Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Maternal contacts across the Baltic Sea – thousands of mitogenomes reveal maternal genetic structure and recent expansions in the Circum-Baltic region

Hovhannes Sahakyan
University of Tartu
Maere Reidla
University of Tartu
Ene Metspalu
University of Tartu
Ivan A. Kuznetsov
University of Tartu
Rodrigo Flores
University of Tartu
Kadri Irdt Rezzakioğlu
University of Tartu
Olga Utevska
University of Tartu
Ajai K. Pathak
University of Tartu
Raimonds Rescenko-Krums
Latvian Biomedical Research and Study Centre
Vita Rovite
Latvian Biomedical Research and Study Centre
Janis Klovins
Latvian Biomedical Research and Study Centre
Adam Ameur
Uppsala University
Ulf Gyllensten
Uppsala University
Estonian Biobank Research Team
University of Tartu

Anne-Mai Ilumäe

ami@ut.ee

University of Tartu

Research Article

Keywords:

Posted Date: July 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10447855/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Present-day European mitochondrial diversity has been interpreted as relatively homogenous, reflecting Mesolithic postglacial expansions from glacial refugia. In contrast, Y-chromosomal diversity shows evidence of strong Neolithic bottlenecks and Bronze Age expansions involving a limited number of successful patrilineages. However, fine-scale maternal demographic patterns have remained unresolved due to small sample sizes and insufficient phylogenetic resolution.

Results

We assembled a large dataset of more than 9,000 complete mitogenomes from six Circum-Baltic populations and integrated ~ 5,000 additional phylogenetically related mitogenomes from published sources. Using these sequences, we reconstructed phylogenetic trees, identified high-frequency terminal branches and estimated their coalescence ages in three Circum-Baltic populations. We identified numerous rare clades that contributed to the rapid increase in maternal effective population size over the past 2,500 years. Temporal changes in maternal effective population size revealed distinct demographic trajectories across populations, with Sweden showing a pronounced decline during the Bronze Age, followed by a steep rise in the late Iron Age. We uncovered elevated sharing of matrilineages between Estonia and Finland, pointing to a maternal genetic component underlying their linguistic affinity that had previously been attributed primarily to male lineages. Phylogenetic links in several subclades suggest long-distance east–west interactions across the East European Forest Belt, involving ancient populations in the forest-steppe contact zones near the Ural Mountains.

Conclusions

Maternal diversity in the Circum-Baltic region was substantially shaped by Bronze and Iron Age demographic processes. Large population-based mitogenome datasets are essential to resolve fine-scale maternal demographic structure and distinguish population-specific histories.

Introduction

Contemporary human genetic diversity has been shaped by complex demographic processes unfolding over millennia. In Europe, autosomal genetic variation has been shown to correlate closely with geography[1, 2], while ancient DNA (aDNA) studies trace the ancestry of modern Europeans to three major sources: Mesolithic hunter-gatherers, Neolithic farmers originating from Anatolia, and early Bronze Age Western Steppe Pastoralists[3, 4]. Studies of uniparental markers in European populations have reported male-biased demographic processes underlying present-day patterns of genetic diversity, including clinal variation in Y-chromosomal lineage frequencies[5], bottlenecks associated with the

Neolithic transition[6], and Bronze Age expansions driven by a limited number of successful patrilineages[7, 8]. In contrast, European mitochondrial DNA (mtDNA) diversity has generally been regarded as relatively homogeneous, primarily reflecting postglacial expansions from refugia. Previous studies of maternal diversity have been constrained by small sample sizes per population[6, 8, 9], reliance on a limited number of polymorphic sites[10, 11], samples restricted to a single population[12] or demographic reconstructions centred on selected haplogroups[13–15]. Taken together, although the broad patterns of mtDNA diversity in Europe are well documented, its fine-scale maternal demographic history has remained largely unresolved.

Situated at the intersection of Northern and Eastern Europe, the Circum-Baltic (Fig. 1A) region represents a meeting ground for two major language families and is widely regarded as an ideal setting for the study of linguistic contact phenomena[16]. Genetic studies, coupled with archaeology, have shown that major cultural transitions – including the arrival of the first hunter-gatherers, the subsequent spread of the first agriculturalists together with steppe-derived groups and later interactions during the Nordic Bronze Age (NBA) and Iron Age (IA) – have left detectable and contrasting genetic signatures in Circum-Baltic (CB) populations, some of which remain discernible today[17–22]. Additionally, a distinctive male-specific feature in the eastern Baltic region is the presence of Y-chromosomal haplogroup (hg) N subclades, which display a declining frequency cline toward Scandinavia and Central Europe and lack any known counterpart among maternal lineages[23].

The overarching goal of this study was to examine the maternal demographic history of the CB region at a large scale. We assessed the effect of sample size on the reconstruction of maternal effective population size (Ne) dynamics. We then resolved the previously uncharacterised internal structure of the regionally common major maternal haplogroups (hgs) by reconstructing phylogenetic trees based on over 9,000 complete mtDNA genomes from populations across the CB region, together with more than 5,000 additional phylogenetically related mitogenomes irrespective of population affiliation. Specifically, we aimed to identify phylogenetically terminal clades that are frequent in populations with contrasting linguistic histories — Finnic-speaking Estonians and Finns and Germanic-speaking Swedes — using geographically neighbouring Baltic Sea populations as phylogenetic context. These terminal clades represent the most recent branches in the phylogeny of the uniparentally inherited parts of the genome and provide insight into localised, sex-specific demographic processes that have shaped the present-day gene pool, including patterns not readily captured by autosomal data[25]. This combination of extensive sampling and fine-scale phylogenetic resolution enabled the estimation of the coalescence ages of the identified terminal clades, allowing us to infer the timing of local maternal expansion events and examine the temporal formation of mtDNA diversity shared among and specific to CB populations. Finally, we integrated published ancient mitogenomes with the reconstructed high-resolution mtDNA trees to gain insight into possible directions of the maternal gene flow.

Results

Frequency of haplogroups and summary statistics in studied populations

The maternal genepool of the studied populations was mainly comprised of western Eurasian hgs. East Eurasian and Siberian hgs were detected at low frequencies, accounting for ~0.5–2% of the total variation, depending on the population. We present the hg frequencies at two phylogenetic resolutions: the phylogenetically deeper level of general hgs and subhaplogroups (subhgs), as defined by PhyloTree mtDNA tree Build 17 (Additional file 1: Table S1A). The most frequent hg H, which describes 37-46% individuals in all studied populations, was also the most diverse, containing 119-324 subhgs (45 in Latvians due to ~5x smaller sample size) at the subhgs level (Additional file 1: Table S1B). The second most frequent hg, U5, contained 19-80 subhgs.

Summary statistics and parameters of genetic diversity based on full mitogenomes are presented in Table S2 (Additional file 2: Table S2). In the Estonian dataset, the ratio of non-synonymous and synonymous mutations of terminal branches (0.58) was significantly higher than that of internal branches (0.47) (Fisher's two-tailed exact test with p=0.026). This observation suggests that purifying selection acted against non-synonymous mutations, consistent with previous studies[26] and stressing the importance of age-correction of the initial coalescence time estimates, as proposed by Soares et al[27]. The Polish population is the most diverse among the sampled populations, as exemplified by the high variety of discovered subhgs, whereas Estonian and Finnish populations were the least diverse. The large sample size of Estonians yielded the fewest unique haplotypes, with only ~57% of haplotypes observed once, in contrast to ~91% of unique haplotypes in the most diverse Polish population. In the case of Estonia, large sample size and a smaller population census are expected to yield more matching haplotypes. More specifically, we found three haplotypes carried by over 20 individuals and altogether over 370 haplotypes shared by at least two individuals in the Estonian population. With at least an order of magnitude larger sample size than Stoljarova et al.[28], the random match probability dropped from 1.2% to 0.18% in the Estonian population. In the Polish population, it is as low as 0.07%, which is substantially lower than the 1.06% previously reported by Malyarchuk et al.[29].

Regionally shared maternal gene pool

We characterised the regional maternal diversity that was either shared among two or three studied CB populations or restricted to a single focal population – Estonia, Finland, or Sweden. We detected a total of 86 shared clades, of which 47 are new compared to PhyloTree version 17 (Fig. 1B), with median coalescence time estimates from the Neolithic to the IA. Additionally, we identified 88 one-population clades (a total of 77 of them novel), several of which represent subclades within phylogenetically ancestral shared clades. Given the large number of mtDNA terminal clades, none can occur at high frequency and their frequencies fluctuate across the examined populations. The exact number of individuals in each identified clade, along with median coalescence age estimates and 95% highest posterior density (HPD) intervals, is presented in Table S3 (Additional file 2: Table S3).

Comparison of effective population size dynamics

In sample subsets of 100 individuals, maternal Ne first diversified around 50,000 BP in all populations, whereas larger sample sizes tend to yield earlier estimates of initial diversification (Additional file 3: Fig.

S2). The earliest diversification, starting at ~80,000 BP, is observed in the Polish dataset (Additional file 3: Fig. S2C), which also displays the highest Ne among our sampled populations (Fig. 2). It is likely that these earliest diversification times – coinciding with or predating the arrival of modern humans in Europe – reflect the ancient split between macrohaplogroups M and N[24]. In all populations, the first pronounced maternal Ne growth occurred during the Paleolithic, starting around ~20,000 BP (Additional file 3: Fig. S2). This expansion took place outside Fennoscandia and the eastern Baltic area, as archaeological evidence of the earliest human presence in North Europe dates to ~10,000 BP[30].

Age distribution of the regionally shared clades and signs of post-Neolithic expansion

To estimate the historical timeframe of the expansion of the most frequent regional clades, we assigned each clade to a chronological period based on its median coalescence age, calculated using two different mutation rates [27,32] and their arithmetic mean as a consensus estimate (Additional file 2: Table S3). Coalescence age estimates obtained using the approach of Rieux et al.[32] were consistently older, as this method lacks the correction for purifying selection proposed by Soares et al.[27]

To evaluate the relationship between clade age and the number of present-day carriers, we generated a scatterplot using the arithmetic mean of clade ages and the total number of extant individuals (Fig. 3). The analysis revealed a moderate positive correlation between the age of the clade and the number of individuals it characterises (Pearson $r>0.51$, Spearman $rho>0.65$, $p < 0.05$, exact values provided in Additional file 2: Table S5). The findings support the hypothesis that older clades have had more time to expand and disperse. However, the small number of one-population clades primarily found in Finland and Sweden reduces statistical power and renders the correlation insignificant (Additional file 2: Table S5).

Using the arithmetic mean as the consensus estimate, the median age of ~51% of all dated clades fell within the IA period, whereas ~38% corresponded to the NBA (Additional file 2: Table S3). The Neolithic period accounted for the smallest proportion of coalescing clades (~11%). The BSAs of clades shared among Estonians, Finnish and Swedish indicate growth in Ne during the early IA, around ~2,000 BP (Additional file 3: Fig. S3). This increase appears earlier and is particularly steep in the Finnish population. A similar pattern was observed for clades shared between Estonia and Finland (Additional file 3: Fig. S4). The proliferation of low-frequency clades with coalescence ages dating to the IA contributed to the late increase in maternal Ne. Extensive sampling is necessary to fully capture their substantial contribution to contemporary maternal genetic diversity.

About 65% of one-population clades coalesced in the IA and ~31% in the NBA (Additional file 2: Table S3). In the case of Swedish one-population clades, all of them began to expand in the IA after ~2,000 BP. A total of 19 one-population clades (11% of all examined clades) with coalescence estimates in the NBA or IA represent subclades of older, phylogenetically higher-level clades that are shared among studied populations. Such clades likely arose in their currently most frequent populations, but have not had sufficient time to disperse geographically. Alternatively, they may occur at low frequencies in

undersampled neighbouring populations, indicating maternal gene flow between geographically proximate populations.

Regarding the total number of individuals, clades with mean coalescence ages dating to the NBA or IA accounted for more than 30% of individuals in both Estonia and Finland, whereas Neolithic-dated clades comprised approximately 9% and 11% of Estonians and Finns, respectively (Additional file 2: Table S6). The Swedish population shared a significantly lower proportion of maternal diversity with the studied CB populations. Nevertheless, post- Neolithic clades accounted for a higher number of Swedish individuals (~10% of the total Swedish sample) than clades with estimated ages in the Neolithic (~1.4% of the total Swedish sample). When only individuals belonging to the examined clades were considered, clades with estimated coalescence ages dating to the NBA or IA accounted for more than 78% of individuals in Sweden, Estonia, and Finland, and more than 57% in Poland and Latvia (Additional file 2: Table S6). This pattern identifies the post-Neolithic era as a major temporal phase from which the most frequent CB terminal maternal lineages trace their expansion.

Maternal legacies of hunter-gatherers and the Corded Ware Culture

The two oldest examined clades, each with a mean estimated coalescence age of ~5,300 BP, belong to hg U5b1 (Fig. 4, Additional file 2: Table S3). Subhg U5b1b2 is one of the most frequent terminal clades and occurs primarily in Finland and Estonia (a total of 50 and 24 individuals, respectively). The oldest published carriers of subhg U5b1b2 were found in hunter-gatherer remains dated to ~5,500 BP from present-day Denmark[19]. Subhg U5b1b1f is much rarer and is mainly shared between Estonia and its southern geographic neighbours Latvia and Poland (Fig. 4). A hunter-gatherer dated to ~7,000 BP and carrying the phylogenetically ancestral U5b1b1-16,192! haplotype was found in modern-day Poland[33] (Fig.4, see Additional file 4: Fig. S5 for the tree topology of hg U5b).

Subclades of varying coalescent ages within haplogroup U5b1b1a are present across all studied populations except the Danish, but are particularly frequent in Finland, Estonia, and Sweden (Fig. 4, Additional file 1: Table S1B). Their detailed inner diversification is shown in the maximum parsimony tree of hg U5 in Fig. S5 (Additional file 4: Fig. S5). These subclades also account for approximately 48% of the Saami maternal gene pool[34]. The earliest published evidence of the U5b1b1a haplotype derive from a ~6,300-year-old individual from Denmark[19] and an individual dated to ~6,100 BP from the Late Mesolithic Narva culture in Lithuania[18] (Fig. 4). The phylogenetically ancestral haplotype U5b1b has been identified in late NBA Estonia[35], but limited aDNA quality prevents resolution of a specific subclade. Overall, hg U5b has become the hallmark of Western hunter-gatherers (WHG). The oldest published carriers of hg U5b1 have been reported from Mesolithic Iberia[36] and their subsequent spread is a likely indicator of northward migrations from the postglacial refugia of southwestern Europe. Subhgs U5b1b1 and U5b1b2 have persisted in modern populations of the eastern Baltic and Fennoscandia and are considerably less frequent in Denmark, echoing the elevated WHG ancestry detected in the autosomal genetic profiles of the former populations[3].

In the Baltics and eastern Fennoscandia, the Corded Ware Culture (CWC) emerged almost simultaneously around 4,900 BP[37], marking the beginnings of animal husbandry and the first attempts at agriculture[38]. In genetic terms, the CWC is generally characterised by admixture between Early European Farmers (EEF) and Yamnaya pastoralists[4]. Several of the most frequent maternal lineages, shared among the studied populations and inferred to have begun expanding during the CWC period, trace their phylogenetic ancestry to Anatolian famers and the EEF. In present-day Estonia and Latvia, the most prolific lineage with its coalescence age falling within the CWC period is hg H5a1a-16483. The oldest aDNA links the phylogenetically ancestral hg H5a to the Neolithic Balkans and the agricultural Cucuteni-Trypillia culture of Southeast Europe[39]. By ~5,300 BP, hg H5a1 appears in the genetically EEF-like Globular Amphora culture in Poland[40], and by ~4,400 BP, in CWC individuals in present-day Estonia[35]. Hg H5a1, or its subclades, which have remained unresolved due to limited aDNA sequence quality, subsequently expanded eastward and appear from the Bronze Age onward in archaeological cultures associated with the descendants of the CWC in the Southern Ural region[41], Western Siberia[42] and Kazakhstan[43].

Within the most frequent West Eurasian haplogroup H, the two most numerous H1 subclades in our dataset are H1f1 and H1n4, both of which show inferred expansion times during the CWC period or the early NBA (the maximum-parsimony tree for haplogroup H1 is presented in Additional File 4: Fig. S6). Both contain local subclades suggestive of longer presence in the region (Fig. 5, Additional file 5: Fig. S7). Hg H1 appears in Central Europe with the first farmers and is considered a sign of demic diffusion of agricultural practices across Neolithic Europe[44,45]. Hg H1n4 lacks an extensive aDNA record, with the only ancient carrier reported from 18th century Finland[12] (Additional file 5: Fig. S7). The earliest published evidence of hg H1f1 dates to the CWC period in southern Estonia[17] and the IA in Poland[46]. Hg H1f1, with an estimated mean coalescence age of ~3,700 BP, is present at low frequencies in all examined populations except the Danish; however, it is the most frequent H1 subclade in modern Finland (~5%, compared with <1% in all other populations), where it has diversified into two predominantly Finnish subclades dated to the late NBA (Fig. 5). It is likely that carriers of clade H1f1 arrived in northeastern Europe with the CWC, with subsequent genetic drift and population growth leading to elevated frequency in Finland.

Among the branches with the oldest age estimates, clade W6a-4164 can be associated with the Western Steppe Herders. Its mean coalescence age is ~4,900 BP, corresponding to the early CWC period. It is one of the most prolific shared clades between Estonia, Latvia, and Poland, while being nearly absent in Fennoscandian populations. The ancestral haplotype, hg W6, has been found in the steppe-associated Yamnaya and Fatyanovo cultures[4,47]. The oldest published aDNA evidence of hg W6a comes from an Eneolithic hunter-gatherer from the Upper Don region of present-day Russia, dated to ~5,500 BP [48]. Subsequently, individuals carrying haplogroup W6a and associated with the CWC were identified in both Central Europe[49] and the Baltic region[18]. In the latter case, the CWC individual from present-day Lithuania displayed elevated Yamnaya-related autosomal ancestry and lacked evidence of genetic affinity with EEF[18]. Taken together, this lineage provides evidence for a maternal genetic contribution from populations migrating from the Pontic–Caspian Steppe during the CWC expansion.

Clade U5a2a1-13563 is the oldest one-population branch and is particularly prolific in Estonia (~2%, absent elsewhere), with a mean coalescence age in the CWC era (Additional file 2: Table S3). This clade belongs to the widely distributed hg U5a2a1, with numerous less frequent terminal clades observed across the studied populations (Fig. 6). The phylogenetically upstream hg U5a2a has been identified in Mesolithic hunter-gatherers from Serbia, Ukraine[39,50] and Iberia[36]. The earliest published carriers of the ancestral U5a2a1 haplotype are individuals from Central Europe associated with the CWC[49]. Subsequently, this haplotype has been found in several individuals across the Baltic region and in Scythian individuals from various sites across the Pontic–Caspian Steppe, all dating to different periods of the NBA (Fig. 6). In modern populations, several deep-rooting and relatively frequent hg U5a2a1 subclades occur in Estonia, Finland, and, to a lesser extent, Poland, but are rare in Sweden. This distribution makes a Scandinavian origin for these subclades less likely and instead suggests continuity and expansion within northeastern European populations (Fig. 6). The presence of numerous locally expanding terminal clades within hg U5a2a1 may indicate multiple episodes of arrival and suggests that this haplogroup has had sufficient time to diversify and expand in northeastern Europe. Additionally, the regional demographic history of hg U5a2a1 differs from that of its sister hg U5a1, whose regional subclades tend to be less frequent and display later coalescence ages (Additional file 2: Table S3).

In summary, maternal lineages with estimated coalescence ages corresponding to the CWC period and with phylogenetic precursors among the EEF appear to be more prevalent in modern Estonians and Finns than those linked to the earliest Western Steppe Herders. This observation supports previous findings from the Baltic region showing that the CWC introduced Steppe ancestry together with the female-biased EEF contribution[18,35], a signal that remains visible in the maternal ancestry composition to the present day. However, some lineages provide evidence for maternal genetic contributions from the Steppe populations during the CWC period.

Maternal hallmarks of the Nordic Bronze and Iron Age

The most prolific maternal subclades with estimated coalescence times in the NBA belong to hgs U5b1b1a1a and U5b1b2a. Both derive from the regionally common clades U5b1b1 and U5b1b2, which were discussed above in connection with the WHG ancestry (Fig. 4). Each further branches into younger monophyletic lineages that underwent expansion during the IA and both are present in at least one of the examined populations (Additional file 2: Table S3). The ancestral U5b1b1a haplotype has been identified in ancient remains from a stone-cist grave in northern Estonia, dated to the late NBA[35](Fig. 4). Stone-cist burials appeared in Estonia and Finland during the NBA and are considered to reflect Scandinavian influence, possibly indicating an influx of newcomers[51]. Together, these observations suggest possible links between the expansion of several U5b1b1a1 subclades during the NBA/early IA (Additional file 2: Table S3) and incomers from Scandinavia. This interpretation is further supported by the elevated WHG and Scandinavian hunter-gatherer autosomal ancestry detected in individuals buried in the stone-cist graves[52]. Consequently, the diversification and expansion of these U5b1b-derived maternal lineages could reflect demographic processes associated with Scandinavian contacts and migration during the late NBA.

Other relatively prolific clades with mean coalescence ages falling within the NBA include V1a1a, H13a1a1d and several subclades of H2a1. Clade V1a1a is one of the most common shared maternal lineages between Estonia and Finland. Its expansion began during the early NBA and resulted in a predominantly Finnish subclade that underwent substantial radiation during the Iron Age (Additional file 2: Table S3). The subclade is also present in a modern Tatar individual from the Volga-Ural region, representing the only sample outside Finland and Estonia in the dataset. Although Neolithic cultures facilitated the spread of haplogroup V in Europe[4], subclade V1a shows evidence of an early eastward migration, appearing in the Bronze Age Afanasievo culture of South Siberia[41]. The distribution of hg V1a therefore suggests a complex demographic history involving both western and eastern Eurasian populations. Its diversification and relatively high frequency in northeastern Europe may reflect local expansion following a possible westward back-migration of some V1a1 subclades. However, the limited availability of full mitogenomes from the Volga–Ural region and Siberia constrains any firm conclusions.

The oldest published ancestral haplotypes, H13a1a1 and H2a1, have both been identified in remains dated to ~7,000 BP from the Chalcolithic Khvalynsk culture of the Middle Volga region[53,54]. Both haplotypes were subsequently detected in multiple Bronze Age cultures across the Pontic–Caspian Steppe, including the Yamnaya[41,53,55–57]. The earliest published occurrence of the now primarily Estonian–Finnish subclade H13a1a1d comes from a steppe-associated ancient individual in Central Europe dating to the early NBA[58]. Other ancient carriers are from later periods – examples include an Etruscan from IA Italy[59] and a Finnish individual from the Middle Ages[60]. Among the dated H2a1 lineages in our dataset, all subclades show expansion times during the late NBA or IA (Additional file 2: Table S3). Subclade H2a1f2 has two aDNA findings linking the ancestral H2a1f haplotype to the Sarmatian culture of the Southern Ural region ~2,200 BP[61] and to an eighth-century Pannonian Avar individual from present-day Hungary with Steppe-related ancestry[62]. The subclade H1a2n is particularly frequent in the modern Swedish population but is rare in the aDNA record; it has been identified in several early Hungarian conquerors[63] with ancestry in the Pontic-Caspian Steppe and in a West Slavic individual from present-day Poland[46], all dated to ~1,000 BP.

The subclade H2a1a-152!-94 is most frequent in the modern Estonian population and has a mean coalescence age estimate in the early IA (~2,300 BP). This subclade has also been identified in an individual from a tarand grave in Ingria, western Russia, dated to the early IA[52]. *Tarand* graves represent a new type of stone burial construction that emerged along the northern coast of present-day Estonia during the transition from the late NBA to the IA. Ancient remains dated to ~2,500 BP from such graves provide the earliest published evidence of Siberian-related autosomal ancestry together with male haplogroup N[52].

In addition, several noteworthy subclades with mean estimated coalescence ages in the early and middle IA belong to hg H1a, including H1a2, H1a7, H1a8 (Fig. 7). Among these, subclade H1a2 stands out as one of the most frequent monophyletic lineage in both Estonia and Finland reaching frequencies of ~1.5% and ~2.6%, respectively. In contrast, it occurs at frequencies below 1% in the Swedish and Latvian populations and is absent from the Polish dataset. Its relatively high prevalence in Estonia and

Finland distinguishes it from other clades with comparable coalescence age estimates (Fig. 3). The elevated frequency of H1a2 in two populations may reflect gene flow from an unsampled third population, expansion in one population followed by recurrent migration into the other, or, less likely, parallel drift leading to elevated frequencies in both populations simultaneously. Subclade H1a2 is also represented by five individuals from different regions of present-day western Russia[29], suggesting possible connections to undersampled populations in the Eastern European plain.

Additional relatively frequent IA subclades include H1a7 and H1a8, both of which are most common in the Estonian and Finnish populations and occur only sporadically elsewhere (Fig. 7). The aDNA record for the ancestral haplotype H1a is sparse, with the earliest occurrence in the Middle Neolithic France[45], followed by an absence from the published record until its reappearance in IA-dated human remains from the CB region[52,64]. Notably, the earliest finding from the CB region was recorded in an individual of non-local origin, as indicated by isotopic comparisons [65], from an early *tarand*-type grave[52]. Consistent with their relatively recent expansion times, the subclades H1a2, H1a7, and H1a8 are absent from the aDNA record prior to the 11th century CE (Fig. 7). However, all three subclades show a high number of identical haplotypes in the Estonian and Finnish datasets. The most prominent example is subclade H1a2, in which 21 unrelated Estonian individuals share an identical haplotype (Additional file 4: Fig. S6). The combination of limited internal differentiation and relatively high frequency may reflect allele surfing during population expansion, whereby lineages at the expansion front rise rapidly in frequency because of stronger genetic drift and subsequently spread into neighbouring areas with minimal internal diversification[66,67].

Discussion

Using an extensive dataset of complete mitogenomes from the CB region, together with phylogenetically related samples, we reconstructed population-specific maternal Ne dynamics and resolved phylogenetic relationships at the highest possible resolution. Our analyses highlight the considerable diversity of mitochondrial phylogeny and the large number of terminal clades within major haplogroups across all examined populations, consistent with a higher Ne than that observed for the Y-chromosome[6,25,68]. The wide geographic spread of older lineages indicates historically high levels of female mobility. Our results suggest that the widely held view of present-day mtDNA variation as mainly reflecting postglacial re-expansions from glacial refugia may, at least in part, be an artefact of limited sample sizes in earlier studies. The estimated coalescence ages of numerous low-frequency terminal clades fall within the last ~2,500 years. Collectively, the large number of such clades contributes to the increase in maternal Ne observed during the IA. As dataset sizes increase, clearer population-specific patterns emerge that were previously obscured by sparse sampling and incomplete mtDNA sequences.

In the Swedish population, the first growth phase of maternal Ne coincides with the expansion of European Neolithic farming populations. This was followed by a decline during the early NBA (~3,800 BP) and a rapid increase in the late IA (Fig. 2). The latter may reflect either demographic expansion or an influx of new maternal lineages. The second interpretation is consistent with recent evidence for the

spread of IA Central European ancestry into Scandinavia [69] and suggests a maternal contribution to this migration. In contrast, the Estonian and Finnish populations display their strongest increase in maternal Ne only in the late IA, without any preceding sharp reduction in diversity. This timing is notable, as it overlaps with the inferred divergence of the western branch of Finnic languages[70] and with the earliest findings from contemporary Estonia of Siberian-related autosomal ancestry together with paternal hg N[52].

Most of the examined clades lack unequivocally identifiable phylogenetic precursors within the examined populations. Younger and rarer lineages are particularly unlikely to be represented in the aDNA record because of limited aDNA sample sizes and the relatively short time available to diversify and expand beyond their initial distribution range. The absence of phylogenetically ancestral haplotypes in the sampled populations may therefore reflect incomplete sampling of local diversity or their historical loss. However, many of these clades likely represent incoming maternal lineages from external sources. In such cases, as discussed by Olivieri et al.[71], the coalescence age should be interpreted as a minimum estimate for the presence of a clade in the region. Conversely, this coalescence estimate may overstate the timing of regional presence if a clade had already diversified elsewhere before entering the study area in an already differentiated form. The latter scenario is less likely for older and more prevalent clades, particularly those exhibiting substantial internal phylogenetic structure, which is more plausibly attributable to local demographic processes. It is possible that some lineages with younger coalescence ages entered the region earlier, during the Neolithic, and subsequently expanded in the NBA or IA. However, archaeological evidence indicates a decline in the number of discovered early NBA sites and suggests a possible reduction in population sizes in both Finland, Estonia and northern Latvia during the early NBA, followed by notable changes in material culture during the later NBA[51,72]. Additionally, recent aDNA studies testify of influx of individuals carrying novel ancestry components into Scandinavia during the early NBA and IA[64,69]. Taken together, multiple lines of evidence suggest that many of the identified subclades reflect incoming post-Neolithic migrations involving a significant female contribution.

Finnic-speaking Estonian and Finnish populations display a relatively high proportion of individuals belonging to terminal subclades shared between the two populations. This pattern of shared recent maternal ancestry was largely obscured in earlier studies based on low-resolution mtDNA haplogroup frequencies. It parallels the high frequency of male hg N in both populations [23] and patterns of autosomal segment sharing associated with a late IA migration event from North Estonia to Finland[73]. The majority of subclades present in both Finland and Estonia began to expand during the NBA or IA, with nearly equal numbers originating in each period regardless of the mutation rate applied (Additional file 2: Table S3). Modern genetic data cannot determine the direction of the migrations underlying this pattern. However, given the relatively small founder population and long-term genetic isolation of Finland[74], it is more likely that frequent contacts across the Gulf of Finland involved gene flow from south to north or that both populations were influenced by similar source populations. It is also possible that the two processes have complemented each other. The demographic isolation of the Finnish population is further reflected in the distribution of clades found primarily in Finland – such clades tend

to be older and more numerous than those observed in the Swedish population (Additional file 5: Fig. S8). The relatively smaller proportion of Estonians belonging to maternal clades shared with the Finns reflects the population's lower degree of isolation and greater maternal input from neighboring, but currently undersampled, populations.

Several examined maternal subclades within hgs H2a1, H13a1a1 and H1a started to expand during the NBA and IA and show phylogenetic connections to either ancient cultures from the steppe or forest-steppe zones near the Uralic Mountains or to sparsely sampled modern populations east of the CB region. Additional subclades from comparable geographic sources likely exist, but given the high diversity of terminal branches in the mitochondrial phylogeny, they remain beyond the scope of this study. Archaeological evidence identifies southern Urals as a major interaction zone connecting steppe cultures with expanding forest-steppe populations, potentially including Uralic-speaking groups[75]. Archaeological evidence indicates cultural change beginning in the late NBA in the northern Baltics, southwestern Finland and central Sweden, marked by the appearance of new pottery types and fortified settlements with parallels in the Upper and Middle Volga regions[51,76]. These developments have been interpreted as reflecting east-to-west movements across the East European Forest Belt, potentially linked to bronze-production and exchange networks involving steppe populations in the southern Ural region.

Our findings support considerable female contributions to the demographic and cultural processes of the NBA and IA periods, including long-distance contacts, that have shaped the contemporary maternal genetic diversity of the CB area.

Conclusions

Our study demonstrates the value of integrating large-scale, population-based resequencing of contemporary mitogenomes with temporally informed phylogenetic and archaeogenetic information for disentangling maternal demographic processes. The complete mitogenome-based analyses indicate that the maternal genetic structure of CB populations has been substantially shaped by demographic processes occurring during the Bronze and Iron Ages. The high number of IA-dated terminal clades, together with population-specific changes in maternal effective population size, indicates that present-day mtDNA variation reflects not only deep postglacial ancestry but, to a greater extent, more recent episodes of regional demographic expansion and migration. Detecting these patterns, however, requires larger sample sizes than comparable analyses of the Y-chromosome. The elevated sharing of terminal clades between Estonia and Finland points to a maternal genetic component of their linguistic affinity – a signal previously obscured by general hg frequency comparisons and small sample sizes. Phylogenetic links in a number of subclades suggest long-distance east–west interactions extending across the East European Forest Belt, involving ancient populations in the forest-steppe contact zones near the Ural Mountains. Results support a model in which maternal gene flow played a substantial role in the demographic and cultural transformations of the NBA and IA periods and has contributed significantly to the formation of contemporary regional maternal genetic diversity.

Materials and Methods

Samples

The sample set comprised 2,232 Estonian, 1,211 Finnish, 1,179 Swedish, 2,839 Danish, 1,616 Polish, and 227 Latvian complete mitochondrial genomes from present-day individuals(Fig. 1A). Full mtDNA sequences from Finnish, Swedish, Danish and Polish populations were obtained through GenBank searches, the 1000 Genomes Project[77] and from published literature (Additional file 6: Table S7), including Li et al.[78], Kaja et al.[79], Behar et al.[24], Piotrowska-Nowak et al.[80], Raule et al.[81] and mitogenomes extracted by Sturk-Andreaggi et al.[82] from the SweGen dataset[83]. Estonian complete mtDNA sequences were extracted from the whole genome sequences of 2,535 Estonian Biobank participants reported in Kals et al.[84]. In the Estonian sample, relatives up to the third degree were removed with the software KING-2.1.6[85] using autosomal variation.

We generated complete mtDNA sequences from a total of 95 Latvian DNA samples obtained from Nelis et al.[1] by Sanger sequencing following published protocols [86, 87]. An additional 129 Latvian complete mtDNA sequences were obtained through collaboration with the Latvian Biobank[88] and three sequences from Behar et al[24]. Sequences were aligned with ChromasPro version 1.7.6 (Technelysium Pty Ltd), and nucleotide mutations were ascertained relative to the revised Cambridge Reference Sequence(rCRS)[89]. All DNA samples were obtained from volunteers who provided informed consent in accordance with the guidelines of the relevant collaborating institutions. To provide phylogenetic context for the terminal maternal clades identified in the studied populations, we retrieved phylogenetically closely related complete mitochondrial genomes from the GenBank database[90].

Haplogroup assignment and phylogenetic analysis

The mtDNA hgs were inferred for all sequences using Haplogrep2[91] based on PhyloTree mtDNA tree Build 17[92]. To explore the terminal clades of the mtDNA phylogeny, we used the PhyloTree mtDNA tree Build 17[92] as the reference tree and manually compiled maximum parsimony trees for each hg present in the studied populations. In addition, we gathered all available phylogenetically close complete mtDNA sequences (a total of 5,182 individuals) from the GenBank database, regardless of their population affiliation and incorporated them into the maximum parsimony trees. When constructing the trees and identifying terminal clades, we excluded unreliable indels at positions 309, 315, 515–522 and 16,193 as well as mutational hotspots at 16,182, 16,183 and 16,519. We then identified locally frequent terminal monophyletic branches within each hg as those in which at least 75% of individuals originate from the CB populations. Next, we selected three focal populations – Estonian, Finnish and Swedish – as populations with sufficiently large sample sizes, well-sampled neighboring regions and contrasting linguistic histories. We examined clades that contained at least five individuals. We classified the clades as shared between examined populations if they contained at least two individuals from a population other than the focal population. If a clade contained only one or no individuals from populations other than the focal population and included at least 5 individuals in total, we classified it as a one-population clade. In the Estonian dataset, characterised by the most uniform sequencing quality, synonymous and

non-synonymous mutations were counted along the branches of maximum-parsimony trees. Descriptive statistical indices for complete mtDNA sequence data were calculated with the Arlequin 3.5.2.2 software package[93].

Coalescence age estimation and Bayesian skyline analyses

We computed the coalescence time estimates for each examined clade using the Bayesian MCMC approach implemented in BEAST v1.7.5[94]. To ensure accurate tree reconstruction, we included mitogenomes from closely related lineages. All analyses were performed using the HKY model of nucleotide substitution[95]. Rate variation among sites was modelled using a Gamma distribution with four rate categories. The strict clock model[96] was used as the clock model. The tree prior was set as piecewise-linear Bayesian skyline model[97]. The calculations were performed considering all substitutions except A16,182C and A16,183C with two BEAST runs for each analysed terminal clade. For the first set, we used the normally distributed prior for mutation rate from Soares et al[27]. As the software assumes a linear mutation rate, we corrected the obtained coalescence time estimates as suggested in Soares et al[27]. For the second set of runs, the normally distributed prior for the mutation rate was taken from Rieux et al.[32]. For each BEAST run, a single MCMC chain was run for 50,000,000 steps, sampling every 1,000 steps, with the first 10% discarded as burn-in. Appropriate effective sample size values (ESS > 200) for each model parameter were checked in Tracer v1.5. We used the median coalescence age estimates from the two runs to calculate the arithmetic mean for the final age estimates of the studied clades used for discussion. Correlation coefficients between the mean age of the analysed clades and the number of individuals characterised by that clade were calculated with the Python Matplotlib[98] library.

We conducted BSAs to estimate maternal Ne dynamics and explored the sample size effect on the BSA results in each population with a sample size exceeding 1,000 individuals. BSAs were conducted by drawing three random subsamples of 100 individuals, followed by subsamples of 500 and 1,179 individuals, and the maximum sample sizes of 1,616 and 2,232 for the Polish and Estonian populations, respectively. The mutation rate prior was set and the resulting time estimates corrected according to Soares et al.[27]. Population group size was set to the default value (10). Due to varying sample sizes, the MCMC chain lengths ranged from 50,000,000 to 800,000,000, until achieving adequate convergence (ESS ≥ 200) after discarding the initial burn-in of 10%. The MCMC recording step was adjusted accordingly. For the Danish sample, when the sample size exceeded 1,000 individuals, the tree-likelihood statistic failed to converge. Given its low value (< 20), increasing the MCMC chain length further was not justified. Skyline plots with the corrected ages were generated in RStudio[99] with basic R Statistical Software packages[100].

Historical periodisation framework

We apply historical periodisation adjusted to Northern Europe. The Neolithic period manifests differently across regions of the CB. In southern Sweden, the Neolithic subsistence pattern emerged with the Funnel Beaker culture around 6,000 BP[101], while in the eastern Baltic and Finland the onset of the

Neolithic was marked by the introduction of Comb Ceramic pottery approximately 5,900 BP, without a pronounced shift in subsistence economy. This was followed by the Corded Ware Culture spanning roughly 4,800 BP to 3,800 BP, which is regarded in the archaeological literature as the Late Neolithic period. This is followed by the NBA, considered to begin around 3,800 BP, and by the IA, which spans 2,500 BP to 800 BP.

Declarations

Ethics approval and consent to participate

All donors have provided informed consent and all experiments were performed in accordance with the relevant guidelines and regulations of collaborating institutions. The activities of the Estonian Biobank are regulated by the Human Genes Research Act, which was adopted in 2000 specifically for the operations of the Estonian Biobank. The research and access to the mitochondrial sequence data in the Estonian Biobank was approved by the Research Ethics Committee of the University of Tartu (permission number 1.1.-12/659).

Consent for publication

All participants provided written informed consent for the use of their sequencing data in scientific research and for publication in aggregated form.

Availability of data and materials

The complete mitochondrial sequences from Estonian individuals used in this study are subject to Estonian Biobank access restrictions stipulated by law and are therefore not publicly available; access may be granted upon reasonable request by contacting the corresponding authors and the Estonian Biobank. The Latvian complete mitochondrial sequences generated in this study will be deposited in GenBank under accession numbers XX–XX (after publication confirmation). The Swedish full mitogenomes are available upon request from the authors of the SweGen project [83]. A total of 598 Polish mitochondrial sequences are available upon request in accordance with the description provided by the Thousand Polish Genomes Project [79]. The accession numbers of all publicly available sequences used in this study are listed in Table S7 (Additional file 6: Table S7). The maximum parsimony trees of examined haplogroups and BEAST log files supporting the conclusions of this article are available in the open-access repository DATADOI, unique persistent identifier and hyperlink to datasets will be provided upon publication confirmation. Temporary links for reviewers:

Phylogenetic trees in xlsx format: https://datadoi.ee/previewurl.xhtml?token=ef17141a-8694-46b5-acfd-6546c6ba41ad

BEAST Log files, viewed with the software BEAUTi: https://datadoi.ee/previewurl.xhtml?token=248cd058-a664-446e-a4d1-50714509fd9e

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by Estonian Research Council grant PRG1071 to Prof. Richard Villems and University of Tartu grants PLTGIARENG21 and PLTGIARENG22. Ajai K. Pathak was supported by the Estonian Research Council grants PUTJD1186 and KOHTO39. Rodrigo Flores was supported by the Horizon Europe programme (grant agreement 101080117). Sequences from the Estonian Biobank were generated through the Estonian Center of Genomics/Roadmap II funded by the Estonian Research Council (project number TT17).

Author contributions

Conceptualisation: H.S., M.R., E.M., and A.-M.I. Methodology: H.S., M.R. and A.-M.I. Formal analysis: H.S., M.R., I.A.K. and A.-M.I. Data curation: M.R., E.M., R.F. and A.-M.I. Resources: M.R., E.M., R.F., R.R.-K., V.R., J.K., A.A., U.G., the EstBB research team and A.-M.I. Visualisation: H.S., M.R., K.I.R., O.U., A.K.P., R.F. and A.-M.I. Writing – original draft: H.S, M.R., E.M. and A.-M.I. Writing – review and editing: H.S., M.R. and A.-M.I., with input from all co-authors. Project administration: A.-M.I. All authors read and approved the final manuscript.

Acknowledgments:

We are grateful to all donors of the genetic material. Parts of this work were written at writing retreats organised by the Institute of Genomics, University of Tartu. Data analysis was carried out in part in the High-Performance Computing Center of the University of Tartu, Estonia. The Estonian Biobank Research Team refers to Andres Metspalu, Lili Milani, Tõnu Esko, Reedik Mägi, Mait Metspalu, Mari Nelis and Georgi Hudjashov. The corresponding authors would like to thank Professor Richard Villems for the invaluable support.

References

1. Nelis M, Esko T, Mägi R, Zimprich F, Zimprich A, Toncheva D et al. Genetic Structure of Europeans: A View from the North–East. Fleischer RC, editor. PLoS ONE. 2009;4:e5472. https://doi.org/10.1371/journal.pone.0005472

2. Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, et al. Correlation between Genetic and Geographic Structure in Europe. Curr Biol. 2008;18:1241–8. https://doi.org/10.1016/j.cub.2008.07.049.
3. 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;513:409–13. https://doi.org/10.1038/nature13673.
4. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015;522:207–11. https://doi.org/10.1038/nature14317.
5. Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, et al. Y-Chromosomal Diversity in Europe Is Clinal and Influenced Primarily by Geography, Rather than by Language. Am J Hum Genet. 2000;67:1526–43. https://doi.org/10.1086/316890.
6. Karmin M, Saag L, Vicente M, Sayres MAW, Järve M, Talas UG, et al. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. Genome Res. 2015;25:459–66. https://doi.org/10.1101/gr.186684.114.
7. Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Wilson Sayres MA, et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nat Genet. 2016;48:593–9. https://doi.org/10.1038/ng.3559.
8. Batini C, Hallast P, Vågene ÅJ, Zadik D, Eriksen HA, Pamjav H, et al. Population resequencing of European mitochondrial genomes highlights sex-bias in Bronze Age demographic expansions. Sci Rep. 2017;7:12086. https://doi.org/10.1038/s41598-017-11307-9.
9. Brandt G, Haak W, Adler CJ, Roth C, Szécsényi-Nagy A, Karimnia S, et al. Ancient DNA Reveals Key Stages in the Formation of Central European Mitochondrial Genetic Diversity. Science. 2013;342:257–61. https://doi.org/10.1126/science.1241844.
10. Lappalainen T, Laitinen V, Salmela E, Andersen P, Huoponen K, Savontaus M-L, et al. Migration Waves to the Baltic Sea Region. Ann Hum Genet. 2008;72:337–48. https://doi.org/10.1111/j.1469-1809.2007.00429.x.
11. Jarczak J, Grochowalski Ł, Marciniak B, Lach J, Słomka M, Sobalska-Kwapis M, et al. Mitochondrial DNA variability of the Polish population. Eur J Hum Genet. 2019;27:1304–14. https://doi.org/10.1038/s41431-019-0381-x.
12. Översti S, Onkamo P, Stoljarova M, Budowle B, Sajantila A, Palo JU. Identification and analysis of mtDNA genomes attributed to Finns reveal long-stagnant demographic trends obscured in the total diversity. Sci Rep. 2017;7:6193. https://doi.org/10.1038/s41598-017-05673-7.
13. Sahakyan H, Hooshiar Kashani B, Tamang R, Kushniarevich A, Francis A, Costa MD, et al. Origin and spread of human mitochondrial DNA haplogroup U7. Sci Rep. 2017;7:46044. https://doi.org/10.1038/srep46044.
14. Kristjansson D, Bohlin J, Nguyen TT, Jugessur A, Schurr TG. Evolution and dispersal of mitochondrial DNA haplogroup U5 in Northern Europe: insights from an unsupervised learning approach to

phylogeography. BMC Genomics. 2022;23:354. https://doi.org/10.1186/s12864-022-08572-y.

15. Pala M, Olivieri A, Achilli A, Accetturo M, Metspalu E, Reidla M, et al. Mitochondrial DNA Signals of Late Glacial Recolonization of Europe from Near Eastern Refugia. Am J Hum Genet. 2012;90:915–24. https://doi.org/10.1016/j.ajhg.2012.04.003.
16. Dahl Ö, Koptjevskaja-Tamm M, editors. Circum-Baltic Languages: Typology and contact Volume 2: Grammar and Typology. Amesterdam Philadelphia: John Benjamins Publishing Company; 2001. https://doi.org/10.1075/slcs.55.
17. Malmström H, Günther T, Svensson EM, Juras A, Fraser M, Munters AR, et al. The genomic ancestry of the Scandinavian Battle Axe Culture people and their relation to the broader Corded Ware horizon. Proc R Soc B. 2019;286:20191528. https://doi.org/10.1098/rspb.2019.1528.
18. Mittnik A, Wang C-C, Pfrengle S, Daubaras M, Zariņa G, Hallgren F, et al. The genetic prehistory of the Baltic Sea region. Nat Commun. 2018;9:442. https://doi.org/10.1038/s41467-018-02825-9.
19. Allentoft ME, Sikora M, Fischer A, Sjögren K-G, Ingason A, Macleod R, et al. 100 ancient genomes show repeated population turnovers in Neolithic Denmark. Nature. 2024;625:329–37. https://doi.org/10.1038/s41586-023-06862-3.
20. Lamnidis TC, Majander K, Jeong C, Salmela E, Wessman A, Moiseyev V, et al. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. Nat Commun. 2018;9:5018. https://doi.org/10.1038/s41467-018-07483-5.
21. Tambets K, Yunusbayev B, Hudjashov G, Ilumäe A-M, Rootsi S, Honkola T, et al. Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. Genome Biol. 2018;19:139. https://doi.org/10.1186/s13059-018-1522-1.
22. Günther T, Malmström H, Svensson EM, Omrak A, Sánchez-Quinto F, Kılınç GM, et al. Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. PLoS Biol. 2018;16:e2003703. https://doi.org/10.1371/journal.pbio.2003703.
23. Ilumäe A-M, Reidla M, Chukhryaeva M, Järve M, Post H, Karmin M, et al. Human Y Chromosome Haplogroup N: A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families. Am J Hum Genet. 2016;99:163–73. https://doi.org/10.1016/j.ajhg.2016.05.025.
24. 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;90:675–84. https://doi.org/10.1016/j.ajhg.2012.03.002.
25. Ilumäe A-M, Post H, Flores R, Karmin M, Sahakyan H, Mondal M, et al. Phylogenetic history of patrilineages rare in northern and eastern Europe from large-scale re-sequencing of human Y-chromosomes. Eur J Hum Genet. 2021;29:1510–9. https://doi.org/10.1038/s41431-021-00897-8.
26. Loogväli E-L, Kivisild T, Margus T, Villems R. Explaining the Imperfection of the Molecular Clock of Hominid Mitochondria. O'Rourke D, editor. PLoS ONE. 2009;4:e8260. https://doi.org/10.1371/journal.pone.0008260
27. Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, et al. Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. Am J Hum Genet. 2009;84:740–59.

https://doi.org/10.1016/j.ajhg.2009.05.001.

28. Stoljarova M, King JL, Takahashi M, Aaspõllu A, Budowle B. Whole mitochondrial genome genetic diversity in an Estonian population sample. Int J Legal Med. 2016;130:67–71. https://doi.org/10.1007/s00414-015-1249-4.
29. Malyarchuk B, Litvinov A, Derenko M, Skonieczna K, Grzybowski T, Grosheva A, et al. Mitogenomic diversity in Russians and Poles. Forensic Sci International: Genet. 2017;30:51–6. https://doi.org/10.1016/j.fsigen.2017.06.003.
30. Tallavaara M, Pesonen P. Human ecodynamics in the north-west coast of Finland 10,000–2000 years ago. Quatern Int. 2020;549:26–35. https://doi.org/10.1016/j.quaint.2018.06.032.
31. Wang RJ, Al-Saffar SI, Rogers J, Hahn MW. Human generation times across the past 250,000 years. Sci Adv. 2023;9:eabm7047. https://doi.org/10.1126/sciadv.abm7047.
32. Rieux A, Eriksson A, Li M, Sobkowiak B, Weinert LA, Warmuth V, et al. Improved Calibration of the Human Mitochondrial Clock Using Ancient Genomes. Mol Biol Evol. 2014;31:2780–92. https://doi.org/10.1093/molbev/msu222.
33. Chyleński M, Juras A, Ehler E, Malmström H, Piontek J, Jakobsson M, et al. Late Danubian mitochondrial genomes shed light into the Neolithisation of Central Europe in the 5th millennium BC. BMC Evol Biol. 2017;17:80. https://doi.org/10.1186/s12862-017-0924-0.
34. Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogväli E-L, et al. The Western and Eastern Roots of the Saami—the Story of Genetic Outliers Told by Mitochondrial DNA and Y Chromosomes. Am J Hum Genet. 2004;74:661–82. https://doi.org/10.1086/383203.
35. Saag L, Varul L, Scheib CL, Stenderup J, Allentoft ME, Saag L, et al. Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe. Curr Biol. 2017;27:2185–e21936. https://doi.org/10.1016/j.cub.2017.06.022.
36. Olalde I, Mallick S, Patterson N, Rohland N, Villalba-Mouco V, Silva M, et al. The genomic history of the Iberian Peninsula over the past 8000 years. Science. 2019;363:1230–4. https://doi.org/10.1126/science.aav4040.
37. Pesonen P, Larsson AM, Holmqvist E. The chronology of Corded Ware Culture in Finland – reviewing new data. Fennoscandia archaeologica. 2019;36:130–41.
38. Oras E, Tõrv M, Johanson K, Rannamäe E, Poska A, Lõugas L, et al. Parallel worlds and mixed economies: multi-proxy analysis reveals complex subsistence systems at the dawn of early farming in the northeast Baltic. R Soc open sci. 2023;10:230880. https://doi.org/10.1098/rsos.230880.
39. Mathieson I, Alpaslan-Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N, Mallick S, et al. The genomic history of southeastern Europe. Nature. 2018;555:197–203. https://doi.org/10.1038/nature25778.
40. Schroeder H, Margaryan A, Szmyt M, Theulot B, Włodarczak P, Rasmussen S, et al. Unraveling ancestry, kinship, and violence in a Late Neolithic mass grave. Proc Natl Acad Sci USA. 2019;116:10705–10. https://doi.org/10.1073/pnas.1820210116.

41. Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, et al. The formation of human populations in South and Central Asia. Science. 2019;365:eaat7487. https://doi.org/10.1126/science.aat7487.
42. Damgaard PDB, Marchi N, Rasmussen S, Peyrot M, Renaud G, Korneliussen T, et al. 137 ancient human genomes from across the Eurasian steppes. Nature. 2018;557:369–74. https://doi.org/10.1038/s41586-018-0094-2.
43. Gnecchi-Ruscone GA, Khussainova E, Kahbatkyzy N, Musralina L, Spyrou MA, Bianco RA, et al. Ancient genomic time transect from the Central Asian Steppe unravels the history of the Scythians. Sci Adv. 2021;7:eabe4414. https://doi.org/10.1126/sciadv.abe4414.
44. Rivollat M, Jeong C, Schiffels S, Küçükkalıpçı İ, Pemonge M-H, Rohrlach AB, et al. Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. Sci Adv. 2020;6:eaaz5344. https://doi.org/10.1126/sciadv.aaz5344.
45. Brunel S, Bennett EA, Cardin L, Garraud D, Barrand Emam H, Beylier A, et al. Ancient genomes from present-day France unveil 7,000 years of its demographic history. Proc Natl Acad Sci USA. 2020;117:12791–8. https://doi.org/10.1073/pnas.1918034117.
46. Stolarek I, Juras A, Handschuh L, Marcinkowska-Swojak M, Philips A, Zenczak M, et al. A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age. Sci Rep. 2018;8:2455. https://doi.org/10.1038/s41598-018-20705-6.
47. Saag L, Vasilyev SV, Varul L, Kosorukova NV, Gerasimov DV, Oshibkina SV, et al. Genetic ancestry changes in Stone to Bronze Age transition in the East European plain. Sci Adv. 2021;7:eabd6535. https://doi.org/10.1126/sciadv.abd6535.
48. Allentoft ME, Sikora M, Refoyo-Martínez A, Irving-Pease EK, Fischer A, Barrie W, et al. Population genomics of post-glacial western Eurasia. Nature. 2024;625:301–11. https://doi.org/10.1038/s41586-023-06865-0.
49. Papac L, Ernée M, Dobeš M, Langová M, Rohrlach AB, Aron F, et al. Dynamic changes in genomic and social structures in third millennium BCE central Europe. Sci Adv. 2021;7:eabi6941. https://doi.org/10.1126/sciadv.abi6941.
50. Mattila TM, Svensson EM, Juras A, Günther T, Kashuba N, Ala-Hulkko T, et al. Genetic continuity, isolation, and gene flow in Stone Age Central and Eastern Europe. Commun Biol. 2023;6:793. https://doi.org/10.1038/s42003-023-05131-3.
51. Lang V. Migrations from the west, south, and east shaped the Proto-Finnic ethnos. FUF. 2025;81–112. https://doi.org/10.33339/fuf.147261.
52. Saag L, Laneman M, Varul L, Malve M, Valk H, Razzak MA, et al. The Arrival of Siberian Ancestry Connecting the Eastern Baltic to Uralic Speakers further East. Curr Biol. 2019;29:1701–e171116. https://doi.org/10.1016/j.cub.2019.04.026.
53. 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.

https://doi.org/10.1038/nature16152.

54. Lazaridis I, Patterson N, Anthony D, Vyazov L, Fournier R, Ringbauer H, et al. The genetic origin of the Indo-Europeans. Nature. 2025;639:132–42. https://doi.org/10.1038/s41586-024-08531-5.
55. Juras A, Chyleński M, Ehler E, Malmström H, Żurkiewicz D, Włodarczak P, et al. Mitochondrial genomes reveal an east to west cline of steppe ancestry in Corded Ware populations. Sci Rep. 2018;8:11603. https://doi.org/10.1038/s41598-018-29914-5.
56. Juras A, Makarowicz P, Chyleński M, Ehler E, Malmström H, Krzewińska M, et al. Mitochondrial genomes from Bronze Age Poland reveal genetic continuity from the Late Neolithic and additional genetic affinities with the steppe populations. Am J Phys Anthropol. 2020;172:176–88. https://doi.org/10.1002/ajpa.24057.
57. Juras A, Ehler E, Chyleński M, Pospieszny Ł, Spinek AE, Malmström H, et al. Maternal genetic origin of the late and final Neolithic human populations from present-day Poland. Am J Phys Anthropol. 2021;176:223–36. https://doi.org/10.1002/ajpa.24372.
58. Furtwängler A, Rebay-Salisbury K, Neumann GU, Kanz F, Ringbauer H, Bianco RA, et al. Tracing social mechanisms and interregional connections in Early Bronze Age Societies in Lower Austria. Nat Commun. 2025;17:131. https://doi.org/10.1038/s41467-025-67906-y.
59. Posth C, Zaro V, Spyrou MA, Vai S, Gnecchi-Ruscone GA, Modi A, et al. The origin and legacy of the Etruscans through a 2000-year archeogenomic time transect. Sci Adv. 2021;7:eabi7673. https://doi.org/10.1126/sciadv.abi7673.
60. Översti S, Majander K, Salmela E, Salo K, Arppe L, Belskiy S, et al. Human mitochondrial DNA lineages in Iron-Age Fennoscandia suggest incipient admixture and eastern introduction of farming-related maternal ancestry. Sci Rep. 2019;9:1–14. https://doi.org/10.1038/s41598-019-51045-8.
61. Unterländer M, Palstra F, Lazaridis I, Pilipenko A, Hofmanová Z, Groß M, et al. Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. Nat Commun. 2017;8:14615. https://doi.org/10.1038/ncomms14615.
62. Maróti Z, Neparáczki E, Schütz O, Maár K, Varga GIB, Kovács B, et al. The genetic origin of Huns, Avars, and conquering Hungarians. Curr Biol. 2022;32:2858–e28707. https://doi.org/10.1016/j.cub.2022.04.093.
63. Neparáczki E, Maróti Z, Kalmár T, Kocsy K, Maár K, Bihari P et al. Mitogenomic data indicate admixture components of Central-Inner Asian and Srubnaya origin in the conquering Hungarians. Caramelli D, editor. PLoS ONE. 2018;13:e0205920. https://doi.org/10.1371/journal.pone.0205920
64. McColl H, Kroonen G, Víctor Moreno-Mayar J, Seersholm FV, Scorrano G, Pinotti T et al. Steppe Ancestry in Western Eurasia and the Spread of the Germanic Languages. Preprint at https://www.biorxiv.org/content/10.1101/2024.03.13.584607v2(2025).
65. Oras E, Lang V, Rannamäe E, Varul L, Konsa M, Limbo-Simovart J, et al. Tracing prehistoric migration: isotope analysis of Bronze and pre-Roman Irgon Age coastal burials in Estonia. EJA. 2016;20:3–24. https://doi.org/10.3176/arch.2016.1.01.

66. Edmonds CA, Lillie AS, Cavalli-Sforza LL. Mutations arising in the wave front of an expanding population. Proc Natl Acad Sci USA. 2004;101:975–9. https://doi.org/10.1073/pnas.0308064100.
67. Hallatschek O, Nelson DR. Gene surfing in expanding populations. Theor Popul Biol. 2008;73:158–70. https://doi.org/10.1016/j.tpb.2007.08.008.
68. Guyon L, Guez J, Toupance B, Heyer E, Chaix R. Patrilineal segmentary systems provide a peaceful explanation for the post-Neolithic Y-chromosome bottleneck. Nat Commun. 2024;15:3243. https://doi.org/10.1038/s41467-024-47618-5.
69. Speidel L, Silva M, Booth T, Raffield B, Anastasiadou K, Barrington C, et al. High-resolution genomic history of early medieval Europe. Nature. 2025;637:118–26. https://doi.org/10.1038/s41586-024-08275-2.
70. Honkola T, Vesakoski O, Korhonen K, Lehtinen J, Syrjänen K, Wahlberg N. Cultural and climatic changes shape the evolutionary history of the Uralic languages. J Evol Biol. 2013;26:1244–53. https://doi.org/10.1111/jeb.12107.
71. Olivieri A, Sidore C, Achilli A, Angius A, Posth C, Furtwängler A, et al. Mitogenome Diversity in Sardinians: A Genetic Window onto an Island's Past. Mol Biol Evol. 2017;34:1230–9. https://doi.org/10.1093/molbev/msx082.
72. Tallavaara M, Pesonen P, Oinonen M. Prehistoric population history in eastern Fennoscandia. J Archaeol Sci. 2010;37:251–60. https://doi.org/10.1016/j.jas.2009.09.035.
73. Kivisild T, Saag L, Hui R, Biagini SA, Pankratov V, D'Atanasio E, et al. Patterns of genetic connectedness between modern and medieval Estonian genomes reveal the origins of a major ancestry component of the Finnish population. Am J Hum Genet. 2021;108:1792–806. https://doi.org/10.1016/j.ajhg.2021.07.012.
74. Martin AR, Karczewski KJ, Kerminen S, Kurki MI, Sarin A-P, Artomov M, et al. Haplotype Sharing Provides Insights into Fine-Scale Population History and Disease in Finland. Am J Hum Genet. 2018;102:760–75. https://doi.org/10.1016/j.ajhg.2018.03.003.
75. Grünthal R, Heyd V, Holopainen S, Janhunen JA, Khanina O, Miestamo M, et al. Drastic demographic events triggered the Uralic spread. DIA. 2022;39:490–524. https://doi.org/10.1075/dia.20038.gru.
76. Lang V. Fortified Settlements in the Eastern Baltic: From Earlier Research to New Interpretations. Archaeologia Lituana. 2018;19:13–33. https://doi.org/10.15388/ArchLit.2018.19.2.
77. The 1000 Genomes Project Consortium, Auton C, Abecasis A, Steering GR, Altshuler committee et al. DM,. A global reference for human genetic variation. Nature. 2015;526:68–74. https://doi.org/10.1038/nature15393
78. Li S, Besenbacher S, Li Y, Kristiansen K, Grarup N, Albrechtsen A, et al. Variation and association to diabetes in 2000 full mtDNA sequences mined from an exome study in a Danish population. Eur J Hum Genet. 2014;22:1040–5. https://doi.org/10.1038/ejhg.2013.282.
79. Kaja E, Lejman A, Sielski D, Sypniewski M, Gambin T, Dawidziuk M, et al. The Thousand Polish Genomes—A Database of Polish Variant Allele Frequencies. IJMS. 2022;23:4532. https://doi.org/10.3390/ijms23094532.

80. Piotrowska-Nowak A, Elson JL, Sobczyk-Kopciol A, Piwonska A, Puch-Walczak A, Drygas W, et al. New mtDNA Association Model, MutPred Variant Load, Suggests Individuals With Multiple Mildly Deleterious mtDNA Variants Are More Likely to Suffer From Atherosclerosis. Front Genet. 2019;9:702. https://doi.org/10.3389/fgene.2018.00702.
81. Raule N, Sevini F, Li S, Barbieri A, Tallaro F, Lomartire L, et al. The co-occurrence of mt DNA mutations on different oxidative phosphorylation subunits, not detected by haplogroup analysis, affects human longevity and is population specific. Aging Cell. 2014;13:401–7. https://doi.org/10.1111/acel.12186.
82. Sturk-Andreaggi K, Ring JD, Ameur A, Gyllensten U, Bodner M, Parson W, et al. The Value of Whole-Genome Sequencing for Mitochondrial DNA Population Studies: Strategies and Criteria for Extracting High-Quality Mitogenome Haplotypes. IJMS. 2022;23:2244. https://doi.org/10.3390/ijms23042244.
83. Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. Eur J Hum Genet. 2017;25:1253–60. https://doi.org/10.1038/ejhg.2017.130.
84. Kals M, Nikopensius T, Läll K, Pärn K, Tõnis Sikka T, Suvisaari J et al. Advantages of genotype imputation with ethnically matched reference panel for rare variant association analyses. Preprint at https://www.biorxiv.org/content/10.1101/579201v2(2019).
85. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26:2867–73. https://doi.org/10.1093/bioinformatics/btq559.
86. Rieder M. Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. Nucleic Acids Res. 1998;26:967–73. https://doi.org/10.1093/nar/26.4.967.
87. Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, et al. Do the Four Clades of the mtDNA Haplogroup L2 Evolve at Different Rates? Am J Hum Genet. 2001;69:1348–56. https://doi.org/10.1086/324511.
88. Rovite V, Wolff-Sagi Y, Zaharenko L, Nikitina-Zake L, Grens E, Klovins J. Genome Database of the Latvian Population (LGDB): Design, Goals, and Primary Results. J Epidemiol. 2018;28:353–60. https://doi.org/10.2188/jea.JE20170079.
89. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999;23:147–147. https://doi.org/10.1038/13779.
90. The Genbank database (http://www.ncbi.nlm.nih.gov/genbank/). Last accessed 15 Jun 2021.
91. Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt H-J, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. 2016;44:W58–63. https://doi.org/10.1093/nar/gkw233.

92. Van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat. 2009;30:E386–94. https://doi.org/10.1002/humu.20921.
93. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10:564–7. https://doi.org/10.1111/j.1755-0998.2010.02847.x.
94. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29:1969–73. https://doi.org/10.1093/molbev/mss075.
95. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 1985;22:160–74. https://doi.org/10.1007/BF02101694.
96. Zuckerkandl E, Pauling L. Molecules as documents of evolutionary history. J Theor Biol. 1965;8:357–66. https://doi.org/10.1016/0022-5193(65)90083-4.
97. Drummond AJ. Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. Mol Biol Evol. 2005;22:1185–92. https://doi.org/10.1093/molbev/msi103.
98. Hunter JD, Matplotlib. A 2D Graphics Environment. Comput Sci Eng. 2007;9:90–5. https://doi.org/10.1109/MCSE.2007.55.
99. Posit Team. RStudio: Integrated Development Environment for R [Internet]. Boston, MA: Posit Software, PBC; 2025. http://www.posit.co/.
100. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2025. https://www.R-project.org/.
101. Price TD. Ancient Scandinavia: an archaeological history from the first humans to the Vikings. New York, NY: Oxford University Press; 2015.

Figures

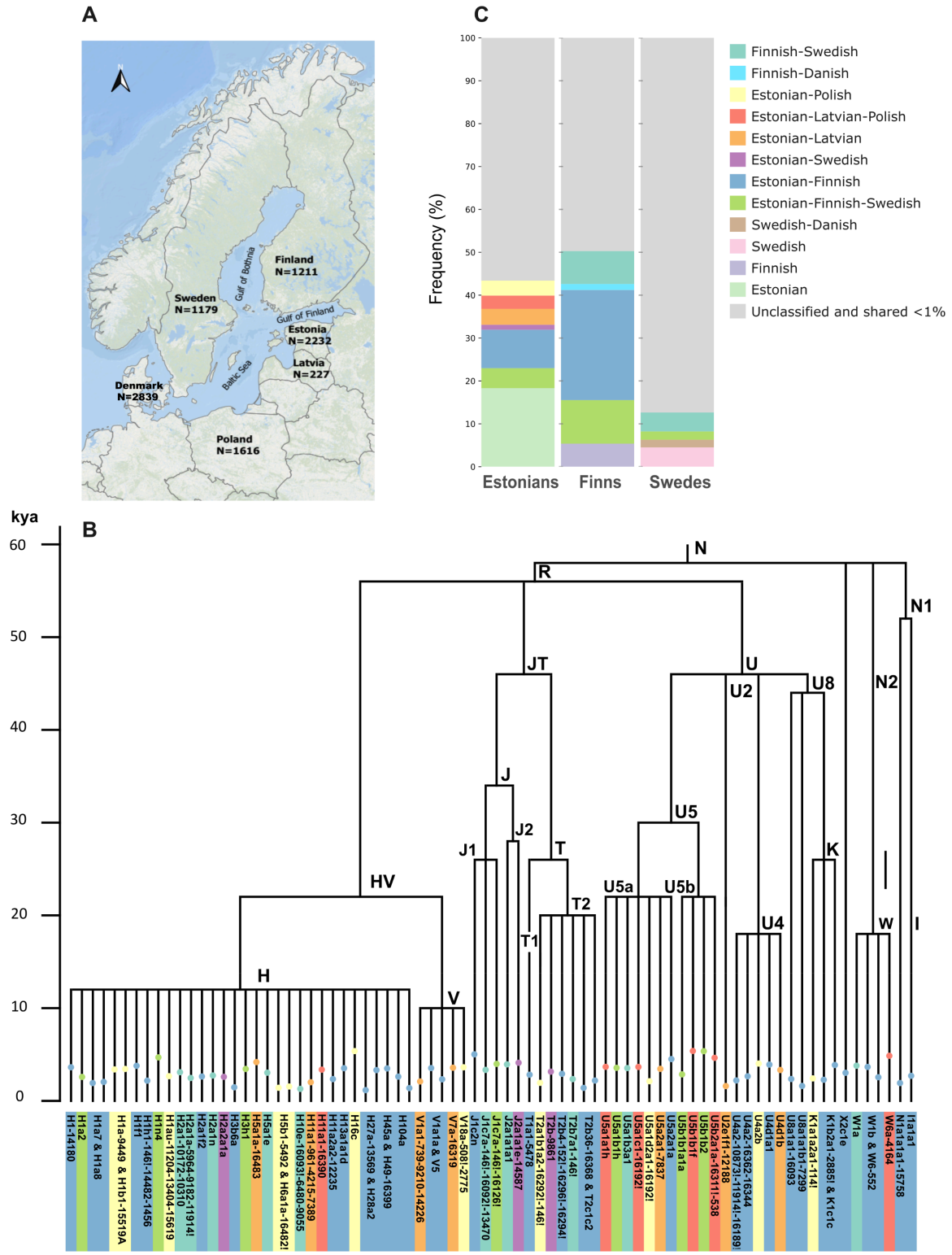


Figure 1

Sample sizes of CB populations, a schematic tree of shared clades and proportions of individuals. A) Numbers indicate sample sizes from the CB populations**. B)** Schematic representation of the phylogenetic tree of 78 shared clades identified in the focal populations. Age estimates for higher-level phylogenetic haplogroups were obtained from Behar et al[24]. **C)** Proportions of individuals belonging to

either shared clades or one-population clades defined according to the set criteria. Colors correspond to the figure key. Abbreviations: kya - thousand years ago.

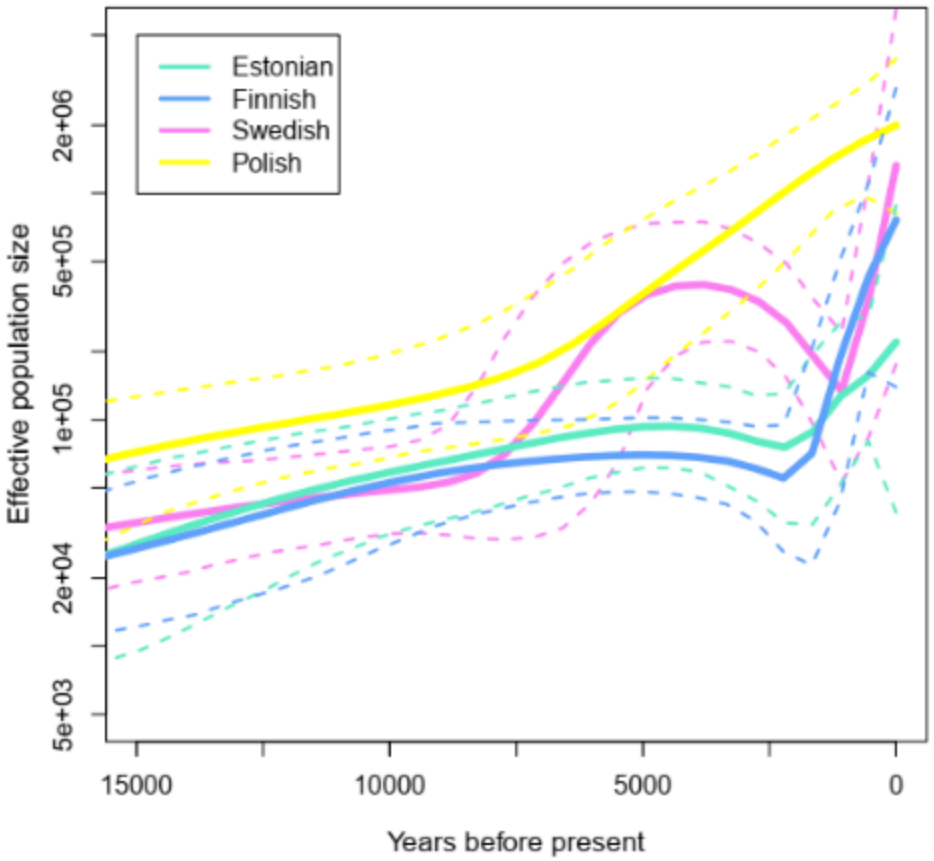


Figure 2

Bayesian Skyline Plots showing female Ne changes over time, based on 1179 complete mitogenomes. The bold solid central lines represent the mean estimates of Ne. Colours correspond to those shown in the figure key. Dotted lines indicate the 95% highest posterior density (HPD) intervals. Ne is plotted on a logarithmic scale, assuming a generation time of 23 years[31].

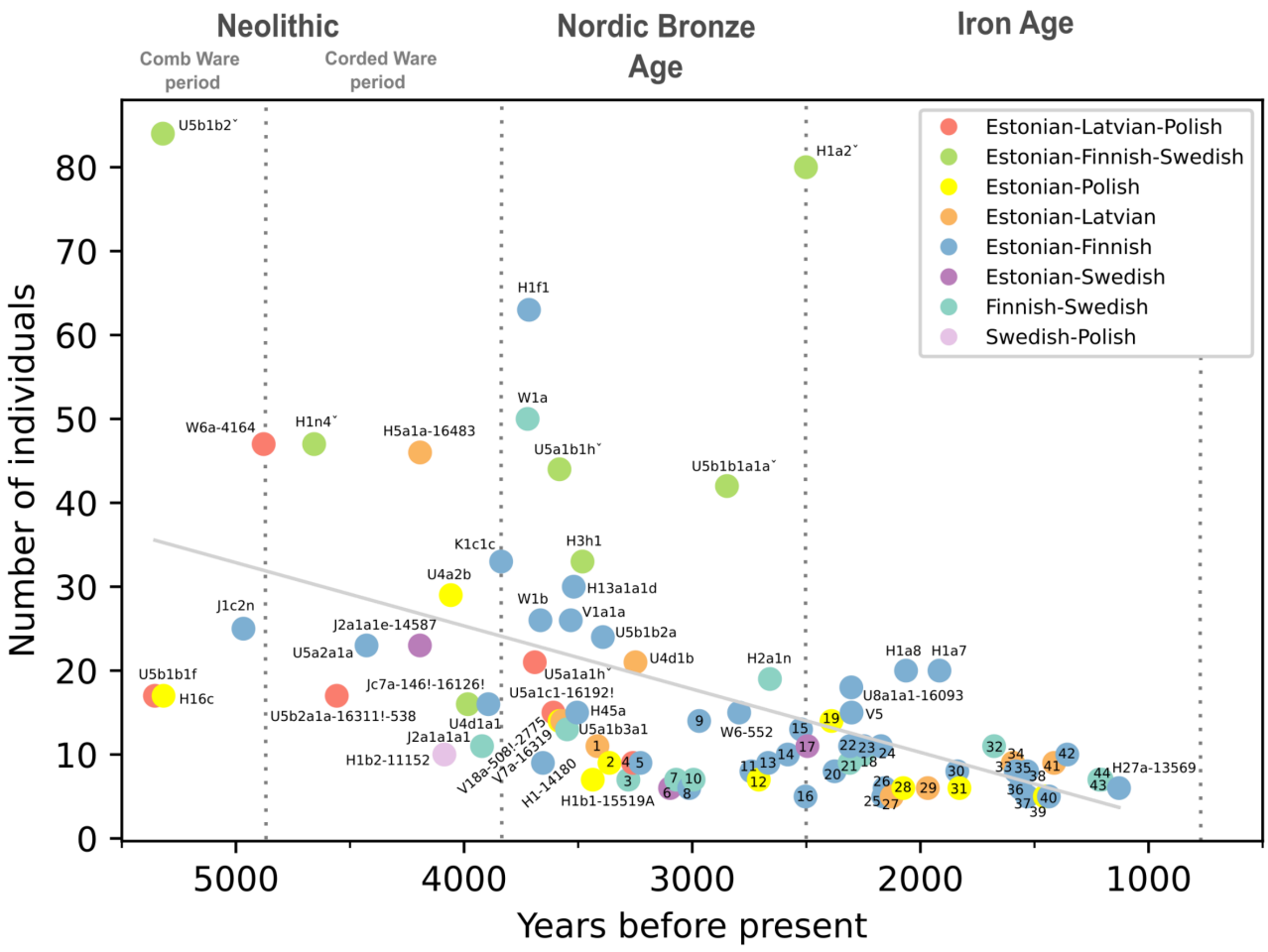


Figure 3

Bayesian Skyline Plots showing female Ne changes over time, based on 1179 complete mitogenomes. The bold solid central lines represent the mean estimates of Ne. Colours correspond to those shown in the figure key. Dotted lines indicate the 95% highest posterior density (HPD) intervals. Ne is plotted on a logarithmic scale, assuming a generation time of 23 years[31].

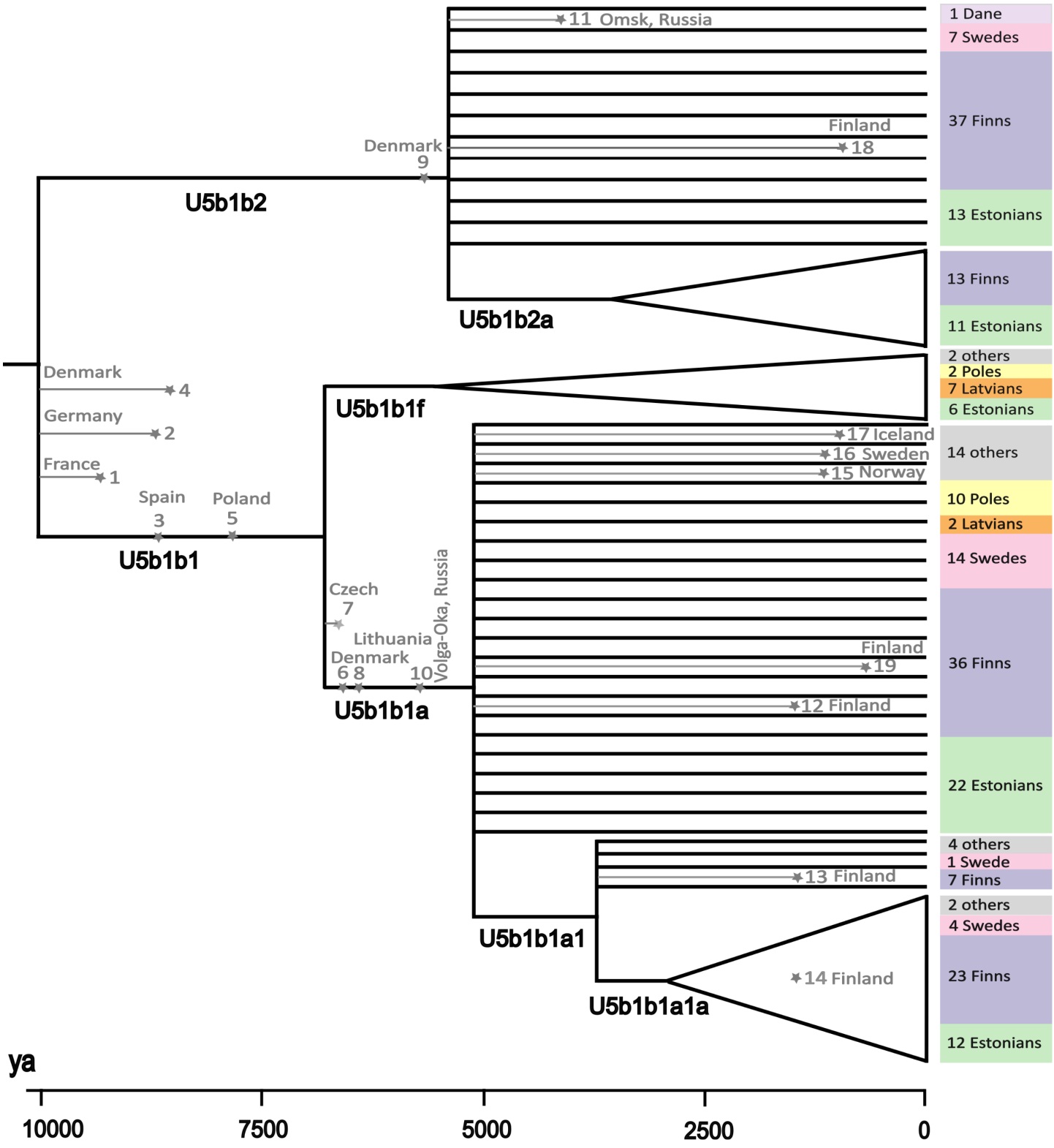


Figure 4

Schematic representation of hgs U5b1b1 and U5b1b2 phylogeny. The phylogenetic tree was reconstructed from a total of 252 modern complete mitogenomes using BEAST v1.7.5. Unsupported dichotomies resulting from the software's strict bifurcation requirement were manually reduced to polytomies. Subclade age estimates were adjusted to the arithmetic mean values reported in Additional File 2: Table S3. Asterisks indicate ancient DNA samples that were manually added to the tree based on

their mean radiocarbon age or inferred cultural-context age, together with their geographic locations. Abbreviations: **ya**– years ago. Numbers correspond to the following publications: **1-2** – Posth C. et al. Palaeogenomics of Upper Palaeolithic to Neolithic European hunter-gatherers, Nature 2023; **3** – Allentoft M.E. et al. Population genomics of post-glacial western Eurasia, Nature 2024; **4, 6, 9** – Allentoft M.E. et al. 100 ancient genomes show repeated population turnovers in Neolithic Denmark, Nature 2024; **5** – Chylenski M. et al. Late Danubian mitochondrial genomes shed light into the Neolithisation of Central Europe in the 5th millennium BC, BMC Evol Biol. 2017; **7** – Olalde I. et al. The Beaker phenomenon and the genomic transformation of northwest Europe, Nature 2018; **8** – Mittnik A. et al. The genetic prehistory of the Baltic Sea region, Nature Communications 2018; **10** – Allentoft M.E. et al. Population genomics of post-glacial western Eurasia, Nature 2024; **11** – Omsk oblast, Russia, Zeng T.C. et al. Ancient DNA reveals the prehistory of the Uralic and Yeniseian peoples, Nature 2025; **12-14, 18-19** – Finland, Översti S. et al. Human mitochondrial DNA lineages in Iron-Age Fennoscandia suggest incipient admixture and eastern introduction of farming-related maternal ancestry, Scientific Reports 2019; **15-17** – Norway, Sweden, Iceland, Margaryan A. et al. Population genomics of the Viking world, Nature 2020.

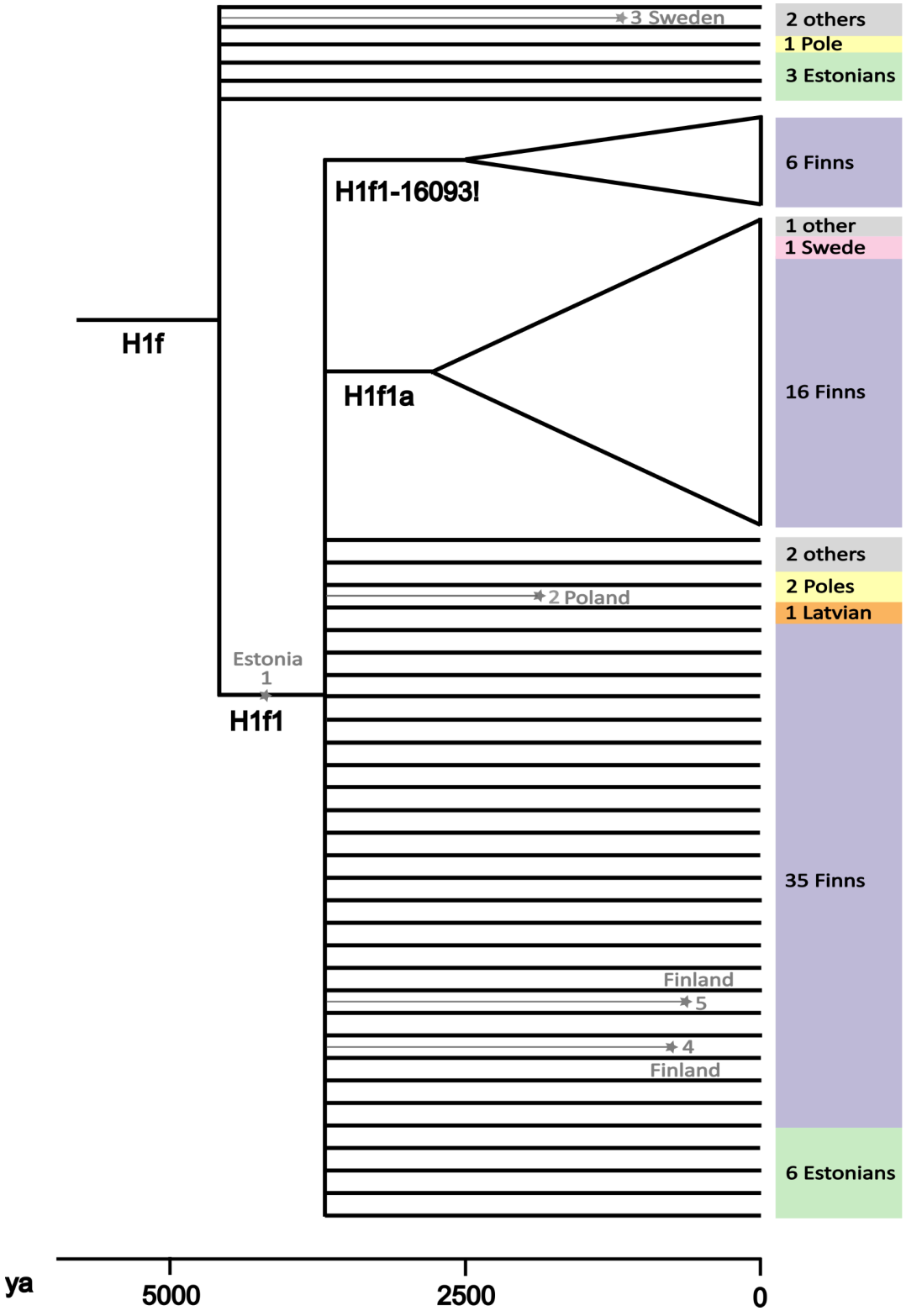


Figure 5

Schematic representation of hg H1f phylogeny. The phylogenetic tree was reconstructed from a total of 76 modern complete mitogenomes using BEAST v1.7.5. Unsupported dichotomies resulting from the software's strict bifurcation requirement were manually reduced to polytomies. Subclade age estimates were adjusted to the arithmetic mean values reported in Additional File 2: Table S3. Asterisks indicate ancient DNA samples that were manually added to the tree based on their mean radiocarbon age or

inferred cultural-context age, together with their geographic locations. Abbreviations: **ya**– years ago. Numbers correspond to the following publications: **1** – Malmström H. et al. The genomic ancestry of the Scandinavian Battle Axe Culture people and their relation to the broader Corded Ware horizon, Proc Biol Sci. 2019; **2** – Stolarek I. et al. A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age, Scientific Reports 2018; **3** – Margaryan A. et al. Population genomics of the Viking world, Nature 2020; **4-5** – Översti S. et al. Human mitochondrial DNA lineages in Iron-Age Fennoscandia suggest incipient admixture and eastern introduction of farming-related maternal ancestry, Scientific Reports2019.

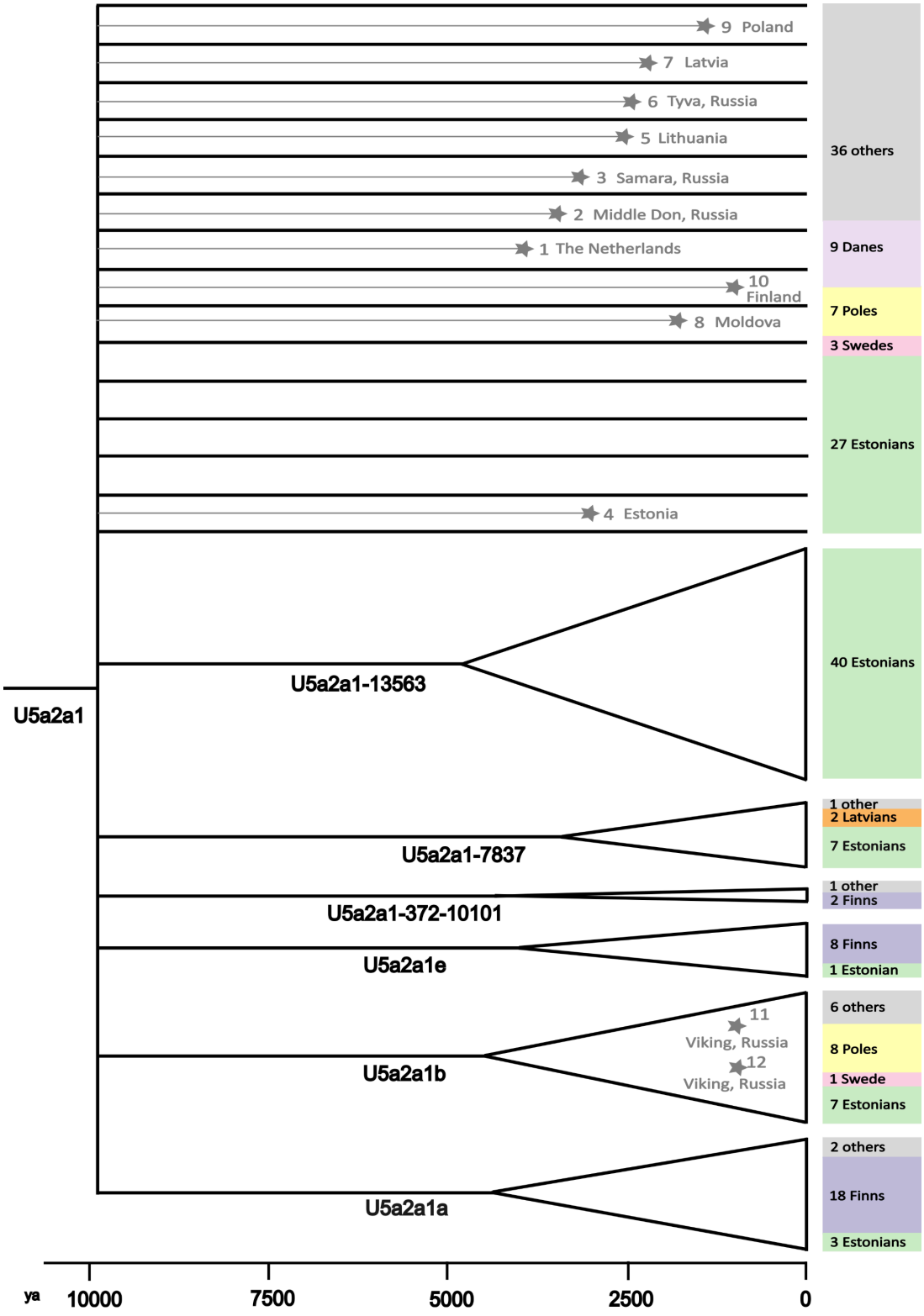


Figure 6

Schematic representation of hg U5a2a1 phylogeny. The phylogenetic tree was reconstructed from a total of 189 modern complete mitogenomes using BEAST v1.7.5. Unsupported dichotomies resulting from the software's strict bifurcation requirement were manually reduced to polytomies. Subclade age estimates were adjusted to the arithmetic mean values reported in Additional File 2: Table S3. Asterisks indicate ancient DNA samples that were manually added to the tree based on their mean radiocarbon

age or inferred cultural-context age, together with their geographic locations. Abbreviations: **ya** – years ago. Numbers correspond to the following publications: **1** – The Netherlands, Olalde I. et al. The Beaker phenomenon and the genomic transformation of northwest Europe, Nature 2018; **2** – Middle Don, Russia, Andreeva T.V. et al. Genetic history of Scythia, Science Advances 2025; **3** – Samara, Russia, Narasimhan V. et al. The formation of human populations in South and Central Asia, Science 2019; **4**– Estonia, Saag L. The Arrival of Siberian Ancestry Connecting the Eastern Baltic to Uralic Speakers further East, Current Biology 2019; **5,7** – Lithuania, Latvia, Mittnik A. et al. The genetic prehistory of the Baltic Sea region, Nature Communications 2018; **6** – Russia, Mary L. et al. Genetic kinship and admixture in Iron Age Scytho-Siberians, Human Genetics 2019; **8** – Moldova, Krzewinska M. et al. Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads, Science Advances 2018; **9** – Poland, Stolarek I. et al. Goth migration induced changes in the matrilineal genetic structure of the central-east European population, Scientific Reports 2019; **10** – Finland, Översti S. et al. Human mitochondrial DNA lineages in Iron-Age Fennoscandia suggest incipient admixture and eastern introduction of farming-related maternal ancestry, Scientific Reports 2019; **11-12** – Russia, Margaryan A. et al. Population genomics of the Viking world, Nature 2020.

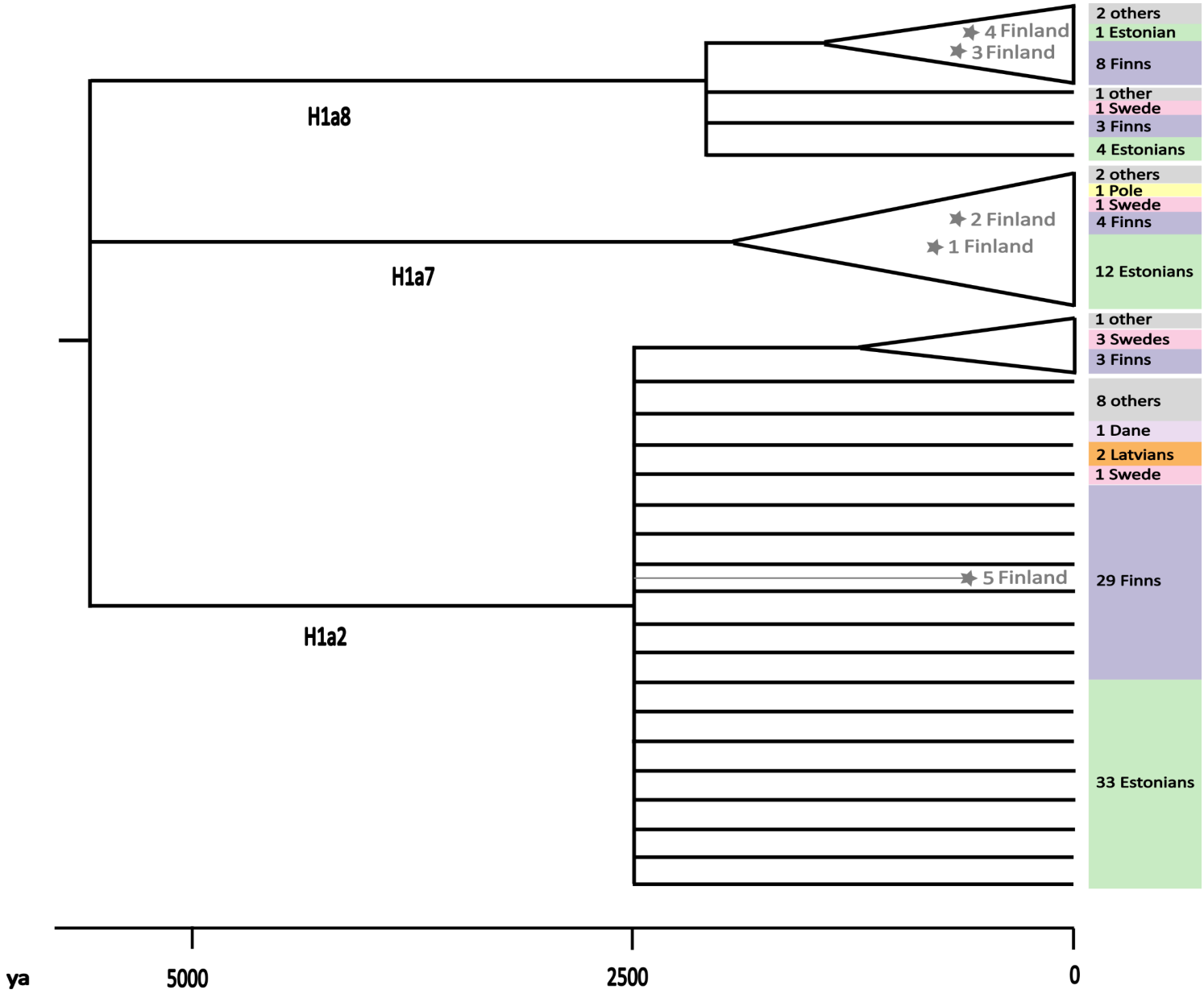


Figure 7

Schematic representation of hgs H1a2, H1a7 and H1a8 phylogeny. The phylogenetic tree was reconstructed from a total of 121 modern complete mitogenomes using BEAST v1.7.5. Unsupported dichotomies resulting from the software's strict bifurcation requirement were manually reduced to polytomies. Subclade age estimates were adjusted to the arithmetic mean values reported in Additional File 2: Table S3. Asterisks indicate ancient DNA samples that were manually added to the tree based on their mean radiocarbon age or inferred cultural-context age, together with their geographic locations. **ya** – years ago. Numbers correspond to the following publications: **1-5** – Översti S. et al. Human mitochondrial DNA lineages in Iron-Age Fennoscandia suggest incipient admixture and eastern introduction of farming-related maternal ancestry, Scientific Reports 2019.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1TableS1.xlsx
- Additionalfile2TableS2TableS6.xlsx
- Additionalfile3.pdf
- Additionalfile4.xlsx
- Additionalfile5.pdf
- Additionalfile6TableS7.xlsx