A 5700 year-old human genome and oral microbiome from chewed birch pitch

Jensen, Theis Z. T.; Niemann, Jonas; Iversen, Katrine Højholt; Fotakis, Anna K.; Gopalakrishnan, Shyam; Vågene, Åshild J.; Pedersen, Mikkel Winther; Sinding, Mikkel-Holger S.; Ellegaard, Martin R.; Allentoft, Morten E.; Lanigan, Liam T.; Taurozzi, Alberto J.; Nielsen, Sofie Holtmark; Dee, Michael W.; Mortensen, Martin N.; Christensen, Mads C.; Sørensen, Søren A.; Collins, Matthew J.; Gilbert, M. Thomas P.; Sikora, Martin; Rasmussen, Simon; Schroeder, Hannes

Published in:
Nature Communications

DOI:
10.1038/s41467-019-13549-9

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Jensen, T. Z. T., Niemann, J., Iversen, K. H., Fotakis, A. K., Gopalakrishnan, S., Vågene, Å. J., Pedersen, M. W., Sinding, M-H. S., Ellegaard, M. R., Allentoft, M. E., Lanigan, L. T., Taurozzi, A. J., Nielsen, S. H., Dee, M. W., Mortensen, M. N., Christensen, M. C., Sørensen, S. A., Collins, M. J., Gilbert, M. T. P., ... Schroeder, H. (2019). A 5700 year-old human genome and oral microbiome from chewed birch pitch. Nature Communications, 10, [5520]. https://doi.org/10.1038/s41467-019-13549-9

Download date: 16. maj. 2024
A 5700 year-old human genome and oral microbiome from chewed birch pitch

Theis Z.T. Jensen, Jonas Niemann, Katrine Højholt Iversen, Anna K. Fotakis, Shyam Gopalakrishnan, Åshild J. Vägene, Mikkel Winther Pedersen, Mikkel-Holger S. Sinding, Martin R. Ellegaard, Morten E. Allentoft, Liam T. Lanigan, Alberto J. Taurozzi, Sofie Holtsmark Nielsen, Michael W. Dee, Martin N. Mortensen, Mads C. Christensen, Søren A. Sørensen, Matthew J. Collins, M. Thomas P. Gilbert, Martin Sikora, Simon Rasmussen & Hannes Schroeder

The rise of ancient genomics has revolutionised our understanding of human prehistory but this work depends on the availability of suitable samples. Here we present a complete ancient human genome and oral microbiome sequenced from a 5700 year-old piece of chewed birch pitch from Denmark. We sequence the human genome to an average depth of 2.3× and find that the individual who chewed the pitch was female and that she was genetically more closely related to western hunter-gatherers from mainland Europe than hunter-gatherers from central Scandinavia. We also find that she likely had dark skin, dark brown hair and blue eyes. In addition, we identify DNA fragments from several bacterial and viral taxa, including Epstein-Barr virus, as well as animal and plant DNA, which may have derived from a recent meal. The results highlight the potential of chewed birch pitch as a source of ancient DNA.
Birch pitch is a black-brown substance obtained by heating birch bark and has been used as an adhesive and hafting agent as far back as the Middle Pleistocene. Small lumps of this organic material are commonly found on archaeological sites in Scandinavia and beyond, and while their use is still debated, they often show tooth imprints, indicating that they were chewed. Freshly produced birch pitch hardens on cooling and it has been suggested that chewing was a means to make it pliable again before using it, e.g. for hafting composite stone tools. Medicinal uses have also been suggested, since one of the main constituents of birch pitch, betulin, has antiseptic properties. This is supported by a large body of ethnographic evidence, which suggests that birch pitch was used as a natural antiseptic for preventing and treating dental ailments and other medical conditions. The oldest examples of chewed pitch found in Europe date back to the Mesolithic period and chemical analysis by Gas Chromatography-Mass Spectrometry (GC-MS) has shown that many of them were made from birch (Betula pendula).

Recent work by Kashuba et al. has shown that pieces of chewed birch pitch contain ancient human DNA, which can be used to link the material culture and genetics of ancient populations. In the current study, we analyse a further piece of chewed birch pitch, which was discovered at a Late Mesolithic/Early Neolithic site in southern Denmark (Fig. 1a; Supplementary Note 1) and demonstrate that it does not only contain ancient human DNA, but also microbial DNA that reflects the oral microbiome of the person who chewed the pitch, as well as plant and animal DNA which may have derived from a recent meal. The DNA is so exceptionally well preserved that we were able to recover a complete ancient human genome from the sample (sequenced to an average depth of coverage of 2.3×), which is particularly significant since, so far, no human remains have been recovered from the site. The results highlight the potential of chewed birch pitch as a source of ancient human and non-human DNA, which can be used to shed light on the population history, health status, and even subsistence strategies of ancient populations.

**Results**

**Radiocarbon dating and chemical analysis.** Radiocarbon dating of the specimen yielded a direct date of 5,858–5,661 cal. BP (GrM-13305; 5,007 ± 11) (Fig. 1b; Supplementary Note 2), which places it at the onset of the Neolithic period in Denmark. Chemical analysis by Fourier-Transform Infrared (FTIR) spectroscopy produced a spectrum very similar to modern birch pitch (Supplementary Fig. 4) and GC-MS revealed the presence of the triterpenes betulin and lupeol, which are characteristic of birch pitch (Fig. 1c; Supplementary Note 3). The GC-MS spectrum also shows a range of dicarboxylic acids and saturated fatty acids, which are all considered intrinsic to birch pitch and thus support its identification.

**DNA sequencing.** We generated approximately 390 million DNA reads for the sample, nearly a third of which could be uniquely mapped to the human reference genome (hg19) (Supplementary Table 2). The human reads displayed all the features characteristic of ancient DNA, including (i) short average fragment lengths, (ii) an increased occurrence of purines before strand breaks, and (iii) an increased frequency of apparent cytosine (C) to thymine (T) substitutions at 5′-ends of DNA fragments (Supplementary Fig. 6) and the amount of modern human contamination was estimated to be around 1–3% (Supplementary Table 3). In addition to the human reads, we generated around 7.3 Gb of

![Fig. 1](https://example.com/fig1.png)
sequence data (68.8%) from the ancient pitch that did not align to the human reference genome.

**DNA preservation and genome reconstruction.** With over 30%, the human endogenous DNA content in the sample was extremely high and comparable to that found in well-preserved teeth and petrous bones. We used the human reads to reconstruct a complete ancient human genome, sequenced to an effective depth-of-coverage of 2.3x, as well as a high-coverage mitochondrial genome (91x), which was assigned to haplogroup K1e (see Methods). To further investigate the preservation of the human DNA in the sample we calculated a molecular decay rate (k, per site per year) and find that it is comparable to that of other ancient human genomes from temperate regions (Supplementary Table 3).

**Sex determination and phenotypic traits.** Based on the ratio between high-quality reads (MAPQ ≥ 30) mapping to the X and Y chromosomes, respectively, we determined the sex of the individual whose genome we recovered to be female. To predict her hair, eye and skin colour we imputed genotypes for 41 SNPs (Supplementary Data 1) included in the HiRipS-X system and find that she likely had dark skin, dark brown hair, and blue eyes (Supplementary Data 2). We also examined the allelic state of two SNPs linked with the primary haplotype associated with lactase persistence in humans and found that she carried the ancestral allele for both (Supplementary Data 1), indicating that she was lactase non-persistent.

**Genetic affinities.** We called 593,102 single nucleotide polymorphisms (SNPs) in our ancient genome that had previously been genotyped in a dataset of >1000 present-day individuals from a diverse set of Eurasian populations, as well as >100 previously published ancient genomes (Supplementary Data 3). Figure 2a shows a principal component analysis (PCA) where she clusters with western hunter-gatherers (WHGs). Allele-sharing from a diverse set of Eurasian populations, as well as >100 SNPs linked with the primary haplotype associated with lactase persistence in humans and found that she carried the ancestral allele for both (Supplementary Data 1), indicating that she was lactase non-persistent.

**Metataxonomic profiling of non-human reads.** To broadly characterise the taxonomic composition of the non-human reads in the sample, we used MetaPhlan, a tool specifically designed for the taxonomic profiling of short-read metagenomic shotgun data (see Methods; Supplementary Data 4). Figure 3a shows a principal coordinate analysis where we compare the microbial composition of our sample to that of 689 microbiome profiles from the Human Microbiome Project (HMP). We find that our sample clusters with modern oral microbiome samples in the HMP dataset. This is also reflected in Fig. 3b which shows the order-level microbial composition of our sample compared to two soil samples from the same site and metagenome profiles of healthy human subjects at five major body sites from the HMP, visualised using MEGAN6.

**Oral microbiome characterisation.** To further characterise the microbial taxon present in the ancient pitch and to obtain species-specific assignments we used MALT, a fast alignment and taxonomic binning tool for metagenomic data that aligns DNA sequencing reads to a user-specified database of reference sequences (see Methods; Supplementary Data 5). As expected, a large number of reads could be assigned to oral taxa, such as Neisseria subflava and Rothia mucilaginosa, as well as several bacteria included in the red complex (i.e. Porphyromonas gingivalis, Tannerella forsythia, and Treponema denticola) (see Table 1). In addition, we recovered 593 reads that were assigned to Epstein–Barr virus (Human gammaherpesvirus 4). We validated each taxon by examining the edit distances, coverage distributions, and post-mortem DNA damage patterns (see Supplementary Note 5).

**Pneumococcal DNA.** We also identified several species belonging to the Mitis group of streptococci (Table 1), including...
Streptococcus viridans and Streptococcus pneumoniae. We reconstructed a consensus genome from the S. pneumoniae reads (Fig. 4) and estimated the number of heterozygous sites (2,597) (see Methods) which indicates the presence of multiple strains. To assess the virulence of the S. pneumoniae strains recovered from the ancient pitch, we aligned the contigs against the full Virulence Factor Database\(^{17}\) in order to identify known S. pneumoniae virulence genes (see Methods). We identified 26 S. pneumoniae virulence factors within the ancient sample, including capsular polysaccharides (CPS), streptococcal enolase (Eno), and pneumococcal surface antigen A (PsaA) (see Supplementary Data 6).

Plant and animal DNA. Lastly, we used a taxonomic binning pipeline specifically designed for ancient environmental DNA\(^{15}\) to taxonomically classify the non-human reads in the sample that mapped to other Metazoa (animals) and Viridiplantae (plants). We only parsed taxa with classified reads accounting for >1% of all reads in each of the two kingdoms and a declining edit distance distribution after edit distance 0 (Supplementary Data 7). We then validated each identified taxon as described above (see Supplementary Note 5). Using these criteria, we identified DNA from two plant species in the ancient sample, including birch (Betula pendula) and hazelnut (Corylus avellana). In addition, we detected over 50,000 reads that were assigned to mallard (Anas platyrhynchos).

Discussion
We successfully extracted and sequenced ancient DNA from a 5700-year-old piece of chewed birch pitch from southern Denmark. In addition to a complete ancient human genome (2.3x) and mitogenome (91x), we recovered plant and animal DNA, as well as microbial DNA from several oral taxa. Analysis of the human reads revealed that the individual whose genome we recovered was female and that she likely had dark skin, dark brown hair and blue eyes. This combination of physical traits has been previously noted in other European hunter-gatherers\(^{19–22}\), suggesting that this phenotype was widespread in Mesolithic Europe and that the adaptive spread of light skin pigmentation in European populations only occurred later in prehistory\(^{23}\). We also find that she had the alleles associated with lactase non-persistence, which fits with the notion that lactase persistence in adults only evolved fairly recently in Europe, after the introduction of dairy farming with the Neolithic revolution\(^{24,25}\).

From a population genetics point of view, the human genome also offers fresh insights into the early peopling of southern Scandinavia. Recent studies of ancient hunter-gatherer genomes from Sweden and Norway\(^{23}\) have shown that, following the retreat of the ice sheets around 12–11 ka years ago, Scandinavia was colonised by two separate routes, one from the south (presumably via Denmark) and one from the northeast, along the coast of present-day Norway. This is supported by the fact that hunter-gatherers from central Scandinavia carry different levels of WHG and EHG ancestry, which reached central Scandinavia from the south and northeast, respectively\(^{23}\). Although we only analysed a single genome, the fact that the Syltholm individual does not carry any EHG ancestry confirms this scenario and suggests that EHG did not reach southern Denmark at this point in prehistory.

The Syltholm genome (5700 years cal. BP) dates to the period immediately following the Mesolithic–Neolithic transition in Denmark. Culturally, this period is marked by the transition from the Late Mesolithic Ertebolle culture (c. 7300–5900 cal. BP) with its flaked stone artefacts and typical T-shaped antler axes, to the early Neolithic Funnel Beaker culture (c. 5900–5300 cal. BP) with its characteristic pottery, polished flint artefacts, and domesticated plants and animals\(^{26}\). In Denmark, the transition from hunting and gathering to farming has often been described as a relatively rapid process, with dramatic shifts in settlement patterns and subsistence strategies\(^{27}\). However, it is still unclear to what extent this transition was driven by the arrival of farming communities as opposed to the local adaptation of farming practices by resident hunter-gatherer populations.
In addition to the human data, we recovered ancient microbial DNA from the pitch which could be shown to have a human oral microbiome signature. Previous studies have demonstrated that calcified dental plaque (dental calculus) provides a robust biomolecular reservoir that allows direct and detailed investigations of ancient oral microbiomes. However, unlike dental calculus, which represents a long-term reservoir of the oral microbiome built up over many years, the microbiota found in ancient mastics are more likely to give a snapshot of the species active at the time. As such, they provide a useful source of information regarding the evolution of the human oral microbiome that can complement studies of ancient dental calculus.

The majority of the bacterial taxa we identified (Table 1) are classified as non-pathogenic, commensal species that are considered to be part of the normal microflora of the human mouth.
and the upper respiratory tract, but may become pathogenic under certain conditions. In addition, we identified three species (*Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola*), included in the so-called red complex, a group of bacteria that are categorised together based on their association with severe forms of periodontal disease. Furthermore, we identified several thousand reads that could be assigned to different bacterial species in the Mitis group of streptococci, including *Streptococcus pneumoniae*, a major human pathogen that is responsible for the majority of community-acquired pneumonia which still causes around 1–2 million infant deaths worldwide, every year.

*Streptococcus pneumoniae* has a remarkable capacity to remodel its genome through the uptake of exogenous DNA from other pneumococci and closely related oral streptococci. Understanding this process and the distribution of pneumococcal virulence factors between different strains can help our understanding of *S. pneumoniae* pathogenesis. We identified 26 *S. pneumoniae* virulence factors within our ancient sample, including several that are involved in host colonisation (e.g. adherence to host cells and tissues, endocytosis) and the evasion and subversion of the host’s immune response (Supplementary Data 6). While more research is needed to fully understand the evolution of this important human pathogen and its ability to cause disease, our capacity to recover virulence factors from ancient samples opens up promising avenues for future research.

In addition to the bacterial taxa, we identified 593 reads that could be assigned to the Epstein–Barr virus (EBV). Previous

![Fig. 4 Streptococcus pneumoniae consensus genome reconstructed from metagenomic sequences recovered from the ancient pitch. From outer to inner ring: *S. pneumoniae* virulence genes (black, shared genes are shown in bold); *S. pneumoniae* coding regions on the positive (blue) and negative (red) strand; mappability (grey); sequence depth for the Sytholm pitch (orange), HOMP sample SRS014468 (light brown), SRS019120 (light blue), SRS013942 (turquoise), SRS015055 (blue), and SRS014692 (dark blue). Sequence depths were calculated by aligning to the *S. pneumoniae* TIGR4 reference genome and visualised in 100 bp windows using Circos.](https://doi.org/10.1038/s41467-019-13549-9)
studies\textsuperscript{43,44} have demonstrated the great potential of ancient DNA for studying the long-term evolution of blood borne viruses. Formally known as Human gammaherpesvirus 4, EBV is one of the most common human viruses infecting over 90% of the world’s adult population\textsuperscript{45}. Most EBV infections occur during childhood and in the vast majority of cases they are asymptomatic or they carry symptoms that are indistinguishable from other mild, childhood diseases. However, in some cases EBV can cause infectious mononucleosis (glandular fever)\textsuperscript{46} and it has also been associated with various lymphoproliferative diseases, such as Hodgkin’s lymphoma and hemophagocytic lymphohistiocytosis, as well as higher risks of developing certain autoimmune diseases, such as dermatomyositis and multiple sclerosis\textsuperscript{47,48}.

Lastly, we identified several thousand reads that could be confidently assigned to different plant and animal species, including birch (\textit{Betula pendula}), hazel (\textit{C. avellana}), and mallard (\textit{A. platyrhynchos}). While the presence of birch DNA is easily explained as it is the source of the pitch, we propose that the hazelnut and mallard DNA may derive from a recent meal. This is supported by the faunal evidence from the site, which is dominated by wild taxa, including \textit{Anas spp.} and hazelnuts\textsuperscript{49}. In addition, there is evidence from many other Mesolithic and Early Neolithic sites in Scandinavia for hazelnuts being gathered in large quantities for consumption\textsuperscript{50}. Together with the faunal evidence, the ancient DNA results support the notion that the people at Syltholm continued to exploit wild resources well into the Neolithic and highlight the potential of ancient DNA analyses of chewed pieces of birch pitch for palaeodietary studies.

In summary, we have shown that pieces of chewed birch pitch are an excellent source of ancient human and non-human DNA. In the process of chewing, the DNA becomes trapped in the pitch where it is preserved due to the aseptic and hydrophobic properties of the pitch which both inhibits microbial and chemical decay. The genomic information preserved in chewed pieces of birch pitch offers a snapshot of people’s lives, providing information on genetic ancestry, phenotype, health status, and even subsistence. In addition, the microbial DNA provides information on the composition of our ancestral oral microbiome and the evolution of specific oral microbes and important human pathogens.

Methods

\textbf{Sample preparation and DNA extraction.} We sampled c. 250 mg from the specimen for DNA analysis. Briefly, the sample was washed in 5% bleach solution to remove any surface contamination, rinsed in molecular biology grade water and left to dry. We tested three different extraction methods using between 20–50 mg of starting material: For method (1), 1 ml of lysis buffer containing 0.45 M EDTA (pH 8.0) and 0.25 mg/ml Proteinase K was added to the sample and left to incubate on a rotor at 56 °C. After 12 h the supernatant was removed and concentrated down to ~150 µl using Amicon Ultra centrifugal rotor at 56 °C. After 12 h the supernatant was removed and concentrated down to ~150 µl using Amicon Ultra centrifugal filter units. For method (2), the sample was digested and purified as above, but with the addition of a phenol-chloroform clean-up step. Briefly, 1 ml phenol (pH 8.0) was added to the lysis mix, followed by 1 ml chloroform:isoamylalcohol. The supernatant was concentrated and purified, as described above. For method (3) the sample was dissolved in 1 ml chloroform:isoamylalcohol. The dissolved sample was then resuspended in 1 ml molecular grade water and purified as described above. DNA extracts prepared using a Proteinase K-based lysis buffer followed by a phenol-chloroform based purification step produced the best results in terms of the endogenous human DNA content (see Supplementary Table 1); however, following phenol-chloroform based purification our single extraction NTC was included, as well as a single library NTC. 10 µl of each library were amplified in 50 µl reactions for between 15 and 28 cycles, using a dual indexing approach\textsuperscript{61}. The optimal number of PCR cycles was determined by qPCR (MxPro 3000, Agilent Technologies). The amplified libraries were purified using SPRI-beads and quantified on a TECAN Infinite M1000 (Agilent Technologies) High Sensitivity qPCR plate. The amplified and indexed libraries were then pooled in equimolar amounts and sequenced on 1/8 of a lane of an Illumina HiSeq 2500 run in SR mode. Following initial screening, additional reads were obtained by pooling libraries #2, #3, and #4 in molar fractions of 0.2, 0.4, and 0.4, respectively and sequencing them on one full lane of an Illumina HiSeq 2500 run in SR mode.

\textbf{Library preparation and sequencing.} 16 µl of each DNA extract were built into double-stranded libraries using a recently published protocol that was specifically designed for ancient DNA\textsuperscript{53}. One extraction NTC was included, as well as a single library NTC. 10 µl of each library were amplified in 50 µl reactions for between 15 and 28 cycles, using a dual indexing approach\textsuperscript{61}. The optimal number of PCR cycles was determined by qPCR (MxPro 3000, Agilent Technologies). The amplified libraries were purified using SPRI-beads and quantified on a 2200 TapeStation (Agilent Technologies) High Sensitivity channel. The amplified and indexed libraries were then pooled in equimolar amounts and sequenced on 1/8 of a lane of an Illumina HiSeq 2500 run in SR mode. Following initial screening, additional reads were obtained by pooling libraries #2, #3, and #4 in molar fractions of 0.2, 0.4, and 0.4, respectively and sequencing them on one full lane of an Illumina HiSeq 2500 run in SR mode.

\textbf{Data processing.} Base calling was performed using Illumina’s bcl2fastq2 conversion software v2.20.0. Only sequences with correct indexes were retained. FastQ files were processed using PALEOMIX v1.2.15\textsuperscript{55}. Adapters and low quality reads (Q < 20) were removed using AdapterRemoval v2.2.6\textsuperscript{56}, only retaining reads >25 bp. Trimmed and filtered reads were then mapped to hg19 (build 37.1) using BWA\textsuperscript{57} with seed disabled to allow for better sensitivity\textsuperscript{58}, as well as filtering out unmapped reads. Only reads with a mapping quality ≥30 were kept and PCR duplicates were removed. MapDamage 2.0\textsuperscript{59} was used to evaluate the authenticity of the retained reads as part of the PALEOMIX pipeline\textsuperscript{55}, using a subsample of 100k reads per sample (Supplementary Fig. 6). For the population genomic analyses, we merged the current sample with individuals from the Human Origins dataset\textsuperscript{31} and >100 previously published ancient genomes (Supplementary Data 1). At each SNP in the Human Origin dataset, we sampled the allele with more reads in the ancient sample, resolving ties randomly, resulting in a pseudohaploid ancient sample.

\textbf{MtDNA analysis and contamination estimates.} We used Schmutzi\textsuperscript{60} to determine the endogenous consensus mtDNA sequence and to estimate present-day human contamination. Reads were mapped to the Cambridge reference sequence (cCRS) and filtered for MAPQ ≥30. Haploid variants were called using the endo-Caller program implemented in Schmutzi\textsuperscript{60} and only variants with a posterior probability exceeding 50 on the PHRED scale (probability of error: 1/100,000) were retained. We then used Haplogrep v2.28\textsuperscript{51} to determine the mtDNA haplogroup, specifying PhyloTree (build 17) as the reference phylogeny\textsuperscript{52}. Contamination estimates were obtained using Schmutzi\textsuperscript{60}’s mtCont program and a database of putative modern contaminant mitochondrial DNA sequences.

\textbf{Genotype imputation.} We used ANGSD\textsuperscript{63} to compute genotype likelihoods in 5 Mb windows around 43 SNPs associated with skin, eye, and hair colour\textsuperscript{10} and lactase persistence into adulthood (Supplementary Data 2). Missing genotypes were imputed using impute2\textsuperscript{64} and the pre-phased 1000 Genome reference panel\textsuperscript{65}, provided as part of the impute2 reference datasets. We used multiple posterior probability thresholds, ranging from 0.95 to 0.50, to filter the imputed genotypes. The imputed genotypes were uploaded to the HaplEx-\textit{R} website\textsuperscript{66} to obtain the predicted outcomes for the pigmentation phenotypes (Supplementary Data 3).

\textbf{Principal component analysis.} Principal component analysis was performed using smartPCA\textsuperscript{66} by projecting the ancient individuals onto a reference panel including >1000 present-day Eurasian individuals from the HO dataset\textsuperscript{11} using the option \textit{lsq} project. Prior to performing the PCA the data set was filtered for a minimum allele frequency of at least 5% and a missingness per marker of at most 50%. To mitigate the effect of linkage disequilibrium, the data were pruned in a 50-SNP sliding window, advanced by 10 SNPs, and removing sites with an R\textsuperscript{2} larger than 0, resulting in a final data set of 593,102 SNPs.

\textbf{D- and f-statistics.} D- and f-statistics were computed using AdmixTool\textsuperscript{67}. To estimate the amount of shared drift between the Syltholm genome and WHG versus EHG and Neolithic farmers, respectively, we computed two sets of f\textsubscript{st}-statistics on the DNA from \textit{B. pendula} (Yoruba, \textit{YRI}, the Barcan, \textit{WHG}) and \textit{D. prunus} (test sample). Standard errors were calculated using a weighted block jackknife. To confirm the absence of EHG and Neolithic farmer gene flow in the Syltholm genome and to contrast this result with those obtained for other Mesolithic and Neolithic individuals from Scandinavia, we computed two sets of D-statistics of the form (D\textsubscript{Yoruba}, EHG/Barcan; X, WHG) testing whether "X" forms a clade to the exclusion of EHG and Neolithic farmers (represented by Barcan), respectively.

\textbf{Negative controls.} We included no template controls (NTC) during the DNA extraction and library preparation steps. The NTCs prepared with the additional phenol-chloroform step were also found to be contaminated with \textit{Delphi} spp., a known laboratory contaminant\textsuperscript{62}. The contaminated libraries were excluded from metagenomic profiling.
**MALTED.** To further characterise the metagenomic reads we performed microbial species identification using MALT v. 0.4.1 (Megan Alignment Tool)\(^1\), a rapid sequence-alignment tool specifically designed for the analysis of metagenomic data. All complete bacterial (\(n = 12,426\)) and viral (\(n = 8094\)) genomes were downloaded from NCBI RefSeq on 13 November 2018, and all complete archaeal (\(n = 280\)) genomes were downloaded from NCBI RefSeq on 17 November 2018 to create a custom database. In an effort to exclude genomes that may consist of composite sequences from multiple organisms, the following entries were excluded:

- GCF\_00094235.1 uncultured phage WW-nAnB
- GCF\_000146025.1 uncultured Termitine group 1 bacterium phylotype Rs-D17

The final MALT reference database contained 33,223 genomes and was created using default parameters in *malt-build* (v. 0.4.1). The sequencing data for the ancient pitch sample, two soil control samples and associated extraction and library blanks were de-enriched for human reads by mapping to the human genome (hg19) using BWAinnamon and excluding all mapping reads. Duplicates were removed with seqkit v.0.7.17\(^1\) using the ‘rmdup’ function with the ‘–by-seq’ flag. The remaining reads were processed with *malt-run* (v. 0.4.1) where BlastN mode and SemiGlobal alignment were used. The minimum percent identity (–minIdent) set to 95, the minSupport parameter set to 10 and the top percent value (–topPercent) set was as 1. Remaining parameters were set to default. MEGAN\(^5\) was used to visualise the output ‘.taxa’ files and to extract the reads assigned to taxonomic nodes of interest for our sample. A taxonomic table of the RAW MALT output for all samples and blanks, as well as species level read assignments to bacteria, archaea and DNA viruses for the ancient pitch sample are shown in Supplementary Data 5, where reads listed are the sum of all reads assigned to the species node, including reads assigned to specific strains within the species. Reads assigned to RNA viruses were not considered for further analyses, since our dataset consisted of DNA sequences only. Due to the limitation of the set of assigned reads to archaeal species (Supplementary Data 5), we did not consider Archaea in downstream analyses of species identification. To validate the microbial taxa, we aligned the assigned reads to their respective reference genomes and examined the edit distances, coverage distributions, and post-mortem DNA damage patterns (see Supplementary Note 5).

**Pneumococcus analysis.** We reconstructed a *S. pneumoniae* consensus genome (Fig. 4) by mapping all reads assigned to *S. pneumoniae* by MALT\(^1\). The *S. pneumoniae* TIGR4 reference genome (NC\_0030283.3) to investigate the presence of multiple strains we estimated the number of heterozygous sites using samtools\(^6\) mpileup function, filtering out transitions, indels, and sites with a depth of coverage below 10. Coverage statistics of the individual isolates were obtained using Bedtools\(^7\) and plotted using Circos\(^8\) in 100 bp windows. Mappability was estimated using GEOM\(^9\) using a k-mer size of 50 and a read length of 42, which is comparable to the average length of the trimmed and mapped reads in the ancient pitch. Virulence genes were identified by assembling the ancient *S. pneumoniae* MALT reads into contigs using megahaplog\(^10\). The contigs were aligned against known *S. pneumoniae* TIGR4 virulence genes in the Virulence Factor Database\(^11\) (downloaded 22/11–2018) using BLASTn\(^12\). Only unique hits with a bitscore >200,\(^13\) \(＞20\)% coverage, and an identity \(＞80\)% were considered as shared genes (Supplementary Data 6).

To identify all streptococcus virulence factors in the ancient pitch, we aligned the contigs against the full Virulence Factor Database\(^12\) (downloaded 22/11–2018) using BLASTn\(^12\) and the same filtering criteria as described above (Supplementary Data 6). To validate the approach we repeated the analysis with five modern oral microbiome samples (SRS014468; SRS191920; SRS013942; SRS015055; SRS014692) from the Human Microbiome Project (HMP)\(^14\) using only the forward read (R1) (Supplementary Data 6). We find that the number of virulence genes we recovered directly correlates with sequencing depth (Supplementary Fig. 16).

**Holi.** For a robust taxonomic assignment of reads aligning to Metazoa (animals) and Viridiplantae (plants), all non-human reads were parsed through the ‘Holi’ pipeline\(^5\), which was specifically developed for the taxonomic profiling of ancient metagenomic shotgun reads. Each read was aligned against the NCBI’s full Nucleotide and RefSeq databases (downloaded November 25th 2018), including a newly sequenced full genome of European hazelnut (*Corylus avellana*, downloaded April 10th 2019)\(^15\). The alignments were then parsed through a naïve lowest common ancestor algorithm (*ngsLCA*) based on the NCBI taxonomic tree. Only taxonomically classified reads for taxa comprising \(＞1\)% of all the reads within the two kingdoms and a declining edit distance distribution after edit distance 0 were parsed for taxonomic profiling and further validation. To validate the assignments, we aligned the assigned reads to their respective reference genomes and examined the edit distances, coverage distributions, and post-mortem DNA damage patterns (see Supplementary Note 5; Supplementary Data 7).

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

**Data availability**

The sequencing reads are available for download from the European Nucleotide Archive under accession number PRJEB30280. All other data are included in the paper or available upon request.

Received: 17 June 2019; Accepted: 15 November 2019; Published online: 17 December 2019

References

1. Mazza, P. P. A. et al. A new Palaeolithic discovery: tar-hafted stone tools in a European Mid-Pleistocene bone-bearing bed. *J. Archaeol. Sci.* 33, 1130–1138 (2006).
2. Kozovyrk, P. R. B., Soressi, M., Pomdstra, D. & Langejans, G. H. J. Experimental methods for the Palaeolithic dry distillation of birch bark: implications for the origin and development of Neanderthal adhesive technology. *Sci. Rep.* 7, 8033 (2017).
3. Aveling, E. M. & Heron, C. Chewing tar in the early Holocene: an archaeological and ethnohistorical evaluation. *Antiquity* 73, 579–584 (1999).
4. Haque, S. et al. Screening and characterisation of antimicrobial properties of semisynthetic betulin derivatives. *PloS ONE* 9, e102699 (2014).
5. Kashuba, N. et al. Ancient DNA from mastic solidifies connection between material culture and genetics of mesolithic hunter-gatherers in Scandinavia. *Commun. Biol.* 2, 185 (2019).
6. Sørensen, S. A. Sylthom: Denmark’s largest Stone Age excavation. *Mesolithic Misc.* 24, 3–10 (2016).
7. Aveling, E. M. & Heron, C. Identification of birch bark tar at the Mesolithic site of Star Carr. *Archaeol. Sci.* 2, 69–80 (1998).
8. Gamba, C. et al. Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* 5, 5257 (2014).
9. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* 40, 4477–4482 (2013).
10. Chaitanya, L. et al. The HrisePlex-5 system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Sci. Int. Genet.* 35, 123–135 (2018).
11. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513, 409–413 (2014).
12. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207–211 (2015).
13. Truong, D. T. et al. MetaPhAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* 12, 902–903 (2015).
14. The Human Microbiome Project Consortium. et al. Structure, function and diversity of the healthy human microbiome. *Nature 486*, 207–214 (2012).
15. Huson, D. H. et al. MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PloS Comput. Biol.* 12, e1004957 (2016).
16. Vägène, Å. J. et al. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* 2, 520–528 (2018).
17. Chen, L. et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 33, D325–D328 (2005).
18. Pedersen, M. W. et al. Postglacial viability and colonization in North America's ice-free corridor. *Nature* 537, 45–49 (2016).
19. Olalde, I. et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature 507*, 225–228 (2014).
20. Skoglund, P. et al. Genomic diversity and admixture differs for stone-age Scandinavian foragers and farmers. *Science* 344, 747–750 (2015).
21. Mathiessen, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
22. Brace, S. et al. Ancient genomes indicate population replacement in Early Neolithic Britain. *Nature Ecol. Evol.* **3**, 765–771 (2019).
23. Günther, T. et al. Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol.* **16**, e2003703 (2018).
24. Marciniak, S. M. et al. Harnessing ancient genomes to study the history of human adaptation. *Nat. Rev. Genet.* **18**, 659–674 (2017).
25. Ségurel, L. & Bon, C. On the Evolution of Lactase Persistence in Humans. *Nature Rev. Genet.* **13**, 178 (2012).
26. Gron, K. J. & Sørensen, L. Cultural and economic negotiation: a new perspective on the Neolithic Transition of Southern Scandinavia. *Antiquity* **92**, 958–974 (2018).
27. Richards, M. P., Price, T. D. & Koch, E. Mesolithic and Neolithic subsistence in Denmark: new stable isotope data. *Curr. Anthropol.* **44**, 288–295 (2003).
28. Becker, C. J. Mosefundne lerkar fra yngre stenalder: studier over trægbægerkulturen i Danmark (Copenhagen, 1948).
29. Brandt, G. et al. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* **342**, 257–261 (2013).
30. Isern, N., Fort, J. & de Roja, V. L. The ancient cline of haplogroup K implies that the Neolithic transition in Europe was mainly demic. *Sci. Rep.* **7**, 11229 (2017).
31. Hofmanová, Z. et al. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl Acad. Sci. USA* **113**, 6886–6891 (2016).
32. González-Fortes, G. et al. Paleogenomic Evidence for Multi-generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin. *Curr. Biol.* **27**, 1801–1810.e10 (2017).
33. Müttäri, A. et al. The early Chalcolithic prehistory of the Baltic Sea region. *Nat. Commun.* **9**, 442 (2018).
34. Bollongino, R. et al. 2000 years of parallel societies in Stone Age Central Europe. *Science* **342**, 479–481 (2013).
35. Lipson, M. et al. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017).
36. Jones, E. R. et al. The neolithic transition in the baltic was not driven by admixture with early European farmers. *Curr. Biol.* **27**, 576–582 (2017).
37. Adler, C. J. et al. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat. Genet.* **45**, 450–455 (2013).
38. Warinner, C. et al. Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46**, 336–344 (2014).
39. Jersike-Christensen, R. R. et al. Quantitative metaproteomics of medieval oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat. Genet.* **49**, 963–974 (2017).
40. Schubert, M. et al. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012).
41. Jorck, M. et al. The HERA (Humanities in the European Research Area) through the joint research programme Prot. 9, 1058 (2016).
42. Schubert, M. et al. Characterization of ancient and modern genomes by SNP detection and phylogenetic and metagenomic analysis using PALEOMIX. *Sci. Rep.* **9**, 1058 (2018).
43. van Oven, M. PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Sci. Int.: Genet. Suppl. Ser.* **5**, e392–e394 (2015).
44. Schubert, M. et al. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012).
45. Schubert, M. et al. HaplotypeGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
46. Weiser, J. N., Ferreira, D. M. & Paton, J. C. Streptococcus pneumoniae: pathogenesis and [endovascular] meningitis. *Clin. Microbiol. Rev.* **21**, 761–842 (2008).
47. van Oven, M. PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Sci. Int.: Genet. Suppl. Ser.* **5**, e392–e394 (2015).
48. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
49. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
50. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
51. Schubert, M. et al. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012).
52. Williams, H. & Crawford, D. H. Epstein-Barr virus: the impact of scientific investigation. *Expert Rev. Mol. Med.* **17**, 1–16 (2012).
53. Rosengren, P. M. Sea-level change in Mesolithic southern Scandinavia. Long- and short-term effects on society and the environment. *Jysk Arkeologisk Selskabs Skrifter* **106** (Jutland Archaeological Society, 2018).
54. Krause-Kyora, B. et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423 (2018).
55. Williams, H. & Crawford, D. H. Epstein-Barr virus: the impact of scientific advances on clinical practice. *Blood* **107**, 862–869 (2006).
56. Henle, G., Henle, W. & Diehl, V. Relation of Burkitt’s tumor-associated herpes-type virus to infectious mononucleosis. *Proc. Natl Acad. Sci. USA* **59**, 94–101 (1968).
57. Toussriott, É. & Roudier, J. Epstein–Barr virus in autoimmune diseases. *Best. Pract. Res. Clin. Rheumatol.* **22**, 883–896 (2008).
58. Richter, C. A., Zhan, X. & Weiss, L. M. Epstein-Barr virus (EBV)-associated lymphoid proliferations, a 2018 update. *Hum. Pathol.* **50**, 19–41 (2018).
59. Zehnder, J. A. et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423 (2018).
60. Williams, H. & Crawford, D. H. Epstein-Barr virus: the impact of scientific advances on clinical practice. *Blood* **107**, 862–869 (2006).
61. Henle, G., Henle, W. & Diehl, V. Relation of Burkitt’s tumor-associated herpes-type virus to infectious mononucleosis. *Proc. Natl Acad. Sci. USA* **59**, 94–101 (1968).
62. Toussriott, É. & Roudier, J. Epstein–Barr virus in autoimmune diseases. *Best. Pract. Res. Clin. Rheumatol.* **22**, 883–896 (2008).
63. Zehnder, J. A. et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423 (2018).
64. Hufnagel, P. et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423 (2018).
65. Williams, H. & Crawford, D. H. Epstein-Barr virus: the impact of scientific advances on clinical practice. *Blood* **107**, 862–869 (2006).
66. Henle, G., Henle, W. & Diehl, V. Relation of Burkitt’s tumor-associated herpes-type virus to infectious mononucleosis. *Proc. Natl Acad. Sci. USA* **59**, 94–101 (1968).
67. Toussriott, É. & Roudier, J. Epstein–Barr virus in autoimmune diseases. *Best. Pract. Res. Clin. Rheumatol.* **22**, 883–896 (2008).
68. Zehnder, J. A. et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423 (2018).
69. Williams, H. & Crawford, D. H. Epstein-Barr virus: the impact of scientific advances on clinical practice. *Blood* **107**, 862–869 (2006).
70. Henle, G., Henle, W. & Diehl, V. Relation of Burkitt’s tumor-associated herpes-type virus to infectious mononucleosis. *Proc. Natl Acad. Sci. USA* **59**, 94–101 (1968).
71. Toussriott, É. & Roudier, J. Epstein–Barr virus in autoimmune diseases. *Best. Pract. Res. Clin. Rheumatol.* **22**, 883–896 (2008).
72. Zehnder, J. A. et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423 (2018).
73. Williams, H. & Crawford, D. H. Epstein-Barr virus: the impact of scientific advances on clinical practice. *Blood* **107**, 862–869 (2006).
74. Henle, G., Henle, W. & Diehl, V. Relation of Burkitt’s tumor-associated herpes-type virus to infectious mononucleosis. *Proc. Natl Acad. Sci. USA* **59**, 94–101 (1968).
Author contributions
T.Z.T.J. and H.S. designed and led the study. S.A.S. provided the sample for analysis. M.C.C. and M.N.M. performed the FTIR and GC-MS analyses. M.W.D. performed the radiocarbon dating. T.Z.T.J., M.H.S.S. and M.R.E. generated the genetic data. T.Z.T.J., J.N., K.H.I., A.K.F., S.G., Å.J.V., M.W.P., S.H.N., M.E.A. and H.S. analyzed the genetic data. T.Z.T.J., J.N., K.H.I., A.K.F., S.G., Å.J.V., M.W.P., M.E.A., L.T.L., A.J.T., M.J.C., M.T.P.G., M.S., S.R., and H.S. interpreted the results. T.Z.T.J. and H.S. wrote the manuscript with input from J.N., K.H.I., and the remaining authors.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-019-13549-9.

Correspondence and requests for materials should be addressed to H.S.

Peer review information Nature Communications thanks Christina Warinner and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

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

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

© The Author(s) 2019