The genetic prehistory of the Baltic Sea region

Mittnik, Alissa

2018-01-30


http://hdl.handle.net/10138/233550
https://doi.org/10.1038/s41467-018-02825-9

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.
The genetic prehistory of the Baltic Sea region

Alissa Mittnik1,2, Chuan-Chao Wang1,3, Saskia Pfrengle2, Mantas Daubaras4, Gunita Zariņa5, Fredrik Hallgren6, Raili Allmäe7, Valery Khartanovich8, Vyacheslav Moiseyev8, Mari Tõrv9, Anja Furtwängler2, Aida Andrades Valtueña1, Michal Feldman1, Christos Economou10, Markku Oinonen11, Andrejs Vasks5, Elena Balanovska12, David Reich13,14,15, Rimantas Jankauskas16, Wolfgang Haak1,17, Stephan Schiffels1 & Johannes Krause1,2

While the series of events that shaped the transition between foraging societies and food producers are well described for Central and Southern Europe, genetic evidence from Northern Europe surrounding the Baltic Sea is still sparse. Here, we report genome-wide DNA data from 38 ancient North Europeans ranging from ~9500 to 2200 years before present. Our analysis provides genetic evidence that hunter-gatherers settled Scandinavia via two routes. We reveal that the first Scandinavian farmers derive their ancestry from Anatolia 1000 years earlier than previously demonstrated. The range of Mesolithic Western hunter-gatherers extended to the east of the Baltic Sea, where these populations persisted without gene-flow from Central European farmers during the Early and Middle Neolithic. The arrival of steppe pastoralists in the Late Neolithic introduced a major shift in economy and mediated the spread of a new ancestry associated with the Corded Ware Complex in Northern Europe.
Recent genome-wide data of Baltic pottery-producing hunter-gatherers revealed genetic continuity with the preceding Mesolithic inhabitants of the same region as well as influence from the more northern EHGs, but did not reveal conclusively whether there was a temporal, geographical or cultural correlation with the affinity to either WHG or EHGs.

The transition from the Late (Final) Neolithic to the Early Bronze Age (LNBA) is seen as a major transformative period in European prehistory, accompanied by changes in burial customs, technology and mode of subsistence as well as the creation of new cross-continental networks of contact seen in the emergence of the pan-European Corded Ware Complex (CWC, ca. 2900–2300 calBCE) in Central and north-eastern Europe. Studies of ancient genomes have shown that those associated with the CWC were closely related to the pastoralists of the Yamnaya Culture from the Pontic-Caspian steppe, introducing a genetic component that was not present in Europe previously. This genetic component is hypothesized to have spread in the subsequent millennia throughout Europe and can be seen in today’s European populations in a decreasing north-east to south-west gradient.

Intriguingly, modern Eastern Baltic populations carry the highest proportion of WHG ancestry of all Europeans, supporting the theory that the hunter-gatherer population of this region left a lasting genetic impact on subsequent populations.

Here, we investigate the modes of cultural and economic transitions experienced by the prehistoric populations surrounding the Baltic Sea. Were the changes seen in the Eastern Baltic Neolithic, which did not involve the introduction of agriculture, driven by contact with neighbouring groups and if so can we identify these? Was the earliest practice of farming in southern Scandinavia a development by a local population or did it involve migration from the South? How did the unique genetic signature of modern Eastern Baltic populations come to be?

We present novel genome-wide data from 38 ancient individuals from the Eastern Baltic, Russia and Sweden spanning 7000 years of prehistory, covering the transition from a mobile hunter-gatherer to a sedentary agricultural lifestyle, as well as the adoption of bronze metallurgy. We show that the settlement of Scandinavia by hunter-gatherers likely took place via at least two routes, and that the first introduction of farming was brought about by the movement of the Central European farmers into the region at around 4000 calBCE. In the Eastern Baltics, foraging remained the dominant economy among interconnected northeastern hunter-gatherer groups that did not experience admixture from European farmers until around 3000 calBCE, when a shift towards agro-pastoralism came about through migrations from the Pontic-Caspian steppe.

Results

Samples and archaeological background. The skeletal remains studied here were recovered from 25 archaeological sites in the territory of modern Lithuania, Latvia, Estonia, Archangelsk Oblast and Karelia (north-western Russia) and Sweden dating from around 7500 to 200 calBCE (Fig. 1, Supplementary Note 2, Supplementary Data 1). In total, we analyzed DNA from 106 human remains. A total of 41 samples with good DNA preservation were selected for deeper shotgun sequencing or SNP capture (Supplementary Data 1). In the latter case, we enriched samples for a panel of around 1.24 million single nucleotide polymorphisms (SNPs) via in-solution capture. After quality control, genome-wide data from 38 individuals, with an average coverage of 0.02–8.8-fold on targeted SNPs, were included in further analysis.
The 38 final samples fall into nine broad groups (Table 1): first, two Mesolithic hunter-gatherers from north-western Russia (ca. 7500–5000 calBCE); secondly, an individual (5720–5560 calBCE) from the lake burial site of Motala, Sweden, adding to the previously published six SHG individuals from this site4; two Mesolithic hunter-gatherers from Lithuania (ca. 6440–5740 calBCE) associated with the Kunda Culture (referred to as Baltic Mesolithic), whose archaeological assemblages found in Lithuania, Latvia, Estonia and adjacent parts of Russia. Twelve individuals were associated with pottery-producing forager cultures. Ten of them from Lithuania and Estonia (ca. 5460–3820 calBCE) were assigned to the Narva Culture that occupied the Eastern Baltic region from the Late Mesolithic to the Middle Neolithic (Baltic EMN Narva) and two individuals from Estonia were associated with the CCC that was spread across the northern part of the Eastern Baltic by the Middle Neolithic (Baltic MN CCC; dated to ca. 3800–3360 calBCE).

Five individuals from Lithuania and Estonia were dated to the Late Neolithic (Baltic LN; ca. 3260–1750 calBCE). For the individual Gyvakarai1, we present genome-wide data at 7.6-fold average coverage from shotgun sequencing.

Fourteen samples from Latvia and Lithuania were attributed to the Baltic Bronze Age (Baltic BA) and date to ca. 1230–230 calBCE.

We supplemented our dataset of ancient Eastern Baltic samples with recently published data from 13 individuals spanning the Mesolithic to Late Neolithic in Latvia21 and Estonia22, and merged data from identical individuals where they overlapped with the latter. We present the first data from the southernmost region of the Eastern Baltic (the territory of modern-day Lithuania), the Early Neolithic of Estonia and the Eastern Baltic Bronze Age.

From southern Sweden, we analyzed one farmer (3950–3650 calBCE) from the EN TRB, the earliest agricultural population in Scandinavia for which there exists no genetic data to date. One sample from northern Sweden (Scandinavia LNBA Olsund, ca. 2570–2140 calBCE) dates to the Late Neolithic but was found without associated archaeological assemblages. The data were analyzed in context with published data from the Scandinavian Middle Neolithic to Bronze Age3, 19 as well as other ancient and modern genome-wide datasets described below.

**Affinities of northern Mesolithic hunter-gatherers.** To gain an overview of the broad genetic affinities of our samples, we projected all 38 ancient genome-wide datasets, as well as previously published ancient samples4, 6, 19, 21, 22, using a principal component analysis (PCA), constructed from 1007 modern individuals from a diverse set of West Eurasian contemporary populations, and used the same individuals to investigate model-based clustering using ADMIXTURE. We see that the Mesolithic foragers of Northern Europe fall into three distinct clusters, associated with EHG, SHG and WHG, respectively, as evidenced by their position on the PCA (Fig. 2a), similar composition of ancestral genetic clusters in ADMIXTURE analysis (Fig. 2b, Supplementary Fig. 4) and in sharing most genetic drift since...
among the highest in EHG (using the X-statistics of the form related to WHG, and populations containing ANE ancestry). A Palaeolithic hunter-gatherer from Switzerland, most closely related to WHG in PCA space, have increased shared genetic drift with both SHG and EHG (Fig. 3), which reveals a difference in mixture proportions between the more northern individuals from the Latvian site (65–76% WHG with 24–35% EHG; Supplementary Table 3) and the samples from the Lithuanian sites to the south (88–100% WHG with 0–12% EHG; Supplementary Table 3).

SHG appear intermediate between WHG/Baltic Mesolithic and EHG in PCA space, have increased shared genetic drift with both shown in outgroup f<sub>p</sub>-statistics (Supplementary Fig. 3) and the statistic D(SHG,WHG; EHG, Mbuti) is strongly significant for excess allele sharing of SHG and EHG (Z = 7.3; Supplementary Table 4). Both EHG and SHG share a non-negligible component in ADMIXTURE analysis which is maximized in some modern Native American populations which points towards ANE ancestry, as represented by the MA1 and AG3 samples from Palaearctic Siberia (maroon component in Fig. 2a, Supplementary Fig. 4). Indeed, D-statistics show that EHG and SHG share significantly more alleles with MA1 and AG3 than WHG (Supplementary Tables 2 and 4).

Additionally, mtDNA divergence from Africa as shown by outgroup f<sub>p</sub>-statistics (Supplementary Fig. 1). Based on these results, our Mesolithic Russian foragers fall within the EHG cluster formed by previously published samples and are grouped as EHG in subsequent analyses.

ADMIIXTURE shows that EHG carry a genetic component (green component in Fig. 2b) that is maximized in hunter-gatherers from the Caucasus (CHG) and shared with Neolithic farmers from Iran and Steppe populations from the Bronze Age, suggesting some common ancestry for these populations, consistent with previous results.

Despite their geographical vicinity to EHG, the two Eastern Baltic individuals associated with the Mesolithic Kunda Culture show a very close affinity to WHG in all our analyses (Fig. 2, Supplementary Figs. 1 and 2), with a significant contribution from ANE, as revealed by negative admixture f<sub>p</sub> results involving a Palaeolithic hunter-gatherer from Switzerland, most closely related to WHG, and populations containing ANE ancestry (Supplementary Table 1). This is additionally confirmed by D-statistics of the form D(Baltic Mesolithic, WHG; X, Mbuti) for populations X with ANE ancestry, which are significant and among the highest in EHG (Z = 6.2; Supplementary Table 2).

Using the apWave/apAdm framework, we modelled the Baltic Mesolithic hunter-gatherers as a two-way mixture between EHG and WHG (Fig. 3), which reveals a difference in mixture proportions between the more northern individuals from the Latvian site (65–76% WHG with 24–35% EHG; Supplementary Table 3) and the samples from the Lithuanian sites to the south (88–100% WHG with 0–12% EHG; Supplementary Table 3).

SHG appear intermediate between WHG/Baltic Mesolithic and EHG in PCA space, have increased shared genetic drift with both shown in outgroup f<sub>p</sub>-statistics (Supplementary Fig. 3) and the statistic D(SHG,WHG; EHG, Mbuti) is strongly significant for excess allele sharing of SHG and EHG (Z = 7.3; Supplementary Table 4). Both EHG and SHG share a non-negligible component in ADMIXTURE analysis which is maximized in some modern Native American populations which points towards ANE ancestry, as represented by the MA1 and AG3 samples from Palaearctic Siberia (maroon component in Fig. 2a, Supplementary Fig. 4). Indeed, D-statistics show that EHG and SHG share significantly more alleles with MA1 and AG3 than WHG (Supplementary Tables 2 and 4). Additionally, mtDNA divergence from Africa as shown by outgroup f<sub>p</sub>-statistics (Supplementary Fig. 1). Based on these results, our Mesolithic Russian foragers fall within the EHG cluster formed by previously published samples and are grouped as EHG in subsequent analyses.

ADMIIXTURE shows that EHG carry a genetic component (green component in Fig. 2b) that is maximized in hunter-gatherers from the Caucasus (CHG) and shared with Neolithic farmers from Iran and Steppe populations from the Bronze Age, suggesting some common ancestry for these populations, consistent with previous results.

Despite their geographical vicinity to EHG, the two Eastern Baltic individuals associated with the Mesolithic Kunda Culture show a very close affinity to WHG in all our analyses (Fig. 2, Supplementary Figs. 1 and 2), with a significant contribution from ANE, as revealed by negative admixture f<sub>p</sub> results involving a Palaeolithic hunter-gatherer from Switzerland, most closely related to WHG, and populations containing ANE ancestry (Supplementary Table 1). This is additionally confirmed by D-statistics of the form D(Baltic Mesolithic, WHG; X, Mbuti) for populations X with ANE ancestry, which are significant and among the highest in EHG (Z = 6.2; Supplementary Table 2).
Fig. 2 PCA and ADMIXTURE analysis reflecting three time periods in Northern European prehistory. a Principal components analysis of 1012 present-day West Eurasians (grey points, modern Baltic populations in dark grey) with 294 projected published ancient and 38 ancient North European samples introduced in this study (marked with a red outline). Population labels of modern West Eurasians are given in Supplementary Fig. 7 and a zoomed-in version of the European Late Neolithic and Bronze Age samples is provided in Supplementary Fig. 8. b Ancestral components in ancient individuals estimated by ADMIXTURE ($k = 11$)
haplogroups found among EHG point towards an eastern influence (Table 1): R1b in U5O077 (this individual had previously been assigned to haplogroup H35) was also found in the Palaeolithic Siberian AG3 and a haplogroup within the C1 clade, which appears today in highest frequencies in north-east Asia and the Americas, was described in several samples from Yuzhny Oleni Ostrov. It was shown that some SHG carry the derived variant of the EDAR allele, which affects hair thickness and tooth morphology among other things, and which is found today in high frequency in East Asians and Native Americans.

Dynamic forager networks in the Eastern Baltic Neolithic. Similarly to the Baltic Mesolithic, the later Eastern Baltic Neolithic hunter-gatherers of the Narva culture exhibit varying proportions of EHG (0–46%) and WHG (54–100%) ancestry (Fig. 3 and Supplementary Table 3). In principle, it is possible that such a pattern is not the result of admixture but a signal of a long-standing geographic cline. However, in this case, it appears to be more compatible with recent admixture between differentially WHG- and EHG-related groups, as we see varying ancestry proportions even within contemporary individuals from the same site or closely located sites (Fig. 3). D-statistics of the form D (Baltic EMN Narva, WHG, X, Mbuti) do not show evidence of admixture with the contemporaneous European farming populations that were related to the Anatolian Neolithic (Supplementary Tables 2 and 5). The later individuals attributed to the Baltic MN CCC exhibit a significantly higher affinity to EHG with the ancestry proportion estimated at 68–99% EHG and 1–32% WHG (Fig. 3, Supplementary Table 3).

Similar to other European Mesolithic hunter-gatherers, our Baltic foragers carry a high frequency of the derived HERC2 allele which codes for light iris colour, and like SHG and EHG they already possess an increased frequency of the derived alleles for SLC45A2 and SLC24A5, coding for lighter skin colour (Supplementary Table 6). The male individuals carry Y-chromosomal haplogroups of the I and R1 clades (Table 1, Supplementary Note 3). Y-haplogroup I has been most commonly found among WHG and SHG, while R1 is found in EHG and other published Eastern Baltic and Romanian hunter-gatherers.

One Narva individual, Spiginas1, dated to ca. 4440–4240 calBCE, belongs to a mitochondrial haplogroup of the H branch, normally associated with the Neolithic expansion into Europe, but shows no evidence of Neolithic farmer ancestry on the nuclear level suggesting that this haplogroup might have been present already in foraging groups (Table 1, Supplementary Table 5). In addition to haplogroup H, the maternal lineages seen in Eastern Baltic samples (n = 35; Supplementary Fig. 5) encompass all of the major haplogroups identified in complete mtDNA genomes from Holocene Scandinavian and Western European hunter-gatherers (n = 21: U2, U5a, U5b) as well as haplogroup U4 which has been found in high frequency in Mesolithic foragers from Russia. We also find mtDNA branch K1, a subclade of U8, in one Baltic Mesolithic forager, adding to the mounting evidence that this lineage was present at low frequency among European hunter-gatherers before the arrival of agriculturalists.

Early farming in Sweden coincides with a shift in ancestry. In contrast to the Eastern Baltic, we see clear evidence for the genetic impact of the Neolithic expansion already around 4000 calBCE in southern Sweden. The individual associated with the EN TRB culture clusters with Middle and Early Neolithic farmers from Europe on the PCA (Fig. 2a) and in the ADMIXTURE analysis exhibits the component maximized in Levantine and West Anatolian early farmers (orange component in Fig. 2b). D-statistics show increased allele sharing of EN TRB with hunter-gatherer populations in comparison to the preceding farmers of the Linearbandkeramik (LBK) culture and no difference to the contemporaneous Middle Neolithic and Chalcolithic (MNChL) farmers of Central Europe (Supplementary Table 4), paralleling the previously described resurgence of WHG ancestry during the European Middle Neolithic. Different models of EN TRB as a linear combination of LBK with SHG, WHG or Baltic hunter-gatherers favour the latter two groups, while SHG is rejected as a source (Supplementary Table 3).

The previously published succeeding farmers of the Middle Neolithic (MN) TRB culture in West Sweden appear as directly descended from the EN TRB, with no significant positive results for D(MN TRB, EN TRB; X, Mbuti) (Supplementary Table 4).

The PWC individuals, who were contemporaneous to the MN TRB but relied mainly on marine resources, appear intermediate between SHG and Middle Neolithic farming cultures on the PCA (Fig. 2a). Indeed, the statistic D(PWC, SHG; X, Mbuti) reaches weak significance when X is MN TRB (Z = 2.94) and a two-way admixture model for PWC involving SHG and TRB farmers is not rejected by qpWave/qpAdm (74 ± 6% SHG and 26 ± 6% EN TRB; Supplementary Table 3).

New networks of contact during the LNBA. The substantial population movement at the beginning of the 3rd millennium calBCE, during the European LNBA, affected the genetic makeup of Eastern and Central Europe and Scandinavia. It also made its mark in the Eastern Baltic region, as seen in our five samples from Lithuania and Estonia dated to this period and the previously published individuals from the Eastern Baltic region. All Baltic LN individuals (ca. 3200–1750 calBCE) fall in PCA Fig. 3 EHG ancestry in SHG and Eastern Baltic hunter-gatherers. Ancestry proportions were estimated with qpAdm, and standard errors are shown as vertical lines. Colours indicate latitude of the site at which the individuals were excavated.
space in the diffuse European LNBA cluster formed by individuals admixed between Early and Middle Bronze Age pastoralists from the Yamnaya culture of the eastern Pontic Steppe and Middle Neolithic European farmers (Fig. 2a). They all carry the genetic component that was introduced into Europe with this pastoralist migration in varying amounts, and the majority also carries the component associated with Anatolian farmers (green and orange, respectively, in Fig. 2b). This genetic impact is furthermore reflected in the uniparental markers where we see novel mitochondrial haplogroups (I, J, T2, W), not found in the preceding foragers, in half of our samples (Supplementary Fig. 5), and I2a Y-chromosomal haplogroups replaced by R1a types (Table 1, Supplementary Note 3).

Computing D-statistics for each individual of the form D(Baltic LN, Yamnaya; X; Mbuti), we find that the two individuals from the early phase of the LN (Plinkaigalis242 and Gyvakarai1, dating to ca. 3200–2600 calBCE) form a clade with Yamnaya (Supplementary Table 7), consistent with the absence of the farmer-associated component in ADMIXTURE (Fig. 2b). Younger individuals share more alleles with Anatolian and European farmers (Supplementary Table 7) as also observed in contemporaneous Central European CWC individuals. The individual Spiginas2, dated to a very late period of the LN (2130–1750 calBCE), stands out in that it shares an excess of alleles with European forager groups when compared to the Yamnaya populations, with the top hits being Switzerland HG, WHG, Baltic Mesolithic and Baltic EMN Narva (Supplementary Table 7).

This result is the earliest evidence for a continuing pattern: we observe that the increased affinity to Baltic hunter-gatherers remains prevalent in the more recent samples from the Baltic BA (dated between ca. 1000 and 230 calBCE) that cluster together on the PCA in the same space occupied by modern Lithuanians and Estonians, shifted from other Europeans to WHG and Baltic hunter-gatherers (Fig. 2a). The statistic D(Baltic BA, Baltic LN; Baltic EMN Narva, Mbuti) is strongly significant (Z = 14.0; Supplementary Table 2) demonstrating the increase in allele sharing with local hunter-gatherers in the Baltic populations after the Late Neolithic. Replacing Baltic EMN Narva with the contemporaneous northern population of Baltic MN CCC does not yield significant results, suggesting that admixture with this population did not play a large role in the ancestry of our studied Bronze Age individuals. Additionally, D-statistics provide significantly positive results for D(Baltic BA, Baltic LN; X, Mbuti) when X was replaced by various agricultural populations of Europe and the Near East (Supplementary Table 2), which suggests that the formation of the Baltic BA gene pool was not completed by admixture between the Baltic LN population and foragers but involved additional gene-flow from outside the Baltic territory. Archaeological evidence supports that the site Kivutkalns, which is represented by 10 of our individuals, was a major bronze-working centre located on a trade route that opened to the Baltic Sea on the west and led inland following the Daugava river29, where contact to surrounding populations might have been common.

The individual from Olsund in north-eastern Sweden was dated to the Late Neolithic, when agriculture had been introduced to the coastal areas of northern Sweden with the Battle Axe Culture, the regional variant of the CWC, while foraging persisted as an important form of subsistence. The remains were found without any associated artefacts, but in close proximity to a site where the assemblage showed a mix between local hunter-gatherer traditions and CWC influence30.

On the PCA plot, this sample falls within the European LNBA cluster (Fig. 2a) and similarly to other individuals from this cluster displays the three components derived from WHG, CHG and Neolithic Levant (Fig. 2b). This provides genomic evidence for the influence of both the early Neolithic and LNBA expansions having reached as far as northern Sweden in the 3rd millennium calBCE, either through a northward expansion from southern Scandinavia or across the Baltic Sea by boat or over the frozen sea31. Assemblages similar to the early Battle Axe Culture of Sweden have been found in south-western Finland, across the Bothnian Sea32, 33 which could be considered a geographically closer source than southern Scandinavia.

Gene-flow into the Eastern Baltic after the Bronze Age. Modern Eastern Baltic populations cluster with Baltic BA on the PCA plot and exhibit among all modern populations the highest shared genetic drift with ancient Baltic populations (Supplementary Fig. 2), but show substantial differences to samples from the Bronze Age. The statistic D(Lithuanian, Baltic BA; X, Mbuti) reveals significantly positive results for many modern Near Eastern and Southern European populations (Supplementary Fig. 6a). Limited gene-flow from more south-western neighbouring regions after the Bronze Age is sufficient to explain this pattern, as nearly all modern populations besides Estonians, especially for Central and Western Europe, have a higher amount of farmer ancestry than Lithuanians.

In contrast, the statistic D(Estonian, BA Baltic; X, Mbuti) gives significant positive hits for East Asian and Siberian populations (Supplementary Fig. 6b).

None of our male Bronze Age individuals carry Y-haplogroup N, which is found in modern Europeans in highest frequencies in Finland and the Baltic states34. Instead, we observe a high frequency of R1a Y-haplogroups.

Discussion

Our analyses support a dynamic population history of the Baltic Sea region, where populations did not remain in ‘genetic stasis’ despite the late adoption of agrarian subsistence strategies when compared to the rest of Europe.

The Mesolithic SHG excavated at Motala, Sweden, owe their genetic signature to an admixture of WHG and EHG and similarly to the latter carry substantial ANE ancestry. In contrast, the two Eastern Baltic Mesolithic Kunda individuals, who predate the SHG, carry comparatively low proportions of ANE ancestry, indicating that this ancestry was never widespread to the south-east of the Baltic Sea and likely reached Scandinavia without traversing the Eastern Baltic. This provides indirect support to the archaeological evidence that Scandinavia was settled by two routes13, suggesting a scenario in which the ANE-related ancestry was brought into Scandinavia with a movement of people via a north-eastern coastal route, where they admixed with a WHG-like population that derived from a migration over the landbridge that connected Denmark and southern Sweden at the time. This scenario is also supported by the finding that three Mesolithic hunter-gatherers excavated at the coast of Norway carry a higher proportion of EHG ancestry compared to the individuals from inland Sweden.

In southern Scandinavia, the sequence of events resembles that seen in Central Europe, albeit several millennia later, in that the earliest agriculture in the region coincides with the appearance of people related to the Anatolian and European Neolithic. However, similar to Middle Neolithic Central and Southern European populations, early Scandinavian farmers are already strongly admixed with hunter-gatherer groups. Without the knowledge of the genetic substrate in Mesolithic southern Scandinavia, these results are consistent with different scenarios; e.g. a movement of a Central European population into southern Scandinavia...
without admixture with local SHG-like populations, or local admixture of LBK-like farmers with a forager population that shows more similarity to WHG or Baltic hunter-gatherers\(^4\). A detailed joint analysis of genetic and archaeological data from hunter-gatherers from northern Germany, Denmark and the southern tip of Sweden is necessary to establish the role of local admixture during the emergence of the TRB culture. Our data support that the Neolithic PWC foragers are largely genetically continuous to SHG, which is congruent with their similarities in subsistence strategies, while continuity between EN TRB and PWC can also be seen in archaeological assemblages\(^36\) and can be attributed to contact between farmers and foragers. Indeed, genetic evidence of admixture between these groups shows that they were not completely isolated from each other but did likely not uphold continuous contact nor intermarry frequently during their prolonged parallel existence in Scandinavia.

In the archaeological understanding, the transition from Mesolithic to Neolithic in the Eastern Baltic region does not coincide with a large-scale population turnover and a stark shift in economy as seen in Central and Southern Europe. Rather, it is signified by a change in networks of contacts and the use of pottery, among other material, cultural and economic change\(^15\). Our results suggest continued admixture between groups in the south of the Eastern Baltic region, who are more closely related to WHG, and northern or eastern groups, more closely related to EHG. Neolithic social networks from the Eastern Baltic to the River Volga could also explain similarities of the hunter-gatherer pottery styles, although morphologically analogous ceramics might also have developed independently due to similar functionality\(^37\). The genetic evidence for a change in networks and possibly even a large-scale population movement is most pronounced in the Middle Neolithic in individuals attributed to the CCC. The distribution of this culture overlaps in the north with the Narva culture and extends further north to Finland and Karelia. Its spread in the Eastern Baltic is linked with a significant change in imported raw materials, artefacts, and the appearance of village-like settlements\(^15\).

We see a further population movement into the regions surrounding the Baltic Sea with the CWC in the Late Neolithic that was accompanied by the first evidence of extensive animal husbandry in the Eastern Baltic\(^15\). The presence of ancestry from the Pontic-Caspiian Steppe among CWC individuals without the genetic component from north-western Anatolian Neolithic Pontic-Caspian Steppe among Baltic CWC individuals without the genetic component from north-western Anatolian Neolithic16 and the southern tip of Sweden is necessary to establish the role of local admixture during the emergence of the TRB culture. Our data support that the Neolithic PWC foragers are largely genetically continuous to SHG, which is congruent with their similarities in subsistence strategies, while continuity between EN TRB and PWC can also be seen in archaeological assemblages\(^36\) and can be attributed to contact between farmers and foragers. Indeed, genetic evidence of admixture between these groups shows that they were not completely isolated from each other but did likely not uphold continuous contact nor intermarry frequently during their prolonged parallel existence in Scandinavia.

Our results suggest continued admixture between groups in the south of the Eastern Baltic region, who are more closely related to WHG, and northern or eastern groups, more closely related to EHG. Neolithic social networks from the Eastern Baltic to the River Volga could also explain similarities of the hunter-gatherer pottery styles, although morphologically analogous ceramics might also have developed independently due to similar functionality\(^37\). The genetic evidence for a change in networks and possibly even a large-scale population movement is most pronounced in the Middle Neolithic in individuals attributed to the CCC. The distribution of this culture overlaps in the north with the Narva culture and extends further north to Finland and Karelia. Its spread in the Eastern Baltic is linked with a significant change in imported raw materials, artefacts, and the appearance of village-like settlements\(^15\).

We see a further population movement into the regions surrounding the Baltic Sea with the CWC in the Late Neolithic that was accompanied by the first evidence of extensive animal husbandry in the Eastern Baltic\(^15\). The presence of ancestry from the Pontic-Caspiian Steppe among CWC individuals without the genetic component from north-western Anatolian Neolithic farmers must be due to a direct migration of steppe pastoralists that did not pick up this ancestry in Central Europe. It suggests import of the new economy by an incoming steppe-like population independent of the agricultural societies that were already established to the south and west of the Baltic Sea. The presence of direct contacts to the steppe could lend support to a linguistic homeland\(^38\) of direct contacts to the steppe could lend support to a linguistic homeland\(^38\). The presence of ancestry from the east that contributed to the male gene pool of Eastern Baltic populations and left linguistic descendants in the Finno-Ugric languages Finnish and Estonian\(^44\)\(^,\)\(^45\). A potential future direction of research is the identification of the proximate population that contributed to the arrival of this eastern ancestry into Northern Europe.

**Methods**

**Sampling and DNA extraction.** DNA was extracted\(^46\) from a total of 81 ancient human samples (teeth and bones) from the Eastern Baltic region, ranging from the Mesolithic Kunda culture to the Late Bronze Age (Supplementary Table 8). From Scandinavia (Sweden), we sampled 22 human remains from Mesolithic, early TRB and LN contexts. Two samples from north-western Russia were associated with Mesolithic contexts. The samples and their archaeological context are described in Supplementary Note 1 and presented in a tabular overview with sequencing results in Supplementary Data 1.

Sampling was performed in the cleanroom facilities at the Institute for Archeological Sciences in Tübingen for the Eastern Baltic material, at the Australian Centre for Ancient DNA for the Popovo sample, at the cleanroom facilities of the Max Planck Institute for the Science of Human History, Jena, for the samples from Olsund and Usthi/Yuzhny Oleni Ostrow, and in the ancient DNA laboratory of the Archaeological Research Laboratory, Stockholm, for the remaining Swedish material. The human remains were treated with ultraviolet (UV) light from all sides for 10 min to reduce surface DNA contamination. Teeth were sawed transversally at the border of root and crown before sampling dentine to position the sample inside of the cusp with a sterile dentistry drill. Bone powder was taken from the inner parts of the bones with a sterile dentistry drill after removing the surface layer of the bone.

Between 30 and 200 mg of powder was used for each DNA extraction (Supplementary Data 1, column M1). The extraction was performed following a silica-column-based protocol optimized for the recovery of small ancient DNA molecules\(^46\) with use of the High Pure Viral Nucleic Acid Large Volume Kit (Roche). Extraction buffer (0.45 M EDTA, pH 8.0 (Life Technologies), 0.25 mg/ml Proteinase K (Sigma-Aldrich)) was added to the bone powder aliquot, and rotated overnight at 37°C. The powder was then pelleted by centrifugation at 14,000 rpm. The supernatant was added to a protein binding buffer (5 M GuHCl (Sigma-Aldrich), 40% Isopropanol (Merck)) with 400 µg sodium acetate, pH 5.5 (Sigma-Aldrich) and mixed. The mixture was then transferred to the High Pure Extender Assembly funnel with a purification column attached and contained in a 50-ml Falcon tube. The tube was then spun at 1500 rpm for at least 8 min with slow acceleration until the powder had mostly passed the purification column. The column was transferred into a new collection tube and the liquid remaining in the funnel was transferred to the column that was then centrifuged at 14,000 rpm. This was followed by a wash step consisting of adding 450 µl of wash buffer (supplied with the High Pure Viral Nucleic Acid Large Volume Kit) to the column and spinning it for 1 min. The column was then washed with 100 µl of wash buffer (2×) and 14,000 rpm for 1 min. The DNA was eluted in a fresh siliconized Eppendorf tube in two elution steps of 50 µl TET (1 mM EDTA, 10 mM Tris-HCl, pH 8.0 (AppliChem), 0.05% Tween-20 (Sigma-Aldrich)) centrifuged for 1 min at 14,000 rpm, resulting in 100 µl of DNA extract for each sample. Negative controls were taken along for each extraction setup.

**Library preparation and targeted enrichment of human mtDNA.** Double-stranded next-generation sequencing libraries were prepared from a 20-µl aliquot of extract following a protocol established for ancient DNA\(^46\). Negative controls were taken along for each library preparation setup. First, a blunt-ending step was performed by adding the template to a mix of 1× NEB buffer 2 (NEB), 100 µM dNTP mix (Thermo Scientific), 0.8 µg/ml BSA (NEB), 0.4 U/µl T4 Polynucleotide Kinase (NEB), 0.024 U/µl T4 Polynucleotide Kinase (NEB) and incubating at 15°C for 15 min. Then, for a second 15 min at 25°C, followed by purification with the MinElute kit (QIAGEN) and elution in 18 µl of TET. The following adapter ligation was performed by adding 1× Quick Ligase Buffer (NEB), 250 nM Illumina Adapters (Sigma-Aldrich) and 0.125 U/µl Quick Ligase (NEB) for a final reaction volume of 40 µl. The mix was incubated for 20 min at room temperature after which another MinElute purification was performed and the DNA is eluted in 20 µl TET. The following fill-in step consisted of adding the 20 µl DNA to 1× Isothermal Buffer (NEB), 125 nM dNTP mix (Thermo Scientific) and 0.4 U/µl Bst

---

**References**

1. **Nature Communications** | DOI: 10.1038/s41467-018-02825-9 | www.nature.com/naturecommunications
Polymerase 2.0 (NEB) for a final reaction volume of 40. The mix was incubated for 20 min at 37°C and for 20 min at 80°C. After the fill-in step, libraries were quantified using a Qubit fluorospectrophotometer. Some DNA extracts showed evidence of inhibition of enzymatic reactions, possibly due to the presence of humic acids or chemicals (glue or hardener) used for bone treatment. To overcome inhibition, library preparation was repeated for samples that had a low DNA yield after initial library preparation or had abnormal extracts (e.g., dark colouring, floating white powder, etc.) using 10-fold less extract as template to dilute the potential inhibiting factors.

Libraries were then barcoded in a PCR-reaction using primers containing sample-specific index sequence combinations and limiting the amount of template molecules to 2 × 10^10 per reaction (0.2 mM of the library-specific P5 and P7 primers, 1× Buffer Pfu Turbo (Agilent), 0.25 mM dNTP mix, 0.3 mg/ml BSA, 0.025 U/µl Pfu Turbo (Agilent) for a total reaction volume of 100 µl). The amplification took place in a modern DNA lab with an initial denaturation of 2 min at 95°C, then 10 cycles of: 30 s at 95°C, 30 s at 58°C, 1 min at 72°C; followed by elongation at 10 min for 72°C. Libraries were enriched for human mitochondrial DNA using a bead-based hybridization protocol, pooling at most five different sample libraries into one capture pool.

**Sequencing for screening.** Libraries and mtDNA-enriched library pools were quantified on an Agilent 2100 Bioanalyzer DNA 1000 chip and pooled at equimolar concentrations. Libraries not enriched for human mtDNA were shotgun sequenced to determine the percentage of endogenous human DNA in every DNA library (Supplementary Methods). Libraries enriched for mtDNA were sequenced separately to allow for reconstruction of the mitochondrial genome of each individual and estimation of modern mitochondrial contamination. Library pools were sequenced according to the manufacturer's protocols on an Illumina HiSeq2500 at the department of Medical Genetics at the University of Tübingen for 2 × 100 cycles to a depth of ~1.5 million reads per sample.

**Data processing for sequencing.** After demultiplexing, resulting sequencing reads were processed using a computational pipeline developed for aDNA that merges paired-end reads (default parameters) and mapping of reads against a user-specified reference genome. Between 326 and 10,039,616 shotgun sequenced reads (Supplementary Data 1, column N) went into mapping with BWA (v0.6.1) against the human genome reference GRCh37/hg19. For mtDNA capture, data between 375 and 7,454,704 reads (Supplementary Data 1, column Q) went into mapping against the human mtDNA reference rCRS (using the circular mapper implemented in the pipeline). The low number of reads mapping for Spiginas3 and Motala213 indicated a failure of reagents during library preparation. The proportion of endogenous human DNA in shotgun sequencing ranged from 0.000% to 59.6% (Supplementary Data 1, column O). Genetic sex could be confidently determined for 55 individuals (Supplementary Data 1, column U).

The mtDNA reconstruction and contamination estimation was performed by an iterative likelihood-based approach, taking into account that the contig from mtDNA sequence should be reconstructed from molecules that originate from a single individual and that show characteristics of DNA. Complete mitochondrial genomes (covered at least 85%) could be reconstructed for 61 individuals and less than 3% mitochondrial contamination (Supplementary Data 1, column S). For these, the percentage of deamination at the molecule ends exceeded 20%. Libraries were then barcoded in a PCR-reaction using primers containing sample-specific index sequence combinations, amplified using a computational pipeline developed for aDNA that merges paired-end reads with same orientation and start and end positions. To avoid an excess of remaining C-to-T and G-to-A transitions at the ends of the reads, two bases of the ends of each read were clipped for each sample except for the non-UDG-treated data for Gyvakarai1, where 10 bases from each end were clipped. For each of the targeted 1240k SNP positions, a read was chosen at random to represent this position using the genotype calls from the 1240k SNP dataset (Supplementary Data 1).

**Nuclear capture and sequencing for genome-wide data.** Forty-one samples (including two previously studied north-western Russian samples) were chosen for nuclear capture, a hybrid shotgun sequencing approach exemplified by the 1240k SNP capture (UDG-hal) libraries that were prepared out of the DNA extracts of these samples by adding the extract to a reaction of total volume 60 µl with 1× Buffer Tango (Thermo Fisher Scientific), 100 µM dNTPs, 1 mM ATP and 0.06 U/µl USER enzyme (NEB) and incubating for 30 min at 37°C. The reaction was then inhibited by adding 0.12 U/µl UGI (NEB). These libraries were then barcoded with sample-specific index sequence combinations, subsequently amplified with Hercule II Fusion (Agilent) and enriched using an in-solution hybridization protocol for a target set of ~12 million nuclear SNPs (1240k SNP set). The enriched libraries from the Eastern Baltic and Swedish samples were paired-end sequenced on a NextSeq500 at the department of Medical Genetics at the University of Tübingen using 2 × 75 bp cycles and a HiSeq4000 at the Max Planck Institute for the Science of Human History in Jena. The UDG-treated library of UCSC was processed at the Max Planck Institute for the Science of Human History in Jena, Germany, and was sequenced there on a HiSeq4000 for 2 × 75 cycles, and the UDG-treated library for Popovo2 was processed at Harvard Medical School, Boston, USA, and sequenced here on a NextSeq500 for 2 × 75 cycles. Additionally, the non-UDG-treated screening library of Gyvakarai1 was paired-end sequenced on two lanes of a HiSeq4000 for 2 × 75 cycles, and on a full run of a NextSeq500 for 2 × 75 cycles. The screening library for Kunitza2 was paired-end sequenced deeper on 80% of the lanes of a HiSeq4000 for 2 × 75 cycles. The three UDG-half libraries for Olsund were single-end sequenced on a HiSeq4000 for 75 cycles. Sequencing strategies and facilities are summarized in Supplementary Data 1, column AE.

After demultiplexing, resulting sequence data were further processed using EAGER. This included mapping with BWA (v0.6.1) against UCSC genome browser human genome reference UCSC (hg19). Mapping heterozygosity on the mtDNA with schmutz56 and heterozygosity on the X chromosome in male samples with ANGSD (Supplementary Data 1), and evaluating the ratio of the reads mapping to X and Y which showed no outliers (see below).

We observe that all our individuals predating the LN appear genetically distinct from any modern-day population that could have contaminated them, and that female samples cluster together with their male counterparts from the same archaeological cultures (Fig. 2), which gives indirect support to the authenticity of our data.

We excluded Saxtorryp5158 from our analysis due to its high degree of contamination on the mtDNA, and Saxtorryp389 as it showed unusual ancestry for its dating and archaeological context, consistent with modern European contamination.

Using the software READ56, we determined individuals Kretoumza2 and Kretoumza5 to be identical twins, consistent with a value of >0.5 for (Kretoumza2, Kretoumza5). Mbutii. We do not include the lower coverage sample Kretoumza5 in ADMIXTURE analysis and other analyses that cluster the individuals into one population, thereby mitigating bias resulting from a defined population consisting of closely related individuals.

We merged our data of Kunitza2 with previously published data of the same individual57 after confirming the identity with outgroup p and READ.

**Sex assignment.** Genetic sex of the 41 selected samples was assigned using shotgun and SNP capture data by calculating the ratio of average X chromosomal and Y chromosomal coverage to average autosomal coverage at the targeted SNPs (X and Y rate, respectively). Samples with an X rate between 0.65 and 1 and a Y rate between 0 and 0.15 were assigned female and those with an X rate between 0.35 and 0.55 and a Y rate between 0.4 and 0.7 were assigned male, supporting the informative value of the Y rate over the X rate using this method (Supplementary Method, Fig. 9), as demonstrated previously.55

**Population genetic analyses.** Due to the nature of ancient DNA research, no pre-determination of sample size by statistical methods was carried out and there was no randomization of experiments or blinding of investigators to allocation during experiments and data analysis.

Reference datasets for ancient populations are taken from the publicly available dataset used in refs. 5, 6, 21, 22 which includes genotypes from samples published earlier in refs. 5, 6, 21, 22, among others, as well as genotyping data from worldwide modern populations (Human Origins or HO dataset) published in the same publications and provided by the David Reich Lab. When analyzing only ancient samples, we make use of the 1,196,358 SNPs targeted by the 1240k SNP capture, using the genotypes of deep shotgun sequenced modern Mbuti as the outgroup. For analyses involving modern populations, we restrict to the intersection of 397,503 SNPs between the 1240k SNP set and the HO dataset.

PCA was performed with smartpca in the EIGENSOFT package by constructing the eigenvectors from modern West Eurasian populations (Supplementary Fig. 7) and projecting ancient individuals on these eigenvectors (Fig. 2a, Supplementary Fig. 8).

Admixture analysis (Fig. 2b) was carried out with ADMIXTURE on 3784 markers per sample and 38 ancient individual for ancestral clusters k = 2 to k = 16 with 100 bootstrap replicates. The SNP dataset was pruned for linkage disequilibrium with PLINK using the parameters –indep-pairwise 200 5 0.5. We considered the cross-
validation (CV) error and report in Fig. 2b and Supplementary Fig. 4 the results of $k = 11$, where the CV error levels out at a minimum.

To quantify population affinities and admixtures suggested in the PCA and ADMIXTURE analysis, we carried out f-statistics using the programs qp3Pop and qpDist in the ADMIXTOOLS suite (https://github.com/DReichLab) for $f_1$- and $f_2$-statistics, respectively. $f_1$-statistics of the form $f_{XY} (0; Outgroup)$ measure the amount of shared genetic drift of populations $X$ and $Y$ after their divergence from an outgroup. Admixture $f_2$-statistics of the form $f_{STestXY}$ indicate when significantly negative that population Test is intermediate in allele frequencies between populations X and Y and could be considered an admixed population. This test was performed with parameter mbfreducing=1E5 and cannot be done for populations with less than two individuals. D-statistics of the form DXXY, Test, Outgroup show if population Test is symmetrically related to X and Y or shares an excess of alleles with either of the two. Results are only reported for statistics based on more than 10,000 SNPs.

To formally test for the number of source populations and the proportion of ancestry these contributed to our studied populations, we used the qpWave and qpAdm programs from ADMIXTOOLS. These programs implement the methodology of using regression of $f_2$-statistics of a Reference population with various outgroups to relate its ancestry to a Test population.6–8 With qpWave, we identified potential source populations for our population under study by testing if a set of Left populations (the Test population under study and its potential proximate source Reference populations) is consistent with being descended from $n$ admixtures of which have unique relationships to the Right outgroup populations (Mbuti, Papuan, Onge, Han, Karitiana, Mto, Ust Ishim, MA1, Villabruna). This is given when rank $n - 1$ cannot be rejected ($p > 0.05$), and rejected (i.e. more than $n$ waves of admixture are needed to explain the deviation from Test and Reference) if rank $n - 1$ can be rejected ($p < 0.05$).

To estimate admixture proportions, we used qpAdm to model the Test population as a mixture of various source populations postulated from the qpWave test, setting as Left populations the Test and source populations and as the Right populations the various outgroups named above.

Data availability. The sequence data reported in this paper are deposited in the Sequence Read Archive (Accession numbers: SAMN08139261-SAMN08139301) and complete mitochondrial consensus sequences are deposited in GenBank (Accession numbers MG428993-MG429049).
Acknowledgements

We thank Isil Kucukkalipci, Cäcilia Freund, Antje Wissgott, Guido Brandt, Nadin Rohland and Swapan Mallick for technical support in DNA analyses, and Herve Bocherens, Vesa Palonen, Anne-Maija Forss and Igor Shevchuk for arranging sample treatment and technical support for radiocarbon analyses. We thank Ben Krause-Kyora for use of the sequencing facilities at the IKMB (Kiel). We thank Henny Piezonka, Cosimo Posth and Iosif Lazaridis for helpful discussion and suggestions. F.H. thanks Maj-Lis Nilsson and Erika Rosengren for support regarding Saxtorp site, and Elise Hoyanta, Anna Larssdotter and Kristina Lindkvist for help regarding the Ohlsund site.

C.-C.W was supported by the Max Planck Society and Nanqiang Outstanding Young Talents Program of Xiamen University. J.K. and A.M. were funded by DFG grant KR 4015/1-1 and the Max Planck Society. E.B. was funded by the RFBR grant 16-06-00303. D.R. is supported by US National Science Foundation HOMINID grant BCS-1032255, US National Institutes of Health grant GM100233, an Allen Discovery Center grant from the Paul Allen Foundation, and is an investigator of the Howard Hughes Medical Institute.

Author contributions


Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-02825-9.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018