

# Repeated plague infections across six generations of Neolithic Farmers

<https://doi.org/10.1038/s41586-024-07651-2>

Received: 1 September 2023

Accepted: 3 June 2024

Published online: 10 July 2024

Open access

 Check for updates

Frederik Valeur Seersholm<sup>1✉</sup>, Karl-Göran Sjögren<sup>2</sup>, Julia Koelman<sup>3</sup>, Malou Blank<sup>2</sup>, Emma M. Svensson<sup>3</sup>, Jacqueline Staring<sup>4</sup>, Magdalena Fraser<sup>3</sup>, Thomaz Pinotti<sup>1,5</sup>, Hugh McColl<sup>1</sup>, Charleen Gaunitz<sup>1</sup>, Tatiana Ruiz-Bedoya<sup>3,6</sup>, Lena Granehäll<sup>3,7</sup>, Berenice Villegas-Ramirez<sup>3</sup>, Anders Fischer<sup>8</sup>, T. Douglas Price<sup>2</sup>, Morten E. Allentoft<sup>1,9</sup>, Astrid K. N. Iversen<sup>10</sup>, Tony Axelsson<sup>2</sup>, Torbjörn Ahlström<sup>11</sup>, Anders Götherström<sup>12,13</sup>, Jan Storå<sup>13</sup>, Kristian Kristiansen<sup>1,2</sup>, Eske Willerslev<sup>1,14</sup>, Mattias Jakobsson<sup>3,15</sup>, Helena Malmström<sup>3,15</sup> & Martin Sikora<sup>1✉</sup>

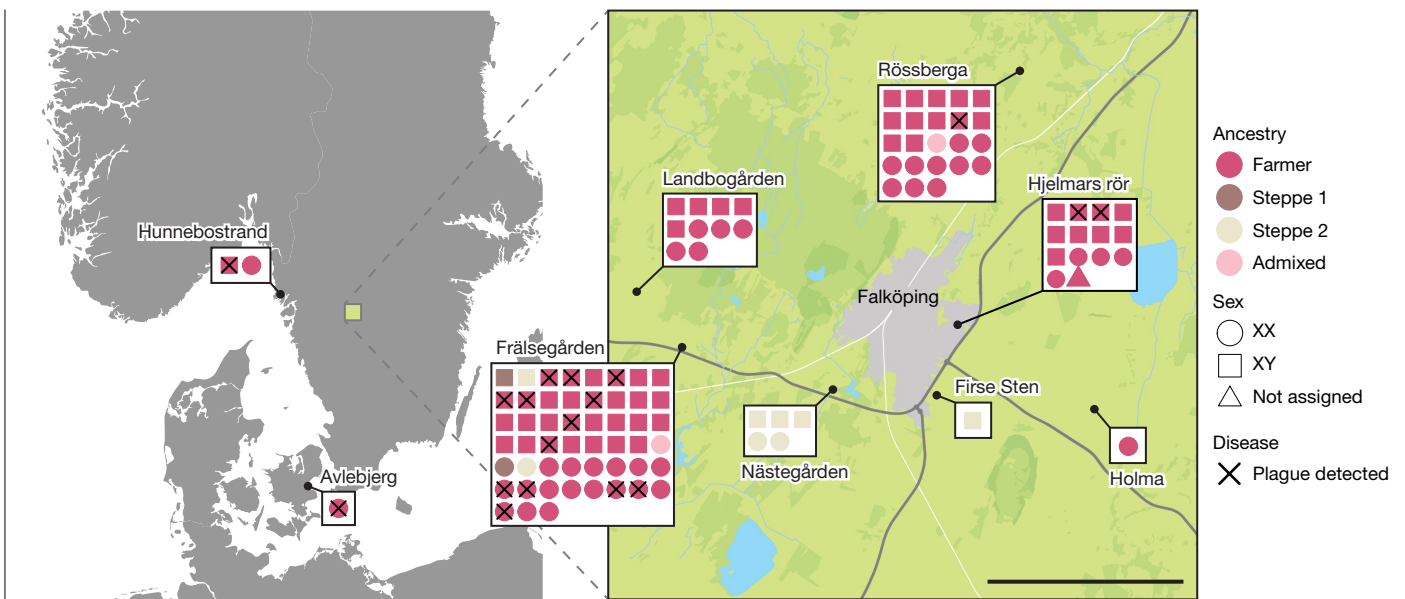
In the period between 5,300 and 4,900 calibrated years before present (cal. BP), populations across large parts of Europe underwent a period of demographic decline<sup>1,2</sup>. However, the cause of this so-called Neolithic decline is still debated. Some argue for an agricultural crisis resulting in the decline<sup>3</sup>, others for the spread of an early form of plague<sup>4</sup>. Here we use population-scale ancient genomics to infer ancestry, social structure and pathogen infection in 108 Scandinavian Neolithic individuals from eight megalithic graves and a stone cist. We find that the Neolithic plague was widespread, detected in at least 17% of the sampled population and across large geographical distances. We demonstrate that the disease spread within the Neolithic community in three distinct infection events within a period of around 120 years. Variant graph-based pan-genomics shows that the Neolithic plague genomes retained ancestral genomic variation present in *Yersinia pseudotuberculosis*, including virulence factors associated with disease outcomes. In addition, we reconstruct four multigeneration pedigrees, the largest of which consists of 38 individuals spanning six generations, showing a patrilineal social organization. Lastly, we document direct genomic evidence for Neolithic female exogamy in a woman buried in a different megalithic tomb than her brothers. Taken together, our findings provide a detailed reconstruction of plague spread within a large patrilineal kinship group and identify multiple plague infections in a population dated to the beginning of the Neolithic decline.

The emergence of agriculture during the Neolithization brought about one of the most profound lifestyle changes in the history of modern humans. The shift in subsistence strategy from hunting, fishing and gathering to farming paved the way for a marked increase in population density and the establishment of larger and more permanent settlements<sup>2</sup>. However, the flourishing economy of the Neolithic came to a sudden halt in many regions of Northern Europe around 5300–4900 calibrated years before present (cal. BP)<sup>5</sup>, in which a marked reduction in the number of human remains radiocarbon-dated to this period suggests a population decline. Coined the Neolithic decline<sup>1</sup>, this demographic bust coincides with the cessation of megalith building in the area and has been suggested to be one of the factors facilitating the Corded Ware expansion into Europe (4800–4400 cal. BP)<sup>6</sup>. Although several scenarios have been put forward, no single driving factor has

hitherto been linked to this decline and this enigma is still heavily debated in the literature<sup>3,5</sup>. Nevertheless, recent findings demonstrating that an ancestral form of *Yersinia pestis* was present in Sweden at this time could potentially solve this debate<sup>4</sup>.

*Yersinia pestis*, the infectious agent of plague, split from its most recent ancestor *Yersinia pseudotuberculosis* some time within the past 50,000 years and has been infecting humans since prehistoric times. The vast majority of prehistoric plague genomes are from Late Neolithic and Bronze Age (LNBA) individuals dating to 4700–2400 cal. BP (refs. 7–9). These genomes fall within two distinct lineages that can be distinguished by the absence (LNBA<sup>-</sup>) or presence (LNBA<sup>+</sup>) of the *ymt* gene<sup>8</sup>. The *ymt* gene is crucial for the bacterium's survival in the flea digestive tract when the source is an infected mouse, black rat or human, and hence for the development of bubonic plague.

<sup>1</sup>Lundbeck Foundation GeoGenetics Centre, Globe Institute, University of Copenhagen, Copenhagen, Denmark. <sup>2</sup>Department of Historical Studies, University of Gothenburg, Gothenburg, Sweden. <sup>3</sup>Human Evolution, Department of Organismal Biology, Uppsala University, Uppsala, Sweden. <sup>4</sup>Lygature, Utrecht, the Netherlands. <sup>5</sup>Laboratório de Biodiversidade e Evolução Molecular (LBEM), Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. <sup>6</sup>Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, Canada. <sup>7</sup>Institute for Mummy Studies Eurac Research, Bolzano, Italy. <sup>8</sup>Sealand Archaeology, Kalundborg, Denmark. <sup>9</sup>Trace and Environmental DNA (TrEnD) Laboratory, School of Molecular and Life Sciences, Curtin University, Perth, Western Australia, Australia. <sup>10</sup>Nuffield Department of Clinical Neurosciences, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. <sup>11</sup>Department of Archaeology and Ancient History, Lund University, Lund, Sweden. <sup>12</sup>Centre for Palaeogenetics, Stockholm University and the Swedish Museum of Natural History, Stockholm, Sweden. <sup>13</sup>Department of Archaeology and Classical Studies, Stockholm University, Stockholm, Sweden. <sup>14</sup>Department of Zoology, University of Cambridge, Cambridge, UK. <sup>15</sup>Palaeo-Research Institute, University of Johannesburg, Johannesburg, South Africa. ✉e-mail: frederikseersholm@gmail.com; martin.sikora@sund.ku.dk



**Fig. 1 | Overview of sample locations.** Each individual is represented by coloured shapes, with squares and circles representing males and females, respectively, and triangles representing sex unknown. Colours indicate genetic ancestry and black crosses designate plague-positive individuals. Scale bar, 5 km.

Until recently, all known prehistoric plague strains fell within these two LNBA clades; however, findings published in Rascovan et al.<sup>4</sup> and Susat et al.<sup>10</sup> demonstrated the presence of earlier diverging *Y. pestis* lineages (pre-LNBA). These two ancestral *Y. pestis* genomes were identified from a Swedish individual with Neolithic Farmer (that is, Anatolian-derived) ancestry (5035–4856 cal. BP)<sup>4</sup> and an individual from Latvia with hunter-gatherer ancestry (5300–5050 cal. BP)<sup>10</sup>, respectively. Although these genomes are of very similar age and ancestral to all other plague genomes available, the two studies arrive at different conclusions: Rascovan et al. argue that their finding supports a role of plague in the Neolithic decline whereas Susat et al. conclude that these early plague forms are probably a result of sporadic zoonotic events.

In Scandinavia the Neolithic decline coincides with the disappearance of the Funnelbeaker/Trichterbecher cultural complex (TRB)<sup>11</sup> and the end of the first wave of megalith building. Despite their ubiquity in the Scandinavian landscape, controversy still exists regarding the mortuary practices and social background of people associated with the megaliths. For example, it has long been assumed that these tombs were used by kinship or family groups<sup>12</sup>, but currently few data exist to substantiate this hypothesis. Osteological analyses show the presence of both sexes in variable proportions, and also of individuals of different ages at death, consistent with a random selection from a living population. By contrast, the number of buried individuals per generation is small and would suggest that only a restricted segment of the population was interred in these tombs<sup>13–15</sup>. It is also debated whether complete bodies were directly introduced and perhaps later rearranged and disturbed, or whether placement in the chamber was only the final act in a longer series of mortuary rituals performed over a longer time and at different locations<sup>14,16–19</sup>. Furthermore it is not known which factors determined burial location, or how different parts of the tombs were typically used.

To elucidate social structure and plague infection frequency in the Scandinavian Middle Neolithic, we analysed ancient human DNA from nine multi-individual burial structures: seven megaliths in Falbygden, inland western Sweden (Landbogården, Frälsegården, Nästegården, Firse sten, Holma, Hjelmars rör and Rössberga), one megalith on the Swedish west coast (Hunnebostrand) and a stone cist in Denmark (Avlebjerg; Supplementary Note 1). We set out to investigate whether the plague cases from ref. 4 were isolated events or whether there was evidence of plague in more individuals at different sites and in different Scandinavian regions during the Neolithic. Furthermore, we aimed to investigate

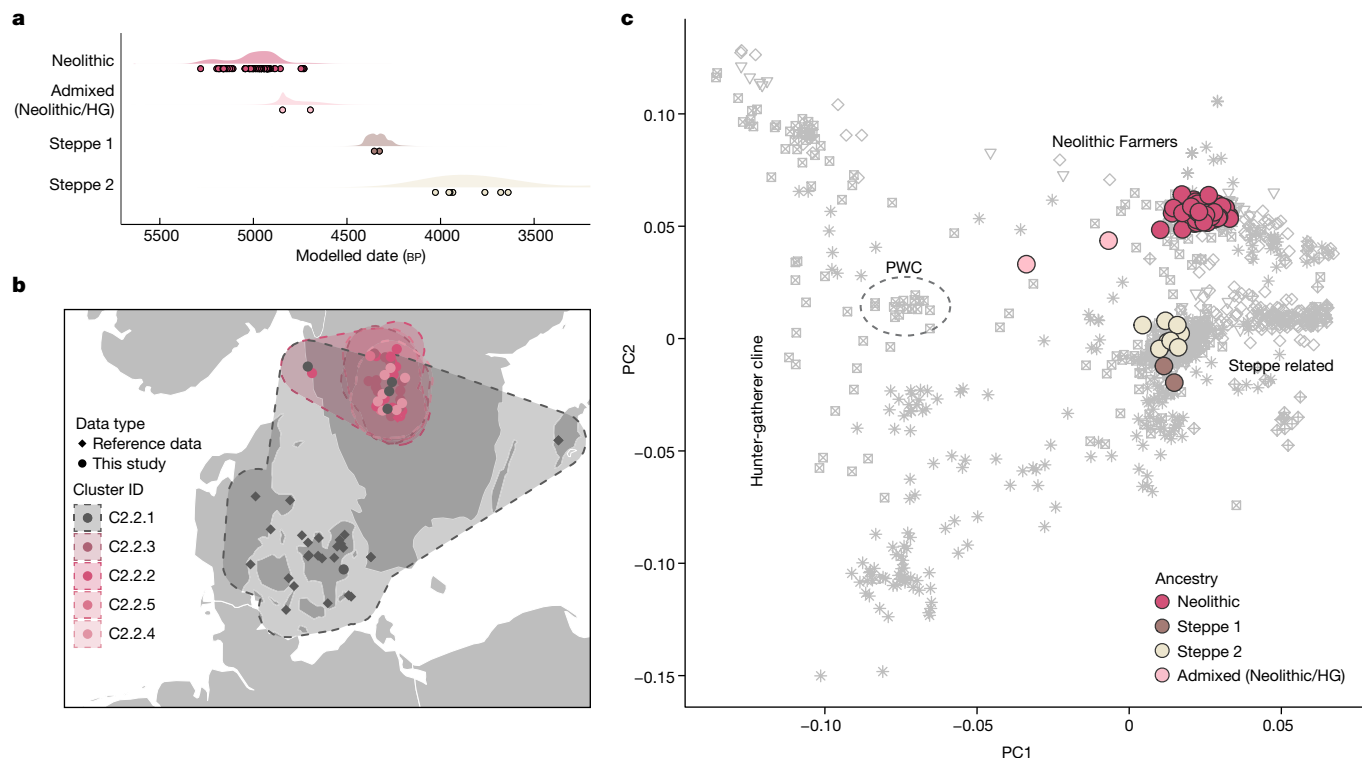
kinship and social relations in several of the best-described megaliths from Sweden to better understand potential disease transmission.

## DNA sequencing and population structure

To investigate the disease frequency and geographical distribution of the Neolithic plague, we sequenced 174 samples from ancient human skeletal remains from across Scandinavia (Supplementary Table 2). After merging data from samples that derived from the same individual and exclusion of low-coverage human data (below 0.01×; Methods), we produced a final dataset representing 108 individuals from nine sites (Fig. 1 and Supplementary Table 1). We found a slight male sex bias in the dataset (58% males; Fig. 1), reflected across all sites except from the Rössberga passage grave, where no sex bias was observed.

To investigate the broader population genetic structure in our data, we merged our dataset with a panel of 1,430 ancient shotgun-sequenced genomes (Supplementary Table 7) and visualized genetic affinities using principal component analysis (PCA; Fig. 2c). We found that the vast majority of individuals analysed ( $n = 96$ ) fell within the broad cluster of European Neolithic populations and Anatolian Farmers. This finding is in agreement with our radiocarbon dating results, which date this group to 5200–4900 cal. BP, associating these individuals with the TRB culture of Southern Scandinavia and Scandinavian MN A (Fig. 2a). Furthermore, we also found evidence of two distinct and slightly younger groups of individuals with Steppe-related ancestry. The first group (Steppe 1,  $n = 2$ ) is dated to around 4400 cal. BP (Scandinavian MN B) and the second group (Steppe 2,  $n = 8$ ) to 4100–3000 cal. BP (Scandinavian LN to LBA). This distinction is corroborated by chromosome Y haplogroup results, which indicate that all Farmer ancestry males have the haplogroup I2 whereas the two Steppe ancestry groups are represented exclusively by haplogroups R1 and I1, respectively (Supplementary Table 1). In general we find that each site is represented by people of similar ancestry (Fig. 1), but at one site (Frälsegården) we find evidence of the continued use of the burial chamber by three temporally and genetically distinct populations.

Lastly, we find that two genomes contain a substantial fraction of hunter-gatherer ancestry, suggesting recent admixture between the TRB group and Scandinavian hunter-gatherers (Fig. 2). One woman from Frälsegården (FRA108) appears to have equal proportions of hunter-gatherer and Neolithic Farmer ancestry; we find that she



**Fig. 2 | Age and ancestry.** **a**, Median radiocarbon date (years cal. BP) for samples, grouped by genetic ancestry (Supplementary Tables 1 and 4). Shaded area represents distribution of kernel density estimation for the group. **b**, Geographical distribution of IBD clusters of Scandinavian

was most probably a first-degree offspring of these two distinct sociocultural groups (Extended Data Fig. 1 and Supplementary Note 5). Similarly for the other individual, a woman from the site Rössberga (ROS027), we find roughly 34% hunter-gatherer ancestry and 66% Neolithic Farmer ancestry, suggesting that she might have lived two or three generations after the admixture event. Although northern groups of hunter-gatherers with Mesolithic ancestry cannot be ruled out<sup>20</sup>, the most likely source of hunter-gatherer DNA in these two admixed individuals is the Pitted Ware Culture (PWC; 5400–4300 cal. BP), which, in Sweden, overlapped with the TRB culture in both time and space<sup>21,22</sup>. It is notable that both women date to the end of the TRB period, perhaps reflecting demographic crisis within the TRB population and/or a loosening of social and cultural boundaries.

### Distinct Neolithic IBD clusters

To disentangle the fine-scale structure between TRB individuals, we compared identity-by-descent (IBD) sharing across the panel and clustered individuals with close affinities into related IBD groups (Methods). We found that the IBD clustering of Northern European Neolithic Farmers was, to a large extent, driven by the close familial relations within Falbygden. Accordingly, we identified four IBD clusters of individuals from Falbygden ( $n = 17, 13, 8$  and  $4$ ) and one cluster of all other Neolithic individuals from Denmark and Sweden ( $n = 32$ ; Fig. 2b). As expected, most of our samples were clustered into the four Falbygden groups whereas our sample from Denmark (Avlebjerg) fell within the unrelated group from Denmark and Sweden (Fig. 2b and Supplementary Table 8). Interestingly, we also found that three TRB individuals from Falbygden clustered with the Danish/Swedish IBD group, suggesting that these individuals came from outside the Falbygden area. This is supported by Sr isotope data for one of them (FRA106,

individuals with Neolithic Farmer ancestry (Supplementary Table 8). **c**, PCA of genomes sequenced in this study (coloured by ancestry) and a reference panel of 810 previously sequenced ancient shotgun genomes from Europe (grey shading; Supplementary Table 7). HG, hunter-gatherer.

adult male, 87 Sr/86 Sr 0.717011), indicating an upbringing outside the Cambro-Silurian bedrock of Falbygden. Moreover, for the two genomes sequenced from the site Hunnebostrand in coastal western Sweden, we found that one genome (HUN002, an adult male) clustered with one of the Falbygden groups whereas the other (HUN001, a female around 70 years of age) clustered with the Danish/Swedish group, suggesting different origins for these two individuals. Sr isotopes in the two individuals confirm different childhood residence (Supplementary Table 5). Both are consistent with childhood in Falbygden, but they could also originate from other places in Scandinavia outside Denmark.

The IBD-sharing results also support the presence of two distinct Steppe-related groups in our dataset discussed above. The early group (chrY haplogroup: R1a1a, age approximately 4400 cal. BP) clusters with a large group of individuals from across Europe of Corded Ware ancestry, including individuals from Battle Axe Culture contexts in Sweden. The later group (chrY haplogroup: I1, age approximately 4100–3000 cal. BP), on the other hand, clusters exclusively with contemporaneous individuals from Eastern Denmark, Sweden and Norway, reflecting results from ref. 23. Using DATES, we were able to date the admixture of ‘Steppe’ and ‘Farmer’ DNA in these two groups<sup>24</sup>. For both groups we found that admixture most probably happened around 4750 cal. BP (Extended Data Fig. 2). In agreement with recent results showing that Steppe-related groups first appeared in Eastern Europe around 4,800 years ago<sup>23</sup>, this finding suggests that admixture occurred in a single pulse before the arrival of Corded Ware complex (CWC) groups in Sweden.

### Social structure across four pedigrees

We investigated close familial relations in the dataset and categorized pairs of individuals as either first-, second- or third-degree relatives (Methods). Using these data we were able to reconstruct one large





the progenitor (FRA011) is found in the northern end and is represented by an isolated mandible. This male belongs to the first generation of the left-side kinship branch. We also identify a young woman (FRA102) unrelated to anyone else in the pedigree in the northern half of the tomb. This individual is an outlier in many ways: her body was given a special treatment because she was buried as a tight package of partially articulated bones (individual C in the excavation report). Moreover, the Sr value (0.717345) in her lower left M1 tooth suggests she spent her early years outside the Falbygden geology although her origin cannot be pinpointed more precisely. Lastly, she appears to have been affected by an ancestral form of the plague (discussed below).

There is a clear division in the placement of the later generations, because the right-side branch is found in the northern half and the left-side branch in the southern half of the chamber (Supplementary Fig. 6). Moreover, genetically female individuals are concentrated in the central part of the chamber whereas males are more evenly distributed (Supplementary Fig. 4).

In the central part of the passage grave the correlation between burial location and generation is not as clear as in the northern part. Nevertheless, there is a general trend of earlier generations being buried in the northern part of the central area whereas later generations are buried more centrally. Furthermore, we find that the four siblings (FRA032–FRA035) from generation five are buried together with their half-brother (FRA030) whereas their half-sister (FRA031) was buried elsewhere in the tomb. This indicates that the male line was more relevant in determining burial location. In contrast to the northern area, the central part holds many complete or partially articulated skeletons, as well as disarticulated bones.

Overall, it is possible to interpret the data from kinship and burial location as evidence for a time-dependent filling of the tomb, starting in the northern (and possibly in the southern) part of the chamber and slowly moving towards the centre. The difference in mortuary treatment can be viewed in different ways. The skulls of the first- and second-generation individuals could have been selected for special treatment and moved to the northern part, presumably from burials in the central area. In this case, postcranial remains of these individuals should be found within the chamber. Another possibility is that selected parts of these individuals were brought in from elsewhere, either as part of a programme of secondary burial or as foundation deposits from another burial.

Close familial relations are also recognized: we identify two nuclear families buried very closely together. Thus, biological kinship and sex seem to have been socially recognized and were used as important tools to categorize Neolithic people in Falbygden, and were determining factors for the placement and handling of dead members of society.

### Three Neolithic plague strains

We carefully examined all the sequenced data for known human pathogens (Methods and Supplementary Table 10). Strikingly, this screening showed that the most frequently found pathogen overall was the plague-causing bacterium *Y. pestis* (18 out of 108 sampled individuals, 17%). We identified five other pathogens that we consider authentic (Supplementary Table 10); of these, two were identified in more than one individual—*Yersinia enterocolitica*, the causative agent of yersiniosis (four of 108, 4%) and *Borrelia recurrentis*, which causes louse-borne relapsing fever (five of 108, 5%)—and indicate the presence of body lice in Falbygden. Interestingly, in one individual (FRA013) we identified a coinfection with *Y. pestis* and *Y. enterocolitica* at 3.9× and 1.8× coverage, respectively (Methods).

We found that plague frequency varied across sites, with the highest rate at Frälsegården (13 of 47, 28%, excluding Steppe-related individuals) and markedly lower rates at all other sites. Plague-positive individuals were found not only in the Falbygden area but as far south as Zealand, Denmark and by the Swedish west coast north of Gothenburg.

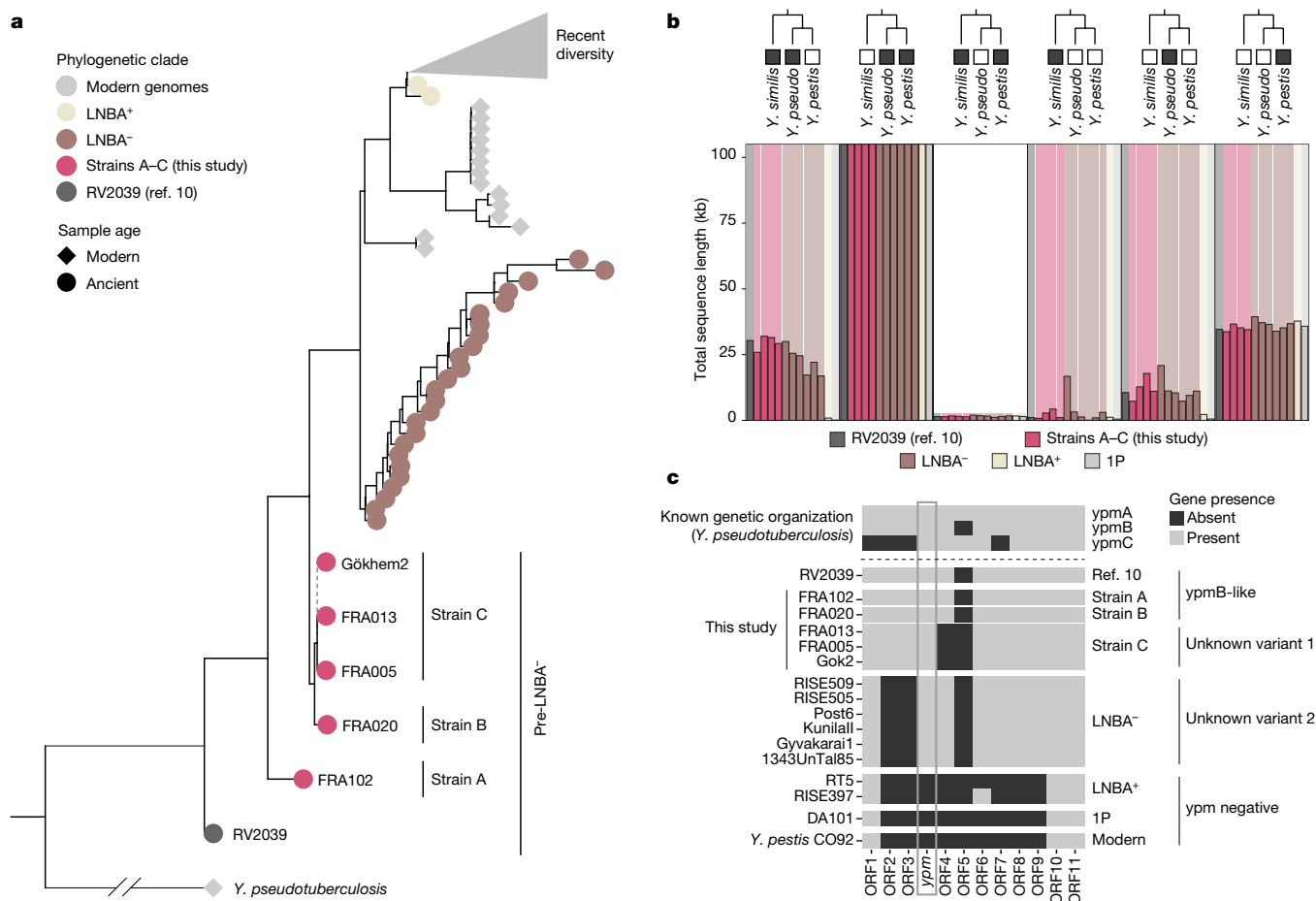
These findings indicate that the plague was widespread in southern Scandinavia 5,000 years ago.

Of the 18 plague-positive individuals, six were classified as tentative detections with coverage below 0.01× whereas seven could be classified as lower-coverage partial genomes (0.01–1.5×) and five were classified as higher-coverage partial genomes (over 1.5×; Supplementary Table 6). For the five higher-coverage genomes (11.5×, 10.5×, 6.4×, 4.5× and 1.8×) there were sufficient data to carry out single-nucleotide polymorphism (SNP) calling and reconstruct the full plague phylogeny, together with previously published ancient and modern genomes (Fig. 4, Extended Data Fig. 7 and Supplementary Fig. 5). In agreement with previous results<sup>4</sup>, all newly sequenced genomes fell basal to the cluster of LNBA plague strains (Fig. 4). We found that the plague genomes from individuals FRA013 and FRA005 were identical to the previously published Gökhem2 genome from the same archaeological site and time period<sup>4</sup>. By contrast, the plague genome from FRA020 differed from the three identical genomes at three positions (Extended Data Fig. 8d)—we designate these as pre-LNBA strains C and B, respectively. Surprisingly, the plague genome from individual FRA102 is markedly different from strains C and B, from which it differs on 29–34 positions (Extended Data Fig. 8a). This genome falls basal to all known plague diversity except for strain RV2039, dated to 5300–5050 cal. BP (ref. 10). We designate this genome pre-LNBA strain A (Fig. 4).

Although no phylogenetic information can be extracted from the tentative plague detections (below 0.01×), lower-coverage genomes (0.01–1.5×) are of sufficient coverage to be placed onto the plague phylogeny (Extended Data Fig. 9). As expected, these placements confirm that all plague genomes sequenced in this study cluster together with the higher-coverage plague genomes from the Frälsegården cluster. In general these placements do not allow us to distinguish between strains A, B and C. There are, however, a few exceptions: the relatively high coverage of the plague genome from FRA106 (1.1×) produced a more accurate placement, which suggests that this strain might belong to strain C. Similarly, the higher number of unique SNPs in the more ancestral genomes from strain A and RV2039 means that lower-coverage genomes similar to ancestral strains are more easily placed onto the phylogeny. Accordingly, we find that individual FRA021 had a plague form similar to strain A whereas the individual from Denmark (AVL001) appears to have a plague form ancestral to strain A (Extended Data Fig. 9).

The distribution of plague-positive individuals in the pedigree presented in Fig. 3 does not readily support a swift and deadly plague epidemic, because plague is detected in all generations except two and six. However, when information on the different plague strains is taken into account, it becomes apparent that plague infections are stratified both chronologically and phylogenetically into two separate clusters. The most ancestral form of the plague (strain A) is also detected in the oldest individual from this study (FRA102), who was buried in the northern part of the chamber with individuals from generations one and two despite being unrelated to everyone in the pedigree. Based on placements of lower-coverage genomes, we also identify a plague form similar to strain A in FRA021, the progenitor of the right-side subfamily. This finding suggests spread of an early form of the disease in the first generation of the pedigree. The mortality rate of this form of the plague is unknown, but the pedigree clearly illustrates that both the family as a whole and the line of FRA021 survived the disease.

The second cluster of plague infections occurs in generations three to five and is caused by strains B and C. Strain C is found in generation four of the left-side family, in which all individuals appear to have been infected by the plague. Strain B on the other hand, is found in a single individual from the right-side subfamily, most probably from generation three. Given the genetic distance between strains B and C and their different distribution across the pedigree, it is possible that the two strains represent separate infection events: one in the right-side subfamily (strain B) and another in the left-side subfamily



**Fig. 4 | Plague genomes.** **a**, Phylogenetic relationship between previously published plague strains and the data produced for this study. Each circle represents one plague genome, coloured by phylogenetic clad. For clarity, all plague strains more divergent than the LNBA clad are collapsed (grey triangle); Supplementary Fig. 5 shows the full tree. **b**, Barplot showing the amount of reference graph gene content covered by ancient *Y. pestis* samples. Reference graph nodes were stratified into groups of presence/absence pattern among modern *Y. pseudotuberculosis* complex species, indicated by cartoon phylogenies. The shaded background in the plot represents the total length of

(strain C). Yet, even though both strains appear to be contemporaneous in Fig. 3, they could be temporally distinct because the error margins of the chronological modelling allow for considerable variation in the modelled dates (Extended Data Fig. 5). Unfortunately, it is not possible to further assess this hypothesis because we are unable to distinguish strains B and C in most of the lower-coverage genomes (Extended Data Fig. 9).

We can only speculate on the impact of the disease on the local population, but we note that all of the hallmarks expected in a swift and deadly epidemic are present for the spread of strain C in the left-hand subfamily: (1) the frequency of plague-positive individuals is exceptionally high, (2) the disease is restricted to the last two generations and (3) all higher-coverage genomes from this subfamily are identical. Taken together, these observations suggest that an outbreak of strain C could have led to the demise of the left-side subfamily, perhaps driven by increased pathogenicity due to recombination around the *ypm* locus in this strain (discussed below).

It is challenging to directly gauge the pathogenicity and route of transmission in ancient plague strains, but the high number of infected individuals analysed here suggests that the disease was able to spread within the population. To detect genetic variation that could hint at the pathogenicity of the Neolithic plague, we first analysed the coverage

reference graph nodes for a given group; coloured bars represent the total length of nodes within that group present in a particular sample. Values on the y axis are capped at 100 kb to aid visualization; groups of ancient strains are indicated by bar colour. **c**, Presence/absence of genes in the unstable genomic region surrounding the superantigen *ypm* gene in *Y. pseudotuberculosis* IP 31758. A gene is defined as being present if the ratio of observed over expected breadth of coverage (given the depth of coverage across the genome) is over 10% for that gene-sample combination (Extended Data Fig. 10). 1P, first pandemic.

of classic plague virulence genes (Supplementary Fig. 6). Using this approach, we found the same virulence genes (*YpfΦ* prophage and *ymt*) to be absent in the Falbygden plague strains, as for both LNBA<sup>-</sup> strains and pre-LNBA strain RV2039. Such analyses of gene content based on read mapping to a single-reference genome are unable to capture the full spectrum of genome variation among ancient *Y. pestis* strains. In particular, genomic variation absent in the reference genome but present in an ancient lineage cannot be detected using this approach. To overcome this limitation we investigated the gene content variation of ancient plague lineages by mapping their sequencing reads to a pangenome variation graph encompassing 82 complete assemblies of the *Y. pseudotuberculosis* complex (56 *Y. pestis*, 24 *Y. pseudotuberculosis* and one *Yersinia similis*). Partitioning the variant graph nodes into groups based on their presence pattern in the three species, we found that early divergent ancient plague lineages (pre-LNBA<sup>-</sup>, Frälsögården and LNBA<sup>-</sup>) harbour up to 50 kb gene content present in some or all *Y. pseudotuberculosis*/*Y. similis* strains that is absent in later strains (LNBA<sup>+</sup>, 1P, modern; Fig. 4b).

Intriguingly, among the genomic regions identified as containing variation absent in later plague strains was a chromosomal locus including the gene encoding for *Y. pseudotuberculosis*-derived mitogen (*ypm*), a superantigenic toxin found in some strains of *Y. pseudotuberculosis*<sup>28</sup>.

All early diverging plague strains showed evidence for the presence of *ypm* whereas the entire locus is absent in all lineages diverging after and including the LNBA<sup>+</sup> type strain RT5. To distinguish the three known *ypm* loci (*ypmA*, *ypmB* and *ypmC*) we characterized gene presence/absence in the genetic environment surrounding this unstable locus (Fig. 4c). We identify a pattern of genetic organization resembling the ancestral locus, *ypmB*, in RV2039 and Falbygden strains A and B, and we identify two hitherto unknown combinations of genes in the *ypm* locus for strain C and later LNBA<sup>-</sup> strains. Although *Y. pseudotuberculosis* strains carrying the *ypmB* allele have been associated with low pathogenicity in humans<sup>29</sup>, the virulence of the unknown *ypm* allele identified in strain C cannot be determined. It is, however, well established that certain *ypm* alleles have epidemic potential. The production of the superantigen toxin YPma, for example, is crucial to the pathogenesis of Far East scarlet-like fever, a severe systemic infectious disease with symptoms including a scarlet-like skin rash and toxic shock syndrome<sup>30</sup>. Far East scarlet-like fever has been described as an ‘epidemic manifestation’ of *Y. pseudotuberculosis* infection in humans and, in 1959, an outbreak in Vladivostok, USSR, caused the hospitalization of 200 out of 300 patients<sup>30,31</sup>.

Given the presence of ancestral *Y. pseudotuberculosis* variation in both pre-LNBA and LNBA<sup>-</sup> strains, it is possible that these plague forms followed a faecal–oral transmission route and showed attenuated pathogenicity. Nevertheless, the presence of the *ypm* gene and, in particular, the finding of a hitherto unknown combination of genes in the *ypm* locus for strain C, could suggest increased virulence. Hence, whereas infectivity, morbidity and mortality might have varied among these early plague strains, it is possible that a recombination event in the *ypm* locus rendered strain C virulent. In combination with other factors, and perhaps other diseases, strain C could thus have played a role in the Neolithic decline. Nonetheless, all individuals from pedigree 1 were buried in the Frälsegården tomb and someone must have survived to bury them. Furthermore, the demographic profile of the graves does not indicate catastrophic mortality<sup>13</sup>, which, together with the high infection rates, could suggest a less severe or chronic disease manifestation. Lastly, it is well established that the later LNBA<sup>-</sup> clade of plague was prevalent and widespread across Eurasia for more than 2,000 years without markedly affecting Bronze Age population sizes<sup>8</sup>.

## Perspectives and conclusions

By the end of the Neolithic at least three main plague lineages had evolved: the most ancestral RV2039 lineage<sup>10</sup>, the Falbygden clade (strains A, B and C) and the lineage that would eventually evolve into the Bronze Age radiation of plague<sup>8</sup>. All three lacked the *ymt* gene, crucial for the bacterium’s survival in the flea digestive tract. Consequently, flea-based transmission of LNBA<sup>-</sup> and pre-LNBA<sup>-</sup> plague strains is unlikely and the manifestation of Neolithic plague unlikely to resemble bubonic plague<sup>32</sup>. It has previously been hypothesized that the pre-LNBA RV2039 lineage and other early forms of the plague were less transmissible and that they represented sporadic zoonotic events<sup>10</sup>. By detecting plague in approximately one in every six sampled individuals, we conclusively show that plague infection was not a rare event in Neolithic Scandinavia. Given the high frequency of the disease, it is possible that it was spreading within the population following human–human, and potentially human–louse–human<sup>33</sup>, transmission. However, it is worth mentioning that the plague rate of 17% reported here does not necessarily reflect the true prevalence of the disease. For example, the plague detection rate might not be representative of the population as a whole because it is a measure of disease frequency within the sampled population, which is restricted to well-preserved individuals buried in tombs. Furthermore, only a fraction of plague-positive cases is expected to carry detectable levels of DNA from *Y. pestis*. In ref. 34, a quantitative PCR screening of known

plague victims showed a detection rate of 5.7% in bones and 37% in teeth, suggesting that the true frequency of the Falbygden plague could be significantly higher than 17%.

The pathogenicity of the ancestral form of plague has been heavily debated in the literature<sup>4,8–10,35</sup> and, without an experimental set-up to directly gauge the virulence of the bacterium, we rely on circumstantial evidence. Our contribution to this discussion is threefold: (1) we track the presence of multiple plague strains across an extended pedigree, suggesting that plague infection was common in the community; (2) we show that the Falbygden plague strains carried a new and a known variant of the *ypm* locus from *Y. pseudotuberculosis*, suggesting that the disease manifestations, morbidity and mortality of these early plague strains might differ substantially; and (3) the Falbygden plague strains carried variable amounts of genetic material from *Y. pseudotuberculosis*, suggesting that the transmissibility and transmission route(s) might have varied somewhat. These results demonstrate that the Neolithic plague was prevalent and potentially lethal. Together with the fact that these plague cases are found in one of the last populations with Neolithic Farmer ancestry observed in Scandinavia, we believe that plague could have been a contributing factor to the Neolithic decline.

In addition to the plague, this study provides new data on familial relations in Middle Neolithic Sweden by reconstructing one large and three smaller pedigrees. We show that the social organization in Falbygden, as represented by the population sampled within these tombs, was patrilineal and patrilocal. This result corroborates findings from the Hazleton North long cairn<sup>36</sup>. However, we note significant divergences from the result reported in ref. 36: we identify four males with multiple reproductive partners, but we do not find any instances in which females produced offspring with different males. This is in contrast to the results from Hazleton North long cairn<sup>36</sup>, in which females with multiple partners far outweigh males (five females versus one male). Furthermore, the subdivision into matrilineal subgroups is not visible at Frälsegården, in which subgroups tend to show a patrilineal structure. Another comparison can be made with the Neolithic cemetery at Gurgy, France<sup>37</sup>, where a patrilineal and patrilocal structure was also found. In contrast to Falbygden, all couples at Gurgy were monogamous.

Lastly, these kinship data provide some of the first concrete examples of social practices commonly discussed in archaeology. We identify a set of three siblings, two males and one female, in which the female appears to have moved to a neighbouring group in which she established her own family. Furthermore, we also report a single instance of third-degree relatives producing offspring resulting in long stretches of ROHs in both of their sons.

Taken together, the data presented here provide a highly detailed and intimate snapshot of what life was like in Neolithic Falbygden, Sweden. The social structure was organized along male kinship lines, and females generally came from other kin groups. Because plague was infecting a significant proportion of the population, excess mortality associated with the disease could have undermined the long-term viability of society, leading to the eventual collapse of this form of Neolithic society.


## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-024-07651-2>.

- Kristiansen, K. in *The Oxford Handbook of Neolithic Europe* (eds Fowler, C. et al.) Ch. 56 (Oxford Univ. Press, 2014).
- Müller, J. in *Environment and subsistence - forty years after Janusz Kruk’s “Settlement studies”* (eds Kadrow, S. & Włodarczak, P.) 493–506 (Dr. Rudolf Habelt GmbH, 2013).
- Colledge, S., Conolly, J., Crema, E. & Shennan, S. Neolithic population crash in northwest Europe associated with agricultural crisis. *Quat. Res.* **92**, 686–707 (2019).

4. Rascovan, N. et al. Emergence and spread of basal lineages of *Yersinia pestis* during the Neolithic Decline. *Cell* **176**, 295–305 (2019).
5. Shennan, S. et al. Regional population collapse followed initial agriculture booms in mid-Holocene Europe. *Nat. Commun.* **4**, 31–34 (2013).
6. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
7. Rasmussen, S. et al. Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell* **163**, 571–582 (2015).
8. Valtueña, A. A. et al. Stone Age *Yersinia pestis* genomes shed light on the early evolution, diversity, and ecology of plague. *Proc. Natl Acad. Sci. USA* **119**, e2116722119 (2022).
9. Andrades Valtueña, A. et al. The Stone Age plague and its persistence in Eurasia. *Curr. Biol.* **27**, 3683–3691 (2017).
10. Susat, J. et al. A 5,000-year-old hunter-gatherer already plagued by *Yersinia pestis*. *Cell Rep.* **35**, 109278 (2021).
11. Blank, M., Sjögren, K. G. & Storå, J. Old bones or early graves? Megalithic burial sequences in southern Sweden based on 14C datings. *Archaeol. Anthropol. Sci.* **12**, 89 (2020).
12. Sánchez-Quinto, F. et al. Megalithic tombs in western and northern Neolithic Europe were linked to a kindred society. *Proc. Natl Acad. Sci. USA* **116**, 9469–9474 (2019).
13. Ahlström, T. Det döda kollektivet: skelettmaterialet från Rössbergagånggriften. in *Falbygden gånggrifter: [D.1] Undersökningar 1985–1998* (eds Persson, P. & Sjögren, K.-G.) (Institutionen för arkeologi, Göteborgs universitet, 2001).
14. Ahlström, T. *Underjordiska Dödsriken. Humanosteologiska Studier Av Neolitiska Kollektivgravar* Vol. 18 (Institutionen för arkeologi, Göteborgs universitet, 2009).
15. Sjögren, K. G. "Mångfalldige Uhrminnes Grafvar..." *Megalitgravar Och Samhälle i Västsverige* (Institutionen för arkeologi, Göteborgs universitet, 2003).
16. Tiley, C. *An Ethnography of the Neolithic* (Cambridge Univ. Press, 1996).
17. Andersen, N. H. & Trolle, A. L. Kult og ritualer i den ældre bondestenalder. *KUML* <https://doi.org/10.7146/kuml.v2000i14.114016> (2000).
18. Sjögren, K.-G. *Fragment av ordning. Undersökning av överplöjda megalitgravar vid Fråsegården, Gökhem socken, Västergötland, 1999–2001*. Report No. Västergötlands museum Rapport 2008-23 (Göteborgs universitet, 2008).
19. Sjögren, K.-G. in *Neolithic Diversities: Perspectives from a Conference in Lund, Sweden* (eds Brink, K. et al.) 200–210 (Department of Archaeology and Ancient History, Lund University, 2015).
20. Fischer, A. et al. Vittrup Man—the life-history of a genetic foreigner in Neolithic Denmark. *PLoS ONE* **19**, e0297032 (2024).
21. Iversen, R., Philippsen, B. & Persson, P. Reconsidering the Pitted Ware chronology: a temporal fixation of the Scandinavian Neolithic hunters, fishers and gatherers. *Prähistorische Z* **96**, 44–88 (2021).
22. Vanhanen, S. et al. Maritime hunter-gatherers adopt cultivation at the farming extreme of Northern Europe 5000 years ago. *Sci. Rep.* **9**, 4756 (2019).
23. Allentoft, M. E. et al. Population genomics of post-glacial western Eurasia. *Nature* **625**, 301–311 (2024).
24. Narasimhan, V. M. et al. The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
25. Ceballos, F. C. et al. Human inbreeding has decreased in time through the Holocene. *Curr. Biol.* **31**, 3925–3934 (2021).
26. Cassidy, L. M. et al. A dynastic elite in monumental Neolithic society. *Nature* **582**, 384–388 (2020).
27. Sedig, J. W., Olalde, I., Patterson, N., Harney, É. & Reich, D. Combining ancient DNA and radiocarbon dating data to increase chronological accuracy. *J. Archaeol. Sci.* **133**, 105452 (2021).
28. Carnoy, C. et al. The superantigen gene *ypm* is located in an unstable chromosomal locus of *Yersinia pseudotuberculosis*. *J. Bacteriol.* **184**, 4489–4499 (2002).
29. Ch'ng, S. L. et al. Population structure and evolution of pathogenicity of *Yersinia pseudotuberculosis*. *Appl. Environ. Microbiol.* **77**, 768–775 (2011).
30. Amphlett, A. Far East scarlet-like fever: a review of the epidemiology, symptomatology, and role of superantigenic toxin: *Yersinia pseudotuberculosis*-derived mitogen A. *Open Forum Infect. Dis.* **3**, ofv202 (2016).
31. Somova, L. M., Antonenko, F. F., Timchenko, N. F. & Lyapun, I. N. Far Eastern scarlet-like fever is a special clinical and epidemic manifestation of *Yersinia pseudotuberculosis* infection in Russia. *Pathogens* **9**, 436 (2020).
32. Hinnebusch, J. B. The evolution of flea-borne transmission in *Yersinia pestis*. *Curr. Issues Mol. Biol.* **7**, 197–212 (2005).
33. Houhamdi, L., Lepidi, H., Drancourt, M. & Raoult, D. Experimental model to evaluate the human body louse as a vector of plague. *J. Infect. Dis.* **194**, 1589–1596 (2006).
34. Schuenemann, V. J. et al. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proc. Natl Acad. Sci. USA* **108**, E746–E752 (2011).
35. Zimble, D. L., Schroeder, J. A., Eddy, J. L. & Latham, W. W. Early emergence of *Yersinia pestis* as a severe respiratory pathogen. *Nat. Commun.* **6**, 7487 (2015).
36. Fowler, C. et al. A high-resolution picture of kinship practices in an Early Neolithic tomb. *Nature* **601**, 584–587 (2022).
37. Rivollat, M. et al. Extensive pedigrees reveal the social organization of a Neolithic community. *Nature* **620**, 600–606 (2023).

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

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

© The Author(s) 2024



## Methods

### Sampling, library building and sequencing

We sampled a total of 133 teeth, 38 petrous bones and three femur samples from the collections at the University of Gothenburg (Gothenburg, Sweden), Statens Historiska Museum (Stockholm, Sweden) and Laboratory of Biological Anthropology (University of Copenhagen, Denmark). Laboratory processing of samples was carried out at the clean laboratory facilities at Globe Institute, University of Copenhagen and at the Human Evolution ancient DNA laboratory at Uppsala University, following strict ancient DNA guidelines. Subsampling and mechanical cleaning of the samples were carried out by removing the surface layer using a drill bit, followed by sampling of the dense bone material of the outer cementum layer for teeth, or the cochlea for petrous bones. DNA extraction was carried out on approximately 100–150 mg of bone material as previously described (ref. 6, modified after ref. 38), followed by a uracil-DNA glycosylase treatment step to remove deaminated cytosines for the majority of samples (Supplementary Table 3). Subsequently, double-stranded DNA-sequencing libraries were built following ref. 39. DNA sequencing was performed on a NovaSeq 6000 instrument (Illumina) in paired-end configuration at the GeoGenetics Sequencing Core (Copenhagen) using the S4 200 cycle kit (v.1.5). Furthermore, for selected samples with promising results but low content of endogenous DNA, single-stranded libraries were built as described in ref. 40 followed by sequencing on the Illumina NovaSeq 6000 platform (Supplementary Table 3).

The Uppsala samples were decontaminated using ultraviolet irradiation, 0.5% sodium hypochlorite and deionized Millipore water. The outer surface layers were removed as described above, and 50–150 mg of either bone powder or bone pieces was collected using drill bits or diamond cutting wheels. DNA was extracted using silica-based protocols<sup>41–43</sup>. Double-stranded DNA libraries were prepared as described above, with some being uracil-DNA glycosylase treated (Supplementary Table 3)<sup>44</sup>. Some libraries were enriched using European MYbaits from MYcroarray following the manufacturer's instructions. Sequencing was performed at SciLifeLab SNP&Seq in Uppsala using Illumina HiSeq 2500, HiSeq X Ten or NovaSeq platforms and 125 or 150 base pair paired-end chemistry.

### Capture enrichment

For libraries in which shotgun-sequencing results suggested relatively high concentrations of plague DNA, in-solution capture enrichment of *Y. pestis* DNA was carried out. We used the myBaits kit from Arbor Sciences described in ref. 45 and carried out hybridization capture following the manufacturer's High Sensitivity protocol, but using only a single round of enrichment and 1.1–4.4  $\mu$ l of baits per reaction. We pooled between one and 11 samples per capture reaction and reamplified the post-capture pools for 16 cycles (Supplementary Table 6).

### Basic bioinformatic processing and PCAs

Demultiplexed fastq files were trimmed for adapter sequences, and overlapping paired-end reads were collapsed using AdapterRemoval (v.2.0)<sup>46</sup>. Next, trimmed reads were mapped to the human reference genome (hg19) using bwa aln (v.0.7.17). The resulting alignments were converted to BAM files and merged, sorted and filtered using Samtools (v.1.12). The resulting BAM files were merged at sample level (Supplementary Table 2), and duplicates were identified using MarkDuplicates (v.2.27.4) from Picard Tools. Next, duplicate reads and reads with mapping quality less than 30 were removed from the sample-level BAM files using Samtools. Lastly, samples deriving from either the same individual or a monozygotic twin were merged based on both an initial run of ngsRelate and results from CrosscheckFingerprints (Picard; Supplementary Table 1). Because it is impossible, based on genetics, to distinguish whether two identical samples are derived from the same individual or from monozygotic twins, we merge all genetically

identical samples. Nevertheless, in the case of FRA029, both samples belonging to this genetic individual are actually derived from the upper right first molar. This observation demonstrates that we have at least one pair of monozygotic twins in our dataset.

Average depth of coverage values was calculated using BEDTools genomecov (v.2.30.0), and mitochondrial haplogroups were assigned using mutserve (v.1.3.0) and haplogrep (v.2.1.25). Chromosome Y haplogroups were assigned using a tool developed in house and described in Supplementary Note 2.

For PCA we called pseudohaploid genotypes by random selection of one of the variants from mpileup (Samtools) that passed mapping and base quality filters (over 30). The resulting panel of pseudohaploid genotypes were merged with a panel of 1,430 previously published ancient shotgun genomes. We used smartpca from eigensoft to carry out PCA on 2,086,279 transversion-only SNPs with minor allele frequency of over 0.1%. All samples were projected onto the variation of the reference panel.

### IBD analysis

For identity-by-descent analysis we imputed the genotypes of 82 individuals with sufficient coverage (over 0.1 $\times$ ) using Glimpse<sup>47</sup> (v.1.1.1; <https://github.com/odelaneau/GLIMPSE/tree/glimpse1>), as described in ref. 23. Imputed genomes were then merged with a panel of imputed shotgun-sequenced genomes (Supplementary Table 7). Genomic segments shared via identity-by-descent were identified in all pairs of ancient samples using IBDseq<sup>48</sup> (v. r1206). Next we excluded the lowest-coverage individual in all pairs of close relatives (shared IBD segments over 1,500 centimorgans (cM), resulting in a total of 54 individuals from this study remaining for clustering (Supplementary Table 8). Genetic clustering was carried out based on IBD-sharing patterns in all pairs of samples, as described in ref. 23. For analysis of ROHs we filtered the imputed panel described above for genotype posterior values over 0.95 and identified segments of homozygosity-by-descent using IBDseq (v. r1206). These segment lengths were converted to cM using the HapMap Phase II genetic map ([https://github.com/johnbowes/CRAFT-GP/tree/master/source\\_data/genetic\\_map\\_HapMapII\\_GRCh37](https://github.com/johnbowes/CRAFT-GP/tree/master/source_data/genetic_map_HapMapII_GRCh37)) and summarized in six bins based on segment length.

### Kinship reconstruction

Initially we tested five different approaches to reconstruct kinship based on ancient DNA data (KING<sup>49</sup>, ngsRelate<sup>50,51</sup>, READ<sup>52</sup>, KIN<sup>53</sup> and the relationship inference algorithm based on pairwise identity-by-state sharing presented in ref. 54). We found that all five approaches produced very similar results, but in some edge cases we observed that ngsRelate outperformed the four other approaches. Accordingly we decided to use ngsRelate for our analysis. We ran ngsRelate on a subset of the imputed panel described above, focusing on ancient Northern Europeans (816 individuals) and on 2,086,279 transversion-only SNPs with minor allele frequency over 0.1% in the panel. Next we classified all pairs of individuals as either zero-, first-, second- or third-degree relatives or as unrelated, based on pairwise relatedness ( $R_{AB}$ ). We used the midpoint between the theoretical estimates of pairwise relatedness to distinguish between each degree of relatedness. For example, for zero-, first- and second-degree relatives the theoretical pairwise relatedness values are 1.00, 0.50 and 0.25, respectively. Accordingly we define first-degree relatives as pairs of individuals with  $R_{AB}$  values between 0.75 (midpoint between 1.0 and 0.5) and 0.375 (midpoint between 0.50 and 0.25; Supplementary Table 9). Lastly, to distinguish relationship type for second-degree relatives we used results from the likelihood files produced by KIN. To run KIN we first ran KINGaroo on filtered BAM files, focusing on 2,086,279 transversion-only SNPs with minor allele frequency over 0.1%. KIN was then run using default settings.

For pedigree reconstruction we used a combination of the automated tool PRIMUS (pedigree reconstruction and identification of a maximum unrelated set<sup>55</sup>) and the manual triangulation approach described in ref. 36 (Supplementary Note 2).

## Pathogen screening

We carried out a screening for hits to all known human pathogens on all samples sequenced. The detailed description of the pathogen-screening pipeline can be found elsewhere<sup>56</sup>. In brief, we classified all reads using KrakenUniq<sup>57</sup> (v.0.5.8) on a custom database of microbial reference genomes, with a focus on human pathogens and common environmental microbes. For genera represented by more than 50 *k*-mer in a sample, traditional mapping was carried out. For a given genus/sample combination we used minimap2 (v.2.17-r941)<sup>58</sup> to map all reads classified to that genus to corresponding reference genomes of all species within that genus. Next we calculated standard mapping statistics for each sample/species combination, such as depth of coverage, ancient DNA damage, average nucleotide identity, mean edit distance and mean read length. We also calculated an estimate of coverage evenness (actual breadth of coverage/expected breadth of coverage given the depth of coverage) and ranked each species within a genus based on its number of unique *k*-mers identified by krakenUniq. We define a positive microbe detection as species represented by more than 50 reads, that are ranked as having the highest number of unique *k*-mers within their genus, have an average number of soft clippings under eight, have an average nucleotide identity of over 0.97 and have a coverage evenness score of over 0.8.

## Plague phylogenetic analysis

For all samples in which *Y. pestis* was detected in the pathogen screening, we mapped trimmed reads from all libraries of that sample to the reference plague genome (CO92; GCA\_000009065.1) using bowtie2 (v.2.3.2) with the following parameters: -D 20 -R 3 -N 1 -L 20 -i S,1,0.50 --end-to-end --no-unal. Next we merged and sorted BAM files, and marked duplicates using MarkDuplicates from Picard Tools. Lastly, duplicate reads and reads of mapping quality under 30 were removed from the BAM files using Samtools.

The reference plague panel was generated by downloading raw read data from 400 modern plague strains, 30 ancient plague strains and data from the *Y. pseudotuberculosis* reference genome (GCF\_000834295.1) at the European Nucleotide Archive. Raw fastq read files were trimmed with AdapterRemoval and mapped to the reference plague genome as described above, followed by filtering and calculation of the average depth of coverage using BEDTools genomcov. We called genotypes for plague samples with an average depth of coverage over ten for modern samples and over four for ancient samples, with the exception of Gok2, which was included despite having lower coverage. Genotype calling was carried out using Genome Analysis Toolkit<sup>59</sup> (v.4.1.9.0). First, sample-wise genomic variant call formats (GVCFs) were generated using HaplotypeCaller with a record at every single site (-ERC BP\_RESOLUTION). Next, GVCFs were merged with CombineGVCFs and genotypes were called jointly with GenotypeGVCFs. Using VariantFiltration, genotypes were set to no-call if genotype quality was below 50 or if depth was over 1,000 or under four or ten, for ancient and modern samples, respectively. Furthermore, genotypes were set to no-call if genotype was supported by less than 90% of the reads at a given position. Lastly we filtered variants using VariantFiltration based on variant confidence/quality by depth, Phred-scaled *P* value, root mean squared mapping quality, *z*-score of Alt versus Ref read mapping qualities, *z*-score of Alt versus Ref read position bias, as follows: --filter-expression "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0".

To generate the phylogenetic tree we converted the vcf file to a multi-fasta using bcftools consensus, replacing missing data with N (flag: -M N) and filtering out regions with high proportions of reads of mapping quality zero (flag: -m [bedfile]; Supplementary Note 2). We used RAXML-NG (v.0.9.0)<sup>60</sup> to generate the phylogenetic tree, applying the GTR + G model and using the *Y. pseudotuberculosis* reference genome (GCF\_000834295.1) as an outgroup. Bootstrap support values were

calculated with raxml using both the transfer bootstrap expectation and Felsenstein's bootstrap proportions metric. Both sets of bootstrap values produced similar support for the main nodes of the phylogeny. Lastly we plotted the best tree generated by RAXML-NG using the ggtree<sup>61,62</sup> (v.2.0.4) package in R (v.4.2.2).

For lower-coverage plague genomes (0.01–1.0×) for which traditional genotype calling was not feasible, we used EPA-ng (v.0.3.8)<sup>63</sup> to place each sample onto the phylogeny described above. For this approach we used major allele genotypes called by selecting the most common allele passing filters (mapping quality (MQ) over 30, base quality (BQ) over 30) from Samtools mpileup, and converted the vcf to fasta format as described above. Next we subset the fasta to the same SNPs as in the reference panel using BEDTools getfasta. We ran the evolutionary placement algorithm on the resulting fasta file, with default parameters supplying the best tree and the associated model from RAXML-NG as reference. Placements were plotted using the ggtree<sup>61,62</sup> (v.2.0.4) package in R (v.4.2.2).

## Pangenome analysis

We used a pangenome variation graph framework to investigate variation in pangenome content among ancient *Y. pestis* lineages. We selected 81 complete reference assemblies of the *Y. pseudotuberculosis* complex (56 *Y. pestis*, 24 *Y. pseudotuberculosis* and one *Y. similis*) from NCBI (assembly level 'chromosome' or 'complete') and built a pangenome variation graph using pggg<sup>64</sup> with the following parameters:

- segment length: 5 kb (-s 5,000)
- pairwise identity: 0.9 (-p 0.9)
- minimum match length: 47 (-k 47)
- target sequence length for POA: 700, 900, 1,110 (-G 700,900,1,100)
- score parameters for POA: asm5 (-P asm5)

We built graphs separately for the chromosome and plasmids and merged them into a single-reference graph using the 'vg combine' function of the vg toolkit<sup>65</sup>. The resulting graph was then indexed for short-read mapping with the Giraffe<sup>66</sup> mapper, with default parameters except *k*-mer length (-k 15) and window length (-w 5) for the minimizer. Shotgun-sequencing reads for each sample were then mapped to this index using 'vg giraffe'. We filtered the resulting graph alignment files (GAM format) for mapping quality of 30 or above (-q 30) and normalized alignment score of 0.9 or above (-r 0.9) and removed reads with tandem repeat motifs at the read ends (-E 3). To obtain coverage statistics we then computed read support for each node in the reference graph using 'vg pack' with the '-d' option. We considered a reference graph node covered in an ancient sample if the proportion of the node length covered by sequencing reads was 0.2 or above. To classify nodes into groups based on their presence or absence among the three *Y. pseudotuberculosis* complex species in the reference graph, we obtained a path coverage haplotype matrix for the reference contigs across all nodes in the graph using 'odgi paths'<sup>67</sup> with the '-H' option. Each node was then labelled according to the combination of *Y. pseudotuberculosis* complex species with coverage observed. For each ancient sample we aggregated coverage across nodes and classes to obtain the total amount of pangenome sequence contained in each class.

To identify the gene content of the *Y. pseudotuberculosis* complex pangenome in ancient *Y. pestis* lineages absent from modern *Y. pestis* strains, we extracted all nodes of the reference graph from classes not containing modern *Y. pestis* (that is, shared by *Y. similis* and *Y. pseudotuberculosis* or observed only within one of the two) in which at least one ancient sample had coverage. The reference graph start/end coordinates for those nodes were then translated into reference path positions using 'odgi position', and nodes were collapsed into regions if their translated position was within 200 base pairs of each other on the same reference contig. From this list of initial candidate regions we selected all of length greater than 500 base pairs, extended them by 10 kb on each side and merged overlapping regions to yield a

final list of genomic regions harbouring putative pangenome content retained in ancient *Y. pestis*. To further investigate coverage across genes in those regions we split the reference graph into region-specific graphs using ‘vg chunk’ and produced single-reference alignments (BAM format) for each ancient sample and region using ‘vg surject’. The resulting BAM files were filtered by removal of duplicates (picard) and retaining only reads with mapping quality over 30, minimum number of aligned bases over 30 and, at most, five soft-clipped bases. Final per-gene coverage statistics were obtained from the filtered BAM files using ‘bedtools coverage’<sup>68</sup>.

## Radiocarbon dating and modelling

For Frälsegården a total of 91 radiocarbon dates are available, of which 46 are produced within the present project. The new datings targeted individuals in the pedigree, those showing plague and those with Steppe ancestry or HG (hunter-gatherer) admixture. Some samples were also re-datings of individuals with anomalous previous dates. Seven new datings were made on samples from the Landbogården passage grave and six from the Nästegården site (Supplementary Table 4).

New datings were performed at the Keck carbon cycle AMS facility, University of California, Irving. The samples were decalcified in 1 N HCl, gelatinized at 60 °C and pH 2.0 and ultrafiltered to select a high molecular weight fraction (over 30 kDa).  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values were measured to a precision level of below 0.1‰ and below 0.2‰, respectively, on aliquots of ultrafiltered collagen, using a Fisons NA1500NC elemental analyser/Finnigan Delta Plus isotope ratio mass spectrometer. Further details can be found at <https://sites.ps.uci.edu/kccams/education/protocols/>.

All samples were dated successfully, and showed C:N ratios well within the accepted range for well-preserved collagen, 2.9–3.6 (ref. 69).  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values indicate a generally terrestrial diet with little or no contribution from marine or freshwater protein, suggesting that reservoir effects are negligible (Supplementary Table 4).

Datings were calibrated in Oxcal 4.4.4 using the Intcal20 calibration curve<sup>70,71</sup>. Because the time period of interest shows a large plateau in the calibration curve, the calibrations result in broad ranges, usually approximately 5250–4850 cal. BP (2 $\sigma$ ). Chronological modelling of individuals within the pedigrees was made, allowing considerably higher precision. The model makes the simple assumption that genetically identified generations can be regarded as sequential phases. Tooth dates were offset by  $20 \pm 5$  years to make them comparable to those from bone, based on formation time. The assumption of a generational interval of  $29 \pm 19$  years, as suggested in ref. 27, was tested but did not improve the result. Four dates were identified as outliers by both low-agreement indices and outlier analysis (HJE012, FRA023, FRA002 and FRA009). These were marked as outliers in the Oxcal codes but not modelled. Detailed results and Oxcal code can be found in Supplementary Note 3 and Supplementary Tables 11–13.

## Strontium isotope analysis

Ten new Sr isotope values were produced within the present project, all on samples from Frälsegården (samples F11500–F11510; Supplementary Table 5). In addition, 24 samples from Frälsegården, Firse sten, Hunnebostrand and Rössberga were previously published<sup>11,72,73</sup>. A further 34 samples were previously analysed and are reported here for the first time (Supplementary Table 5). The measurements are bulk values taken on small pieces of tooth enamel, representing an average over the time of enamel formation. All measurements of  $^{87}\text{Sr}/^{86}\text{Sr}$  ratios were conducted at the Geochronology and Isotope Geochemistry Laboratory, Department of Geological Sciences, University of North Carolina, Chapel Hill. Samples were dissolved in nitric acid and the strontium fraction purified by ion selective chromatography (Eichrom Sr resin) before analysis by thermal ionization mass spectrometry (TIMS) on a VG Sector 54 mass spectrometer run in dynamic mode. Internal precision in the laboratory is consistently around 0.0007% standard

error (or  $1\sigma = 0.00006$  in the ratio of a particular sample). Long-term, repeated measurements of SRM-987 are around 0.710260—an acceptable difference from the recognized value of 0.710250—and raw sample values from individual runs are standardized to the recognized value of SRM-987. Baseline values from Falbygden and surrounding areas in western Sweden were published in ref. 74.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

Raw fastq files from this study have been deposited in the European Nucleotide Archive under accession number PRJEB76142; see Supplementary Table 15 for an overview of European Nucleotide Archive accession numbers and their corresponding sample and library identifiers. Furthermore, human (hg19) and *Y. pestis* (GCF\_000009065) BAM alignment files and genotype data in vcf file format are available from <https://doi.org/10.17894/ucph.d2098f40-1263-4e28-afed-a0d50ff25c0c>. The map of Scandinavia in Fig. 1 was created with the rworldmap package in R using data from the Natural Earth dataset, and the insert map of Falbygden was created with data from OpenStreetMap. The map of Scandinavia in Fig. 2b was created with the R package ‘maps’ using data from the Natural Earth dataset.

- Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* **13**, 2447–2461 (2018).
- Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5448 (2010).
- Gansauge, M. T., Aximu-Petri, A., Nagel, S. & Meyer, M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat. Protoc.* **15**, 2279–2300 (2020).
- Yang, D. Y., Eng, B., Wayne, J. S., Dudar, J. C. & Saunders, S. R. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol.* **105**, 539–543 (1998).
- Svensson, E. M. et al. Tracing genetic change over time using nuclear SNPs in ancient and modern cattle. *Anim. Genet.* **38**, 378–383 (2007).
- Svensson, E. et al. Genome of Peştera Muierii skull shows high diversity and low mutational load in pre-glacial Europe. *Curr. Biol.* **31**, 2973–2983 (2021).
- Briggs, A. W. & Heyn, P. Preparation of next-generation sequencing libraries from damaged DNA. *Methods Mol. Biol.* **840**, 143–154 (2012).
- Wagner, D. M. et al. *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. *Lancet Infect. Dis.* **14**, 319–326 (2014).
- Lindgreen, S. AdapterRemoval: easy cleaning of next generation sequencing reads. *BMC Res. Notes* **5**, 337 (2012).
- Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J. & Delaneau, O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* **53**, 120–126 (2021).
- Browning, B. L. & Browning, S. R. Detecting identity by descent and estimating genotype error rates in sequence data. *Am. J. Hum. Genet.* **93**, 840–851 (2013).
- Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
- Hanghøj, K., Moltke, I., Andersen, P. A., Manica, A. & Korneliussen, T. S. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *GigaScience* **8**, giz034 (2019).
- Korneliussen, T. S. & Moltke, I. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics* **31**, 4009–4011 (2015).
- Kuhn, J. M. M., Jakobsson, M. & Günther, T. Estimating genetic kin relationships in prehistoric populations. *PLoS ONE* **13**, e0195491 (2018).
- Popli, D., Peyrègne, S. & Peter, B. M. KIN: a method to infer relatedness from low-coverage ancient DNA. *Genome Biol.* **24**, 10 (2023).
- Schroeder, H. et al. Unraveling ancestry, kinship, and violence in a Late Neolithic mass grave. *Proc. Natl Acad. Sci. USA* **166**, 10705–10710 (2019).
- Staples, J. et al. PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *Am. J. Hum. Genet.* **95**, 553–564 (2014).
- Sikora, M. et al. The landscape of ancient human pathogens in Eurasia from the Stone Age to historical times. Preprint at [bioRxiv](https://doi.org/10.1101/2023.10.06.561165) <https://doi.org/10.1101/2023.10.06.561165> (2023).
- Pockrandt, C., Zimin, A. V. & Salzberg, S. L. Metagenomic classification with KrakenUniq on low-memory computers. *J. Open Source Softw.* **7**, 4908 (2022).
- Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
- Van der Auwera, G. & O’Connor, B. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (O’Reilly Media, 2020).
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).

61. Yu, G., Lam, T. T. Y., Zhu, H. & Guan, Y. Two methods for mapping and visualizing associated data on phylogeny using GGTree. *Mol. Biol. Evol.* **35**, 3041–3043 (2018).
62. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
63. Barbera, P. et al. EPA-ng: massively parallel evolutionary placement of genetic sequences. *Syst. Biol.* **68**, 365–369 (2019).
64. Garrison, E. et al. Building pangene graphs. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.04.05.535718> (2023).
65. Garrison, E. et al. Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat. Biotechnol.* **36**, 875–879 (2018).
66. Sirén, J. et al. Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science* **374**, abg8871 (2021).
67. Guarracino, A., Heumos, S., Nahnsen, S., Prins, P. & Garrison, E. ODGI: understanding pangene graphs. *Bioinformatics* **38**, 3319–3326 (2022).
68. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
69. van Klinken, G. J. Bone collagen quality indicators for palaeodietary and radiocarbon measurements. *J. Archaeol. Sci.* **26**, 687–695 (1999).
70. Ramsey, C. B. Dealing with outliers and offsets in radiocarbon dating. *Radiocarbon* **51**, 1023–1045 (2009).
71. Reimer, P. J. et al. The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757 (2020).
72. Hollund, H. I., Blank, M. & Sjögren, K.-G. Dead and buried? Variation in post-mortem histories revealed through histotaphonomic characterisation of human bone from megalithic graves in Sweden. *PLoS ONE* **13**, e0204662 (2018).
73. Sjögren, K.-G., Price, D. & Ahlström, T. Megaliths and mobility in south-western Sweden. Investigating relationships between a local society and its neighbours using strontium isotopes. *J. Anthropol. Archaeol.* **28**, 85–101 (2009).
74. Blank, M., Sjögren, K.-G., Knipper, C., Frei, K. M. & Storå, J. Isotope values of the bioavailable strontium in inland southwestern Sweden—a baseline for mobility studies. *PLoS ONE* **13**, e0204649 (2018).

**Acknowledgements** This study was supported by the Lundbeck Foundation (no. R322-2019-2610 to F.V.S.), L. J. Häggglunds Stiftelse (to F.V.S) and Riksbankens Jubileumsfond (no. M 21-0018). Furthermore, the research at the University of Copenhagen was carried out under the

Lundbeck Foundation GeoGenetics Centre, which is supported by the the Lundbeck Foundation (nos. R302-2018-1799 2155 and R155-2013-16338), the Novo Nordisk Foundation (no. NNF18SA0035006), the Wellcome Trust (no. UNS69906), the Carlsberg Foundation (no. CF18-0024), the Danish National Research Foundation (no. 44113220) and the University of Copenhagen (KU2016 programme). The research at Uppsala University was supported by The Swedish Research Council (nos. 2013-1905 to M.J., A.G. and J. Storå, and 2017-02503 to H. Malmström), Riksbankens Jubileumsfond (no. M13-0904:1 to M.J., A.G. and J. Storå) and the Knut and Alice Wallenberg foundation (Atlas of Ancient Human Genomes in Sweden project to M.J., A.G. and J. Storå). Computations and data handling were enabled by resources provided by the Swedish National Infrastructure for Computing at the Uppsala Multidisciplinary Center for Advanced Computational Science, partially funded by the Swedish Research Council through grant agreement no. 2018-05973. Sequencing was performed by The National Genomics Infrastructure Uppsala and at GeoGenetics Sequencing Core, Copenhagen. Investigations at Fråsegården were funded by Länsstyrelsen Västra Götaland and Riksbankens Jubileumsfond, grant no. P2006-0011:1-E.

**Author contributions** F.V.S., K.-G.S. and M.S. led the study. A.G., E.W., M.J., H. Malmström and M.S. supervised research. K.-G.S., M.B., A.F., M.E.A., T. Ahlström and J. Storå collected samples. F.V.S., E.M.S., J. Staring, M.F., T.P., H. McColl, C.G., T.R.B., L.G. and B.V.-R. conducted laboratory work. F.V.S., K.-G.S., J.K. and M.S. analysed data and produced the figures. K.-G.S., A.F., T. Axelsson, T. Ahlström, J. Storå, K.K. and H. Malmström provided interpretation of the archaeological context. K.-G.S., T.D.P. and T. Ahlström collected and curated bioarchaeological data. T.D.P. analysed strontium samples. A.K.N.I. provided interpretation of pathogen host interactions. F.V.S., K.-G.S. and M.S. wrote the manuscript with input from all co-authors.

**Competing interests** The authors declare no competing interests.

#### Additional information

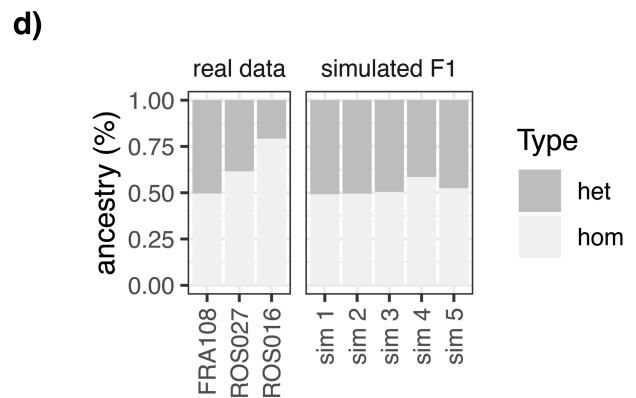
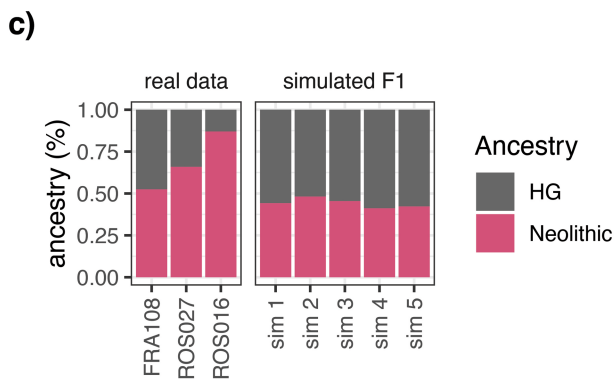
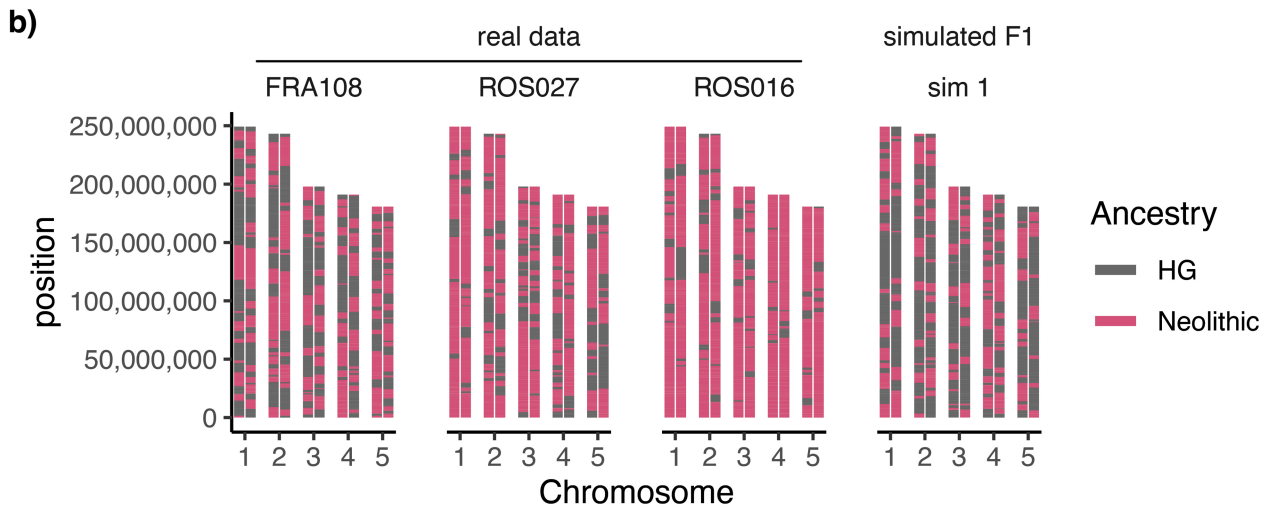
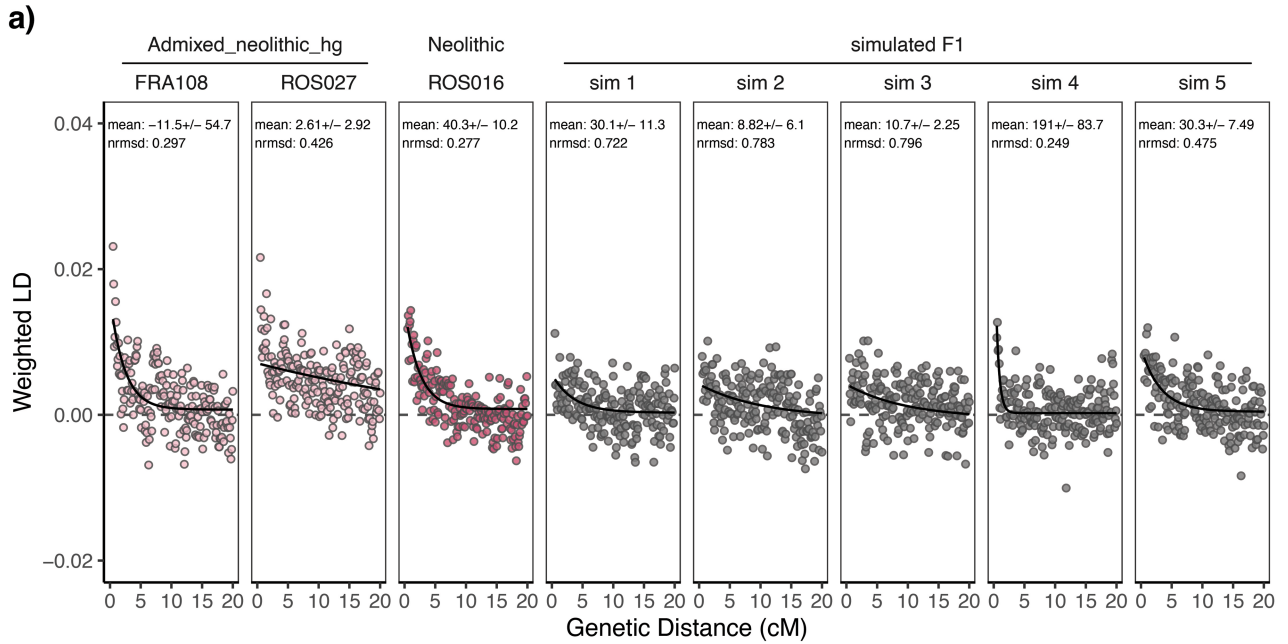
**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-024-07651-2>.

**Correspondence and requests for materials** should be addressed to Frederik Valeur Seersholm or Martin Sikora.

**Peer review information** Nature thanks Duncan Garrow, Ben Krause-Kyora, Ludovic Orlando and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

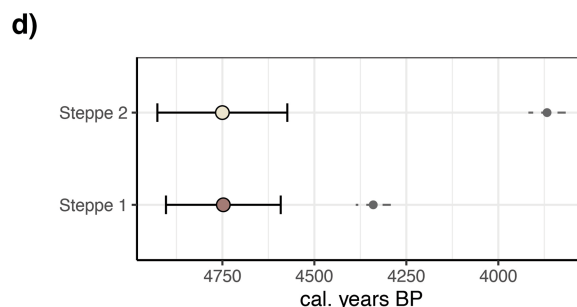
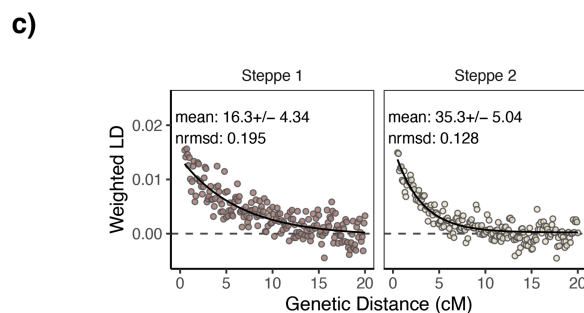
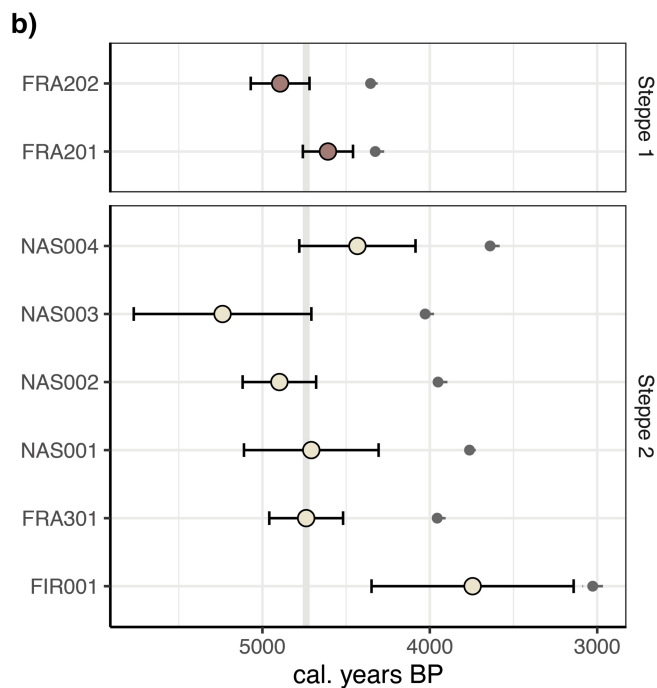
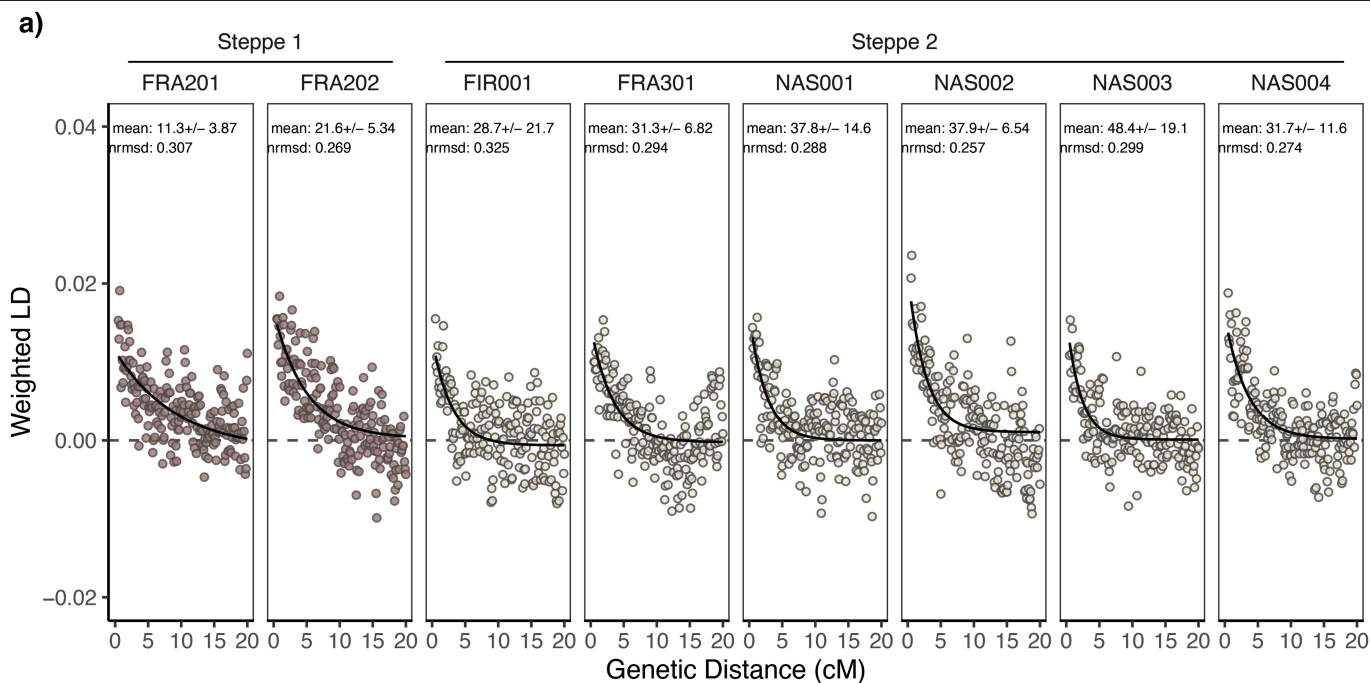




**Extended Data Fig. 1 | An investigation of the two admixed individuals.**

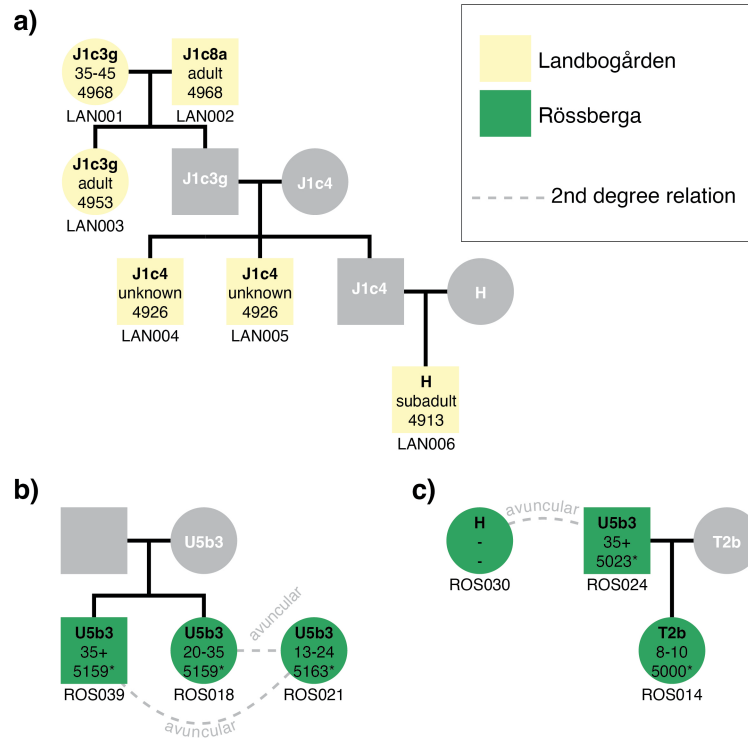
**a)** Estimates of admixture timing in admixed and simulated F1-individuals. Admixture time estimates were calculated with DATES using Swedish Neolithic farmer related individuals and PWC individuals as source groups (see methods). One un-admixed individual (ROS016) was included as a control. **b)** Local ancestry inference of admixed individuals across chromosomes one to five. RFmix estimates of hunter-gatherer (PWC) and Neolithic Swedish ancestries for admixed individuals (FRA108 and ROS027), a Neolithic individual with no

evidence of recent admixture (ROS016) and simulated F1 individuals with equal parts HG and Neolithic DNA. To simplify the plot, only chromosomes one to five and one simulated individual is shown. **c)** Total fraction of each ancestry type across the genome of each individual. **d)** Total fraction of genotype class across the genome of each individual, where 'het' indicates regions where one allele is of HG ancestry while the other is of Neolithic ancestry, while 'hom' indicates that both alleles are of the same ancestry.



**Extended Data Fig. 2 | Estimates of admixture timing in Steppe related Individuals.** Admixture time estimates were calculated with DATES using Swedish Neolithic farmer related individuals and Yamnaya individuals as source groups (see methods). **a)** Sample-wise weighted linkage disequilibrium measures against genetic distance in centimorgan. In the top right corner of each plot, mean relative admixture time and standard error is shown in generations ago. **b)** Estimates of absolute admixture time using a generation

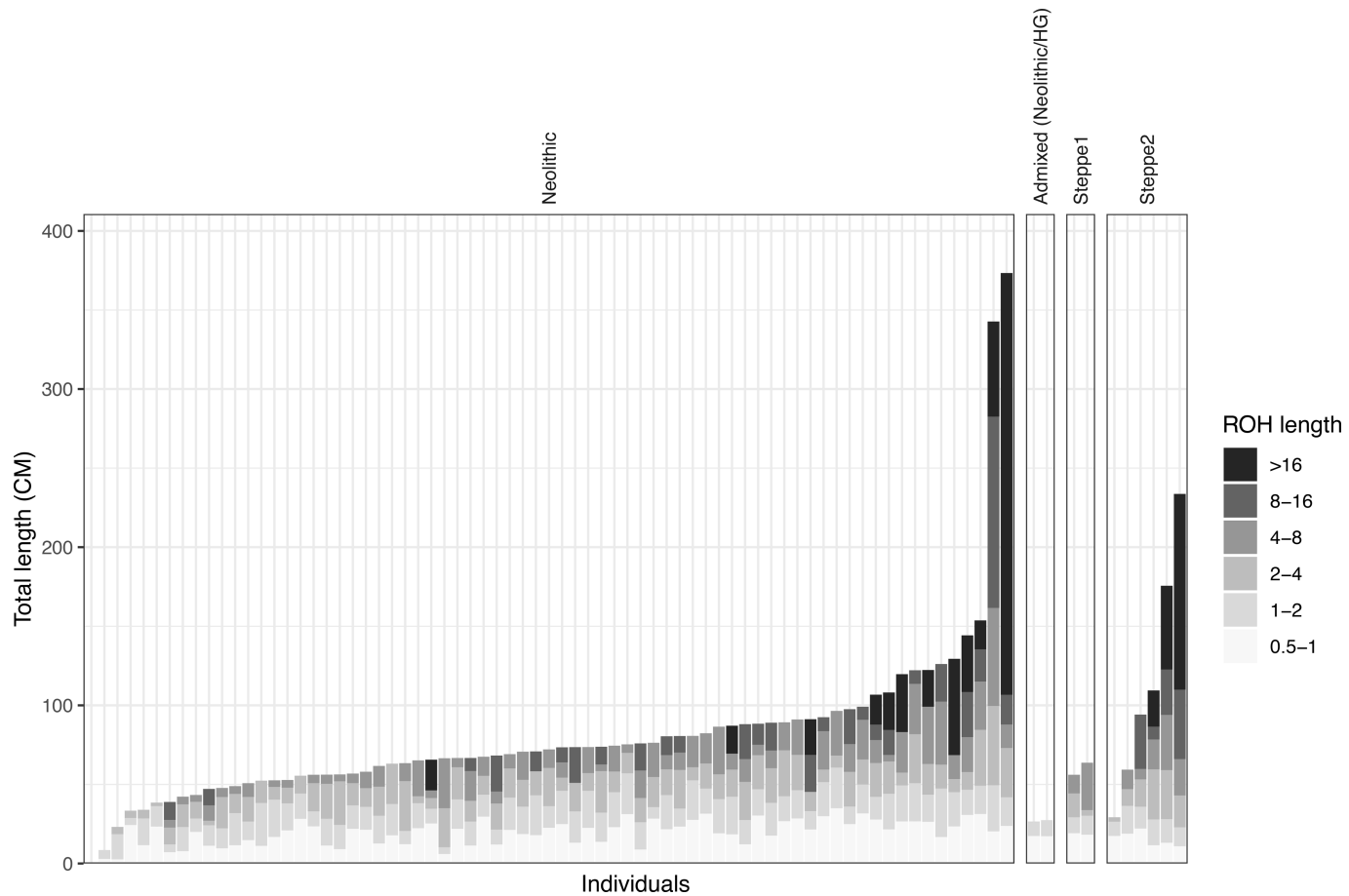
time of 25 years and the calibrated ages of each sample (Supplementary Table 4; shown in grey). Highlighted in grey is the time period between 4,718 and 4,758 cal. BP where confidence intervals of all samples except FIR001 overlap. **c)** Weighted linkage disequilibrium measures against genetic distance estimated for the two Steppe related groups. **d)** Estimates of absolute admixture time for the two Steppe related groups using the average age of samples in each group (shown in grey) as the group age.



**Extended Data Fig. 3 | Pedigrees from Landbogården and Rössberga.**

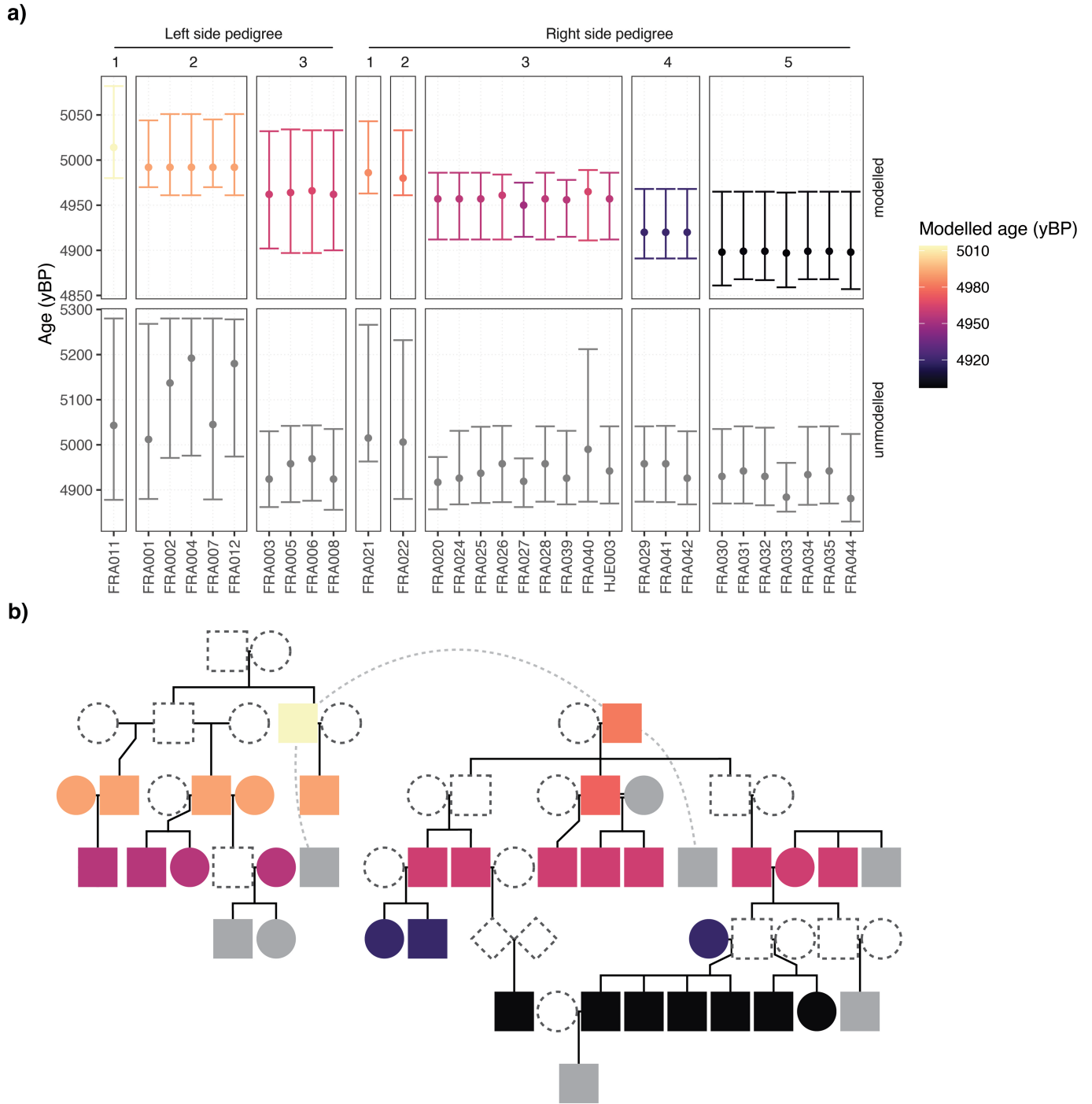
**a)** Pedigree 2 from Landbogården. **b)** Pedigree 3 from Rössberga. **c)** Pedigree 4 from Rössberga. Squares and circles represent males and females, respectively, and information on mitochondrial haplogroup, osteological age estimate, and radiocarbon date for each individual is indicated inside each shape. Yellow and green colours indicate the sites Landbogården and Rössberga, respectively,

while grey colour represents unsampled individuals. Solid black lines indicate first degree relations. Dashed grey lines signify unknown 2<sup>nd</sup> degree relationships. \*For individuals from Rössberga radiocarbon dates could not be modelled, accordingly dates for these individuals are just reported as the median of the calibrated date.



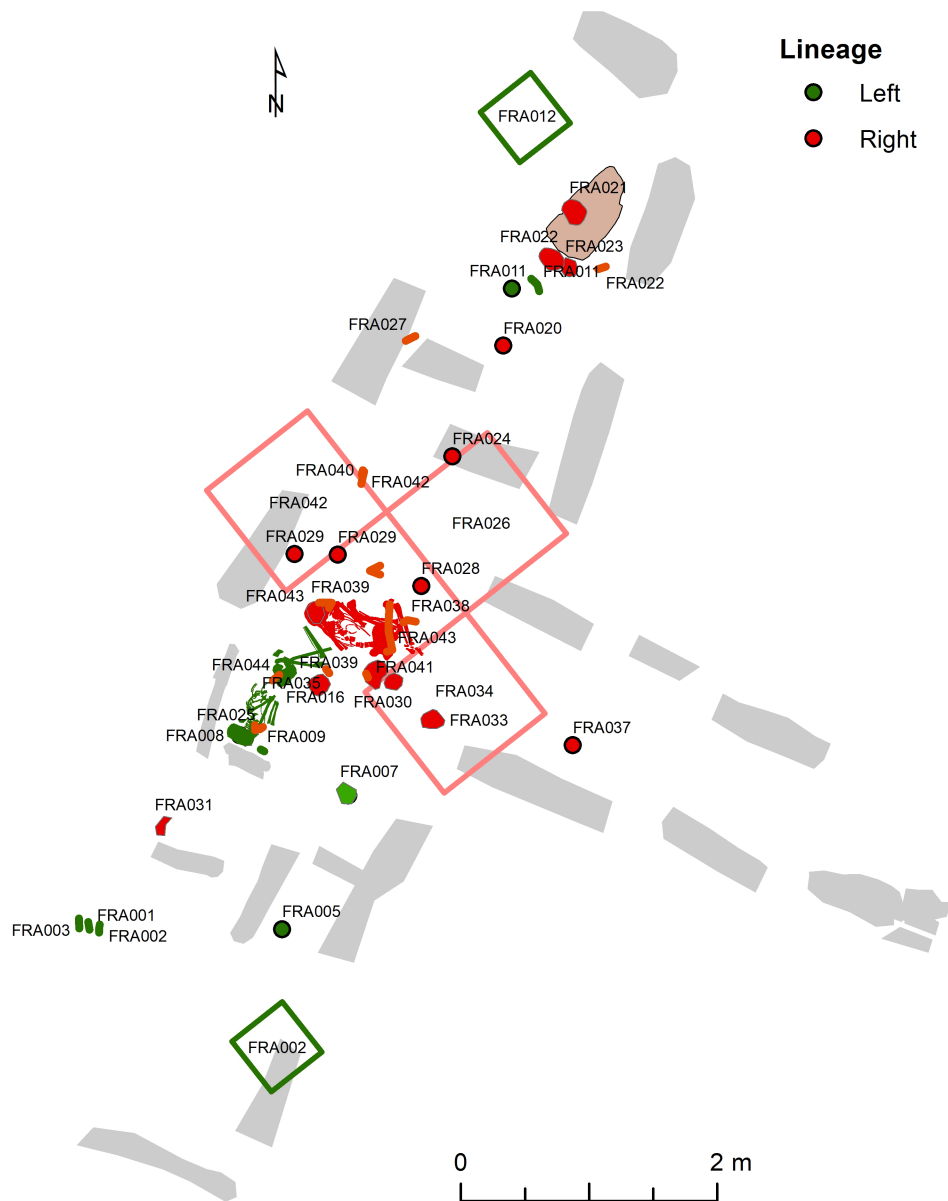
**Extended Data Fig. 4 | Runs of Homozygosity.** Total length of long Runs of Homozygosity (>0.5 cM) for each individual coloured by ROH size. Plot is split into four panels representing each ancestry group (see Fig. 2).



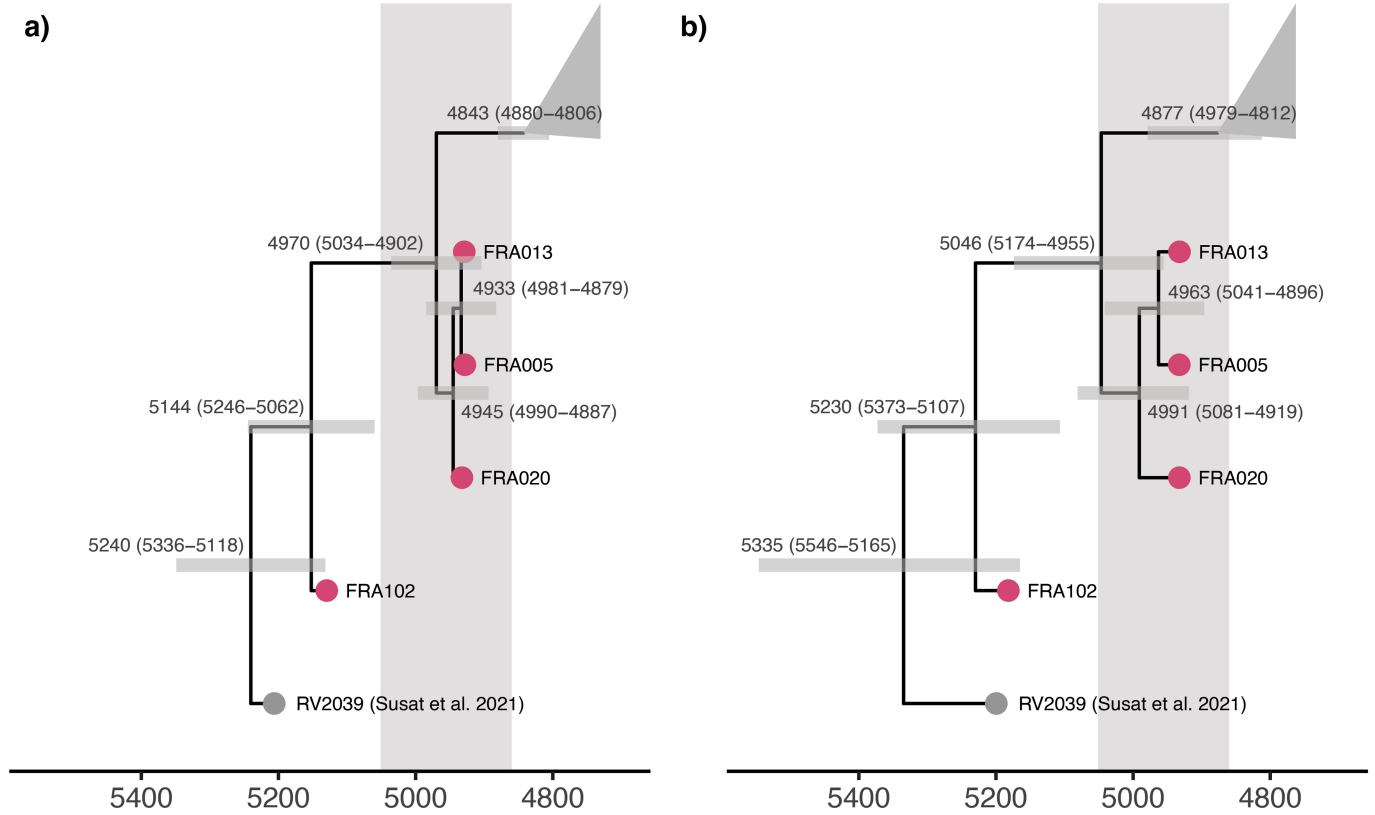


**Extended Data Fig. 5 | Chronological modelling.** **a)** Median ages and  $2\sigma$  age ranges for modelled and unmodelled dates. Plot is stratified by pedigree branch (left side and right side, respectively) and generation in pedigree. **b)** Pedigree 1 from Frälsegården coloured by modelled age, following the color scheme of a. Grey color indicates individuals where no modelled age is available.

Such individuals with missing data are comprised by 1) individuals that cannot be placed in the pedigree (stippled lines), 2) individuals with dates classified as outliers (Supplementary Table 4), or 3) individuals that are linked to the pedigree through unknown second degree relationships.

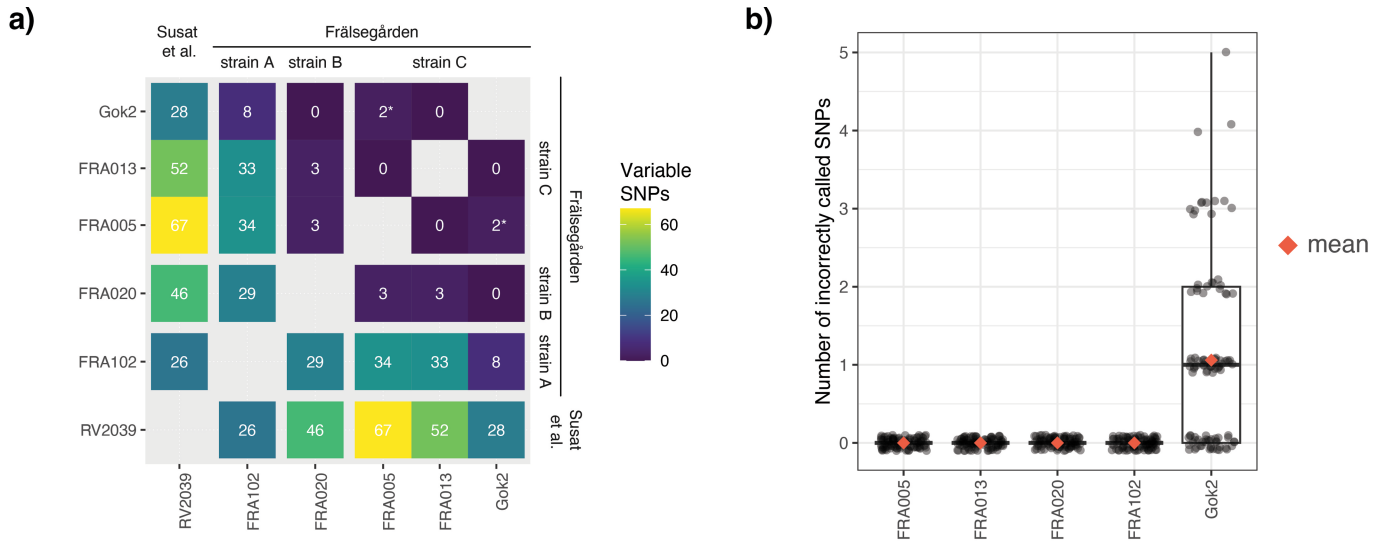


**Extended Data Fig. 6 | Burial locations within the Frälsegården passage grave coloured by lineage.** Green colour represents the left side pedigree from Fig. 3, while red colour represents the right side pedigree. For samples without exact coordinates, the quadrant from which the sample was excavated is plotted instead.



**Extended Data Fig. 7 | Molecular dating of ancient plague strains.** Dating analysis were carried out on preLNBA strains (grey and pink color) and LNBA-strains (collapsed in grey triangle). The modelled chronological span of the pedigree from Frälsegården is highlighted in grey. **a)** BEAST analysis using the

Coalescent Bayesian Skyline demographic model and assuming an Optimised Relaxed Clock and the GTR substitution model with four gamma categories and empirical frequencies. **b)** BactDating analysis using the 'relaxedgamma' model.



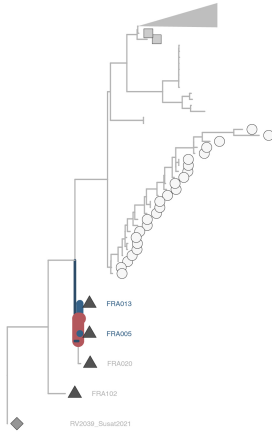
**Extended Data Fig. 8 | Variation within preLNBA plague strains.** **a)** Heatmap of pairwise SNP diversity of preLNBA plague strains. Plague genomes identified in this study (from Frålsögården) are divided into strains A, B, and C (see Fig. 4). \*Unreliable SNPs (discussed in Supplementary Note 4). **b)** Number of incorrectly called SNPs from simulated data. For each sample, simulated plague data was

generated with the same damage profile and coverage as the real data in 100 replicates. After variant calling the total number of incorrectly called SNPs were counted for each replicate. **c)** SNPs unique to Gok002. **d)** SNPs unique to FRA020.



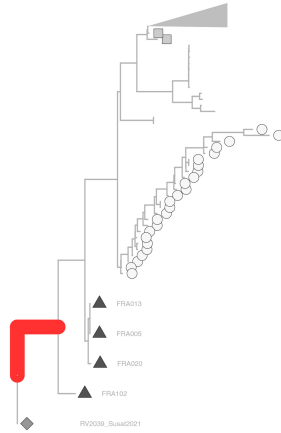
## FRA106

coverage: 1.1x  
site: Frålsegården, Sweden



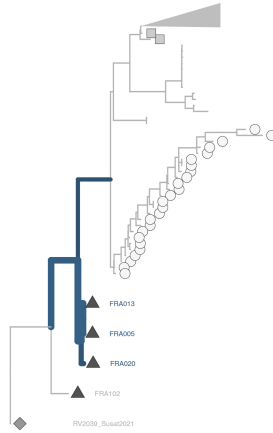
## AVL001

coverage: 0.2x  
site: Avlebjerg, Stroby, Denmark



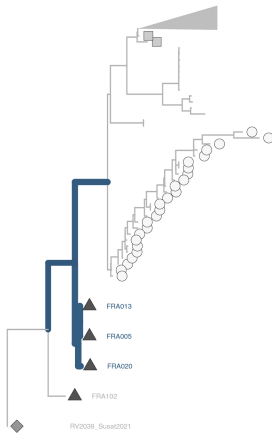
## FRA003

coverage: 0.13x  
site: Frålsegården, Sweden



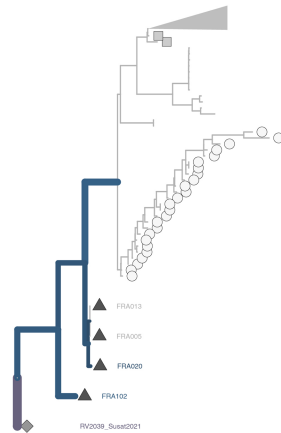
## FRA042

coverage: 0.074x  
site: Frålsegården, Sweden



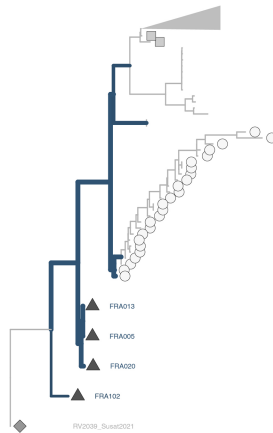
## HJE003

coverage: 0.05x  
site: Hjelms rår, Sweden

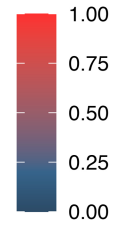


## Gok4

coverage: 0.023x  
site: Frålsegården, Sweden



Likelihood weight ratio

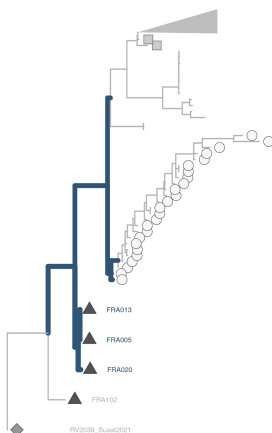


hgMajor

- LNBAminus
- LNBAplus
- ◆ preLNBA
- ▲ test

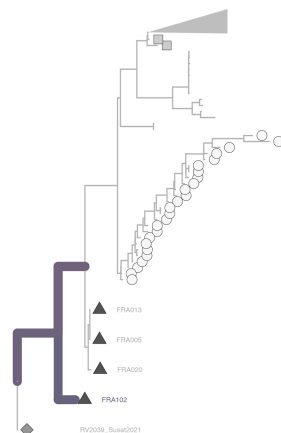
## NA (CGG106001)

coverage: 0.023x  
site: Landbogården, Sweden

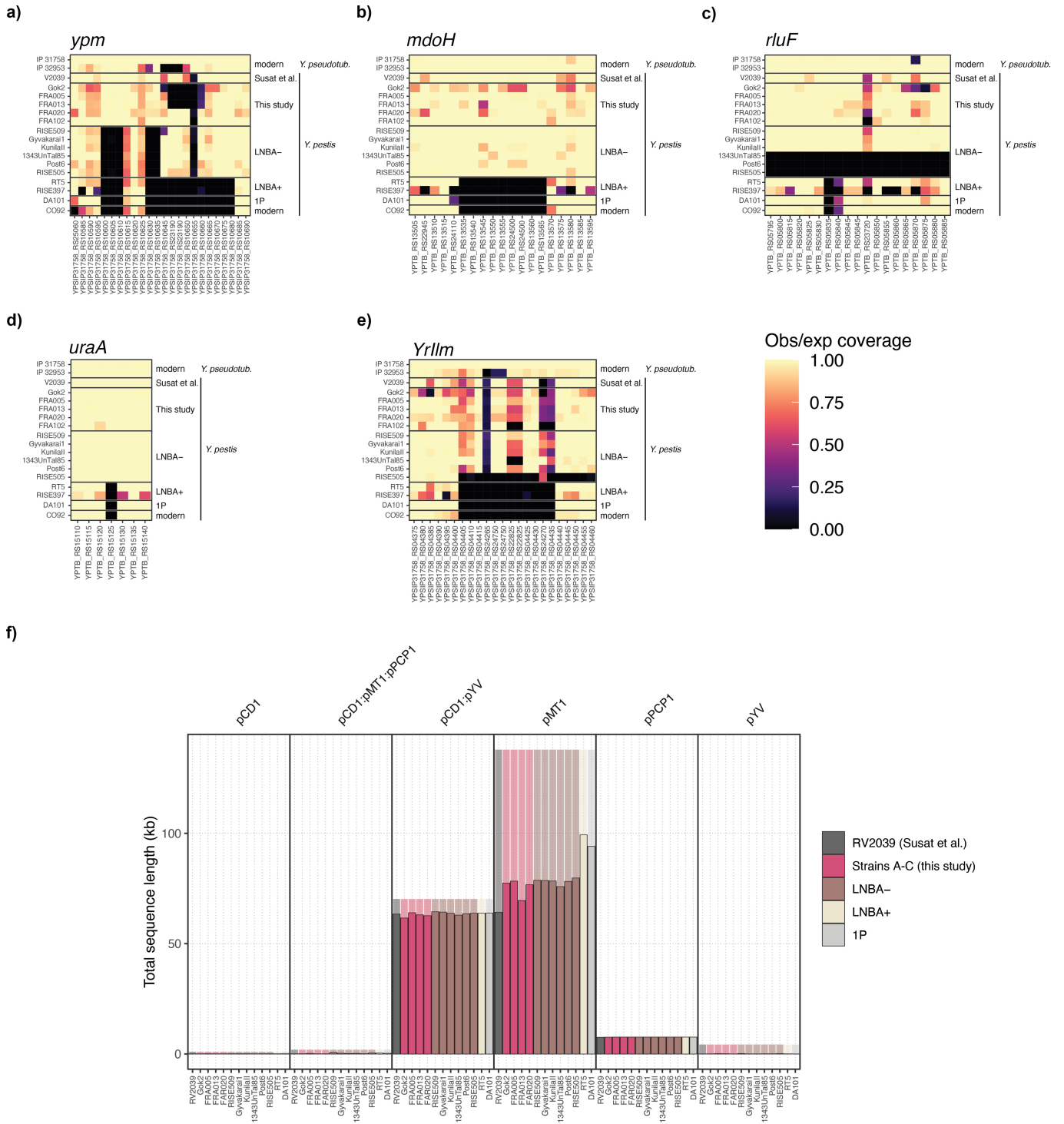


## FRA021

coverage: 0.015x  
site: Frålsegården, Sweden



**Extended Data Fig. 9 | Phylogenetic placements of lower-coverage partial genomes (0.01-1x).** Branch colouring and stroke indicates likelihood of placement on a given branch.



**Extended Data Fig. 10 | Pangenomic analysis of ancient plague genomes.**  
**a-e)** Per-gene coverage statistics from *Y. pseudotuberculosis* complex pangenome variation graphs. Heatmaps showing presence/absence of genes in five genomic regions where ancient *Y. pestis* lineages harbour *Y. pseudotuberculosis* complex pangenome gene content absent in modern *Y. pestis*. Fill colour indicates the ratio of observed over expected breadth of coverage given the depth of coverage across the genome for each gene and ancient sample. **f)** Variations in plasmid gene content among select *Y. pestis* strains. Barplot showing total

length of reference graph nodes covered in a given sample. The plot is stratified into categories of reference plasmid presence/absence at a given node, and only nodes covered by any of the three *Y. pestis* plasmids are shown. I.e. 'plasmid pCD1:plasmid pYV' refers to gene content present in both pCD1 (from *Y. pestis*) and pYV (from *P. pseudotuberculosis*). Shaded background indicates the total amount of sequence within each class and coloured bars represent the total length of nodes within that category present in a particular sample. Groups of ancient strains are indicated by bar colour (1 P: first pandemic).

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

```

primus (v1.9.0)
raxml-ng (1.2.0)
samtools (1.17)
seqtk (1.3-r106)
smartpca (eigensoft v. 8.0.0)
yhaplo (1.1.2)
KIN (0.1.0)
KINgaroo (0.1.0)

```

```

R packages:
viridis_0.6.2
viridisLite_0.4.1
treeio_1.22.0
tidyr_1.3.0
stringr_1.5.0
scales_1.2.1
readr_2.1.4
RColorBrewer_1.1-3
purrr_1.0.1
kinship2_1.9.6
quadprog_1.5-8
Matrix_1.5-4
igraph_1.4.2
googlesheets4_1.1.0
ggtree_3.6.2
ggh4x_0.2.4
ggplot2_3.4.2
forcats_1.0.0
dplyr_1.1.1
doParallel_1.0.17
iterators_1.0.14
foreach_1.5.2

```

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Fastq files will be made available on the European Nucleotide Archive upon publication of the manuscript.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Reporting on race, ethnicity, or other socially relevant groupings

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No tests were carried out to predetermine sample size. Sample size was determined by the availability of archaeological material, and on the DNA preservation in these samples.
Data exclusions	7 DNA libraries were excluded from this study based on high contamination estimates as determined by ContamMix (v1.0.10; please see Supplementary Table 3 for details). Furthermore, another 3 libraries were flagged as suspected sample swaps and excluded because they differed fundamentally with other libraries from the same sample. Lastly all samples with a final depth of coverage under 0.01X were excluded from downstream analyses.
Replication	Out of the 109 genetic individuals analysed in this study, 95 individuals are represented by more than one sequencing library. Having multiple sequencing libraries for each sample serves to validate sequencing results and to pinpoint potential sample swaps. Apart from this type of replication, replication of experimental findings is generally not applicable for this kind of ancient DNA study because of the unique nature of ancient human remains.
Randomization	Not applicable
Blinding	Not applicable

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input type="checkbox"/>	<input checked="" type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

## Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

# Palaeontology and Archaeology

Specimen provenance	Hunnebostrand (SHM 7532:107a+b) was sampled in 2006 at Statens Historiska Museum (SHM) in Stockholm by verbal agreement with Leena Drenzel, SHM. Frälsögården (VGM 1M16-107047) is stored in-house at the Department of Historical Studies, Gothenburg university. Sampling for DNA and other analyses is ongoing by verbal agreement with Maria Vretemark, Västergötlands museum, Skara. Avleberg (NM A 37692-701) is stored at the National Museum, Copenhagen. It was sampled and published by a previous project (Allentoft et al 2022, 2023). The samples from Rössberga were analysed as part of the Atlas project in 2013 under permit number 33-696-2013, issued by SHM (Statens Historiska Museum). The tooth from Firse sten was sampled in 2019-10-30 as part of the project Megalitgravar på Falbygden with permission from Västergötlands museum.
Specimen deposition	Leftover DNA digests, extract and sequencing libraries are stored at the DNA laboratory facilities at Globe Institute, Copenhagen. Upon completion of this project, leftover bone material will be returned to respective museum or university collections from which they were sampled (see above).
Dating methods	Datings were performed at the Keck carbon cycle AMS facility, University of California, Irving. The samples were decalcified in 1N HCl, gelatinized at 60°C and pH 2, and ultrafiltered to select a high molecular wt fraction (>30kDa). $\delta^{13}C$ and $\delta^{15}N$ values were measured to a precision of <0.1‰ and <0.2‰, respectively, on aliquots of ultrafiltered collagen, using a Fisons NA1500NC elemental analyzer/Finnigan Delta Plus isotope ratio mass spectrometer. Datings were calibrated in Oxcal 4.4.4 using the Intcal20 calibration curve.
<input checked="" type="checkbox"/>	Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.
Ethics oversight	No ethical approval was required for this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.