Human mitochondrial DNA lineages in Iron-Age Fennoscandia suggest incipient admixture and eastern introduction of farming-related maternal ancestry

Sanni Översti1, Kerttu Majander1,2,3, Elina Salmela1,2, Kati Salo4, Laura Arppe5, Stanislav Belskiy6, Heli Etu-Sihvola6, Ville Laakso7, Esa Mikkola6, Saskia Pfencgle1, Mikko Putkonen9, Jussi-Pekka Taavitsainen7, Katja Vuoristo8, Anna Weissman6,7, Antti Sajantila9, Markku Oinonen6, Wolfgang Haak2, Verena J. Schuenemann3,10, Johannes Krause2, Jukka U. Palo9,11 & Päivi Onkamo1,12

Human ancient DNA studies have revealed high mobility in Europe’s past, and have helped to decode the human history on the Eurasian continent. Northeastern Europe, especially north of the Baltic Sea, however, remains less well understood largely due to the lack of preserved human remains. Finland, with a divergent population history from most of Europe, offers a unique perspective to hunter-gatherer way of life, but thus far genetic information on prehistoric human groups in Finland is nearly absent. Here we report 103 complete ancient mitochondrial genomes from human remains dated to AD 300–1800, and explore mtDNA diversity associated with hunter-gatherers and Neolithic farmers. The results indicate largely unadmixed mtDNA pools of differing ancestries from Iron-Age on, suggesting a rather late genetic shift from hunter-gatherers towards farmers in North-East Europe. Furthermore, the data suggest eastern introduction of farmer-related haplogroups into Finland, contradicting contemporary genetic patterns in Finns.

Genetic studies on anthropological remains have exceedingly helped to shed light on various human populations as well as past events and processes. These include for instance the initial colonization of Europe by modern humans c. 40 kya1,2, the Holocene hunter-gatherer communities1,3, the major population turnover associated with the Neolithic spread of agriculture from Anatolia3–5, and the massive Bronze-Age influx of genes, culture and customs of the Yamnaya-related people into Europe from the Pontic-Caspian steppe6,7.

In terms of mitochondrial DNA (mtDNA), these cultural turnovers and population migrations in Europe involved also changes in the haplogroup composition. In hunter-gatherer populations the dominating mitochondrial lineage has been U, especially its subgroups U4, U5a and U5b3. In the advent of the Neolithic revolution, these U subgroups were largely supplanted by farmer-associated haplogroups H, HV, J, K, N1a, T2 and W4,5. The
subsequent spread of Yamnaya-related people and Corded Ware Culture in the late Neolithic and Bronze Age were accompanied with the increase of haplogroups I, U2 and T1 in Europe (See4 and references therein).

Whereas the ancient DNA (aDNA) composition and its changes in mainland Europe are increasingly well understood, northeastern Europe has been far less studied. The oldest human DNA analyzed in this region derive from the Mesolithic burial sites in Huseby Klev, western Sweden (9800 calBP14), Himmervikholmen in Norway (9300 calBP15) and Yuzny Oleminj Ostrov on Lake Onega, Russia (~8400 calBP11). In addition, Mesolithic to Bronze-Age DNA data from several sites in the Baltic countries have been published6,12-14.

Despite the relatively close geographical proximity, little is known about the ancient DNA diversity in regions immediately north of the Baltics. This is largely due to the scarcity of preserved anthropological remains. In the hemiboreal forest zone the soil pH, together with annual freeze-thaw cycles has highly detrimental effects on bone material to the extent to which no unburnt remains older than ~2000 years exist15. However, archaeological evidence strongly suggests that the most notable colonization events in the region have taken place much earlier16,17. Consequently, the lack of archaeological bone material greatly limits the capability of aDNA studies in resolving the human population history of the Taiga belt and the processes that have shaped present-day diversity. Despite these shortcomings, aDNA has recently been recovered from c. 1500 year-old bones from Levänluhta in western central Finland16,17. Genomic data from these samples show a Siberian ancestry component still prominently present today, particularly in the indigenous Saami people, and to a lesser extent in modern Finns. Although these data suggest a widespread presence of genetically Saami-like people around eastern Fennoscandia during the Iron Age, more wide-spread sampling in space and time is necessary for understanding the past population dynamics, and emerging of the contemporary genetic diversity in Finland.

In terms of both genetics and culture, modern Finns show a unique combination of eastern and western European elements, which most likely reflects the settlement history. The first archaeological evidence of human presence in Finland dates relatively late to c. 11000–9000 years ago, soon after the continental ice sheet retreated. According to the archaeological record, the region has since supported a continuous human occupation until today17. Several influential waves of material culture have been shown to extend into Finland: the first one brought the Sperrings or Särässiemi pottery to the area c. 7500 years ago20, the second presented the typical Comb-Ceramic at 6000 years ago16,17, one of the most influential prehistoric cultures in the wider region. Finally, around 4700 years ago, the Corded-Ware culture reached Finland16,17,21. The most significant cultural changes, possibly driven by expansions, proceeded from east/south-east and extended into most of today’s territory of Finland, while the Corded-Ware culture influence spread from south occupying only the southwestern part of Finland. Later, Bronze Age brought an increase of human activity17 and coinciding advance of cereal cultivation22. Finland also saw a pronounced Scandinavian impact along the western coast, while the inland was dominated by eastern influences, seen e.g. in the arrival of ferrous metallurgy c. 500 BC and stylistic features of Bronze-Age and Iron-Age items21. Iron Age in Finland starts around 500 BC and continues until the end of Crusade period (c. 1200 AD in west and 1300 AD in east). Unlike most of the Europe, the Middle Ages starts as late as 1200 AD in western parts of country and 1300 AD in east and shifts into Early modern period in the beginning of 16th century (for a review of archaeological and historical periods in Finland see23).

Linguistically, contemporary Finns and Saami differ from most other Europeans in speaking a Uralic language, unrelated to the majority of European languages, which belong to the Indo-European language family. Finns are also genetically distinct from their neighboring populations and form outliers in the genetic variation within Europe24. This genetic uniqueness derives from much reduced genetic diversity24,25 and an Asian influence to the gene pool24. Within Finland, an unusually strong genetic border bisects the population along a northwest to southeast axis24,26,27, and is interpreted to reflect an ancient boundary between hunter-gatherer and farmer populations28. The expanse of agriculture north-east of this border was probably limited by environmental factors, especially the length of the growing season. Later, this border has most likely acted in demarcating the spread of western and eastern political and cultural impacts influencing the placement of first political border between Sweden and Novgorod through the middle of Finland (Treaty of Noteborg 1323 AD).

In order to gain better insight into the genetic history of Finns, we here describe 103 complete mitochondrial genomes reconstructed from bone samples from ten burial sites in southern Finland and the Republic of Karelia, Russia (former Finnish territory; Fig. 1). The main focus is on the 70 complete mitochondrial genomes from five archaeological burial sites in Finland spanning spatially from western coast to Lake Ladoga, and Late Roman Iron Age (300 AD) to the Middle Ages (1500 AD) (Tables 1, S1 and Supplementary Material S1). In addition, we include 33 mitochondrial genomes from later, mainly historical burials (1400–1800 AD) from five sites across southern Finland. While mtDNA genomes of Iron-Age and Early-Medieval Finland cannot be used to directly target questions about the colonization of Finland, they provide a spatio-temporal transect to the maternal ancestry of the early inhabitants in this region, and help to understand patterns observed in Finland's modern mtDNA diversity.

Results

Authenticity of ancient-DNA results. Based on the shotgun sequencing, out of the total of 141 individuals sampled, 134 were included in mitochondrial capture. Mitochondrial genomes for 103 individuals passed the quality control thresholds, while 31 samples were excluded from further analyses due to insufficient data (less than fivefold mitochondrial coverage) or high contamination levels (Supplementary Table S1). Ancient-DNA yield for all 103 samples was studied with several criteria of authentication. All samples showed fragment sizes ranging between 40–250 bp, as expected for ancient DNA29. Fragments under 30 bp were filtered out as a mapping quality control. All samples had an average fragment length of 47 to 95 bp. The authentic ancient DNA is often fragmented compared to the modern DNA, and fragments as short as 50–65 bp are common. The samples included in the downstream analyses yielded between 1426 and 395345 unique human mitochondrial fragments with an average coverage ranging from 5-fold to 1683-fold. The first-base damage on the fragments...
varied between 5–36% on the 3′-end and 4–34% on the 5′-end. Previous studies have proven that cytosine deamination is influenced by the age of the sample and the mean temperature of the site. Considering the climatic conditions in Finland, e.g., low mean temperature, and the relatively young age especially for the post-medieval samples, 3′ and 5′ damage values below 5% are plausible. No samples were therefore omitted from the study based on these criteria.

The contamination rates of the 103 samples were further evaluated by Schmutzi. 36 samples had Schmutzi contamination estimates exceeding 5% and were excluded (Supplementary Table S1). The remaining samples were then analyzed with ContamMix; the resulting crude contamination estimates as well as the a posteriori estimates of contamination along with their 95% confidence intervals (CI) from the MCMC are reported in Supplementary Table S1. The CIs ranged from 0% to 17.2%; in ten cases they exceeded 10%, even though estimates by Schmutzi had remained below 5%. These cases were visually inspected with Geneious 11.0.3 (www.geneious.com). For each of them, the majority call supported the previously assigned haplogroup.

Table 1. Sites presented in this study. *Dating based on the burial contexts. **Defined based on individual’s 14C dates with Oxcal commands ‘Phase/Boundary’. Phase boundaries provide estimate for starts and ends of the site usages. Phase boundaries are here defined for sites with four or more individuals with 14C dates. The mean values are given. Radiocarbon dates include datings presented in this study and datings published elsewhere. See Section 2.2, Supplementary Material S2 and Supplementary Table S2 for more detailed information.
Radiocarbon datings. For this study, we report new 14C dates for 42 individuals (Supplementary Table S1). Radiocarbon dates for nine individuals were determined previously (see Supplementary Table S1). Based on radiocarbon dates and/or dating of the context, the studied burial sites cover the timespan from the Roman Iron Age (300 AD) to historical times (19th century). For sites Levänluhta, Luistari, Hollola, Hiitola, Tuukkala, Pälkäne and Porvoo the highest posterior densities (HPD) for site’s start and end boundaries were determined. The mean values for obtained phase boundaries are presented in Table 1, and 68% and 95% HPD regions are presented in Supplementary Table S2 and in Supplementary Fig. S1. Intervals for mean values of boundaries obtained based on radiocarbon dates were in accordance with dates determined based on the archaeological context (Table 1).

MtDNA data and haplotype variation. A total of 95 unique complete-mitogenome haplotypes were observed among the 103 complete sequences retrieved: three haplotypes were shared between sampling sites and five within a site. In the latter cases, the placement of the skeletal samples suggests that the shared haplotypes have been carried by different individuals, who may have been maternally related: identical haplotypes (haplogroup U5a2a1e) were obtained from remains of a c. 5-year-old child (grave 18, TU666) and an older woman (grave 7, TU655) from Hollola. Identical haplotypes (haplogroup H85) were also observed in a middle-aged adult (grave 6, TU661) and a c. 18-month-old child (grave 15, TU668) from Hollola. At the Hiitola site, identical haplotypes (haplogroup W6) were shared between two individuals from distinct graves (individual TU566 from grave 80 and individual TU675 from grave 30). At the Tuukkala site, two individuals showed identical haplotypes (haplogroup H10e, individuals TU631 and TU645). At Turku, two adults shared the same haplotype belonging to the basal haplogroup H (samples TU582 and TU588). Haplotypes for 103 individuals are presented in Supplementary Table S3.

As the subsequent statistical methods assume that samples derive from unrelated individuals, five samples - one of each identical haplotype pairs within sites (TU666, TU668, TU675, TU645 and TU588) - were removed from the subsequent analyses due to their possible maternal relatedness.

The mean number of pairwise differences, calculated from complete mitochondrial genomes, was highest within Porvoo (MNPD = 33.7 ± 16.8) and lowest within Renko (MNPD = 21.8 ± 10.8) (Supplementary Table S4). Due to the small number of individuals per site and utilization of unique complete mtDNA sequences, haplotype diversities (H) were relatively high (with mean 1.0 and standard deviation ranking from 0.0202 to 0.1768).

MtDNA haplogroup composition at the ancient sites. Burial site-specific haplogroup frequencies of the 98 complete mitochondrial sequences showed considerable between-site variation (Fig. 2 and Supplementary Table S5). The observed frequencies of the main haplogroups in the whole dataset resembled the prevalence among contemporary Finns. As today, haplogroups U and H were the most common, yet with slightly higher overall frequencies than today (U 33.7% vs. 24.1%, and H 41.8% vs. 33.2%). However, when grouped temporally into Iron-Age and medieval sites (IAM) and early-modern and modern sites (EMM), differences were observed: the IAM sites (i.e., Levänluhta, Luistari, Hollola, Hiitola and Tuukkala) demonstrated significantly higher overall prevalence of haplogroup U (40.9%) than the EMM sites (i.e., Pälkäne, Porvoo, Renko, Turku and Hamina, 18.8%) but also high inter-site variability. Among the EMM samples haplogroup H dominated (U 18.8%, H 46.9%). This inter-site variability of the haplogroup U/H ratio had a clear spatial pattern also among the IAM samples. The western cluster (IAM south-west: Levänluhta, Luistari and Hollola) had average U and H frequencies of 58.3% and 27.8%, respectively, whereas the corresponding values in the eastern cluster (IAM east: Hiitola,
Tuukkala) were 20.0% and 53.3%. In IAM east the highest frequency for an individual subhaplogroup was 30.0% obtained for H1. Strikingly, this U/H ratio is the opposite compared to contemporary eastern and western Finns.

**Differences in haplogroup composition between the sites.** Among the 12 Levänluhta samples, five individuals carried haplogroup (hg) U5b, four of which belonged to the sub-hg U5b1b1a. Additionally, the Levänluhta site included three individuals with hg U5a, resulting in a total frequency of 66.7% for hg U5. In contrast, with only two haplotypes of sub-hg H1, the frequency of hg H was well below values observed in modern European populations. The high U5b1b1a frequency resembles that observed today in Saami populations of northern Europe. This actually corresponds well to a related recent study that is showing the close genetic affinity between Levänluhta individuals and modern Saami. However, the Levänluhta individuals also carried mtDNA haplogroups that are absent or rare among the Saami population today, U5a, H1 (0.0–4.0%) and haplogroups K and T. The Levänluhta site clearly showed a unique composition, which resulted in significant genetic distances to all other ancient sites at sequence level, with ΦST values of >10% (see below).

Individuals from the Hollola site, 14C dated to 955–1390 calAD (Table 1), also showed a high overall frequency of hg U (64.3%), similar to Levänluhta. However, differences in subhaplogroup distribution between Hollola and Levänluhta suggest a possible non-modern-Saami-like hunter-gatherer ancestry in this region. Interestingly, subhaplogroup U5b1b1a, typical among contemporary Saami, was not observed in Hollola. In contrast, most of the Hollola U haplotypes belong to haplogroups U4 and U5a (frequencies in Hollola 28.6% and 14.3%, respectively), which are rare or absent in Saami today. Moreover, U4 is also rare in modern Finns while the frequency for U5a is around 6%. Haplotypes belonging to different subhaplogroups of hg H were more common in Hollola than in Levänluhta, occurring in altogether five samples. Haplogroups K and T were absent in the Hollola sample.

A rather different picture emerged from the Luistari samples, showing a substantial genetic distance to Levänluhta (ΦST = 0.134, p < 0.01). Haplogroup U5b1 was entirely absent, and the U haplotypes observed belong to subhaplogroup U4, U5b2 and U2. Lineage U2 is prevalent in some Uralic speaking groups today. The overall haplogroup distribution in Luistari was more similar to the modern European populations dominated by agriculture-associated Neolithic haplogroups H and occurrences of T2 and W1 (see Introduction), than in Levänluhta and Hollola sites.

The two easternmost sites, Hiitola and Tuukkala, proved genetically distant from the western Levänluhta and Hollola sites, despite being approximately contemporaneous with the Hollola individuals. The Neolithic signal in the mtDNA gene pool of ancient Finns in general was much stronger in the east. Both Hiitola and Tuukkala samples showed high frequencies of hg H (61.5% and 47.1%, respectively), together with other Neolithic haplogroups J, K, W and X. Notably, these eastern sites shared three haplogroups: H1a7, H1a8a and H10g. According to GenBank searches these three haplogroups are rare in modern populations: for H1a7 four modern sequences were found, two in Finnish (KY682072 and MF686118), one in Swedish (KJ487971) and one in British (GU797829) populations. For haplogroup H1a8a only two matches were found, one among Finnish (I9153203) and one of an unknown origin (JQ019444), whereas three modern sequences were found for haplogroup H10g: two Finnish (KR732275 and MF497508) and one from Russia (GU122976). Notably, H1a7 is known to be common in modern Karelia. The eastern sites also comprise rare subhaplogroups U1 (hg U1b2 in Hiitola) and U8 (hg U8b1a2b in Tuukkala), which are atypical for contemporary Finns.

Early modern and moderna sites represents similar frequencies of U and H as the combined Iron Age and Medieval East (18.8% and 46.9%, respectively). Contrasting IAM sites and contemporary Finns, EMM sites harbors high prevalence of haplogroup T; frequency in EMM is as high as 21.9%, while in other Finnish populations the frequency is less than 8%.

With the IAM sites, we observed that Levänluhta differed significantly from all the other sites, except Hollola (ΦST = 0.05042, p = 0.02441) and Tuukkala (ΦST = 0.04387, p = 0.06055) (Fig. 3a and Supplementary Table S6). The largest distance from Levänluhta was to the eastern Hiitola site (ΦST = 0.15468). The distance between Levänluhta and contemporary Finns was smaller but still significant, with a distance to contemporary north-east (NE) ΦST = 0.04077 and to contemporary south-west (SW) slightly higher ΦST = 0.06473. While Luistari differed only from Levänluhta, the Hollola site differed both from Hiitola and the EMM (ΦST = 0.05205 and ΦST = 0.05135, respectively, p < 0.05 for both, not from Levänluhta (ΦST = 0.06445, p > 0.05) (Fig. 3a and Supplementary Table S6). Hiitola differed, in addition to Levänluhta and Hollola, from EMM and from both groups of contemporary Finns (ΦST = 0.04111 for NE and ΦST = 0.03437 for SW). When considering the genetic distances between individual sites, it has to be noted that the relatively low sample sizes might affect the ΦST values and the results should be interpreted with caution. However, for pooled IAM and EMM sites (see Fig. 2b), for which the sample sizes are ≥30, the genetic distance calculations should not be that sensitive for bias caused by small sample sizes.

Clustering the IAM sites further roughly according to their geographical location to IAM south-west (hg U more prevalent) and IAM east (hg H more prevalent) further demonstrated the pattern opposite to modern mtDNA diversity distribution (Fig. 3b). IAM south-west differed statistically significantly from contemporary SW (ΦST = 0.01670, p = 0.00488) and EMM (ΦST = 0.05350, p = 0.00098) but not from contemporary NE (ΦST = 0.00036, p = 0.41895). In addition, EMM and contemporary SW differed from each other (ΦST = 0.01140,
p = 0.04102). Conversely, IAM east differed from the contemporary NE ($\Phi_{ST} = 0.00849$) more than from contemporary SW ($\Phi_{ST} = 0.00514$).

Haplotype level median-joining network (Supplementary Fig. S3) demonstrates that ancient and contemporary Finns exhibit in principle same main haplogroups, whereas the most notable differences are within the haplogroup frequencies between the ancient populations. Individuals from IAM eastern sites are more prevalent in the haplogroup H cluster, while individuals from IAM southwestern sites are more concentrated on the haplogroup U cluster. Contemporary Finns are in both clusters, indicating possible mixture of IAM southwestern and IAM eastern populations.

**Figure 3.** (a) Pairwise $\Phi_{ST}$ distances for ancient and contemporary Finns. Early modern and modern Finns consists of individuals from Pälkäne, Porvoo, Renko, Julin and Hamina sites. Contemporary Finns are divided into south-west (SW) and north-east (NE) subpopulations according to $\Phi_{ST}$ values and $p$-values are presented in Supplementary Table S6). (b) Pairwise $\Phi_{ST}$ distances for ancient and contemporary Finns. Iron-Age and medieval (IAM) sites are grouped into subpopulations: IAM south-west consists of Levänluhta, Luistari and Hollola; IAM east consists of Hiitola and Tuukkala. Early modern and modern Finns contain individuals from Pälkäne, Porvoo, Renko, Julin and Hamina sites. Contemporary Finns are divided into south-west (SW) and north-east (NE) subpopulations according to Palo et al.$^{27}$ and Neuvonen et al.$^{28}$.

**Main haplogroup frequencies in space and time.** To evaluate the possible impact of spatial and temporal factors on the distributions of haplogroup U, largely associated with European hunter-gatherers, and farmer-associated haplogroup H within the IAM sites, we performed multinomial logistic regression analyses. In a stepwise forward analysis, the only statistically significant independent variable explaining the differences in the haplogroup composition was the distance from eastern reference point Lahdenpohja (compared to ‘H’ and ‘Others’ significance for Lahdenpohja was 0.013 and 0.103, respectively) (Supplementary Table S7). Neither the ages of the samples nor distance from the southern and western reference points were requisite for the best-fit model. However, the addition of the eastern reference point significantly improved the fit between model and data ($p = 0.027$). Based on the odds ratios, it is less likely that an individual from southwest belongs to haplogroups ‘H’ or ‘Others’ than an individual from an eastern archaeological site. Similar results were obtained when using hunter-gatherer associated haplogroups (U and V), farmer associated haplogroups (H, J, K and T) and ‘Others’ as categorically distributed dependent variables. We chose to include the haplogroup V as ‘hunter-gatherer’ while there is no direct evidence for association of hg V with the hunter-gatherers. This is assumed here because of V’s northern distribution and its high prevalence (up to 58%) among the Saami, the archetypal nomadic population lacking many farmer-associated haplogroups$^{34,38}$. Distance from the eastern reference point was the only predictor included in the model (with significance of 0.031 for farmer associated haplogroups and 0.082 for other haplogroups). Assuming that haplogroups U and H can be associated to hunter-gatherers and farmers, respectively, the results suggest a spread of the more central European like, farmer-related haplogroups spreading from the east. However, as mentioned above, association of hg V is unclear. Omitting V from the hunter-gatherer group does not change results noteworthy (Supplementary Table S7).

**Genetic affinities of ancient Finns to other ancient and contemporary populations.** To further explore the affinity of ancient Finns to other ancient and contemporary populations, we carried out principal component analysis (PCA) based on haplogroup frequencies. We plotted the first two components of the PCA plot for ancient Finns, 31 other ancient populations, contemporary Finns and Saami, which account for 55% of the total variance (Figs 4, S4). Interestingly, southwestern Iron-Age sites Levänluhta and Hollola fall close to hunter-gatherer populations from Baltic, Central and Southern Europe. In addition Levänluhta is located in proximity to modern day Saami. This suggests the hunter-gatherer type of maternal ancestry in these two sites. In contrast, eastern IAM sites Hiitola and Tuukkala, EMM sites and contemporary SW Finns clustered with European Neolithic, Bronze-Age and Iron-Age populations. The southwestern site Luistari, as well as the contemporary NE Finns, were located roughly between these two clusters, indicating a possible mix of maternal ancestry
Discussion

Here we report 103 human mitochondrial DNA genomes from approx. 300 AD to 1800 AD, a transect both in time and space, which represents thus far the largest collection of individuals with ancient human DNA analyzed from Finland. Analysis of the prehistoric samples from Iron-Age and medieval sites from western and eastern Finland revealed a high overall prevalence of haplogroup U in southwestern sites, in stark contrast with a high frequency of haplogroup H in the east, which is opposite to what is observed in modern day Finland. Moreover, there is relatively high differentiation between the ancient sites.

Genetic layers of mitochondrial variation among the Iron Age and Medieval Finns. Unexpectedly high variation in maternal lineages could be observed between the southwestern Levänluhta, Luistari and Hollola Iron-Age sites. Especially the distribution of U subhaplogroups differed clearly between sites: The oldest site, Levänluhta, represented a high frequency of U5a and the modern Saami-related haplogroup U5b1b1a, which is present in contemporary Finns only in moderate frequency of around 3.0%\(^35\). Indeed, recent studies considering Levänluhta, in which nuclear genomes have been retrieved, confirm the genetic continuation with the modern Saami population\(^18,19\). The strong drift experienced by the Saami groups of present day, shown by their high levels of LD throughout genome and low diversity in uniparental markers (See\(^38\) and references therein) could explain why some mtDNA lineages, such as U5a, would have vanished from present day Saami.

Both Luistari and Hollola lacked the U5b1b1a, but instead Hollola displayed a wider variety of other U subhaplogroups, such as U5a1, U5a2 and U5b2. In addition, Luistari and Hollola sites showed relatively high frequencies of different subhaplogroups of U4 (i.e., U4a, U4b and U4d), which are rare in contemporary Finns and absent from modern Saami. Instead, in contemporary populations, U4 exists in high frequencies in Volga-Ural region (up to 24% in Komi-Zyryans)\(^36\) and with lower frequencies around the Baltic Sea, such as in Latvians and Tver Karelians (both around 8%)\(^37\). Taking into account that U4 have been prevalent in neighboring areas among Scandinavian\(^10,39–41\), Baltic Comb Ceramics Culture\(^12,13,44\), Baltic Comb Ware Culture\(^12,14\) and in Siberia during the Early metal period\(^11\), we might be observing ancestries belonging to an earlier layer of ancient inhabitants of the region.
Taking these different distributions of mtDNA haplogroups from the Iron-Age and medieval sites into consideration, our results suggest three different streams of mitochondrial ancestry: Saami-like haplogroups (U5b1b1a, possible also U5a), non-Saami-like hunter-gatherer related haplogroups (especially U4) and haplogroups associated with Neolithic farmers (H, J, K and T). In this context we use ‘Saami-like’ as a term that shows genetic continuity with modern-day Saami groups. Different proportions of these ancestries could be observed both in later EMM sites and also modern-day southwestern and northeastern Finns. This suggests a fluctuation of each of these mitochondrial ancestry proportions over space and time.

The ancient distribution of mtDNA lineages contradicts the contemporary east-west divergence.

The Finnish population has been a subject of multitude medical genetic studies for many decades. Assessments of genetic diversity have revealed a number of idiosyncrasies in the modern Finnish gene pool. These include, for instance, the enrichment of c. 40 rare genetic diseases and the absence of some major ones in the rest of European metapopulation, as a clear distinction from the largely clinal differences observed in most of Europe. Furthermore, these studies have demonstrated the existence of notable genetic differentiation between southwestern and northeastern parts of Finland. This differentiation is especially pronounced in Y-chromosomes, showing opposite frequency trends of haplogroups N1c (25% SW, 75% NE) and I (56% SW, 24% NE).

The modern mitochondrial DNA diversity in Finland resembles that observed in the Central Europe, but holds a relatively high overall frequency of haplogroup U, and also a notable proportion of subhaplogroups which have frequency peaks in or are exclusive to Finland. The genetic structure within Finland is minimal when at the level of mtDNA haplotypes are considered, but pronounced in the frequencies of haplogroups assumed Paleolithic (here U and V) or Neolithic (H, J, K, T) in Europe: the paleolithic haplogroups are more common in the north-east (“Contemporary NE”) subdivisions in the south-west (“Contemporary SW”). This, together with the Y-chromosomal subdivision, has been interpreted to reflect an ancient border between populations relying on farming (south-west Finland) and foraging (north-east). The observed genetic border running diagonally from north-west to south-east coincides with differences in a number of linguistic and cultural differences all the way to folk traditions. It also coincides with the first medieval political border, the Treaty of Nöteborg, between the Swedish and Novgorodian spheres of influence agreed in 1323 AD (see48).

The ancient mitochondrial genomes analyzed here show a notable pattern opposite to the modern variation: mtDNA types usually associated with the hunter-gatherer communities were significantly more common in the ancient western cluster (Levänluhta, Luistari and Holliola) than in the east (Hiitolä, Tuukkala), with the haplogroup U frequency as high as 58.3%. In contrast, the farming-related lineages were observed in particular in the ancient eastern cluster. This pattern of division between the ancient sites, and the contradictions with their respective local modern population frequencies emerged also in formal testing of pairwise \( \Phi_{ST} \) values: the western cluster was closer to the modern NE subpopulation than to the modern SW subpopulation whereas the eastern cluster showed closer affinity with the modern mtDNA variation in southwestern Finland.

Bidirectional expansion of agriculturally oriented populations into Finland?. Assuming that the haplogroup composition has correlated with the mode of subsistence, the observed pattern of east to west transect suggests a bidirectional spreading of agricultural human groups into Finland. Although there is evidence of sporadic small-scale cultivation in southeastern Finland already during the Neolithic Stone Age (c. 5300–4000 BC), the start of agriculture in Finland has been traditionally associated with the Corded-Ware Culture (CWC) arriving across the Baltic Sea approximately 4700 years ago. Indeed, there are scattered findings of animal husbandry from southwestern parts of country starting from c. 2500 calBC, but in general archaeological evidence supporting transition to agriculture as a consequence of introduction of Corded-Ware culture, are sparse (for discussion see49). Some independent observations of animal domestication and cultivation (see and references there in) are identifiable during the Bronze Age, but documentation remains still very limited. This suggests that cultivation has probably been relatively uncommon and local for centuries, as little direct evidence for cereal cultivation in Finland prior Iron Age exists. Polen records show notable increase of cereals starting only at 100 AD and reaching maximum as late as 1300 AD overlapping the time span of Iron Age and Medieval sites presented in this study.

As a support for the late introduction of farming populations in to Finland we do not see strong affinities of western IAM to for example the CWC maternal gene pools from Estonia and Lithuania, suggesting either that the mtDNA gene flow between these two regions has been low or that shared mtDNA variation had dissolved before the Iron-Age in Finland. Alternatively, the CWC expansion may have been largely male-driven as suggested by50. However, we observe a strong Neolithic signal in the Iron-Age mtDNA pool in Eastern Finland, thus rather suggesting a southeastern/eastern arrival route of an agro-pastoralist population into the country. Interestingly, their maternal genetic legacy also corresponded to the contemporary modern Finland, especially in SW. We therefore propose that either there has been east-to-west directed gene flow during the Middle Ages, after the introduction of agricultural haplogroups into the east, or that the late change in SW maternal gene pool may reflect recent immigration from more western/southern sources, such as the migration from Sweden during the Swedish reign in Finland (from 1200s–1809). Iron-Age has evidenced high mobility around the Baltic Sea, as evidenced by the genetic and isotope analyses of human remains from 10th to 12th century in Sigtuna, eastern Sweden.

The multinomial logistic regression analysis lent support to the eastern introduction of agriculturally related maternal ancestry. The likely migration routes for the observed ancestral elements were investigated through different combinations of factor dependencies as the multinomial logistic regression. The test revealed the distance from Lahdenpohja on the eastern border of Finland as the only statistically significant variable explaining the differences in the haplogroup composition. Neither time scale, nor distance from the southwestern locations
The Neolithic farmer-related signal in the mtDNA diversity in the Iron-Age samples is mainly found in the southeast, whereas in contemporary population it predominates in the southwest. The reasons for this discrepancy are likely diverse, and could be affected by such recent events as the evacuation of nearly 0.5 M inhabitants during the World War II and their resettling into the area of current Finland. However, these evacuees were resettled rather evenly across southern Finland and should not create the observed pattern. It rather suggests that the division between SW and NE Finland had still been more substantial in the early 1900s. Another, more fundamental explanation for the genetic subdivision comes from the environmental demands of sedentary farming. In southwestern Finland the soil is more amenable to field-farming and due to the warming Atlantic effect that gradually shades into more arid continental climate, the growing season in Finland is the longest in the southwestern coast. These environmental differences follow the NW-SW border similarly to the genetic distances. As the country has been sparsely inhabited until modern times, it is plausible that farming oriented populations, in search of more favorable conditions, have over the centuries concentrated into the SW parts of the country.

The mitochondrial DNA genomes from Iron-Age Finland show variation that can be linked to either hunter-gatherer or agricultural human groups. These elements are still present in the mitochondrial gene pool of contemporary Finns but relatively evenly distributed throughout the country. In contrast, the Iron-Age mtDNA variation show significant differences between sampling sites, with hunter-gatherer and farmer-associated elements dominating in different regions than today. Rather surprisingly, the agricultural population signal has been stronger in eastern Finland in the past, which might reflect a bidirectional arrival of farming-associated populations into Finland.

Materials and Methods

Sample selection. The human skeletal remains used in this study were collected from five archaeological sites and five historical cemeteries (for more detailed information of the sites and references for original publications in Supplementary Material S1). Archaeological sites include Levänluhta, used as a burial place from Roman Iron Age until the end of Merovingian Period (archaeological dating 300–800 AD), Luistari, consisting of graves from Merovingian to Crusade Period (archaeological dating 600–1200 AD) and Hollola, Hiitola and Tuukkala, largely Christian-style cemeteries spanning from Crusade Period to Early Middle Ages (archaeological datings 1050–1400 AD, 1200–1500 AD and 1200–1400 AD, respectively) (Fig. 1, Table 1 and Supplementary Material S1). The five historical sites include the cemetery of the Church of St. Michael in Päkkäne, the cemetery of the Church of St. Jacob in Renko, the Cathedral site in Porvoo, the Julin's site in Turku, and the Ryazan regimental church cemetery in Hamina. Samples were obtained from the archaeological collections of the Finnish Heritage Agency, Department of Archaeology in the University of Turku, Department of Anatomy in the University of Helsinki, and the Peter the Great Museum of Anthropology and Ethnography (Kunstkamera), Russian Academy of Sciences.

Contextual archaeological evidence, such as grave goods and burial customs, together with radiocarbon analyses were used to confirm the dating of each site and/or individual. Details of dating, sample sizes and number of genomes obtained are presented in Table 1. Detailed information on individual burial sites and samples are given in Supplementary Material S1 and Supplementary Table S1.

Reference populations used in comparative analyses. To evaluate possible changes in the Finnish mitochondrial gene pool during the past thousand years, samples were compared to HVR1 + HVR2 (16024–16385, 72–340) data from 832 modern Finns for whom the county-level geographical origin is known. In order to compare the mitochondrial profile of ancient Finns to other ancient nearby populations, haplogroup frequencies were collected from 31 ancient populations (Supplementary Table S8).

Sampling. All samples were processed in dedicated aDNA facilities with regularly UV:d and bleach-treated laminar hoods, inside a clean room space. For Levänluhta, Porvoo, Renko, Päkkäne and Hamina sites, sampling was conducted at the University of Tübingen, Germany from the start. For the Luistari, Hiitola, Hollola and Tuukkala sites, the bone powder was produced in a clean-room space for small-scale ancient DNA work at the Helsinki University Department of Forensic Medicine, and then transferred to Tübingen University facilities, stored in plastic tubes. All subsequent laboratory work with the bone powder was conducted at the facilities in Tübingen. A protective overall, facemask, hair net and two layers of disposable gloves were used at all times when handling the samples. Decontamination was carried out by ultraviolet light exposure of the plastic ware, reagents and samples, and by removal of the immediate surface at the point of sampling before drilling into the bone. Teeth were sawed in two, with the exception of the Levänluhta samples, where dentine was already exposed due to heavy fragmentation. Dental pulp and the surrounding dentine were used for DNA sampling, drilling into the crown or the root. For the petrous part of the temporal bone, a wedge was cut off to reveal the inner ear channels and sample was taken from the inside of the channel, as previously described85. A dentist’s drill together with cooled-down drill heads for minimal heat exposure were used at the University of Tübingen facility and a “field kit” with Dremel or dentist’s drill were used at the Helsinki University Department of Forensic Medicine.

Extraction of ancient DNA. The extraction was performed according to a modified version of the original protocol83. For each sample, ~50 mg of bone powder was used for the extraction by eluting it in 100μl of TET...
of 10–16 samples. As well as two negative laboratory controls to measure the levels of contamination were carried along for every batch.

A positive control extracted from a cave bear bone, to confirm the success of extraction and library preparation, as well as amplification of PCR, with heating cycles chosen individually per library according to the copy number after indexing. Amplified libraries were purified using MinElute spin columns with the standard protocol provided by the manufacturer (Qiagen). A qPCR together with Agilent Bioanalyzer 2100 device, and a DNA1000LabChip were subsequently used on the amplified libraries to measure the concentration of DNA, as well as fragment size distribution. A successful library composition and admissible indexing efficiency. The indexed libraries were amplified using PCR, with heating cycles chosen individually per library according to the copy number after indexing.

Mitochondrial capture and sequencing. Mitochondrial genomes were achieved using a mitochondrial in-solution capture as described in Maricic et al. Complete human mitochondrial DNA sequence was used to produce in-house made baits, which were ligated to adapters. The bait DNA was then purified and denatured to the single stranded form, and attached to streptavidin-coated magnetic beads. Pools of 4–6 samples, combined in equal mass ratios for altogether 2 μg of DNA, were captured with the above mentioned beads, and sequenced on the Illumina platforms: for samples from sites Turku, Hiitola, Tuukkala and Hollola, along with four samples from the Luistari site (TU619, TU621, TU622, TU623), single-end sequencing data was produced on HiSeq4000 run for 75 + 8 + 8 cycles, whereas samples from all the other sites, including the rest of the samples from Luistari, paired-end data was produced with NextSeq500 for 2 × 150 + 8 + 8 cycles, at the Max Planck Institute for the Science of Human History, Jena.

Processing of the sequence data. Raw-read sequencing data were processed using the EAGER-pipeline for aDNA sequencing data. Reads were demultiplexed using both indices, and the adapters were clipped off with the AdapterRemoval program integrated in EAGER, with minimum overlap set to 1 bp. Short reads reaching a minimum of 10 bp overlap were merged, validated by the paired-end fragment compatibility. For the single-end-sequenced samples this step was omitted. Minimum read length of 30 bp and base quality of 20 was required for all final reads. The reads were aligned to the complete human reference genome Hg19 with the BWA mapping algorithm for the shotgun sequence data, whereas CircularMapper, a custom-made tool for circular genomes included in EAGER, was used to map the enriched mitochondrial reads to the human mitochondrial reference genome (updated Cambridge reference sequence, rCRS).

Authentication of ancient DNA and haplogroup assignments. In order to evaluate the authenticity of the reconstructed mitochondrial genome sequences, contamination rates, read-length distributions and deamination patterns at the 5′ and 3′ ends of DNA fragments were inspected for each sample with the program MapDamage integrated in EAGER. Further contamination estimation was carried out with Schmutzi and ContamMix programs.

The mitochondrial consensus-calling in Schmutzi was used to produce the consensus sequences of both the endogenous source and the most likely single contaminating source. Complete consensus sequences were called against the rCRS with a filter value of q20. For the Schmutzi-based consensus sequences a manual correction was performed for position 3107 to correspond to the “N” embedded in rCRS. The mtDNA haplogroups were determined using HaploGrep with respect to PhyloTree version 17.

For contamination estimation in ContamMix, consensus sequences created by Schmutzi were combined with a reference dataset of 311 mitochondrial genomes from worldwide populations (provided by ContamMix) and aligned with mafft version 7.305. The untrimmed mitochondrial reads from the Eager pipeline were then extracted from BAM files into fastQ files and mapped back to the assembly. ContamMix then evaluates whether the reads assign more probably to their respective consensus or one of the worldwide mitochondrial genomes, i.e. possible contaminant source. The ContamMix was run with trimming of seven bases of each side of the read to remove the accumulated damage typical for ancient DNA.

The mitochondrial genome sequences with highest ContamMix estimates were further visually inspected in Geneious 11.0.3 (www.geneious.com). In this inspection, the majority call support for relevant diagnostic mutations against the rCRS reference genome was compared to the PhyloTree version 17. We applied the automated variant caller in Geneious to the alignments with minimum support of 3x coverage and variant frequency of 66.6% for diagnostic SNPs to confirm the authenticity of the haplogroup assignments.

Radiocarbon dating. Radiocarbon dates were produced by the Laboratory of Chronology, Finnish Museum of Natural History Luomus in Helsinki, Finland (Hela) and Klaus-Tschira C14-laboratory in Mannheim, Germany (MAMS). Bone collagen was extracted with the modified Longin method. The collagen samples were combusted, graphitized and measured by using Accelerator Mass Spectrometry (AMS). The results are provided as conventional radiocarbon dates without potential reservoir effect corrections. The radiocarbon dates were calibrated using the OxCal program version 4.3, IntCal 13 as the calibration curve. For sites with four or more individuals with 14C dates also the boundaries for phase’s start and end were determined with OxCal 4.3. Example of the OxCal code is given in Supplementary Material S2. Timescale discussed throughout the text is defined as calendar years.
level analysis. Further to evaluate the relation of ancient individuals to contemporary Finns on a haplotype level, the best-fit models for different datasets were estimated with jModelTest 70. The substitution model used was Tamura & Nei71 with gamma correction (shape parameter 'H

To evaluate the impact of grouping the haplogroups into hunter-gatherer and farmer related hgs' also categories 'U

Stepwise forward method was used with entry probability 0.05, and probability was tested with likelihood ratios. To

Overview of statistical analyses and datasets used in this study.

| Analysis | Data type | Populations used | Results presented in |
|----------|-----------|------------------|----------------------|
| Basic diversity indices | Complete sequences | Ancient Finns | Supplementary Table S4 |
| \(\Phi_{ST}\) | HVR1 + HVR2 | Ancient and contemporary Finns | Fig. 3 and Supplementary Table S6 |
| Multinomial logistic regression | Presence/absence of haplogroup | Ancient Finns | Supplementary Table S7 |
| Principal component analysis (PCA) | Haplogroup frequencies | Ancient Finns, other ancient and contemporary populations (Supplementary Table S8) | Figs 4 and S4 |
| Evaluation of the possible sampling bias in the observed haplogroup frequencies | Haplogroup frequencies | Ancient and contemporary Finns | Supplementary Fig. S2 |
| Network analysis | HVR1 + HVR2 sequence data | Ancient and contemporary Finns | Supplementary Fig. S3 |
| Evaluation the impact of small sample size on PCA | Haplogroup frequencies | Contemporary Finns, ancient Finns and other ancient populations (reference populations presented in Supplementary Table S8) | Supplementary Fig. S5 |

Table 2. Overview of statistical analyses and datasets used in this study.

Statistical analyses. A summary of statistical analyses performed in this study for different sets of populations and dataset (i.e., complete sequence, HVR1 + HVR2, haplogroup frequencies) is presented in Table 2.

For the statistical analysis, sequences were aligned with Muscle 3.8.3168. To explore the genetic diversity within ancient populations, basic diversity indices such as haplotype diversity, mean pairwise distance and nucleotide diversity were calculated for the complete mitochondrial genomes with Arlequin 3.5.2.290. To determine genetic distances between ancient and modern Finns, pairwise \(\Phi_{ST}\) values based on the sequence data were calculated with Arlequin 3.5.2.2. Genetic distances, were calculated by utilizing HVR1 + HVR2, as the contemporary reference data with detailed geographical origin were restricted to HVR regions only27,28. To estimate the significance for the \(\Phi_{ST}\) values, permutation tests with 10000 permutations were used. The visualization of the \(\Phi_{ST}\) values was done by R heatmap.2 function with hierarchical clustering based on the Euclidean distance. The best-fit models for different datasets were estimated with jModelTest9. The substitution model used was Tamura & Nei71 with gamma correction (shape parameter \(\alpha = 0.67\)) and Tamura & Nei with gamma correction (\(\alpha = 0.44\)) for the complete sequence data and HVR1 + HVR2 data, respectively. Poly-C region (positions 309–315), AC indels (positions 515–522) and mutational hotspot at position 16519 were masked for the population level analysis. Further to evaluate the relation of ancient individuals to contemporary Finns on a haplotype level, median-joining network analysis52 was performed with PopArt23. Positions bearing more than 5% of missing data were masked for the network analysis.

To statistically test the impact of possible factors affecting the spatial and temporal distribution of haplogroups U and H in our ancient samples, we conducted a multinomial logistic regression analysis. Each individual was considered as separate occurrence and haplogroups ‘U’, ‘H’ and ‘Others’ were set as categorical dependent variables. Because the main interest was to estimate the impact of time and geography on the occurrence of the haplogroups, the following independent variables were chosen: (1) Median age of sample (or the mean age of the site if radiocarbon dates were not available) (2) site’s distance (in km) from southern reference point Hanko, (3) site’s distance (in km) from western reference point Uusikaupunki and (4) site’s distance (in km) from eastern reference point Lahdenpohja (see Fig. 1). These geographical points were chosen to represent the most plausible entry points of different migration routes to the study area, south (Hanko on the southern coast), west (Uusikaupunki on the western coast) and east/south-east (Lahdenpohja). The last mentioned represents the eastern migration route both along the Karelian Isthmus and north of Lake Ladoga, strongly supported by archaeological evidence21. For multinominal logistic regression a stepwise forward method was used with entry probability 0.05, and probability was tested with likelihood ratios. To evaluate the impact of grouping the haplogroups into hunter-gatherer and farmer related hgs’ also categories ‘U + V’, ‘H + J + K + T’ and ‘Others’ were tested. In addition, categories ‘U’, ‘H + J + K + T’ and ‘Others’ were tested, due to controversial definition of haplogroup V (See Section 2.7). Analyses were performed with IBM SPSS version 25 (IBM Corp. Released 2017, IBM SPSS Statistics for Windows, Version 25.0 Armonk, NY: IBM Corp.).

To trace genetic affinities between ancient Finns and other ancient populations, we visualized haplogroup composition of each site (based on haplogroup frequencies) using principal component analysis (PCA). PCA was computed using MATLAB and Statistical Toolbox Release 2015b (The MathWorks, Inc., Natick, Massachusetts, United States). Populations and haplogroup frequencies used are presented in Supplementary Table S8.

Data Availability

Complete mitochondrial sequences will be deposited in GenBank under accession numbers MN540463-MN540565.

References

1. Pooth, C. et al. Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a Late Glacial population turnover in Europe. Current Biology 26, 827–833 (2016).
2. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. Nature 524, 216 (2015).
3. Bramanti, B. et al. Genetic discontinuity between local hunter-gatherers and central Europe’s first farmers. Science 326, 137–140 (2009).
47. Herva, V., Mökkönen, T. & Nordqvist, K. A northern Neolithic? Clay work, cultivation and cultural transformations in the boreal

49. Lahtinen, M. & Rowley-Conwy, P. Early farming in Finland: was there cultivation before the Iron Age (500 BC)?

36. Bermisheva, M., Tambets, K., Villems, R. & Khusnutdinova, E. Diversity of mitochondrial DNA haplotypes in ethnic populations of

37. Lappalainen, T.

33. Fu, Q.

32. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus

31. Kistler, L., Ware, R., Smith, O., Collins, M. & Allaby, R. G. A new model for ancient DNA decay based on paleogenomic meta-

30. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA

11. Der Sarkissian, C.

18. Lamnidis, T. C.

15. Ahola, M., Salo, K. & Mannermaa, K. Almost Gone: Human Skeletal Material from Finnish Stone Age Earth Graves.

22. Lahtinen, M., Oinonen, M., Tallavaara, M., Walker, J. W. & Rowley-Conwy, P. The advance of cultivation at its northern European

23. Lao, O.

24. Salmela, E.

10. Günther, T.

9. Kashuba, N.

8. Stolarek, I.

7. Haak, W.

6. Allentoft, M. E.

5. Haak, W. et al. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. Science 310, 1016–1018 (2005).

5. Haak, W. et al. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. PLoS Biology 8, e1000536 (2010).

4. Haak, W. et al. Population genomics of Bronze Age Eurasia. Nature 522, 167–172 (2015).

7. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522, 207–211 (2015).

3. Stolarz, I. et al. A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age. Scientific

28. Cramp, L. et al. Neolithic dairy farming at the extreme of agriculture in northern Europe. Proceedings of the Royal Society B: Biological

27. Mathieson, I. et al. The genomic history of southeastern Europe. Nature 555, 197 (2018).

26. Norio, R. Finnish disease heritage I. Hum. Genet. 112, 441–456 (2003).

25. Alenius, T., Mökkönen, T. & Lelhamä, A. Early Farming in the Northern Boreal Zone: Reassessing the History of Land Use in

24. Herva, V., Mökkönen, T. & Nordqvist, K. A northern Neolithic? Clay work, cultivation and cultural transformations in the boreal

23. Skoglund, P.

22. Skoglund, P. et al. Ancient DNA reveals prehistoric gene-flow from Siberia in the complex human population history of North

21. Tallavaara, M., Pesonen, P. & Oinonen, M. Prehistoric population history in eastern Fennoscandia. Journal of Archaeological Science

20. Carpelan, C. Käännekohtia Suomen esihistoriassa aikavälillä 5100–1000 eKr.

19. Sikora, M.

18. Pesonen, P., Oinonen, M., Carpelan, C. & Onkamo, P. Early Subneolithic ceramic sequences in eastern Fennoscandia—a Bayesian

17. Lahtinen, M., Oinonen, M., Tallavaara, M., Walker, J. W. & Rowley-Conwy, P. The advance of cultivation at its northern European

16. Carpelan, C. Käännekohtia Suomen esihistoriassa aikavälillä 5100–1000 eKr. (Gaudeamus, 2015).

15. Ahola, M., Salo, K. & Mannermaa, K. Almost Gone: Human Skeletal Material from Finnish Stone Age Earth Graves. Fennoscandia

14. Mathieson, I. et al. The genomic history of southeastern Europe. Nature 555, 197 (2018).

13. Kistler, L., Ware, R., Smith, O., Collins, M. & Allaby, R. G. A new model for ancient DNA decay based on paleogenomic meta-

12. Der Sarkissian, C. et al. Ancient DNA reveals prehistoric gene-flow from Siberia in the complex human population history of North

11. Der Sarkissian, C. et al. Ancient DNA reveals prehistoric gene-flow from Siberia in the complex human population history of North
50. Bläuer, A. & Kantanen, J. Transition from hunting to animal husbandry in Southern, Western and Eastern Finland: new dated osteological evidence. Journal of Archaeological Science 40, 1646–1666 (2013).
51. Goldberg, A., Günther, T., Rosenberg, N. A. & Jakobsson, M. Ancient X chromosomes reveal contrasting sex bias in Neolithic and Bronze Age Eurasian migrations. Proceedings of the National Academy of Sciences, 201616392 (2017).
52. Krzewinska, M. et al. Genomic and Strontium Isotope Variation Reveal Immigration Patterns in a Viking Age Town. Current Biology 28, 2730–2738, e10 (2018).
53. Pinhasi, R. et al. Optimal ancient DNA yields from the inner ear part of the human petrous bone. PloS One 10, e0129102 (2015).
54. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010, pdb.prot5448 (2010).
55. Kircher, M. In Ancient DNA. Human Evolution 197–228 (Springer, 2012).
56. Maricic, T., Whitten, M. & Paabo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PloS One 5, e14004 (2010).
57. Peltre, A. et al. EAGER: efficient ancient genome reconstruction. Genome Biology 17, 60 (2016).
58. Andrews, R. M. et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature Genetics 23, 147–147 (1999).
59. Ginolhac, A., Rasmussen, M., Gilbert, M. T. F., Willerslev, E. & Orlando, L. mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics 27, 2153–2155 (2011).
60. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. (2016).
61. Van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Human Mutations 30, E386–E394 (2009).
62. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30, 772–780 (2013).
63. Katoh, K. et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30, 3059–3066 (2002).
64. Longin, R. New method of collagen extraction for radiocarbon dating. Nature 230, 241 (1971).
65. Bocherens, H. et al. Paleobiological implications of the isotopic sigNature (13 C, 15 N) of fossil mammal collagen in Scladina Cave (Sclayn, Belgium). Quatern. Res. 48, 370–380 (1997).
66. Ramsey, C. B. Bayesian analysis of radiocarbon dates. Radiocarbon 51, 337–360 (2009).
67. Reimer, P. J. et al. Selection and treatment of data for radiocarbon calibration: an update to the International Calibration (IntCal) criteria. Radiocarbon 55, 1923–1945 (2013).
68. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792–1797 (2004).
69. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources 10, 564–567 (2010).
70. Darrida, D., Tabosada, G. L., Doallo, R. & Posada, D. [ModelTest 2: more models, new heuristics and parallel computing. Nature methods 9, 772 (2012).
71. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology Evolution 10, 512–526 (1993).
72. Bandelt, H., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology Evolution 16(1), 37–48 (1999).
73. Leigh, J. W. & Bryant, D. PopART: full-feature software for haplotype network construction. Methods Ecological Evolution 6(9), 1110–1116 (2015).
74. Talve, I. Finnish Folk Culture. Finnish Literature Society; English edition. pp 349 (2000)

Acknowledgements

We are grateful for the Leväänluhta project and Elämän historiat hampaissa project for providing us samples from Leväänluhta and Luistari, respectively. For the Julini’s site samples we would like to thank Benito Casagrande, Liisa Seppänen, Sirkku Pihlman, Heikki Vuorinen and Juha Varrela. For the samples from Renko and Porvoo, professor Helena Ranta and University of Helsinki are acknowledged. We would also like to thank Markku Niskanen for providing us the 14C dating for one Renko individual (Renko H29, JK1927). Ella Reiter and Shweta Niskanen for providing us the 14C dating for one Renko individual (Renko H29, JK1927). Ella Reiter and Shweta Niskanen for providing us the 14C dating for one Renko individual (Renko H29, JK1927). Ella Reiter and Shweta Niskanen for providing us the 14C dating for one Renko individual (Renko H29, JK1927). Ella Reiter and Shweta Niskanen for providing us the 14C dating for one Renko individual (Renko H29, JK1927). Ella Reiter and Shweta Niskanen for providing us the 14C dating for one Renko individual (Renko H29, JK1927). Ella Reiter and Shweta Niskanen for providing us the 14C dating for one Renko individual (Renko H29, JK1927).

Author Contributions

K.S., L.A., H.E.-S., V.L., E.M., J.-P.T., K.V., A.W. provided archaeological material and related information. K.M., S.O., M.P. performed the sampling. K.M., S.O., S.P. performed the laboratory work. K.M., S.O. processed the sequence reads and generated the mtDNA genotypes. S.O., J.U.P., E.S. performed the statistical analyses. M.O., L.A., H.E.-S. performed the radiocarbon dating analysis. S.O., M.O. post-processed the 14C datings (i.e. phase start and end boundaries). S.O., J.U.P., K.M., P.O., E.S., K.S. wrote the manuscript with input from all co-authors. P.O., J.K., J.U.P., V.J.S., W.H., A.S. supervised the study.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-51045-8.

Competing Interests: The authors declare no competing interests.

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