

ARTICLE

<https://doi.org/10.1038/s41467-019-13549-9>

OPEN

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

Theis Z.T. Jensen <sup>1,2,10</sup>, Jonas Niemann<sup>1,2,10</sup>, Katrine Højholt Iversen <sup>3,4,10</sup>, Anna K. Fotakis <sup>1</sup>, Shyam Gopalakrishnan <sup>1</sup>, Åshild J. Vågene<sup>1</sup>, Mikkel Winther Pedersen <sup>1</sup>, Mikkel-Holger S. Sinding <sup>1</sup>, Martin R. Ellegaard <sup>1</sup>, Morten E. Allentoft<sup>1</sup>, Liam T. Lanigan<sup>1</sup>, Alberto J. Taurozzi<sup>1</sup>, Sofie Holtzmark Nielsen<sup>1</sup>, Michael W. Dee<sup>5</sup>, Martin N. Mortensen <sup>6</sup>, Mads C. Christensen<sup>6</sup>, Søren A. Sørensen<sup>7</sup>, Matthew J. Collins<sup>1,8</sup>, M. Thomas P. Gilbert <sup>1,9</sup>, Martin Sikora <sup>1</sup>, Simon Rasmussen <sup>4</sup> & Hannes Schroeder <sup>1\*</sup>

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.

<sup>1</sup>The Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 1353, Denmark. <sup>2</sup>BioArch, Department of Archaeology, University of York, York YO10 5DD, UK. <sup>3</sup>Department of Bio and Health Informatics, Technical University of Denmark, Kongens Lyngby 2800, Denmark. <sup>4</sup>Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 2200, Denmark. <sup>5</sup>Centre for Isotope Research, University of Groningen, Groningen 9747 AG, The Netherlands. <sup>6</sup>The National Museum of Denmark, I.C. Modewegs Vej, Brede, Kongens Lyngby 2800, Denmark. <sup>7</sup>Museum Lolland-Falster, Frisegade 40, Nykøbing Falster 4800, Denmark. <sup>8</sup>McDonald Institute for Archaeological Research, University of Cambridge, Cambridge CB2 3ER, UK. <sup>9</sup>University Museum, NTNU, 7012 Trondheim, Norway. <sup>10</sup>These authors contributed equally: Theis Z. T. Jensen, Jonas Niemann, Katrine Højholt Iversen \*email: [hschroeder@bio.ku.dk](mailto:hschroeder@bio.ku.dk)

**B**irch 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<sup>1,2</sup>. 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<sup>3</sup>. 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<sup>4</sup>. 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<sup>3</sup>. 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*)<sup>3</sup>.

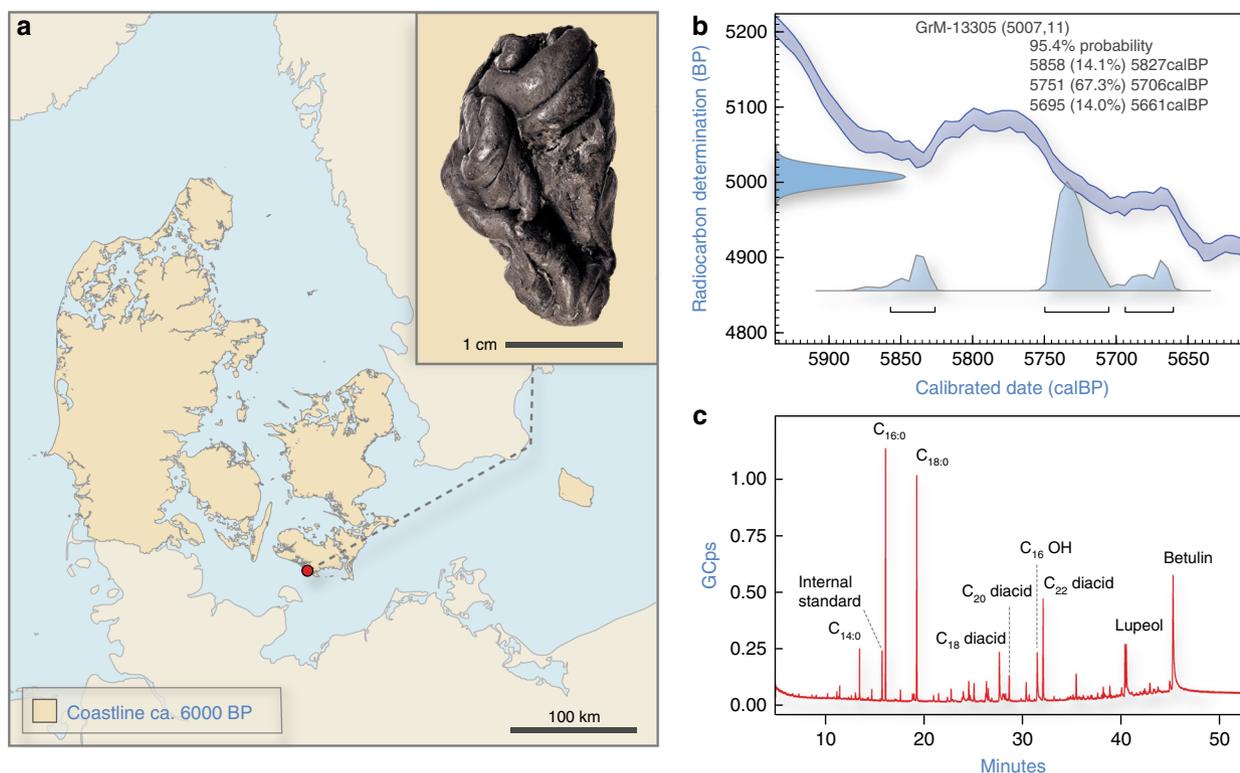
Recent work by Kashuba et al.<sup>5</sup> has shown that pieces of chewed birch pitch contain ancient human DNA, which can be used to link the material culture and genetics of ancient populations. In the current study, we analyse a further piece of chewed birch pitch, which was discovered at a Late Mesolithic/Early 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<sup>6</sup>. 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;  $5,007 \pm 11$ ) (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)<sup>3</sup>. 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<sup>7</sup>.

**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** A chewed piece of birch pitch from southern Denmark. **(a)** Photograph of the Syltholm birch pitch and its find location at the site of Syltholm on the island of Lolland, Denmark (map created using data from Astrup<sup>78</sup>). **(b)** Calibrated date for the Syltholm birch pitch (5,858–5,661 cal. BP;  $5,007 \pm 7$ ). **(c)** GC-MS chromatogram of the Syltholm pitch showing the presence of a series of dicarboxylic acids ( $C_{xx}$  diacid) and saturated fatty acids ( $C_{xx}:0$ ) and methyl 16-Hydroxyhexadecanoate ( $C_{16}OH$ ) together with the triterpenes betulin and lupeol, which are characteristic of birch pitch<sup>3</sup>.

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<sup>8</sup>. We used the human reads to reconstruct a complete ancient human genome, sequenced to an effective depth-of-coverage of 2.3×, as well as a high-coverage mitochondrial genome (91×), 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<sup>9</sup>, 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 HIrisPlex-S system<sup>10</sup> 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.

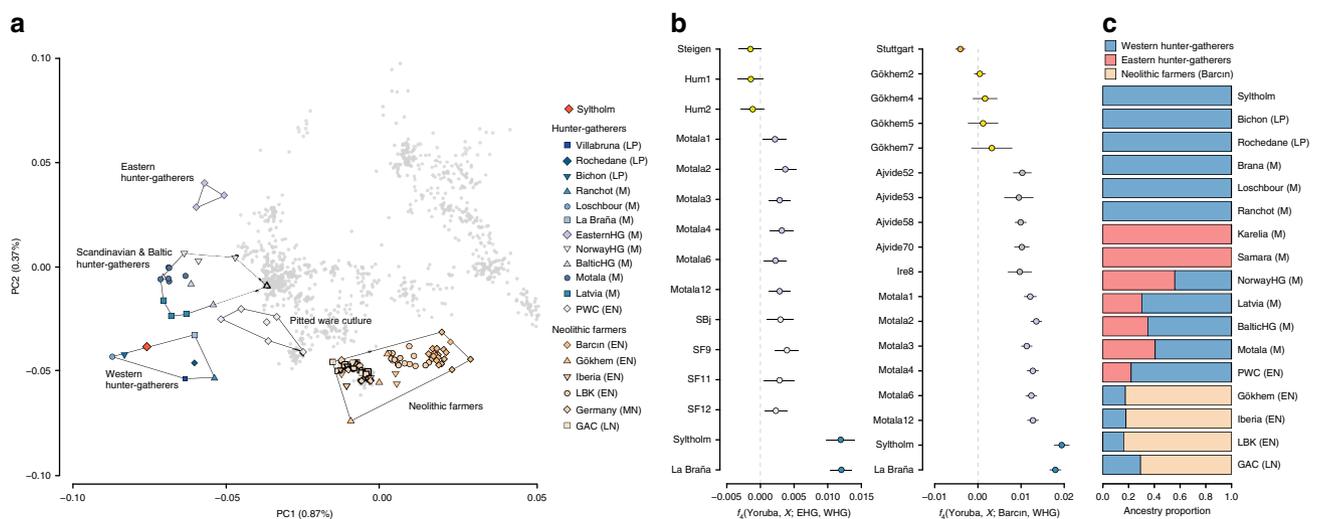
**Genetic affinities.** We called 593,102 single nucleotide polymorphisms (SNPs) in our ancient genome that had previously been genotyped in a dataset of >1000 present-day individuals from a diverse set of Eurasian populations<sup>11</sup>, as well as >100 previously published ancient genomes (Supplementary Data 3). Figure 2a shows a principal component analysis (PCA) where she clusters with western hunter-gatherers (WHGs). Allele-sharing estimates based on *f<sub>d</sub>*-statistics show the same overall affinity to WHGs (Fig. 2b). This is also reflected in the *qpAdm* analysis<sup>12</sup> (see Methods) which demonstrates that a simple one way model assuming 100% WHG ancestry cannot be rejected in favour of

more complex models (Fig. 2c; Supplementary Table 6). To formally test this result we computed two sets of *D*-statistics of the form *D*(Yoruba, EHG/Barcin; test, WHG) and find no evidence for significant levels of EHG or Neolithic farmer gene flow (Supplementary Fig. 7; Supplementary Tables 7, 8).

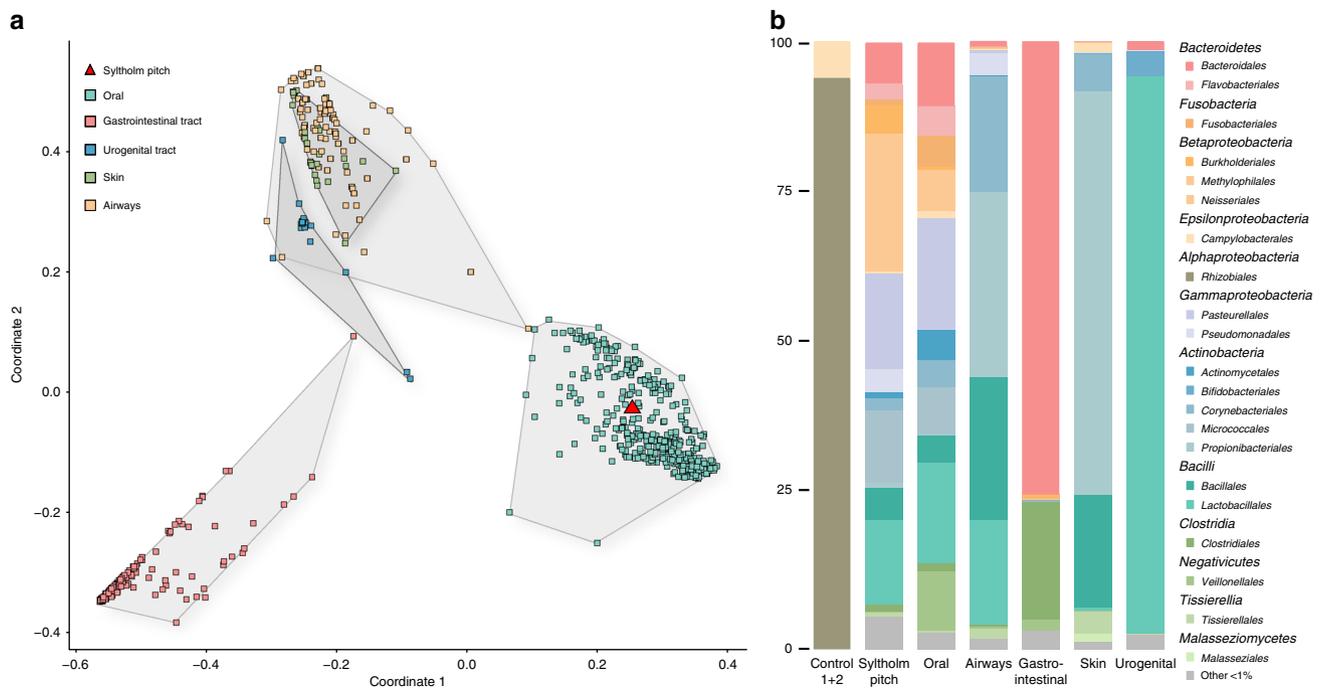
**Metataxonomic profiling of non-human reads.** To broadly characterise the taxonomic composition of the non-human reads in the sample, we used MetaPhlan2<sup>13</sup>, a tool specifically designed for the taxonomic profiling of short-read metagenomic shotgun data (see Methods; Supplementary Data 4). 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)<sup>14</sup>. 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<sup>14</sup>, visualised using MEGAN6<sup>15</sup>.

**Oral microbiome characterisation.** To further characterise the microbial taxa present in the ancient pitch and to obtain species-specific assignments we used MALT<sup>16</sup>, 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



**Fig. 2** Genetic affinities of the Sylltholm individual. **a** Principal component analysis of modern Eurasian individuals (in grey) and a selection of over 100 previously published ancient genomes, including the Sylltholm genome. The ancient individuals were projected on the modern variation (see Methods). **b** Allele-sharing estimates between the Sylltholm individual, other Mesolithic and Neolithic individuals, and WHGs versus EHG and Neolithic farmers, respectively, as measured by the statistic *f<sub>d</sub>*(Yoruba, X; EHG/Barcin, WHG). **c** Ancestry proportions based on *qpAdm*<sup>12</sup>, specifying WHG, EHG, and Neolithic farmers (Barcin) as potential ancestral source populations. PWC Pitted Ware Culture, LBK Linearbandkeramik, GAC Globular Amphora Culture, LP Late Paleolithic, M Mesolithic, EN Early Neolithic, MN Middle Neolithic, LN Late Neolithic. Data are shown in Supplementary Tables 4–6.



**Fig. 3** Metagenomic profile of the Syltholm birch pitch. **a** PCoA with Bray-Curtis at genera level with 689 microbiomes from HMP<sup>14</sup> and the Syltholm sample (see Methods). **b** Order-level microbial composition of the Syltholm sample compared to a control sample (soil) and metagenome profiles of healthy human subjects at five major body sites from the HMP<sup>14</sup>, visualised using MEGAN<sup>15</sup>.

*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<sup>17</sup> 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<sup>18</sup> 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 (2.3×) and mitogenome (91×), 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<sup>19–22</sup>, 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<sup>23</sup>. 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<sup>24,25</sup>.

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<sup>23</sup> 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<sup>23</sup>. 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 Ertebølle 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<sup>26</sup>. 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<sup>27</sup>. 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.

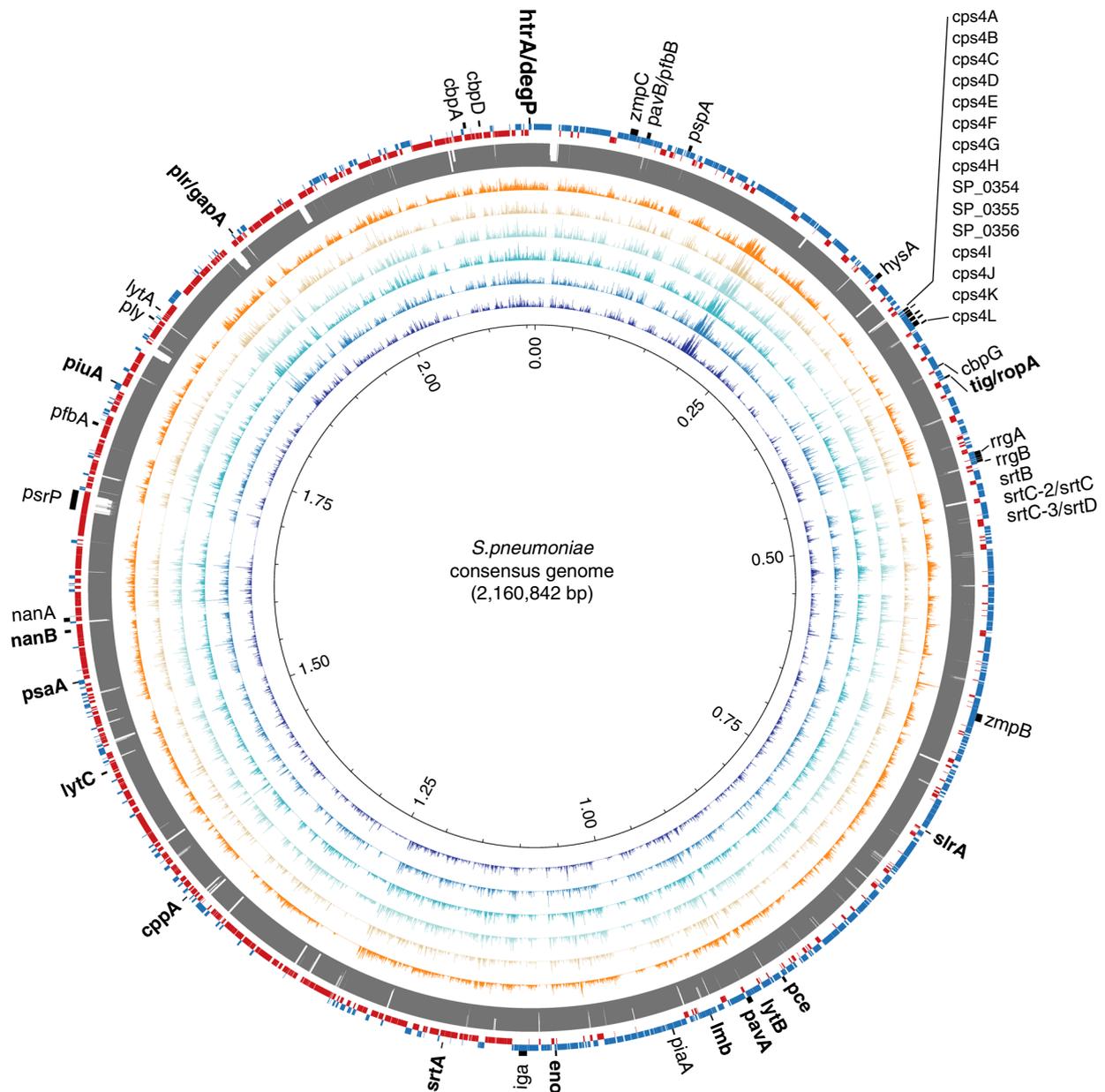
**Table 1** List of non-human taxa identified in the Syltholm pitch, including the 40 most abundant oral bacterial taxa, viruses, and eukaryotes. Bacteria in the red complex are denoted with an asterisk. Depth (DoC) and breadth of coverage (>1x) were calculated using BEDTools<sup>72</sup>. Deamination rates at the 5' ends of DNA fragments were estimated using mapDamage 2.0.9<sup>59</sup>. -Δ% refers to the negative difference proportion introduced by Hübler et al<sup>79</sup>. (see Supplementary Note 5).

Species	Reads	Fragment length (bp)	DoC	SD DoC	>1x (%)	C-T 5' (%)	-Δ%
Bacteria							
<i>Neisseria subflava</i>	308,732	56	7.5	6.2	83.7	14.5	0.9
<i>Rothia mucilaginosa</i>	296,610	52	6.9	5.6	82.3	14.0	0.9
<i>Streptococcus pneumoniae</i>	176,782	57	4.7	6.3	65.7	13.8	0.9
<i>Neisseria cinerea</i>	153,683	58	4.9	5.1	71.7	15.1	1.0
<i>Lautropia mirabilis</i>	117,040	53	2.0	1.9	71.9	13.0	1.0
<i>Neisseria meningitidis</i>	100,540	51	2.3	4.3	42.4	14.9	0.9
<i>Aggregatibacter segnis</i>	95,670	58	2.8	2.8	73.3	14.5	0.9
<i>Neisseria elongata</i>	68,407	54	1.6	1.9	67.6	15.1	0.9
<i>Prevotella intermedia</i>	65,324	56	1.2	1.4	55.0	16.2	0.9
<i>Streptococcus</i> sp. ChDC B345	52,614	61	1.6	2.7	50.3	13.8	0.9
<i>Streptococcus</i> sp. 431	43,787	59	1.2	1.9	47.5	13.6	0.8
<i>Aggregatibacter aphrophilus</i>	43,231	56	1.1	1.6	50.4	15.0	0.8
<i>Streptococcus pseudopneumoniae</i>	38,832	61	1.1	2.4	34.9	14.4	0.9
<i>Capnocytophaga leadbetteri</i>	36,461	59	0.9	1.1	49.8	14.0	0.8
<i>Corynebacterium matruchotii</i>	36,070	52	0.7	0.9	44.0	13.0	1.0
<i>Gemella morbillorum</i>	32,284	63	1.2	1.5	56.4	16.3	1.0
<i>Streptococcus viridans</i>	27,840	60	0.8	1.5	36.5	14.5	1.0
<i>Neisseria gonorrhoeae</i>	27,704	53	0.7	2.0	21.3	15.0	1.0
<i>Neisseria sicca</i>	27,290	57	0.6	1.4	22.5	13.7	0.9
<i>Fusobacterium nucleatum</i>	26,783	64	0.8	1.1	47.8	14.1	0.9
<i>Prevotella fusca</i>	26,295	57	0.5	0.7	34.6	15.7	1.0
<i>Kingella kingae</i>	25,811	55	0.7	1.0	44.2	14.4	1.0
<i>Ottowia</i> sp. 894	25,425	52	0.5	0.7	34.6	14.4	1.0
<i>Streptococcus</i> sp. NPS 308	24,937	59	0.8	1.4	37.5	14.3	0.8
<i>Actinomyces oris</i>	24,029	52	0.4	0.7	29.8	12.7	1.0
<i>Streptococcus australis</i>	23,777	60	0.7	1.3	31.5	13.8	1.0
<i>P. propionicum</i>	22,864	50	0.3	0.6	26.8	13.2	0.9
<i>Haemophilus</i> sp. O36	19,707	62	0.7	1.5	28.4	14.5	1.0
<i>Porphyromonas gingivalis</i> *	17,651	55	0.4	0.7	32.2	17.2	1.0
<i>Capnocytophaga gingivalis</i>	16,734	58	0.3	0.6	27.1	15.0	1.0
<i>Neisseria polysaccharea</i>	14,442	57	0.4	1.4	15.0	15.8	1.0
<i>Tannerella forsythia</i> *	14,187	55	0.2	0.5	19.8	15.3	1.0
<i>Streptococcus</i> sp. A12	13,232	59	0.4	0.9	24.9	14.6	0.9
<i>Capnocytophaga sputigena</i>	12,587	58	0.2	0.5	19.9	14.7	0.9
<i>Neisseria lactamica</i>	11,971	56	0.3	1.0	14.2	14.2	0.8
<i>Treponema denticola</i> *	11,379	59	0.2	0.5	19.5	14.0	0.8
<i>Rothia dentocariosa</i>	10,944	54	0.2	0.5	20.0	13.6	1.0
<i>Tannerella</i> sp. HOT-286	10,397	53	0.2	0.5	15.7	14.0	1.0
<i>Actinomyces meyeri</i>	10,105	51	0.3	0.5	21.3	14.0	1.0
<i>Filifactor alocis</i>	9,948	61	0.3	0.6	25.6	15.0	1.0
Viruses							
<i>Epstein-Barr virus</i>	593	51	0.2	0.4	13.3	17.8	1.0
Eukaryotes							
<i>Anas platyrhynchos</i>	55,986	51	<0.1	0.05	0.2	15.6	1.0
<i>Corylus avellana</i>	8,615	55	<0.1	0.04	0.1	19.7	1.0
<i>Betula pendula</i>	3,291	54	<0.1	0.02	<0.1	16.1	1.0

Our analyses have shown that the Syltholm individual does not carry any Neolithic farmer ancestry, suggesting that the genetic impact of Neolithic farming communities in southern Scandinavia might not have been as instant or pervasive as once thought<sup>28</sup>. While the mtDNA we recovered belongs to haplogroup K1e, which is more commonly associated with early farming communities<sup>29–31</sup>, there is mounting evidence to suggest that this lineage was already present in Mesolithic Europe<sup>32–34</sup>. Overall, the lack of Neolithic farmer ancestry is consistent with evidence from elsewhere in Europe, which suggests that genetically distinct hunter-gatherer groups survived for much longer than previously assumed<sup>35–37</sup>. These WHG “survivors” might have triggered the resurgence of hunter-gatherer ancestry that is proposed to have occurred in central Europe between 7000 and 5000 BP<sup>12</sup>.

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<sup>38–40</sup> 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



**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 Syltholm 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<sup>73</sup>.

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<sup>41</sup>. 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<sup>42</sup>.

*S. pneumoniae* has a remarkable capacity to remodel its genome through the uptake of exogenous DNA from other

pneumococci and closely related oral streptococci<sup>42</sup>. 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

studies<sup>43,44</sup> 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<sup>45</sup>. 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)<sup>46</sup> 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<sup>47,48</sup>.

Lastly, we identified several thousand reads that could be confidently assigned to different plant and animal species, including birch (*B. pendula*), hazelnut (*C. avellana*), and mallard (*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 *Anas* sp. and hazelnuts<sup>6,49</sup>. In addition, there is evidence from many other Mesolithic and Early Neolithic sites in Scandinavia for hazelnuts being gathered in large quantities for consumption<sup>50</sup>. 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

**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 filters (MWCO 30 kDa), mixed 1:10 with a PB-based binding buffer<sup>51</sup>, and purified using MinElute columns, eluting in 30 µl EB. 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:isoamyl alcohol. 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 metagenomic profiling the extracts were found to be contaminated with *Delftia* spp., a known laboratory contaminant<sup>52</sup>. The contaminated libraries were excluded from metagenomic profiling.

**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 *Delftia* spp., suggesting that the contaminants were introduced during this step. In addition, we included two soil samples from the site, weighing c. 2 g each, as negative controls. DNA was extracted as described above using 3 ml EDTA-based lysis buffer followed by 9 ml 25:24:1 phenol:chloroform:isoamyl alcohol mixture to account for the larger amount of starting material. The sequencing results are reported in Supplementary Table 1.

**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<sup>53</sup>. 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<sup>54</sup>. 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) using High Sensitivity tapes. 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.

**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 v1.2.12<sup>55</sup>. Adapters and low quality reads (Q < 20) were removed using AdapterRemoval v2.2.0<sup>56</sup>, only retaining reads >25 bp. Trimmed and filtered reads were then mapped to hg19 (build 37.1) using BWA<sup>57</sup> with seed disabled to allow for better sensitivity<sup>58</sup>, as well as filtering out unmapped reads. Only reads with a mapping quality ≥30 were kept and PCR duplicates were removed. MapDamage 2.0.9<sup>59</sup> was used to evaluate the authenticity of the retained reads as part of the PALEOMIX pipeline<sup>55</sup>, using a subsample of 100k reads per sample (Supplementary Fig. 6). For the population genomic analyses, we merged the ancient sample with individuals from the Human Origin dataset<sup>11</sup> 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.

**MtDNA analysis and contamination estimates.** We used Schmutzi<sup>60</sup> to determine the endogenous consensus mtDNA sequence and to estimate present-day human contamination. Reads were mapped to the Cambridge reference sequence (rCRS) and filtered for MAPQ ≥ 30. Haploid variants were called using the *endoCaller* program implemented in Schmutzi<sup>60</sup> 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 v2.2.6<sup>61</sup> to determine the mtDNA haplogroup, specifying PhyloTree (build 17) as the reference phylogeny<sup>62</sup>. Contamination estimates were obtained using Schmutzi's *mtCont* program and a database of putative modern contaminant mitochondrial DNA sequences.

**Genotype imputation.** We used ANGS<sup>63</sup> to compute genotype likelihoods in 5 Mb windows around 43 SNPs associated with skin, eye, and hair colour<sup>10</sup> and lactase persistence into adulthood (Supplementary Data 2). Missing genotypes were imputed using impute2<sup>64</sup> and the pre-phased 1000 Genome reference panel<sup>65</sup>, 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 HIRISplex-S website<sup>10</sup> to obtain the predicted outcomes for the pigmentation phenotypes (Supplementary Data 3).

**Principal component analysis.** Principal component analysis was performed using smartPCA<sup>66</sup> by projecting the ancient individuals onto a reference panel including >1000 present-day Eurasian individuals from the HO dataset<sup>11</sup> using the option *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<sup>2</sup> larger than 0, resulting in a final data set of 593,102 SNPs.

**D- and f-statistics.** D- and f-statistics were computed using *AdmixTools*<sup>67</sup>. 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<sub>4</sub>*-statistics of the form *f<sub>4</sub>*(Yoruba, X; EHG/Barcin, WHG) where "X" stands for the 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*(Yoruba, EHG/Barcin; X, WHG) testing whether "X" forms a clade to the exclusion of EHG and Neolithic farmers (represented by Barcin), respectively.

**qpAdm.** Admixture proportions were modeled using *qpAdm*<sup>12</sup>, specifying Mesolithic Western European hunter-gatherers (WHG), Eastern hunter-gatherers (EHG) and early Neolithic Anatolian farmers (Barcin), as possible ancestral source populations. We present the model with the lowest number of source populations that fits the data, as well as the model with all three admixture components (see Supplementary Table 6). When estimating the admixture proportions for WHGs and EHG, the test sample was excluded from their respective reference populations.

**MetaPhlan.** We used MetaPhlan2<sup>13</sup> to create a metagenomic profile based on the non-human reads (Supplementary Data 4). The reads were first aligned to the MetaPhlan2 database<sup>13</sup> using Bowtie2 v2.2.9 aligner<sup>68</sup>. PCR duplicates were removed using PALEOMIX filteruniquebam<sup>58</sup>. For cross-tissue comparisons 689 human microbiome profiles published in the Human Microbiome Project Consortium<sup>14</sup> were initially used, comprising samples from the mouth ( $N = 382$ ), skin ( $N = 26$ ), gastrointestinal tract ( $N = 138$ ), urogenital tract ( $N = 56$ ), airways and nose ( $N = 87$ ). The oral HMP samples consist of attached/keratinised gingiva ( $N = 6$ ), buccal mucosa ( $N = 107$ ), palatine tonsils ( $N = 6$ ), tongue dorsum ( $N = 128$ ), throat ( $N = 7$ ), supragingival plaque ( $N = 118$ ), and subgingival plaque ( $N = 7$ ). Pairwise ecological distances among the profiles were computed at genus and species level using taxon relative abundances and the `vegdist` function from the `vegan` package in R<sup>69</sup>. These were used for principal coordinate analysis (PCoA) of Bray–Curtis distances in R using the `pcor` function included in the `APE` package<sup>70</sup>. Subsequently, we calculated the average relative abundance of each genus for each of the body sites present in the Human Microbiome Project and visualised the abundance of microbial orders of our sample and the HMP body sites with MEGAN6<sup>15</sup>.

**MALT.** To further characterise the metagenomic reads we performed microbial species identification using MALT v. 0.4.1 (Megan ALignment Tool)<sup>16</sup>, 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\_000922395.1 uncultured crAssphage  
GCF\_000954235.1 uncultured phage WW-nAnB  
GCF\_000146025.2 uncultured Termite group 1 bacterium phylotype Rs-D17

The final MALT reference database contained 33,223 genomes and was created using default parameters in *alt-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 BWA aln and excluding all mapping reads. Duplicates were removed with `seqkit v.0.7.1`<sup>71</sup> using the `rmDup` function with the `-by-seq` flag. The remaining reads were processed with *alt-run* (v. 0.4.1) where BlastN mode and SemiGlobal alignment were used. The minimum percent identity (`-minPercentIdentity`) was set to 95, the minimum support (`-minSupport`) parameter was set to 10 and the top percent value (`-topPercent`) was set as 1. Remaining parameters were set to default. MEGAN6<sup>15</sup> was used to visualise the output `.rma6` files and to extract the reads assigned to taxonomic nodes of interest for our sample. A taxon 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 limited number of reads assigned 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<sup>16</sup> to the *S. pneumoniae* TIGR4 reference genome (NC\_003028.3). To investigate the presence of multiple strains we estimated the number of heterozygous sites using `samtools`<sup>57</sup> `mpileup` function, filtering out transitions, indels, and sites with a depth of coverage below 10. Coverage statistics of the individual alignments ( $MQ \geq 30$ ) were obtained using `Bedtools`<sup>72</sup> and plotted using `Circos`<sup>73</sup> in 100 bp windows. Mappability was estimated using `GEM2`<sup>74</sup> 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 extracts into contigs using `megahit`<sup>75</sup>. The contigs were aligned against known *S. pneumoniae* TIGR4 virulence genes in the Virulence Factor Database<sup>17</sup> (downloaded 22/11–2018) using `BLASTn`<sup>76</sup>. Only unique hits with a bitscore >200, >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<sup>17</sup> (downloaded 22/11–2018) using `BLASTn`<sup>76</sup> 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; SRS019120; SRS013942; SRS015055; SRS014692) from the Human Microbiome Project (HMP)<sup>14</sup> 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<sup>18</sup>, 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)<sup>77</sup>. The alignments were then parsed through a naive lowest common ancestor algorithm (ngsLCA) based on the NCBI taxonomic tree. Only taxonomically classified reads for taxa comprising  $\geq 1\%$  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

- 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**, 1310–1318 (2006).
- Kozowyk, P. R. B., Soressi, M., Pomstra, D. & Langejans, G. H. J. Experimental methods for the Palaeolithic dry distillation of birch bark: implications for the origin and development of Neandertal adhesive technology. *Sci. Rep.* **7**, 8033 (2017).
- Aveling, E. M. & Heron, C. Chewing tar in the early Holocene: an archaeological and ethnographic evaluation. *Antiquity* **73**, 579–584 (1999).
- Haque, S. et al. Screening and characterisation of antimicrobial properties of semisynthetic betulin derivatives. *PLoS ONE* **9**, e102696 (2014).
- Kashuba, N. et al. Ancient DNA from mastics solidifies connection between material culture and genetics of mesolithic hunter-gatherers in Scandinavia. *Commun. Biol.* **2**, 185 (2019).
- Sørensen, S. A. Syltholm: Denmark’s largest Stone Age excavation. *Mesolithic Misc.* **24**, 3–10 (2016).
- Aveling, E. M. & Heron, C. Identification of birch bark tar at the Mesolithic site of Star Carr. *Ancient Biomolecules* **2**, 69–80 (1998).
- Gamba, C. et al. Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
- 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).
- Chaitanya, L. et al. The HliisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Sci. Int. Genet.* **35**, 123–135 (2018).
- Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
- Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
- Truong, D. T. et al. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**, 902–903 (2015).
- The Human Microbiome Project Consortium. et al. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
- Huson, D. H. et al. MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* **12**, e1004957 (2016).
- Vågæne, Å. J. et al. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2**, 520–528 (2018).
- Chen, L. et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* **33**, D325–D328 (2005).
- Pedersen, M. W. et al. Postglacial viability and colonization in North America’s ice-free corridor. *Nature* **537**, 45–49 (2016).
- Olalde, I. et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).
- Skoglund, P. et al. Genomic diversity and admixture differs for stone-age Scandinavian foragers and farmers. *Science* **344**, 747–750 (2014).

21. Mathieson, 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. *Nat. 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. & Perry, G. H. 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. *Annu. Rev. Genomics Hum. Genet.* **18**, 297–319 (2017).
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 tragtbøgerkulturen i Danmark* (Copenhagen, 1948).
29. Bramanti, B. et al. Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers. *Science* **326**, 137–140 (2009).
30. Brandt, G. et al. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* **342**, 257–261 (2013).
31. Isern, N., Fort, J. & de Rioja, V. L. The ancient cline of haplogroup K implies that the Neolithic transition in Europe was mainly demic. *Sci. Rep.* **7**, 11229 (2017).
32. Hofmanová, Z. et al. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl Acad. Sci. USA* **113**, 6886–6891 (2016).
33. González-Forbes, 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).
34. Mittnik, A. et al. The genetic prehistory of the Baltic Sea region. *Nat. Commun.* **9**, 442 (2018).
35. Bollongino, R. et al. 2000 years of parallel societies in Stone Age Central Europe. *Science* **342**, 479–481 (2013).
36. Lipson, M. et al. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017).
37. 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).
38. 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).
39. Warinner, C. et al. Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46**, 336–344 (2014).
40. Jersie-Christensen, R. R. et al. Quantitative metaproteomics of medieval dental calculus reveals individual oral health status. *Nat. Commun.* **9**, 4744 (2018).
41. Suzuki, N., Yoneda, M. & Hirofujii, T. Mixed red-complex bacterial infection in periodontitis. *Int. J. Dent.* **2013**, 587279 (2013).
42. Weiser, J. N., Ferreira, D. M. & Paton, J. C. *Streptococcus pneumoniae*: transmission, colonization and invasion. *Nat. Rev. Microbiol.* **16**, 355–367 (2018).
43. Krause-Kyora, B. et al. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *Elife* **7**, e36666 (2018).
44. Mühlemann, B. et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423 (2018).
45. Williams, H. & Crawford, D. H. Epstein-Barr virus: the impact of scientific advances on clinical practice. *Blood* **107**, 862–869 (2006).
46. 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).
47. Toussiro, E. & Roudier, J. Epstein-Barr virus in autoimmune diseases. *Best. Pract. Res. Clin. Rheumatol.* **22**, 883–896 (2008).
48. Rezk, S. A., Zhao, X. & Weiss, L. M. Epstein-Barr virus (EBV)-associated lymphoid proliferations, a 2018 update. *Hum. Pathol.* **79**, 18–41 (2018).
49. Bangsgaard, P. *Report on the faunal remains from MLF 00906-II Syltholm II*. (Zoological Museum, University of Copenhagen, 2015).
50. Regnell, M. Plant subsistence and environment at the Mesolithic site Tågerup, southern Sweden: new insights on the 'Nut Age'. *Veg. Hist. Archaeobot.* **21**, 1–16 (2012).
51. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
52. Salter, S. J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).
53. Caroe, C. et al. Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* **9**, 410–419 (2018).
54. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2011).
55. Schubert, M. et al. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9**, 1056 (2014).
56. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
57. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
58. Schubert, M. et al. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012).
59. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
60. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
61. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
62. van Oven, M. PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Sci. Int.: Genet. Suppl. Ser.* **5**, e392–e394 (2015).
63. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinforma.* **15**, 356 (2014).
64. 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).
65. Consortium, The 1000 Genomes Project. et al. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2011).
66. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
67. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
68. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
69. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
70. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
71. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS ONE* **11**, e0163962 (2016).
72. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
73. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
74. Marco-Sola, S., Sammeth, M., Guigó, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat. Methods* **9**, 1185–1188 (2012).
75. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultrafast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
76. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinforma.* **10**, 421 (2009).
77. Rowley, E. R. et al. A Draft Genome and High-Density Genetic Map of European Hazelnut (*Corylus avellana* L.). Preprint at <https://www.biorxiv.org/content/https://doi.org/10.1101/469015v1> (2018).
78. Astrup, P. M. *Sea-level change in Mesolithic southern Scandinavia. Long- and short-term effects on society and the environment*. Jysk Arkæologisk Selskabs Skrifter **106** (Jutland Archaeological Society, 2018).
79. Huebler, R., Key, F. M. M., Warinner, C., Bos, K. I. & Krause, J. HOPS: Automated detection and authentication of pathogen DNA in archaeological remains. Preprint at <https://www.biorxiv.org/content/https://doi.org/10.1101/534198v2> (2018).

## Acknowledgements

We thank the Museum Lolland-Falster for access to the sample and the staff at the Danish National High-Throughput Sequencing Center for technical assistance. We also thank Miren Iraeta Orbegozo, Oliver Smith and Kristine Bohmann for their input and helpful discussion. This research was funded by a research grant from VILLUM FONDEN (grant no. 22917) awarded to H.S. T.Z.T.J. and J.N. were supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 676154 (ArchSci2020). K.H.I. was supported by the Danish Heart Foundation. M.J.C., A.J.T., and L.T.L. were funded by Danish National Research Foundation (DNRF128). M.D. was supported by a European Research Council grant (ECHOES, 714679). S.R. was supported by the Novo Nordisk Foundation grant NNF14CC0001 and the Jorck Foundation Research Award. H.S. was supported in part by HERA (Humanities in the European Research Area) through the joint research programme "Uses of the Past" and the European Union's Horizon 2020 research and innovation programme under grant agreement no. 649307 (CitiGen).

### 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