopause
TGM6	Stroke
EMR2	Blood Pressure Determination

EMR3	Blood Pressure Determination
GPC5	Serum metabolites, multiple sclerosis, height, Coronary Artery Disease, Glucose, Monocyte Chemoattractant Protein-1, Mental Competency, Cholesterol, HDL, Echocardiography, Lung Neoplasms, Nephrotic Syndrome
DAO	Erythrocyte Count, Hemoglobins
SF1	Gout
ACTN1	Arteries
TRIM16	Hemoglobin A, Glycosylated
COX10	Echocardiography
MTTP	Plasma cholesterol levels and body mass index, ApoB-48, lipid metabolism disorders, diabetes, type 2, blood pressure, arterial, steatohepatitis, non-alcoholic, body mass; cholesterol, LDL; cholesterol, total; insulin; apoB, atherosclerosis, coronary; lipoprotein; lipids, blood pressure, arterial diabetes, type 2 glucose insulin, Fatty Liver| Hepatitis C, Chronic
TEC	Inflammatory Bowel Diseases
NRAS	Erythrocytes
MPP7	Iron, Body Mass Index, Echocardiography, Cardiovascular Diseases, Electrocardiography, Alzheimer Disease, Asthma
TTLL5	Body Height
CPN1	Iron, Alkaline Phosphatase
LIPA	Coronary Artery Disease
AGK	Dehydroepiandrosterone Sulfate
COL4A2	Coronary Artery Disease, Vascular Calcification
COLQ	Alcoholism, Body Height, Iron
DCHS2	C-Reactive Protein, Lipoproteins, Blood Coagulation Factors, Erythrocytes, Lipids, Triglycerides, Blood Pressure, Fibrinogen, Alzheimer Disease

Table S10.3 Results of the GO-term enrichment analysis.

GO-ID	Total number of genes	Expected number of genes among outliers	Observed number of genes among outliers	Nominal p-value	FDR	Description of GO term
GO:0060603	37	1.707	9	0.000018	0.029482	mammary gland duct morphogenesis
GO:0060443	53	2.48	10	0.000061	0.0506862	mammary gland morphogenesis
GO:0022612	109	4.591	14	0.000099	0.0506862	gland morphogenesis
GO:0060444	25	0.905	6	0.000134	0.0506862	branching involved in mammary gland duct morphogenesis
GO:0021536	63	2.437	9	0.000146	0.0506862	diencephalon development
GO:0061180	68	2.919	10	0.000299	0.0849803333	mammary gland epithelium development
GO:0071514	22	0.701	5	0.000466	0.1161662857	genetic imprinting
GO:0030879	127	5.523	14	0.000735	0.1490481111	mammary gland development
GO:0048732	266	10.015	21	0.00074	0.1490481111	gland development
GO:0048589	265	12.449	24	0.000866	0.1562793	developmental growth
GO:0033135	62	2.321	8	0.001343	0.2200139091	regulation of peptidyl-serine phosphorylation
GO:0050432	30	0.935	5	0.001592	0.2226172632	catecholamine secretion
GO:0072077	18	0.617	4	0.001716	0.2226172632	renal vesicle morphogenesis
GO:0035023	165	7.987	17	0.001935	0.2226172632	regulation of Rho protein signal transduction
GO:0006885	46	1.415	6	0.002005	0.2226172632	regulation of pH
GO:0045740	48	1.368	6	0.00209	0.2226172632	positive regulation of DNA replication
GO:0051926	21	0.599	4	0.002166	0.2226172632	negative regulation of calcium ion transport
GO:0006655	10	0.295	3	0.002262	0.2226172632	phosphatidylglycerol biosynthetic process
GO:0040019	8	0.431	3	0.002278	0.2226172632	positive regulation of embryonic development
GO:0048754	143	4.78	12	0.002602	0.24296675	branching morphogenesis of an epithelial tube

S10.2 Testing selection on known pigmentation SNPs

To complement the genome-wide scan above, we specifically looked into signals of selection in known pigmentation-associated SNPs as pigmentation is one of the major traits under selection pressure, especially in high latitudes [273]. Pigmentation is a trait well studied in populations of European descent (see also section S8.1). Here we focus on three major-effect SNPs in the genes *OCA2/HERC2* affecting eye pigmentation, and *SLC45A2* as well as *SLC24A5* affecting skin pigmentation. We observe (Figure 3B) that the allele frequencies of the derived allele at all three SNPs is higher in SHGs than expected based on their genome-wide admixture proportions (qpAdm estimates; section S6.4) and the allele frequencies in EHGs and WHGs. To test whether these allele frequency changes are significant, we performed simulations. For each SNP and each SHG individual, we randomly sampled the alleles from the two source populations based on the individual's genome-wide qpAdm admixture proportions and the allele frequencies in the source populations. Before calculating allele frequencies in the admixed SHGs, we randomly sample the same number of SHGs for which data was available in the empirical study to account for noise due to missing data. This simulation is assumed to provide a null distribution of SHG allele frequencies without selection. After 1,000,000 simulations, we find that the allele frequencies in *OCA2/HERC2* (p=0.021573) and *SLC45A2* (p=0.015819) are significantly elevated while the change seen in *SLC24A5* is not significant (p=0.1690196). The results of this simulation are shown in Figure 3B.

These results suggest that high latitude conditions exhibited a selection pressure on pigmentation phenotypes in SHGs. The polygenic architecture of skin pigmentation as well as the occurrence of different combinations of depigmentation mutations in different parts of the world suggests that selection on skin pigmentation is mainly due to physiological advantages of light pigmentation in high latitudes [274]. Hair and eye-color pigmentation on the other hand could have been affected by drift and sexual selection as less mutations need to be involved [274].

S11 Genetic testing of the post-glacial migration routes into Scandinavia

The genomic affinities of the Mesolithic Scandinavians analyzed in this study suggest a complex migration pattern. Based on geography, genetics, the position of the ice-sheet on the Scandinavian peninsula and archaeology (see section S1, Figure S1.2), we hypothesize different migration routes into Scandinavia around 10,000 BP:

a) a migration of WHGs from the south,
b) a migration of EHGs from the east across the Baltic Sea,
c) a migration of EHGs from the east and along the north-Atlantic coast,
d) a migration of EHGs from the east and south of the Baltic Sea,

as well as combinations of these four migration routes (Figure S11.1). These hypothetical scenarios allow us to formulate expectations of the genetic pattern seen in the different individuals. A migration of only WHGs would cause strong affinities of all SHGs to that group, whereas an exclusively EHG-like migration would make all SHGs very similar to EHGs. If there were combinations of migrations from both sources, a pattern of geographic structure would emerge: different SHGs would be characterized by differential affinities to WHGs and EHGs. We formulate these affinities in terms of expected tendencies for f_4(Chimp, X; EHG, WHG) which are shown in Figure S11.1. We only give qualitative expectations as a lot of additional information regarding population splits, drift parameters and admixture proportions would be required to formulate quantitative expectations for f_4. The qualitative expectations still allow us to test rank correlations between expected values and observed values of f_4(Chimp, X; EHG, WHG) to evaluate how the different hypotheses fit the data. These rank correlations are not informative for single migrations but the data suggests that SHGs are in fact a mixture of EHGs and WHGs. The only scenario that produces a positive rank correlation (rho=0.677, p=0.004) between observed and expected values is a combination of (a) and (c) – one migration of WHGs from the south and another of EHGs from the east following the Norwegian coast. Such a scenario explains the higher affinities of Norwegian SHGs with EHGs as well as the higher affinities of Swedish SHGs with WHGs. The qpAdm admixture proportions (Figure 1A) are also significantly different between Norwegian and Swedish Mesolithic individuals (Wilcoxon test; p=0.01399). These affinities are consistently observed across all population genomic analyses in this study (PCA, ADMIXTURE, f_4, qpAdm, TreeMix) and they are not correlated with the age of the samples (Figure S11.2). Therefore, we conclude that two migrations into Scandinavia fit the genetic data: a migration from central Europe into southern Sweden and a second migration along the Norwegian coast and around the ice sheet covering the center of the Scandinavian

peninsula. The populations spread across Scandinavia and mixed creating the geographic population structure seen in the genomic data.

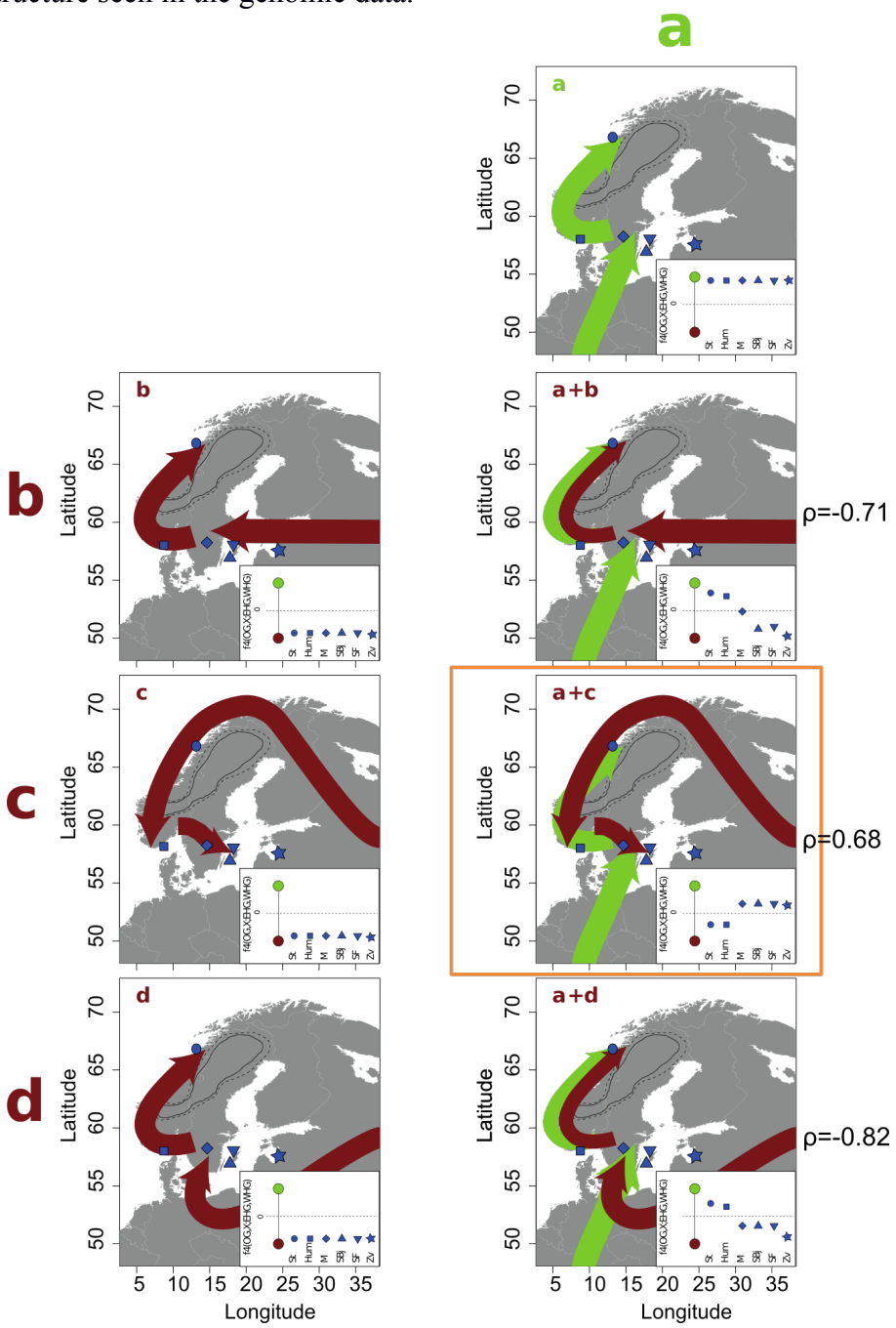


Figure S11.1 Hypotheses for the colonization of Scandinavia. Migration from the south is shown in green as (a) and different routes from the east are shown in red-brown as (b-d). Different combinations of the migration routes are shown on the right. Insets show the expected genetic affinities of Mesolithic individuals from the different sites shown on the maps assuming the respective scenario (see Fig. 1C for the pattern observed with empirical data). Spearman's rank correlation coefficient between observed and expected f4 is shown next to the maps. The

scenario compatible with the observations (a combined with c) is encircled with an orange square.

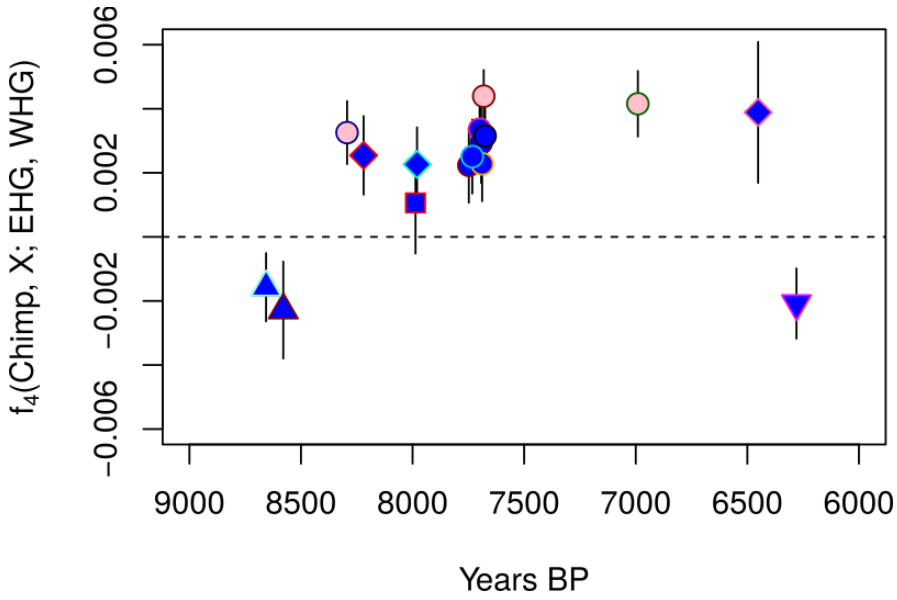


Figure S11.2 Affinities of SHGs to the two source populations as a function of time (compare to Fig. 1C).

References

1. Björck S. A review of the history of the Baltic Sea, 13.0-8.0 ka BP. Quaternary international. 1995;27: 19–40.

2. Andrén T, Björck J, Johnsen S. Correlation of Swedish glacial varves with the Greenland (GRIP) oxygen isotope record. Journal of Quaternary Science. 1999;14: 361–371.

3. Mayewski PA, Rohling EE, Curt Stager J, Karlén W, Maasch KA, David Meeker L, et al. Holocene climate variability in Northernmost Europe. Quaternary Research. 2004;62: 243–255. doi:10.1016/j.yqres.2004.07.001

4. Allen JRM, Long AJ, Ottley CJ, Graham Pearson D, Huntley B. Holocene climate variability in northernmost Europe. Quaternary Science Reviews. 2007;26: 1432–1453. doi:10.1016/j.quascirev.2007.02.009

5. Manninen MA. Culture, Behaviour, and the 8200 cal BP Cold Event: Organisational Change and Culture Environment Dynamics in Late Mesolithic Northern Fennoscandia [Internet]. 2014. Available: https://helda.helsinki.fi/handle/10138/42470

6. Skar B, Breivik HM. Introduction: Environment and adaptation of forager pioneers in the North-western regions of Europe - The Ecology of Early Settlement in Northern Europe. In: Persson P, Riede F, Skar B, editors. Ecology of early settlement in Northern Europe: conditions for subsistence and survival. Sheffield, UK ; Bristol, CT: Equinox Publishing Ltd; 2017.

7. Larsson L. The earliest settlement of Scandinavia and its relationship with neighbouring areas. Almquist & Wiksell International; 1996.

8. Mörner N-A. The Baltic Ice Lake-Yoldia Sea transition. Quaternary International. 1995;27: 95–98. doi:10.1016/1040-6182(94)00065-D

9. Jensen J., Bennike O, Witkowski A, Lemke W, Kuijpers A. Early Holocene history of the southwestern Baltic Sea: the Ancylus Lake stage. Boreas. 1999;28: 437–453.

10. Riede F. The Laacher See-eruption (12,920 BP) and material culture change at the end of the Aller{ø}d in Northern Europe. Journal of Archaeological Science. 2008;35: 591–599.

11. Riede F. Climate and demography in early prehistory: using calibrated 14C dates as population proxies. Human Biology. 2009;81: 309–337.

12. 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. Available: https://books.google.se/books?id=M4lSAwAAQBAJ

13. Gyllencreutz R. Late Glacial and Holocene paleoceanography in the Skagerrak from high-resolution grain size records. Palaeogeography, Palaeoclimatology, Palaeoecology. 2005;222: 344–369.

14. Knutsson K, Knutsson H. The postglacial colonization of humans, fauna and plants in northern Sweden. Arkeologi i norr. 2012;13: 004618–461003.

15. Sørensen R. Late Weichselian deglaciation in the Oslofjord area, south Norway. Boreas. 1979;8: 241–246. doi:10.1111/j.1502-3885.1979.tb00806.x

16. Svendsen JI, Mangerud J. Late Weichselian and holocene sea-level history for a cross-section of western Norway. Journal of Quaternary Science. 2010;2: 113–132. doi:10.1002/jqs.3390020205

17. Prøsch-Danielsen L, Høgestøl M. A Late Weichselian site found at Galta, Rennesøy, southwest Norway. In: Fischer A, editor. Man and Sea in the Mesolithic Coastal Settlement Above and Below Present Sea Level. Oxford: Oxbow Books Limited; 1995. pp. 123–130.

18. Prøsch-Danielsen L. New light on the Holocene shore displacement curve on Lista, the southernmost part of Norway. Norsk Geografi sk Tidsskrift. 1997;51: 83–101.

19. 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.

20. 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.

21. Bang-Andersen S. The Myrvatn group, a Preboreal find-complex in southwest Norway. Contributions to the Mesolithic in Europe Papers Presented at the Fourth International Symposium‘ The Mesolithic in Europe. Leuven University Press; 1990. pp. 215–226. Available: https://www.google.com/books?hl=en&lr=&id=SpMTT_E9bzIC&oi=fnd&pg=PA215&dq=The+Myrvatn+Group,+a+Preboreal+find-complex+in+southwest+Norway&ots=6WSsJHOB14&sig=X0pv30veyYMriF_h8fbPfvyluyc https://books.google.se/books?hl=en&lr=&id=SpMTT_E9bzIC&oi=fnd&pg=PA21

22. Thommessen T. The Early Settlement of Northern Norway. In: Larsson L, editor. The Earliest Settlement of Scandinavia. Stockholm: Almquist & Wiksell International; 1996. pp. 233–240. Available: https://www.researchgate.net/profile/Anders_Fischer2/publication/273461348_At_the_Border_of_Human_Habitat._The_Late_Palaeolithic_and_Early_Mesolithic_in_Scandinavia/links/5503f9760cf2d60c0e650f46.pdf

23. Åhrberg ES, Pauler I. En Tidligmesolitisk boplats. In: Jaksland L, editor. E18 Brunlanesprosjektet, Bind II: Undersøkte lokaliteter fra tidligmesolitikum og senere. Kulturhistorisk museum, Fornminneseksjonen, Universitet i Oslo; 2012.

24. 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.

25. Kleppe J. Desolate landscapes or shifting landscapes? Late glacial/ early postglacial settlement of northernmost Norway in the light of new data from Eastern Finnmark. In: Riede F, Tallavaara M, editors. Lateglacial and Postglacial Pioneers in Northern Europe BAR International Series 2599. Oxford: Oxford Archaeopress; 2014. pp. 121–145.

26. Fuglestvedt I. The Ahrensburgian Galta 3 site in SW Norway. Dating, technology and cultural affinity. Acta Archaeologica. 2007;78: 87–110.

27. 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.

28. 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.

29. Schmitt L, Svedhage K. Chronological aspects of the Hensbacka–a group of hunter-gatherers/fishers on the west coast of Sweden during the Pleistocene/Holocene transition: an example of early coastal colonization. Danish Journal of Archaeology. 2015;4: 75–81.

30. Bjerck HB. Norwegian Mesolithic trends: a review. In: Bailey GN, Spikins, editors. Mesolithic Europe. Cambridge University Press; 2008. pp. 60–106. Available: https://www.google.com/books?hl=en&lr=&id=dsDwpU0L0BwC&oi=fnd&pg=PA60&dq=Norwegian+Mesolithic+Trends:+A+review&ots=_eGpYizaLV&sig=XerUxgMKBwSBUwQLH9nmFupx8RQ https://books.google.se/books?hl=en&lr=&id=dsDwpU0L0BwC&oi=fnd&pg=PA60&dq=Norwegian+Mesolithic+Tr

31. Glørstad H akon. Where are the missing boats? The pioneer settlement of Norway as long-term history. Norwegian Archaeological Review. 2013;46: 57–80.

32. Damlien H. Eastern pioneers in westernmost territories? Current perspectives on Mesolithic hunter–gatherer large-scale interaction and migration within Northern Eurasia. Quaternary International. 2014;X: 1–12.

33. Damlien H. Between Tradition and adaptation. Long-term trajectories of lithic tool-making in South Norway during the postglacial colonization and its aftermath (c. 9500-7500 cal BC). 2016.

34. Skar B, Lidén K, Eriksson G, Sellevold B. A submerged Mesolithic grave site reveals remains of the first Norwegian seal hunters. In: Bjerck HB, al E, editors. Marine Ventures Archaeological perspectives on Human–Sea Relations. Sheffield: Equinox Publishing; 2016. Available: http://www.diva-portal.org/smash/record.jsf?pid=diva2:1058702

35. Möller P, Östlund O, Barnekow L, Sandgren P, Palmbo F, Willerslev E. Living at the margin of the retreating Fennoscandian Ice Sheet: The early Mesolithic sites at Aareavaara, northernmost Sweden. The Holocene. 2013;23: 104–116. doi:10.1177/0959683612455546

36. Tallavaara M, Pesonen P, Oinonen M. Prehistoric population history in eastern Fennoscandia. Journal of Archaeological Science. 2010;37: 251–260.

37. Tallavaara M, Seppä H. Did the mid-Holocene environmental changes cause the boom and bust of hunter-gatherer population size in eastern Fennoscandia? The Holocene. 2011;22: 215–225.

38. Rankama T, Kankaanpää J. First evidence of eastern Preboreal pioneers in arctic Finland and Norway. Quartär. 2011;58: 183–206.

39. Tallavaara M, Manninen M, Pesonen P, Hertell E. Radiocarbon dates and postglacial dynamics in eastern Fenneoscandia. In: Riede F, Tallavaara M, editors. Lateglacial and Postglacial Pioneers in Northern Europe. Oxford: Archaeopress; 2014. pp. 161–175.

40. Veski S, Heinsalu A, Klassen V, Kriiska A, Lõugas L, Poska A, et al. Early Holocene coastal settlements and palaeoenvironment on the shore of the Baltic Sea at Pärnu, southwestern Estonia. Quaternary International. 2005;130: 75–85.

41. 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.

42. Zvelebil M. Innovating hunter-gatherers: the Mesolithic in the Baltic. In: Bailey G, Spikins P, editors. Mesolithic Europe. Cambridge: Cambridge University Press; 2008. pp. 18–59. Available: https://www.google.com/books?

hl=en&lr=&id=dsDwpU0L0BwC&oi=fnd&pg=PA18&dq=Innovating+Hunter-Gatherers:+The+Mesolithic+in+the+Baltic&ots=_eGpYjz4FR&sig=eeETLKwCJiKlnYrc6dn_fOIMgXI

43. Pitulko VV, Nikolsky PA, Girya EY, Basilyan AE, Tumskoy VE, Koulakov SA, et al. The Yana RHS site: humans in the Arctic before the last glacial maximum. Science. 2004;303: 52–56.

44. Stroeven AP, Hättestrand C, Kleman J, Heyman J, Fabel D, Fredin O, et al. Deglaciation of Fennoscandia. Quaternary Science Reviews. 2016;147: 91–121.

45. Östlund O. Aareavaara and the Pioneer Period in Northern Sweden. In: Blankholm HP, editor. The early economy and settlement in Northern Europe: pioneering, resource use, coping with change. 2017.

46. Gurina NN. The main stages in the cultural development of the ancient population of the Kola peninsula. Fennoscandia Archaeologica. 1987;IV: 35–49.

47. 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. doi:10.1080/00293652.2013.770416

48. Sørensen M. The arrival and development of pressure blade technology in southern Scandinavia. In: Desrosier P, editor. The emergence of pressure blade making: from origin to modern experimentation. Montreal: Springer; 2012. pp. 237–259. Available: http://link.springer.com/chapter/10.1007/978-1-4614-2003-3_9

49. Bjerck HB, Fisher A. The North Sea Continent and the pioneer settlement of Norway. Man and Sea in the Mesolithic. 1995. p. 1.

50. 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. doi:10.1016/j.quaint.2015.10.062

51. Goebel T. The “microblade adaptation” and recolonization of Siberia during the late Upper Pleistocene. In: Elston RG, Kuhn S.-, editors. Archeological Papers of the American Anthropological Association. Virginia: Archaeological Papers of the American Anthropological Association; 2002. pp. 117–131. Available: http://onlinelibrary.wiley.com/doi/10.1525/ap3a.2002.12.1.117/full

52. Inizan M-L. Pressure Débitage in the Old World: Forerunners, Researchers, Geopolitics–Handing on the Baton. In: Desrosier PM, editor. The emergence of pressure blade making: from origin to modern experimentation. New York: Springer; 2012. pp. 11–42. Available: http://link.springer.com/chapter/10.1007/978-1-4614-2003-3_2

53. Yi M, Barton L, Morgan C, Liu D, Chen F, Zhang Y, et al. Microblade technology and the rise of serial specialists in north-central China. Journal of Anthropological Archaeology. 2013;32: 212–223.

54. Hartz S, Terberger T, Zhilin M. New AMS-dates for the Upper Volga Mesolithic and the origin of microblade technology in Europe. Quartär. 2010;57: 155–169.

55. Manninen MA, Knutsson K. Lithic raw material diversification as an adaptive strategy—Technology, mobility, and site structure in Late Mesolithic northernmost Europe. Journal of Anthropological Archaeology. 2014;33: 84–98.

56. Manninen MA, Hertell E, Rankama T. Few and far between–an archive survey of Finnish blade finds. Mesolithic Interfaces Variability in Lithic Technologies in Eastern Fennoscandia. Finland: The

Archaeological Society of Finland; 2011. pp. 112–141. Available: http://www.academia.edu/download/12705287/Manninen_Hertell_2011_MI_05_Archive_Survey_of_Finnish_Blade_Finds.pdf

57. Pelegrin J. Long blade technology in the Old World: an experimental approach and some archaeological results. In: Apel J, Knutsson K, editors. Skilled production and Social Reproduction. Uppsala: Societas Archaeologica Upsaliensis Stone Studies Uppsala; 2006. pp. 37–68. Available: https://uu.diva-portal.org/smash/get/diva2:556648/FULLTEXT01.pdf#page=38

58. Sellevold BJ, Skar B. The first lady of Norway. NIKU 1994-1999 Kulturminneforskningens mangfold NIKU Temahefte 31. Oslo: Norsk institutt for kulturminneforskning; 1999. pp. 6–11.

59. Nymoen P, Skar B. The Unappreciated Cultural Landscape: indications of submerged Mesolithic settlement along the Norwegian southern coast. In: Benjamin J, Bonsall C, Pickard C, Fischer A, editors. Submerged Prehistory. Oxford: Oxbow Books; 2011. pp. 38–54.

60. Bennike P, Alexandersen V. Danmarks urbefolkning. Nationalmuseets arbejdsmark. 1997; 143–156.

61. Stutz LN. The way we bury our dead. Reflections on mortuary ritual, community and identity at the time of the Mesolithic-Neolithic transition. Documenta praehistorica. 2010;37: 33–42.

62. Petersen EB, Jønsson JH, Juel C, Kj\a er A. Diversity of Mesolithic Vedb\a ek. Acta Archaeologica. 2015;86: 7–13.

63. Larsson L. Ethnicity and traditions in Mesolithic mortuary practices of southern Scandinavia. In: Shennan S, editor. Archaeological Approaches to Cultural Identity, Unwin Hyman, London. London: Routledge; 1989. pp. 210–218.

64. Jørgensen R. Måløyhatten. Undersøkelse av funnsted for menneske-kjeve I grotte på Måløyhatten, Måløya, gnr. 30. Steigen commune, Nordland fylke; 2015.

65. Schnittger B, Rydh H, Kolmodin L, Stolpe H. Grottan Stora förvar på Stora karlsö. Wahlström & Widstrand; 1940.

66. Apel J, Storå J. The Pioneer Settlements of Gotland–Early Holocene Maritime Relations in the Baltic Sea Area. In: Persson, editor. The Early Settlement of Northern Europe: An Ecological Approach. Sheffield: Equinox Publishing; p. (in press). Available: http://www.archaeology.su.se/polopoly_fs/1.172481.1395910712!/menu/standard/file/PM_140402_Jan_Apel-Jan_Stora.pdf

67. Lindqvist C, Possnert G, Burenhult G. The subsistence economy and diet at Jakobs/Ajvide, Eksta parish and other prehistoric dwelling and burial sites on Gotland in long-term perspective. Remote sensing. 1997. pp. 29–90.

68. Lindqvist C, Possnert G. The first seal hunter families on Gotland : on the Mesolithic occupation in the Stora Förvar cave. Current Swedish archaeology. Current Swedish archaeology Stockholm : Swedish Archaeological Society [Svenska arkeologiska samfundet], [1993]-; 1999.

69. Sørensen M. Teknologiske traditioner i Maglemosekulturen. En diakron analyse af Maglemosekulturens fl\a ekkeindustri. In: Eriksen V, editor. \AArhus: Jysk Ark\a eologisk Selskab; 2006. pp. 19–75. Available: http://www.forskningsdatabasen.dk/en/catalog/2289806542

70. Arwidsson G. Stenåldersmannen från Stora Bjärs i Stenkyrka. Arkeologi på Gotland, Gotlandica 14: 17. 1979;25.

71. Eriksson G, Frei KM, Howcroft R, Gummesson S, Molin F, Lidén K, et al. Diet and mobility among Mesolithic hunter-gatherers in Motala (Sweden)-The isotope perspective. Journal of Archaeological Science: Reports. 2016; Available: http://www.sciencedirect.com/science/article/pii/S2352409X16302097

72. Nordqvist B. Coastal adaptations in the Mesolithic: a study of coastal sites with organic remains from the Boreal and Atlantic periods in Western Sweden [Internet]. Gothenburg University. 2000. Available: http://swepub.kb.se/bib/swepub:oai:services.scigloo.org:51008?tab2=abs&language=en

73. Andersson S, Wigforss J. Senmesolitikum i Göteborgs-och Alingså asområdena (GOTARC. Serie C. Arkeologska skrifter, 58) [Internet]. Göteborg University,; 2004. Available: http://swepub.kb.se/bib/swepub:oai:services.scigloo.org:50017?vw=short&tab2=abs

74. Glørstad H akon. The structure and history of the Late Mesolithic societies in the Oslo fjord area 6300-3800 BC. Lindome: Bricoleur Press; 2010.

75. Sørensen SA. Kongemosekulturen i Sydskandinavien. Frederikssund: Egnsmuseet F\a ergeg\a arden; 1996.

76. Johansson AD. Ældre stenalder i Norden. Farum: SDA; 2000.

77. Åkerlund A. Human responses to shore displacement: living by the sea in Eastern Middle Sweden during the Stone Age. UV Stockholm, Avd. för arkeologiska undersökningar, Riksantikvarieämbetet; 1996.

78. Lindgren C. Människor och kvarts: sociala och teknologiska strategier under mesolitikum i östra Mellansverige [Internet]. 2004. Available: http://www.diva-portal.org/smash/record.jsf?pid=diva2:200748

79. Pettersson M, Wikell R. Tidigmesolitiska säljägare i Tyresta för 10 000 år sedan. Späckbetong, gråsäl och tomtning på en kobbe i Ancylussjön 120 km från fastlandet. Fornvännen. 2013;108: 73–92.

80. Carlsson T, Gruber G, Molin F. The Mesolithic in Östergötland. An Introduction. In: Gruber G, editor. Identities in Transition Mesolithic Strategies in the Swedish Province of Östergötland. Linköping: Riksantikvarieämbetet; 2005. pp. 8–23.

81. 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. doi:10.1038/nature13673

82. 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; doi:10.1038/nature14317

83. Brown TA, Nelson DE, Vogel JS, Southon JR. Improved collagen extraction by modified Longin method. Radiocarbon. 2006;30: 171–177.

84. Chu PP. Dietary variation among the prehistoric Asiatic Eskimo [Internet]. Simon Fraser University. 1998. Available: http://www.nlc-bnc.ca/obj/s4/f2/dsk3/ftp04/mq37502.pdf

85. Honch NV, McCullagh JS, Hedges RE. Variation of bone collagen amino acid δ13C values in archaeological humans and fauna with different dietary regimes: developing frameworks of dietary discrimination. American journal of physical anthropology. 2012;148: 495–511.

86. Eriksson G, Lidén K. Dietary life histories in Stone Age northern Europe. Journal of Anthropological Archaeology. 2013;32: 288–302.

87. Mangerud J, Bondevik S, Gulliksen S, Karin Hufthammer A, Høisæter T. Marine 14C reservoir ages for 19th century whales and molluscs from the North Atlantic. Quaternary Science Reviews. 2006;25: 3228–3245. doi:10.1016/j.quascirev.2006.03.010

88. Eriksson G. Part-time farmers or hard-core sealers? Västerbjers studied by means of stable isotope analysis. Journal of Anthropological Archaeology. 2004;23: 135–162.

89. Ramsey CB, Others. Bayesian analysis of radiocarbon dates. Radiocarbon. 2009;51: 337–360.

90. Ramsey CB. OxCal Program v3. 10. Online at: http://www rlaha ox ac uk/orau. 2005;

91. Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Bronk Ramsey C, et al. IntCal13 and Marine13 radiocarbon age calibration curves 0-50,000 years cal BP. 2013;

92. Skoglund P, Malmstrom H, Omrak A, Raghavan M, Valdiosera C, Gunther T, et al. Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. Science. 2014;344: 747–750. doi:10.1126/science.1253448

93. 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. doi:10.1002/(SICI)1096-8644(199804)105:4<539::AID-AJPA10>3.0.CO;2-1

94. 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. doi:10.1093/molbev/msm015

95. 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. Proc Natl Acad Sci USA. 2015;112: 11917–11922. doi:10.1073/pnas.1509851112

96. 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. doi:10.1073/pnas.1314445110

97. Meyer M, Kircher M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harbor Protocols. 2010;2010: pdb.prot5448-pdb.prot5448. doi:10.1101/pdb.prot5448

98. 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. doi:10.1038/ncomms6257

99. Briggs AW, Heyn P. Preparation of next-generation sequencing libraries from damaged DNA. Methods in Molecular Biology. 2012;840: 143–154. doi:10.1007/978-1-61779-516-9_18

100. Kircher M. Analysis of high-throughput ancient DNA sequencing data. Methods Mol Biol. 2012;840: 197–228. doi:10.1007/978-1-61779-516-9_23

101. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324

102. Kuhn JMM, Jakobsson M, Günther T. Estimating genetic kin relationships in prehistoric populations. bioRxiv. 2017; 100297.

103. 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. doi:10.1016/j.cell.2008.06.021

104. 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. doi:10.1093/bioinformatics/btp352

105. Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. Lalueza-Fox C, editor. PLoS ONE. 2012;7: e34131. doi:10.1371/journal.pone.0034131

106. 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. doi:10.1038/nature15393

107. Skoglund P, Malmstrom H, Raghavan M, Stora J, Hall P, Willerslev E, et al. Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. Science. 2012;336: 466–469. doi:10.1126/science.1216304

108. Broad Institute. Picard tools. https://broadinstitute.github.io/picard/. 2016; Available: http://broadinstitute.github.io/picard/

109. Schmidt S. (GATK) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Proceedings of the International Conference on Intellectual Capital, Knowledge Management & Organizational Learning. 2009;20: 254–260. doi:10.1101/gr.107524.110.20

110. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

111. 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: 43–9. doi:10.1038/nature12886

112. 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. doi:10.4161/fly.19695

113. 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. doi:10.1126/science.1211177

114. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 2014;15: 356. doi:10.1186/s12859-014-0356-4

115. 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. doi:10.1016/j.ajhg.2012.09.004

116. Schiffels S, Haak W, Paajanen P, Llamas B, Popescu E, Loe L, et al. Iron Age and Anglo-Saxon genomes from East England reveal British migration history. Nature communications. 2016;7: 10408. doi:10.1038/ncomms10408

117. van Oven M. PhyloTree Build 17: Growing the human mitochondrial DNA tree. Forensic Science International: Genetics Supplement Series. 2015;5: 9–11. doi:10.1016/j.fsigss.2015.09.155

118. Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, et al. The genetic history of Ice Age Europe. Nature. 2016;534: 200–5. doi:10.1038/nature17993

119. Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, et al. The genetic legacy of paleolithic Homo sapiens sapiens in extant europeans: A Y chromosome perspective. Science. 2000;290: 1155–1159. doi:10.1126/science.290.5494.1155

120. Karmin M, Saag L, Vicente M, Wilson Sayres MA, Järve M, Talas UG, et al. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. Genome Research. 2015;25: 459–466. doi:10.1101/gr.186684.114

121. 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. doi:10.1038/ncomms9912

122. 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. doi:10.1038/nature16152

123. Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, et al. An early modern human from Romania with a recent Neanderthal ancestor. Nature. 2015;524: 216–9. doi:10.1038/nature14558

124. 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. doi:10.1126/science.aaa0114

125. 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. doi:10.1038/nature12960

126. 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. doi:10.1016/j.cub.2016.12.060

127. 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. doi:10.1038/nature14507

128. Rootsi S, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, et al. Phylogeography of Y-Chromosome Haplogroup I Reveals Distinct Domains of Prehistoric Gene Flow in Europe. The American Journal of Human Genetics. 2004;75: 128–137. doi:10.1086/422196

129. Karlsson AO, Wallerström T, Götherström A, Holmlund G. Y-chromosome diversity in Sweden - a long-time perspective. European journal of human genetics : EJHG. 2006;14: 963–70. doi:10.1038/sj.ejhg.5201651

130. Passarino G, Cavalleri GL, Lin A a, Cavalli-Sforza LL, Børresen-Dale A-L, Underhill P a. Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. European journal of human genetics : EJHG. 2002;10: 521–529. doi:10.1038/sj.ejhg.5200834

131. Lappalainen T, Laitinen V, Salmela E, Andersen P, Huoponen K, Savontaus ML, et al. Migration waves to the baltic sea region. Annals of Human Genetics. 2008;72: 337–348. doi:10.1111/j.1469-1809.2007.00429.x

132. 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. doi:10.1093/bioinformatics/btr330

133. Vianello D, Sevini F, Castellani G, Lomartire L, Capri M, Franceschi C. HAPLOFIND: A New Method for High-Throughput mtDNA Haplogroup Assignment. Human Mutation. 2013;34: 1189–1194. doi:10.1002/humu.22356

134. Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli E-L, Silva NM, et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet. 2012;90: 675–684. doi:10.1016/j.ajhg.2012.03.002

135. Marinov GK, Wang YE, Chan D, Wold BJ. Evidence for site-specific occupancy of the mitochondrial genome by nuclear transcription factors. PLoS ONE. 2014;9. doi:10.1371/journal.pone.0084713

136. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, et al. Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet. 2000;67: 1251–1276.

137. Malyarchuk B, Grzybowski T, Derenko M, Perkova M, Vanecek T, Lazur J, et al. Mitochondrial DNA phylogeny in eastern and western Slavs. Molecular Biology and Evolution. 2008;25: 1651–1658. doi:10.1093/molbev/msn114

138. Posth C, Renaud G, Mittnik A, Drucker DG, Rougier H, Cupillard C, et al. Pleistocene mitochondrial genomes suggest a single major dispersal of non-africans and a late glacial population turnover in Europe. Current Biology. 2016;26: 827–833. doi:10.1016/j.cub.2016.01.037

139. 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. doi:10.1038/nature12736

140. Bollongino R, Nehlich O, Richards MP, Orschiedt J, Thomas MG, Sell C, et al. 2000 Years of Parallel Societies in Stone Age Central Europe. Science. 2013;342: 479–481. doi:10.1126/science.1245049

141. Sánchez-Quinto F, Schroeder H, Ramirez O, Ávila-Arcos MC, Pybus M, Olalde I, et al. Genomic Affinities of Two 7,000-Year-Old Iberian Hunter-Gatherers. Current Biology. 2012;22: 1494–1499. doi:10.1016/j.cub.2012.06.005

142. Richards MB, Macaulay VA, Bandelt HJ, Sykes BC. Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet. 1998;62: 241–260. doi:10.1046/j.1469-1809.1998.6230241.x

143. Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, et al. Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. American Journal of Human Genetics. 2009;84: 740–759. doi:10.1016/j.ajhg.2009.05.001

144. Malyarchuk B, Derenko M, Grzybowski T, Perkova M, Rogalla U, Vanecek T, et al. The Peopling of Europe from the Mitochondrial Haplogroup U5 Perspective. Gilbert MTP, editor. PLoS ONE. 2010;5: e10285. doi:10.1371/journal.pone.0010285

145. Fu Q, Mittnik A, Johnson PLF, Bos K, Lari M, Bollongino R, et al. A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. Current Biology. 2013;23: 553–559. doi:10.1016/j.cub.2013.02.044

146. Der Sarkissian C, Brotherton P, Balanovsky O, Templeton JEL, Llamas B, Soubrier J, et al. Mitochondrial genome sequencing in mesolithic North East Europe unearths a new sub-clade within the broadly distributed human haplogroup C1. PLoS ONE. 2014;9. doi:10.1371/journal.pone.0087612

147. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient Admixture in Human History. Genetics. 2012;192: 1065–1093. doi:10.1534/genetics.112.145037

148. 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. doi:10.1038/nature13810

149. 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. doi:10.1086/519795

150. 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.

151. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. PLoS Genetics. 2006;2: e190. doi:10.1371/journal.pgen.0020190

152. Skoglund P, Northoff BH, Shunkov MV, Derevianko AP, Pääbo S, Krause J, et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proceedings of the National Academy of Sciences. 2014;111: 2229–2234.

153. 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: 820–2. doi:10.1126/science.aad2879

154. 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.

155. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19: 1655–1664. doi:10.1101/gr.094052.109

156. 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. doi:10.1093/bioinformatics/btm233

157. Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, et al. Reconstructing Native American population history. Nature. 2012;488: 370–374. doi:10.1038/nature11258

158. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 2012;8: e1002967. doi:10.1371/journal.pgen.1002967

159. 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. 2015; 1–6. doi:10.1073/pnas.1518445113

160. 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 Genetics. 2011;7. doi:10.1371/journal.pgen.1001373

161. 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. doi:10.1534/genetics.112.147330

162. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. Nature genetics. 2014;46: 919–25. doi:10.1038/ng.3015

163. Ralph P, Coop G. The Geography of Recent Genetic Ancestry across Europe. PLoS Biology. 2013;11. doi:10.1371/journal.pbio.1001555

164. Browning BL, Browning SR. Improving the Accuracy and Efficiency of Identity by Descent Detection in Population Data. Genetics. 2013; genetics.113.150029. doi:10.1534/genetics.113.150029

165. Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. mapDamage: Testing for damage patterns in ancient DNA sequences. Bioinformatics. 2011;27: 2153–2155. doi:10.1093/bioinformatics/btr347

166. Jónsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L. mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 2013;29: 1682–1684.

167. Walsh S, Lindenbergh A, Zuniga SB, Sijen T, De Knijff P, Kayser M, et al. Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. Forensic Science International: Genetics. 2011;5: 464–471. doi:10.1016/j.fsigen.2010.09.008

168. Walsh S, Wollstein A, Liu F, Chakravarthy U, Rahu M, Seland JH, et al. DNA-based eye colour prediction across Europe with the IrisPlex system. Forensic Science International: Genetics. 2012;6: 330–340. doi:10.1016/j.fsigen.2011.07.009

169. Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A, et al. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Science International: Genetics. 2013;7: 98–115. doi:10.1016/j.fsigen.2012.07.005

170. Spichenok O, Budimlija ZM, Mitchell AA, Jenny A, Kovacevic L, Marjanovic D, et al. Prediction of eye and skin color in diverse populations using seven SNPs. Forensic Science International: Genetics. 2011;5: 472–478. doi:10.1016/j.fsigen.2010.10.005

171. Branicki W, Liu F, Van Duijn K, Draus-Barini J, Pośpiech E, Walsh S, et al. Model-based prediction of human hair color using DNA variants. Human Genetics. 2011;129: 443–454. doi:10.1007/s00439-010-0939-8

172. Hart KL, Kimura SL, Mushailov V, Budimlija ZM, Prinz M, Wurmbach E. Improved eye- and skin-color prediction based on 8 SNPs. Croatian medical journal. 2013;54: 248–56. doi:10.3325/cmj.2013.54.248

173. Pneuman A, Budimlija ZM, Caragine T, Prinz M, Wurmbach E. Verification of eye and skin color predictors in various populations. Legal Medicine. 2012;14: 78–83. doi:10.1016/j.legalmed.2011.12.005

174. Walsh S, Chaitanya L, Clarisse L, Wirken L, Draus-Barini J, Kovatsi L, et al. Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Science International: Genetics. 2014;9: 150–161. doi:10.1016/j.fsigen.2013.12.006

175. Hofmanová. Early farmers from across Europe directly descended from Neolithic Aegeans. PNAS. 2016; 201523951. doi:10.1073/pnas.1523951113

176. Lamason RL, Mohideen M-APK, Mest JR, Wong AC, Norton HL, Aros MC, et al. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science. 2005;310: 1782–1786. doi:10.1126/science.1116238

177. Stokowski RP, Pant PVK, Dadd T, Fereday A, Hinds D a, Jarman C, et al. A genomewide association study of skin pigmentation in a South Asian population. American journal of human genetics. 2007;81: 1119–1132. doi:10.1086/522235

178. Norton HL, Kittles RA, Parra E, McKeigue P, Mao X, Cheng K, et al. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. Molecular Biology and Evolution. 2007;24: 710–722. doi:10.1093/molbev/msl203

179. Beleza S, Santos AM, McEvoy B, Alves I, Martinho C, Cameron E, et al. The timing of pigmentation lightening in Europeans. Molecular Biology and Evolution. 2013;30: 24–35. doi:10.1093/molbev/mss207

180. Soejima M, Tachida H, Ishida T, Sano A, Koda Y. Evidence for recent positive selection at the human AIM1 locus in a European population. Molecular Biology and Evolution. 2006;23: 179–188. doi:10.1093/molbev/msj018

181. Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, Yen A, et al. Identifying recent adaptations in large-scale genomic data. Cell. 2013;152: 703–713. doi:10.1016/j.cell.2013.01.035

182. 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. doi:10.1007/s00439-011-1110-x

183. Marcus JH, Novembre J. Visualizing the Geography of Genetic Variants. Bioinformatics (Oxford, England). 2016; doi:10.1093/bioinformatics/btw643

184. Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, et al. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. Human Genetics. 2008;123: 177–187. doi:10.1007/s00439-007-0460-x

185. Kılınç GM, Omrak A, Özer F, Günther T, Büyükkarakaya AM, Bıçakçı E, et al. The Demographic Development of the First Farmers in Anatolia. Current Biology. 2016;26: 2659–2666. doi:10.1016/j.cub.2016.07.057

186. Nakagome S, Alkorta-Aranburu G, Amato R, Howie B, Peter BM, Hudson RR, et al. Estimating the ages of selection signals from different epochs in human history. Molecular biology and evolution. 2015;33: 657–669.

187. Basu Mallick C, Iliescu FM, Möls M, Hill S, Tamang R, Chaubey G, et al. The Light Skin Allele of SLC24A5 in South Asians and Europeans Shares Identity by Descent. PLoS Genetics. 2013;9. doi:10.1371/journal.pgen.1003912

188. Yuasa I, Umetsu K, Harihara S, Kido A, Miyoshi A, Saitou N, et al. Distribution of the F374 allele of the SLC45A2 (MATP) gene and founder-haplotype analysis. Annals of Human Genetics. 2006;70: 802–811. doi:10.1111/j.1469-1809.2006.00261.x

189. Canfield V a, Berg A, Peckins S, Wentzel SM, Ang KC, Oppenheimer S, et al. Molecular phylogeography of a human autosomal skin color locus under natural selection. G3 (Bethesda, Md). 2013;3: 2059–67. doi:10.1534/g3.113.007484

190. Adhikari K, Fontanil T, Cal S, Mendoza-Revilla J, Fuentes-Guajardo M, Chacón-Duque J-C, et al. A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. Nature communications. 2016;7: 10815.

191. Kamberov YG, Wang S, Tan J, Gerbault P, Wark A, Tan L, et al. Modeling recent human evolution in mice by expression of a selected EDAR variant. Cell. 2013;152: 691–702. doi:10.1016/j.cell.2013.01.016

192. Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G, et al. Common Variants in the Trichohyalin Gene Are Associated with Straight Hair in Europeans. American Journal of Human Genetics. 2009;85: 750–755. doi:10.1016/j.ajhg.2009.10.009

193. Li R, Brockschmidt FF, Kiefer AK, Stefansson H, Nyholt DR, Song K, et al. Six novel susceptibility loci for early-onset androgenetic alopecia and their unexpected association with common diseases. PLoS Genetics. 2012;8. doi:10.1371/journal.pgen.1002746

194. Jobling M, Hollox E, Hurles M, Kivisild T, Tyler-Smith C. Human Evolutionary Genetics, Second Edition. Garland Science; 2013.

195. Jobling M, Hollox E, Hurles M, Kivisild T, Tyler-Smith C. Human Evolutionary Genetics [Internet]. 1989. doi:10.1017/CBO9781107415324.004

196. Hein DW. Molecular genetics and function of NAT1 and NAT2: Role in aromatic amine metabolism and carcinogenesis. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis. 2002;506–507: 65–77. doi:10.1016/S0027-5107(02)00153-7

197. Sabbagh A, Darlu P, Crouau-Roy B, Poloni ES. Arylamine N-Acetyltransferase 2 (NAT2) Genetic Diversity and Traditional Subsistence: A Worldwide Population Survey. PLoS ONE. 2011;6. doi:10.1371/journal.pone.0018507

198. Patin E, Barreiro LB, Sabeti PC, Austerlitz F, Luca F, Sajantila A, et al. Deciphering the ancient and complex evolutionary history of human arylamine N-acetyltransferase genes. American Journal of Human Genetics, The. 2006;78: 423–436. doi:10.1086/500614

199. Peng Y, Shi H, Qi X, Xiao C, Zhong H, Ma RZ, et al. The ADH1B Arg47His polymorphism in east Asian populations and expansion of rice domestication in history. BMC evolutionary biology. 2010;10: 15. doi:10.1186/1471-2148-10-15

200. Kuznetsov IB, Mcduffie M, Moslehi R. A web server for inferring the human N-acetyltransferase-2 (NAT2) enzymatic phenotype from NAT2 genotype. Bioinformatics. 2009;25: 1185–1186. doi:10.1093/bioinformatics/btp121

201. He YJ, Shapero MH, Mcleod HL. Novel tagging SNP rs1495741 and 2-SNPs (rs1041983 and rs1801280) yield high prediction of NAT2 genotype in HapMap samples. Pharmacogenetics and Genomics. 2012;22: 322–324. doi:10.1097/FPC.0b013e3283510a51

202. Selinski S, Blaszkewicz M, Lehmann M-L, Ovsiannikov D, Moormann O, Guballa C, et al. Genotyping NAT2 with only two SNPs (rs1041983 and rs1801280) outperforms the tagging SNP rs1495741 and is equivalent to the conventional 7-SNP NAT2 genotype. Pharmacogenetics and genomics. 2011;21: 673–8. doi:10.1097/FPC.0b013e3283493a23

203. Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. Evidence for archaic adaptive introgression in humans. Nature reviews Genetics. 2015;16: 359–371. doi:10.1038/nrg3936

204. Wang TJ, Zhang F, Richards JB, Kestenbaum B, Van Meurs JB, Berry D, et al. Common genetic determinants of vitamin D insufficiency: A genome-wide association study. The Lancet. 2010;376: 180–188. doi:10.1016/S0140-6736(10)60588-0

205. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010;466: 707–13. doi:10.1038/nature09270

206. Mathias RA, Sergeant S, Ruczinski I, Torgerson DG, Hugenschmidt CE, Kubala M, et al. The impact of FADS genetic variants on $ømega$6 polyunsaturated fatty acid metabolism in African Americans. BMC genetics. 2011;12: 50. doi:10.1186/1471-2156-12-50

207. Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jørgensen ME, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science (New York, NY). 2015;349: 1343–7. doi:10.1126/science.aab2319

208. Tosi F, Sartori F, Guarini P, Olivieri O, Martinelli N. Delta-5 and delta-6 desaturases: Crucial enzymes in polyunsaturated fatty acid-related pathways with pleiotropic influences in health and disease. 2014. doi:10.1007/978-3-319-7320-0_7

209. Wooding S, Kim U-K, Bamshad MJ, Larsen J, Jorde LB, Drayna D. Natural selection and molecular evolution in PTC, a bitter-taste receptor gene. American journal of human genetics. 2004;74: 637–46. doi:10.1086/383092

210. Risso DS, Mezzavilla M, Pagani L, Robino A, Morini G, Tofanelli S, et al. Global diversity in the TAS2R38 bitter taste receptor: revisiting a classic evolutionary PROPosal. Scientific reports. 2016;6: 25506. doi:10.1038/srep25506

211. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. Nature genetics. 2007;39: 1256–1260.

212. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, et al. Personalized copy number and segmental duplication maps using next-generation sequencing. Nature genetics. 2009;41: 1061–1067. doi:10.1038/ng.437

213. Breese MR, Liu Y. NGSUtils: A software suite for analyzing and manipulating next-generation sequencing datasets. Bioinformatics. 2013;29: 494–496. doi:10.1093/bioinformatics/bts731

214. Benson G. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Research. 1999;27: 573–580. doi:10.1093/nar/27.2.573

215. Inchley CE, Larbey CD, Shwan NA, Pagani L, Saag L, Antão T, et al. Selective sweep on human amylase genes postdates the split with Neanderthals. Scientific reports. 2016;6: 37198.

216. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. The New England journal of medicine. 2008;359: 2208–2219. doi:10.1056/NEJMoa0804742

217. Cornelis MC, Qi L, Zhang C, Kraft P, Manson JA, Cai T, et al. Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. Annals of Internal Medicine. 2009;150: 541–550. doi:10.1059/0003-4819-150-8-200904210-00008

218. Mielke JH, Konigsberg LW, Relethford J. Human biological variation. Oxford University Press; 2011.

219. Melzer D, Perry JRB, Hernandez D, Corsi A-M, Stevens K, Rafferty I, et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). PLoS genetics. 2008;4: e1000072. doi:10.1371/journal.pgen.1000072

220. Mondry A, Loh M, Liu P, Zhu A-L, Nagel M. Polymorphisms of the insertion / deletion ACE and M235T AGT genes and hypertension: surprising new findings and meta-analysis of data. BMC Nephrol. 2005;6: 1. doi:10.1186/1471-2369-6-1

221. Zeron-Medina J, Wang X, Repapi E, Campbell MR, Su D, Castro-Giner F, et al. XA polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. Cell. 2013;155. doi:10.1016/j.cell.2013.09.017

222. Yoshiura K, Kinoshita A, Ishida T, Ninokata A, Ishikawa T, Kaname T, et al. A SNP in the ABCC11 gene is the determinant of human earwax type. Nat Genet. 2006;38: 324–330. doi:10.1038/ng1733

223. Crabb DW, Edenberg HJ, Bosron WF, Li TK. Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity. The inactive ALDH2(2) allele is dominant. J Clin Invest. 1989;83: 314–316. doi:10.1172/JCI113875

224. Quillen EE, Chen X-D, Almasy L, Yang F, He H, Li X, et al. ALDH2 is associated to alcohol dependence and is the major genetic determinant of "daily maximum drinks" in a GWAS study of an isolated rural Chinese sample. Am J Med Genet B Neuropsychiatr Genet. 2014;165B: 103–110. doi:10.1002/ajmg.b.32213

225. Brooks PJ, Enoch M-A, Goldman D, Li T-K, Yokoyama A. The alcohol flushing response: an unrecognized risk factor for esophageal cancer from alcohol consumption. PLoS Med. 2009;6: e50. doi:10.1371/journal.pmed.1000050

226. Asberg a, Hveem K, Thorstensen K, Ellekjter E, Kannelønning K, Fjøsne U, et al. Screening for hemochromatosis: high prevalence and low morbidity in an unselected population of 65,238 persons. Scandinavian journal of gastroenterology. 2001;36: 1108–1115. doi:10.1080/003655201750422747

227. Whittington CA. Was the C282Y mutation an Irish Gaelic mutation that the Vikings help disseminate? Medical Hypotheses. 2006;67: 1270–1273. doi:10.1016/j.mehy.2006.06.013

228. Milman N, Pedersen P. Evidence that the Cys282Tyr mutation of the HFE gene originated from a population in Southern Scandinavia and spread with the Vikings. Clinical Genetics. 2003;64: 36–47. doi:10.1034/j.1399-0004.2003.00083.x

229. Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. Nature. 2007;447: 279–283. doi:10.1038/nature05775

230. Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. Nature Reviews Genetics. 2010;11: 17–30.

231. Claes P, Hill H, Shriver MD. Toward DNA-based facial composites: preliminary results and validation. Forensic Science International: Genetics. 2014;13: 208–216.

232. Claes P. Method for Predicting a Phenotype from a Genotype [Internet]. WO/2015/173435, 2015. Available: https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2015173435

233. Claes P, Liberton DK, Daniels K, Rosana KM, Quillen EE, Pearson LN, et al. Modeling 3D Facial Shape from DNA. Luquetti D, editor. PLoS Genetics. 2014;10: e1004224. doi:10.1371/journal.pgen.1004224

234. Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, et al. The international HapMap project. Nature. 2003;426: 789–796.

235. Claes P, Walters M, Clement J. Improved facial outcome assessment using a 3D anthropometric mask. International journal of oral and maxillofacial surgery. 2012;41: 324–330.

236. Snyders J, Claes P, Vandermeulen D, Suetens P. Development and comparison of non-rigid surface registration algorithms and extensions. 2014; Available: https://lirias.kuleuven.be/handle/123456789/451862

237. Shrimpton S, Daniels K, De Greef S, Tilotta F, Willems G, Vandermeulen D, et al. A spatially-dense regression study of facial form and tissue depth: towards an interactive tool for craniofacial reconstruction. Forensic science international. 2014;234: 103–110.

238. Makinen TM. Different types of cold adaptation in humans. Frontiers in bioscience. 2010;2: 1047–1067. doi:10.2741/S117

239. 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. doi:10.1126/science.1190371

240. 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. doi:10.1038/ncomms10775

241. Cooper GM. Distribution and intensity of constraint in mammalian genomic sequence. Genome Research. 2005;15: 901–913. doi:10.1101/gr.3577405

242. Ullah AZD, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). Nucleic Acids Research. 2012; gks364. doi:10.1093/nar/gks364

243. 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. doi:10.1186/1471-2350-8-S1-S2

244. Newton-Cheh C, Guo C-Y, Wang TJ, O'donnell CJ, Levy D, Larson MG. Genome-wide association study of electrocardiographic and heart rate variability traits: the Framingham Heart Study. BMC medical genetics. 2007;8: 1.

245. Gottlieb DJ, T O'Connor G, Wilk JB. Genome-wide association of sleep and circadian phenotypes. BMC medical genetics. 2007;8: 1.

246. Lasky-Su J, Neale BM, Franke B, Anney RJ, Zhou K, Maller JB, et al. Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2008;147: 1345–1354.

247. Shi M, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, et al. Genome wide study of maternal and parent-of-origin effects on the etiology of orofacial clefts. American Journal of Medical Genetics Part A. 2012;158: 784–794.

248. Rose JE, Behm FM, Drgon T, Johnson C, Uhl GR. Personalized smoking cessation: interactions between nicotine dose, dependence and quit-success genotype score. Molecular Medicine (Cambridge, Mass). 2010;16: 247–253. doi:10.2119/molmed.2009.00159

249. Salyakina D, Ma DQ, Jaworski JM, Konidari I, Whitehead PL, Henson R, et al. Variants in several genomic regions associated with asperger disorder. Autism Research. 2010;3: 303–310.

250. Fox CS, Heard-Costa N, Cupples LA, Dupuis J, Vasan RS, Atwood LD. Genome-wide association to body mass index and waist circumference: the Framingham Heart Study 100K project. BMC medical genetics. 2007;8: 1.

251. Bailey CJ, Gross JL, Pieters A, Bastien A, List JF. Effect of dapagliflozin in patients with type 2 diabetes who have inadequate glycaemic control with metformin: a randomised, double-blind, placebo-controlled trial. The Lancet. 2010;375: 2223–2233.

252. Levy D, Larson MG, Benjamin EJ, Newton-Cheh C, Wang TJ, Hwang S-J, et al. Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. BMC medical genetics. 2007;8: 1.

253. Lunetta KL, D'Agostino RB, Karasik D, Benjamin EJ, Guo C-Y, Govindaraju R, et al. Genetic correlates of longevity and selected age-related phenotypes: a genome-wide association study in the Framingham Study. BMC medical genetics. 2007;8: 1.

254. Gieger C, Geistlinger L, Altmaier E, De Angelis MH, Kronenberg F, Meitinger T, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS Genet. 2008;4: e1000282.

255. Kathiresan S, Manning AK, Demissie S, D'agostino RB, Surti A, Guiducci C, et al. A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. BMC medical genetics. 2007;8: 1.

256. Benjamin EJ, Dupuis J, Larson MG, Lunetta KL, Booth SL, Govindaraju DR, et al. Genome-wide association with select biomarker traits in the Framingham Heart Study. BMC medical genetics. 2007;8: 1.

257. Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. Human Molecular Genetics. 2009;18: 767–778. doi:10.1093/hmg/ddn388

258. Mero I-L, Lorentzen $\backslash$AAslaug R., Ban M, Smestad C, Celius EG, Aarseth JH, et al. A rare variant of the TYK2 gene is confirmed to be associated with multiple sclerosis. European journal of human genetics. 2010;18: 502–504.

259. Gourraud P-A, McElroy JP, Caillier SJ, Johnson BA, Santaniello A, Hauser SL, et al. Aggregation of multiple sclerosis genetic risk variants in multiple and single case families. Annals of neurology. 2011;69: 65–74.

260. Cleynen I, John JMM, Henckaerts L, van Moerkercke W, Rutgeerts P, van Steen K, et al. Molecular reclassification of Crohn's disease by cluster analysis of genetic variants. PLoS ONE. 2010;5. doi:10.1371/journal.pone.0012952

261. Okada Y, Kamatani Y, Takahashi A, Matsuda K, Hosono N, Ohmiya H, et al. A genome-wide association study in 19 633 Japanese subjects identified LHX3-QSOX2 and IGF1 as adult height loci. Human molecular genetics. 2010; ddq091.

262. Fujimoto A, Nishida N, Kimura R, Miyagawa T, Yuliwulandari R, Batubara L, et al. FGFR2 is associated with hair thickness in Asian populations. Journal of human genetics. 2009;54: 461–465.

263. Wheeler HE, Metter EJ, Tanaka T, Absher D, Higgins J, Zahn JM, et al. Sequential use of transcriptional profiling, expression quantitative trait mapping, and gene association implicates MMP20 in human kidney aging. PLoS Genet. 2009;5: e1000685.

264. Suzuki J, Yamazaki Y, Li G, Kaziro Y, Koide H, Guang L. Involvement of Ras and Ral in chemotactic migration of skeletal myoblasts. Mol Cell Biol. 2000;20: 4658–4665.

265. Eynon N, Alves AJ, Sagiv M, Yamin C, Sagiv M, Meckel Y. Interaction between SNPs in the NRF2 gene and elite endurance performance. Physiol Genomics. 2010;41: 78–81. doi:10.1152/physiolgenomics.00199.2009

266. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. The American Journal of Human Genetics. 2006;78: 629–644.

267. Paradis E. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics. 2010;26: 419–420.

268. Gautier M, Klassmann A, Vitalis R. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. Mol Ecol Resour. 2017;17: 78–90. doi:10.1111/1755-0998.12634

269. Tang K, Thornton KR, Stoneking M. A new approach for using genome scans to detect recent positive selection in the human genome. PLoS biology. 2007;5: e171.

270. Kofler R, Schlötterer C. Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. Bioinformatics. 2012;28: 2084–2085.

271. Ahlström T. Den exogama gränsen: Kring interaktionen mellan jägare-samlare och bönder-boskapsskötare under mellanneolitisk tid. Till Gunborg: arkeologiska samtal. 1997;

272. Ahlström T. Pitted ware skeletons and boreal temperatures. Lund Archaeological Review,(3). 1997; 37–48.

273. Fan S, Hansen MEB, Lo Y, Tishkoff SA. Going global by adapting local: A review of recent human adaptation. Science. 2016;354: 54–59. doi:10.1126/science.aaf5098

274. 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. doi:10.1098/rstb.2016.0349 372(1724):20160349.