



OPEN

# Metagenomic analysis of Mesolithic chewed pitch reveals poor oral health among stone age individuals

Emrah Kırdök<sup>1✉</sup>, Natalija Kashuba<sup>2</sup>, Hege Damlien<sup>3</sup>, Mikael A. Manninen<sup>4</sup>, Bengt Nordqvist<sup>5</sup>, Anna Kjellström<sup>6</sup>, Mattias Jakobsson<sup>7</sup>, A. Michael Lindberg<sup>8</sup>, Jan Storå<sup>6</sup>, Per Persson<sup>3</sup>, Björn Andersson<sup>9</sup>, Andrés Aravena<sup>10</sup> & Anders Götherström<sup>11,12</sup>

Prehistoric chewed pitch has proven to be a useful source of ancient DNA, both from humans and their microbiomes. Here we present the metagenomic analysis of three pieces of chewed pitch from Huseby Klev, Sweden, that were dated to 9,890–9,540 before present. The metagenomic profile exposes a Mesolithic oral microbiome that includes opportunistic oral pathogens. We compared the data with healthy and dysbiotic microbiome datasets and we identified increased abundance of periodontitis-associated microbes. In addition, trained machine learning models predicted dysbiosis with 70–80% probability. Moreover, we identified DNA sequences from eukaryotic species such as red fox, hazelnut, red deer and apple. Our results indicate a case of poor oral health during the Scandinavian Mesolithic, and show that pitch pieces have the potential to provide information on material use, diet and oral health.

The Scandinavian Peninsula (hereafter referred to as Scandinavia) gradually became accessible to humans when the Weichsel ice sheet melted away after the late glacial maximum (LGM) peaked between 27,000 and 21,000 years before present (BP). There is material evidence for sporadic human presence in ice-free areas around 16,000 BP, but from around 12,800 BP there was a proper expansion into Scandinavia<sup>1</sup>, and from 11,700 BP there was a continuous human presence in the southwestern part of the Peninsula<sup>2–4</sup>. However, the earliest human DNA from the Scandinavian Peninsula is only slightly younger than 10,000 years old<sup>5</sup>.

The early Mesolithic sites exhibit a rich material culture (lithics, bone and antler) that has been used to infer demography, mobility, social relations, use of technology, and the subsistence strategies<sup>6</sup>. While this has provided an understanding of the Early Mesolithic people in Scandinavia, we still don't have enough molecular data to contextualise their oral microbiome profile, pathogen burden, diet related information, and paleoenvironmental utilisation.

Recently, metagenomic analysis of ancient DNA reads sequenced from ancient human samples have enabled researchers to obtain information about the historical human populations with a great depth. Ancient dental calculus has become the predominately used material to investigate oral microbiome, due to its structural integrity<sup>7</sup>. Lately, chewed pitch materials are also proved to be useful to access historical oral microbiome<sup>8</sup>.

Ancient dental calculus studies provide information about the oral microbial species in dental biofilm formation and maturation, and its association with dental health<sup>9–11</sup>. Moreover, changes in the oral microbiome through

<sup>1</sup>Department of Biotechnology, Faculty of Science, Mersin University, 33100Yenişehir Mersin, Turkey. <sup>2</sup>Department of Archaeology and Ancient History, Uppsala University, Engelska Parken, Thunbergsvägen 3H Box 626, 751 26 Uppsala, Sweden. <sup>3</sup>Museum of Cultural History, University of Oslo, St. Olavs Plass, P.O. Box 6762, NO-0130 Oslo, Norway. <sup>4</sup>PAES, Ecosystems and Environment Research Programme, Faculty of Biological and Environmental Sciences and Helsinki Institute of Sustainability Science, University of Helsinki, Viikinkaari 1, P.O. Box 65, Helsinki, Finland. <sup>5</sup>Foundation War-Booty Site Finnestorp, Klarinettvägen 75, 434 75 Kungälv, Sweden. <sup>6</sup>Department of Archaeology and Classical Studies, Osteoarchaeological Research Laboratory, Stockholm University, Stockholm, Sweden. <sup>7</sup>Department of Organismal Biology, Human Evolution, Uppsala University, Evolutionsbiologiskt Centrum EBC Norbyvägen 18 A, Uppsala, Sweden. <sup>8</sup>Department of Chemistry and Biomedical Sciences, Faculty of Health and Life Sciences, Linnaeus University, Hus Vita, 44018 Kalmar, Sweden. <sup>9</sup>Department of Cell and Molecular Biology (CMB), Karolinska Institutet, P.O. Box 285, 171 77 Stockholm, Sweden. <sup>10</sup>Department of Molecular Biology and Genetics, Faculty of Science, Istanbul University, Vezneciler, 34134 Istanbul, Turkey. <sup>11</sup>Centre for Palaeogenetics, Svante Arrhenius Väg 20C, 106 91 Stockholm, Sweden. <sup>12</sup>Department of Archaeology and Classical Studies, Archaeological Research Laboratory, Stockholm University, Stockholm, Sweden. ✉email: emrahkirdok@gmail.com

archaeological time periods can also be investigated. For example, the transition from hunter-gatherer societies to farmers brought a significant increase in human oral health pathologies, which are associated with dietary and lifestyle changes<sup>12</sup>. Several ancient DNA studies had investigated changes in oral microbiome composition from the Mesolithic period to modern day and their relation to dietary changes and population movements<sup>13–16</sup>, as well as pre and post Columbian contact<sup>17</sup>. It is also possible to identify dietary remains and DNA reads related to the paleoenvironment through ancient DNA and microscopic studies<sup>18–20</sup>.

Another promising source of ancient microbiome DNA investigated here are masticates, or pitch pieces that have been chewed by humans and therefore contain materials from the oral cavity. Previous studies had investigated the oral microbiome and dietary components from 5,700 years old chewed pitch from Denmark and successfully proved the potential usage in ancient DNA research domain<sup>5,8</sup>.

To better understand the Mesolithic community in Huseby Klev, we provide a metagenomic analysis of the pitch mastic pieces that are more than 9,500 years old, and that already have yielded information on Mesolithic demography<sup>5</sup>. This DNA represents the oldest known genomic and metagenomic material to date from humans and human activities in Scandinavia.

Our analysis shows the oral microbiome profile in three chewed pitch pieces along with high coverage genomes of human commensal oral microbes. By comparing our data with modern oral microbiomes, we demonstrate the increased abundance of oral microbes that could cause pathological conditions such as periodontitis. Moreover, we provide DNA evidence on dietary components and possible paleoenvironmental eukaryotes that represents the Huseby Klev archaeological site<sup>21</sup>.

## Results

### A snapshot of the Mesolithic oral microbiome

To have a broad overview of the metagenomic composition in our ancient pitch materials, we first compared the metagenomic profile of the samples (ble004, ble007, and ble008) against the taxonomic composition reported in the modern human oral cavity ( $N = 725$ ), modern human nasal cavity ( $N = 93$ ), modern human skin ( $N = 42$ ), modern human stool ( $N = 500$ ), modern human vagina ( $N = 86$ ), ancient human dental calculus ( $N = 395$ ), and chewed pitch ( $N = 1$ ) samples. The last sample is also chewed pitch but from a later period (5,858–5,661 cal. BP)<sup>8–11,14–17,19,22–31</sup>.

Phylum profiles of the ancient samples show a pattern closely related to oral microbiome samples. Also, the phylum-level profile of the pitch samples closely resembles the Syltholm pitch profile (Figure S1). In addition, the species' relative abundances show a remarkable presence of commensal human bacteria (Figure S2).

Subsequently, we calculated Bray–Curtis distances using species taxonomic composition to compare microbial diversity between the samples<sup>32</sup> and visualised the distance matrix with the non-Metric Multidimensional Scaling (NMDS) method. The NMDS plot and Bray–Curtis distances show that microbial compositions in the ancient pitch material are closer to oral samples than to other parts of the human body. The ancient ble pitch samples cluster with modern oral microbiome samples in the NMDS plot along with the Syltholm chewed birch pitch (Fig. 1, Figure S3). In addition to these analyses, we quantified the potential oral microbiome contribution to the ancient mastic samples by using the same external dataset along with environmental sediment samples and we observed 39–57% of oral microbiome and ancient dental calculus contribution to ble chewed pitch pieces (Figure S4)<sup>33–36</sup>.

These results strongly suggest that the microbial profile of the pitch material originates from the oral cavity, corroborating that the mastic pieces were orally processed.

### Validation of bacterial species presence

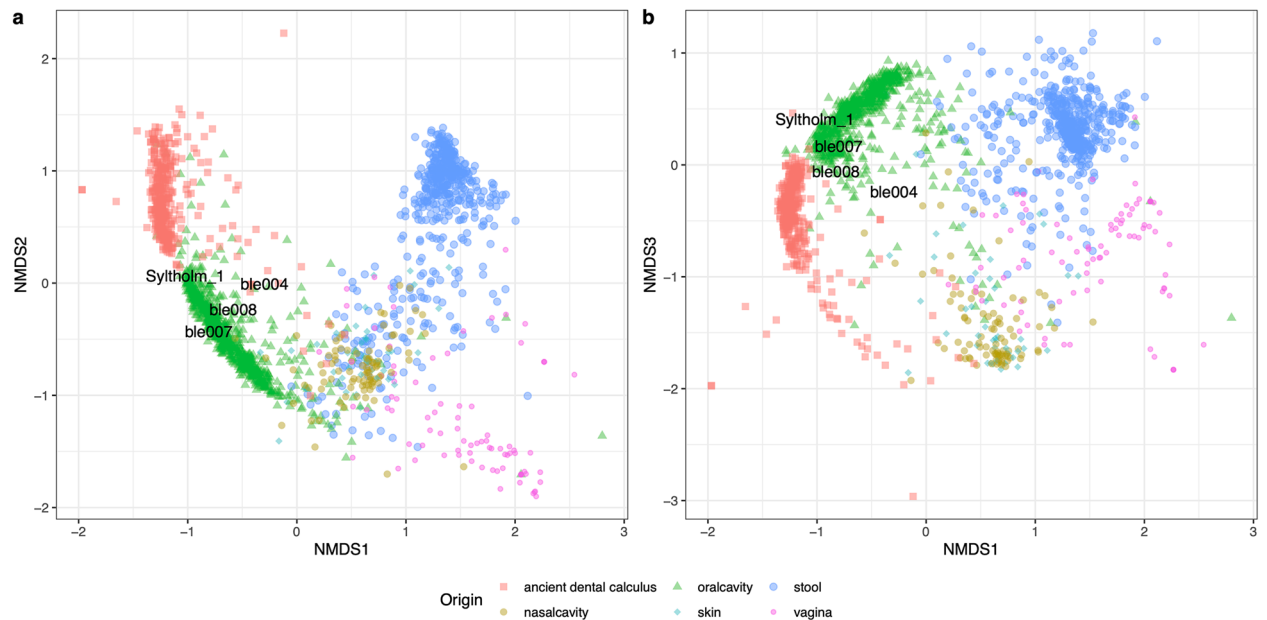
The taxonomic profiling mapped DNA reads in our libraries to 281 bacterial species (Table S1). To validate the presence of these bacteria in our ancient chewed pitch material, we aligned all reads to their corresponding reference genome. A homogeneous coverage across the genomes distinguishes bacterial presence from false positives.

We found 51 individual bacterial genomes with a minimum of  $1 \times$  depth of coverage and 40% breadth of coverage (Table 1). Most of these genomes correspond to pathogenic or opportunistic oral bacterial species. For example, we identified sequences from *Porphyromonas gingivalis* and *Treponema denticola* that belong to the red complex (a group of species that are found predominantly in periodontitis cases), *Aggregatibacter actinomycetemcomitans* (predominantly in sample ble004), and *Streptococcus mutans*. Furthermore, in all samples and predominantly in sample ble004 we note the presence of *Haemophilus influenzae*, *Aggregatibacter actinomycetemcomitans*, *Cardiobacterium valvarum*, *Eikenella corrodens*, and *Kingella denitrificans*.

### Authentication of ancient oral bacteria

We used three established criteria to authenticate the antiquity of the reads<sup>37,38</sup>: edit distance distribution, length distribution, and deamination patterns. It is expected that most of the sequences have edit distance at most 1. Also, we expect the majority of the sequences to be shorter than 100 nucleotides, and they should have deamination patterns towards the ends (Table S1, Figure S5). Edit distance describes the number of substitutions between a reference genome and aDNA reads thus providing an estimate for the evolutionary distance.

Table 1 shows the summary statistics after the genome alignment steps with at least  $1 \times$  depth of coverage. Note that several libraries of the ble004 sample were damage-repaired for human population genetics analysis, and thus shows a lower deamination rate. However, the non-damage repaired libraries of ble004 show the authentic deamination patterns.



**Figure 1.** Non-metric multidimensional projection of the Bray–Curtis distance matrix based using the MetaPhlan3 species abundance matrix. In this figure, we used in modern human oral cavity ( $N = 725$ ), modern human nasal cavity ( $N = 93$ ), modern human skin ( $N = 42$ ), modern human stool ( $N = 500$ ), modern human vagina ( $N = 86$ ), ancient human dental calculus ( $N = 395$ ), and chewed pitch ( $N = 1$ ) to characterise the microbiome composition in the ancient chewed materials. We can see that ancient chewed pitch materials are similar to the oral samples.

### De novo assembly of ancient bacteria

The metagenomic profile indicates presence of ancient oral bacteria, several of them likely opportunistic pathogens. Next, we reconstructed large portions of the ancient genomes by assembling de novo to further authenticate our data (Table S2). This strategy enabled us to assemble contigs that could be associated to a total of 49 unique bacterial species in our data.

### Differentially abundant oral bacteria in ancient pitch materials

The types of bacterial species present in ancient chewing gum appears to be useful to explore the oral microbiome. This is in concordance with recent studies where some 15% of the salivary microbiome seems to get encapsulated in modern chewing gums<sup>39</sup>. Thus, the microbiome profiles from the pitch material should make it possible to identify the signs of dysbiosis in the microbiomes of the ancient individuals.

Using modern periodontitis, caries, and healthy salivary datasets<sup>40,41</sup>, we explored the presence of differentially abundant microbes that are specific to dysbiosis conditions. The analysis indicated 27 differentially abundant bacteria that could be used as markers for periodontitis and caries conditions (Fig. 2, Table S3). We found increased abundances of *Actinomyces* species along with *Treponema denticola*, *Streptococcus anginosus*, *Slackia exigua*, and *Fusobacterium nucleatum*, which are common in modern periodontitis cases. Concerning dental caries, we found increased abundance of *Streptococcus sobrinus* and *Parascardovia denticolens* compared to healthy oral microbiome profile.

With this method, we identified differentially abundant bacterial species in libraries ble004, ble007, and ble008 (Table S4). Several of those marker species were found in ancient masticated pitch samples such as *Treponema denticola*, *Actinomyces johnsonii*, *Actinomyces timonensis*, and *Streptococcus anginosus* (Table 2).

Moreover, the results suggest a significant increase in the oral commensal bacterial species, in respect to healthy salivary samples such as *Rothia dentocariosa*, *Gemella sanguinis*, and *Streptococcus sanguinis* as well as *Haemophilus influenzae*, *Aggregatibacter actinomycetemcomitans*, *Cardiobacterium valvarum*, *Eikenella corrodens*, and *Kingella denitrificans*.

The ble004 sample is especially interesting, as eleven periodontitis-specific marker species show abundance patterns linked to disease. On the other hand, we found only one marker species (*Actinomyces timonensis*) linked to dental caries.

### Machine learning predicts periodontitis-like oral microbiome in chewed pitch materials

We used a machine learning method to find the species that correlate with dysbiosis, and to predict sign of dysbiosis in ancient samples.

Using modern salivary microbiome data, we trained several Random Forests<sup>42</sup> models for periodontitis and caries conditions, and selected the features (species) that contribute to the overall accuracy of the models. The best models classified the correct health status of the samples with an estimated error rate under 14%. The

Sample	Bacteria	Read Count	Mean length	DOC	BOC (%)	$\delta S$	SD
ble004	<i>Rothia dentocariosa</i> ATCC 17,931 X	802 082	52.90	16.93	92	0.20	0.01
ble004	<i>Corynebacterium durum</i> F0235	253 338	50.83	4.58	91	0.33	0.03
ble004	<i>Lautropia mirabilis</i> ATCC 51,599	1 196 572	50.07	18.96	90	0.99	0.01
ble004	<i>Actinomyces johnsonii</i> F0542	535 182	50.93	8.19	90	0.93	0.06
ble004	<i>Rothia mucilaginosa</i> DY-18	1 446 426	53.05	33.84	87	0.98	0.02
ble004	<i>Abiotrophia defectiva</i> ATCC 49,176	792 762	52.25	20.27	86	0.14	0.00
ble004	<i>Gemella sanguinis</i> ATCC 700,632	543 902	60.69	18.38	85	0.18	0.01
ble004	<i>Actinomyces oris</i>	600 330	48.90	9.65	85	0.80	0.15
ble004	<i>Pseudomonas fluorescens</i> F113	680 499	84.30	8.38	85	0.12	0.23
ble004	<i>Streptococcus sanguinis</i> SK36	623 791	53.65	14.01	83	0.12	0.00
ble004	<i>Eubacterium nodatum</i> ATCC 33,099	178 607	55.68	5.43	81	0.15	0.01
ble004	<i>Granulicatella adiacens</i> ATCC 49,175	376 560	54.45	10.53	81	0.12	0.00
ble004	<i>Peptostreptococcus stomatis</i> DSM 17,678	301 582	58.23	8.83	80	0.13	0.00
ble004	<i>Neisseria mucosa</i> C102	1 227 593	54.03	30.75	79	0.15	0.00
ble004	<i>Streptococcus gordonii</i> str. Challis substr. CH1	245 194	54.45	6.08	79	0.13	0.01
ble004	<i>Neisseria elongata</i> subsp. glycolytica ATCC 29,315	254 896	52.56	5.94	77	0.22	0.01
ble004	<i>Veillonella parvula</i> DSM 2008	168 862	53.53	4.24	63	0.13	0.01
ble004	<i>Parvimonas micra</i>	120 248	56.76	4.19	59	0.14	0.01
ble004	<i>Streptococcus thermophilus</i> JIM 8232	196 068	52.48	5.33	51	0.09	0.01
ble004	<i>Streptococcus salivarius</i>	247 717	52.05	5.89	46	0.09	0.00
ble007	<i>Rothia dentocariosa</i> ATCC 17,931	71 783	64.89	1.86	71	0.64	0.01
ble007	<i>Rothia mucilaginosa</i> DY-18	115 316	61.53	3.13	67	0.65	0.01
ble007	<i>Gemella sanguinis</i> ATCC 700,632	62 974	70.56	2.47	67	0.73	0.02
ble008	<i>Rothia dentocariosa</i> ATCC 17,931	107 407	71.61	3.07	80	0.51	0.01
ble008	<i>Corynebacterium durum</i> F0235	76 522	70.79	1.93	76	0.53	0.02
ble008	<i>Eggerthia cateniformis</i> OT 569 = DSM 20,559	41 568	NA	1.82	72	0.57	0.02
ble008	<i>Rothia mucilaginosa</i> DY-18	125 238	66.01	3.65	71	0.55	0.01
ble008	<i>Lautropia mirabilis</i> ATCC 51,599	95 446	68.10	2.06	65	0.66	0.02
ble008	<i>Gemella sanguinis</i> ATCC 700,632	45 607	80.51	2.04	64	0.56	0.03
ble008	<i>Streptococcus sanguinis</i> SK36	72 704	69.91	2.13	61	0.40	0.01
ble008	<i>Streptococcus mutans</i> UA159	25 788	79.79	1.01	59	0.52	0.02
ble008	<i>Neisseria mucosa</i> C102	52 113	70.37	1.69	57	0.49	0.02
ble008	<i>Abiotrophia defectiva</i> ATCC 49,176	41 683	67.71	1.38	52	0.52	0.02
ble008	<i>Gemella haemolysans</i> ATCC 10,379	26 027	77.24	1.05	49	0.39	0.02
ble008	<i>Streptococcus salivarius</i>	51 019	67.50	1.57	46	0.45	0.02
ble008	<i>Streptococcus mitis</i> B6	40 508	67.63	1.28	43	0.42	0.02
ble008	<i>Granulicatella adiacens</i> ATCC 49,175	31 470	69.88	1.13	42	0.45	0.03
ble008	<i>Streptococcus thermophilus</i> JIM 8232	47 626	67.62	1.67	41	0.41	0.02

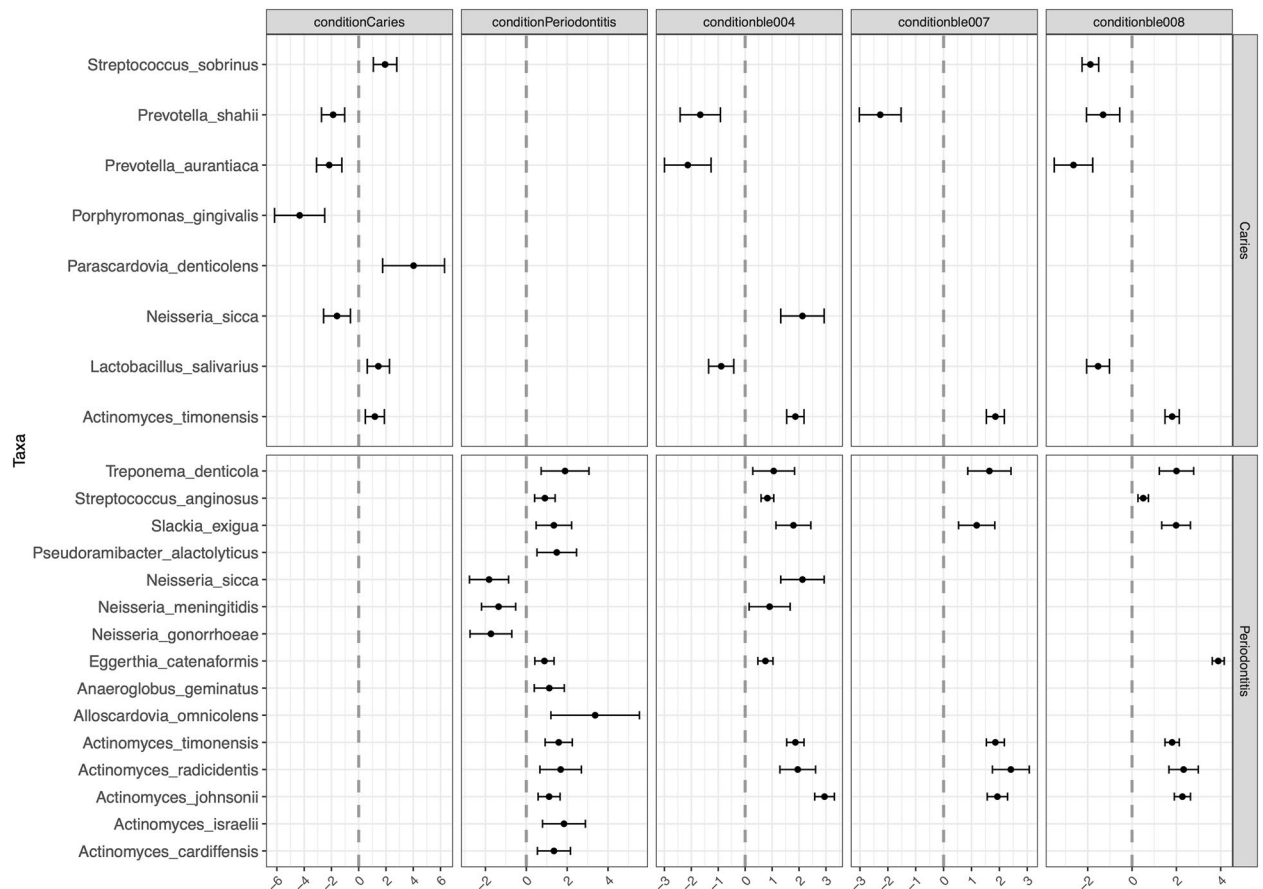
**Table 1.** Genome reconstruction of ancient pathogens. Single stranded deamination rates were low for ble004 samples since damage-repaired libraries were included in the analysis. However, higher deamination rates observed in non-damage repaired ble004 DNA libraries confirm their authenticity. DOC: Mean depth of coverage, BOC: Breadth of coverage,  $\delta S$ : Single stranded deamination rate, SD: Standard deviation for delta S.

selected models identified 38 bacterial species (Table S6 for the periodontitis and caries models. They included most of the marker species that are differentially abundant in dysbiosis cases, and several additional species.

The clear results, including the low estimated error rates, allow us to infer the microbiome profile in ancient pitch materials. This was carried out by comparing the microbial abundances in the periodontitis and dental caries models, and calculating the probability values for concordance (Fig. 3). The probability values for the dental caries models were not high enough to conclude the presence of caries with sufficient certainty. However, the periodontitis models suggest the occurrence of periodontitis based on the oral microbiome from ancient pitch samples, with more than 70% probability. In the ble004 sample, the probability of the presence of a periodontitis-like microbiome is 84%. These results supported the conclusions from the differential abundance analysis.

### Eukaryotic species in ancient pitch pieces

Using five Kraken2 confidence interval thresholds, we authenticated DNA reads from 15 possible eukaryotic species (Table S7–S11). In Table 3, only results from threshold 0.5 are shown. DNA reads mapped to hazelnut (*Corylus avellana*), red deer (*Cervus elaphus*), and brown trout (*Salmo trutta*) in ble004 and ble007; wolf (*Canis lupus*), hazelnut, red fox (*Vulpes vulpes*), arctic fox (*Vulpes lagopus*) in ble008 were present in all five threshold

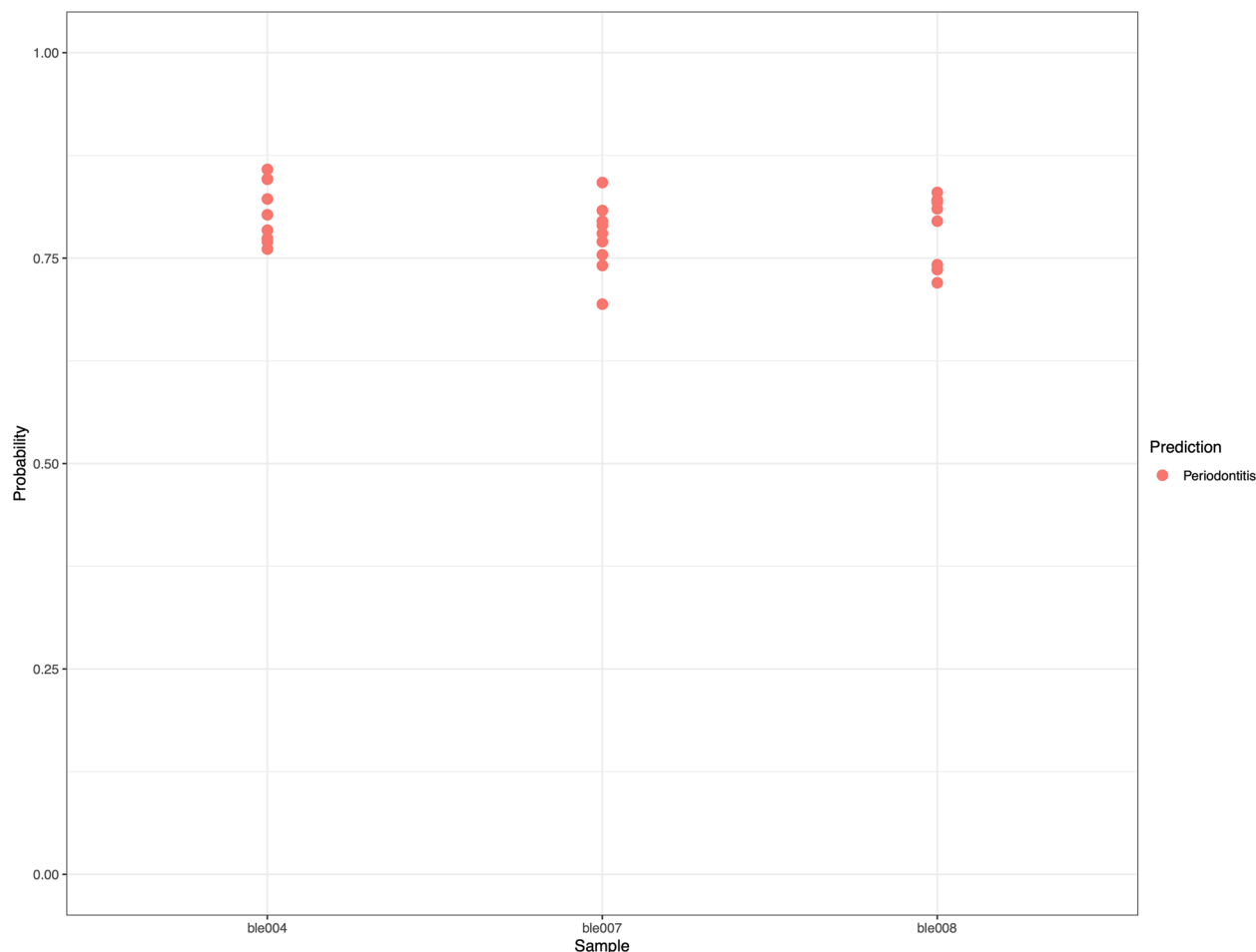


**Figure 2.** The Beta Binomial method applied to ancient mastics. Rows (Caries and Periodontitis) describing the two dysbiosis conditions tested. Each column represents a comparison condition. The condition control column represents the marker species that are significantly changed in each dysbiosis condition. Condition ble004, ble007 and ble008 columns represent the microbial species that were significant changes in ancient materials, compared to healthy cases.

	Species	Markers	ble004	ble007	ble008
Periodontitis	<b><i>Treponema_denticola</i></b>	<b>1.9 ± 1.2</b>	<b>1.06 ± 0.8</b>	<b>1.64 ± 0.8</b>	<b>2.01 ± 0.8</b>
	<b><i>Actinomyces_israelii</i></b>	<b>1.84 ± 1.1</b>	-	-	-
	<b><i>Actinomyces_timonensis</i></b>	<b>1.59 ± 0.7</b>	<b>1.86 ± 0.3</b>	<b>1.86 ± 0.3</b>	<b>1.81 ± 0.3</b>
	<b><i>Actinomyces_johnsonii</i></b>	<b>1.11 ± 0.5</b>	-	<b>1.93 ± 0.4</b>	<b>2.27 ± 0.4</b>
	<b><i>Streptococcus_anginosus</i></b>	<b>0.9 ± 0.5</b>	<b>0.83 ± 0.2</b>	-	<b>0.5 ± 0.2</b>
	<b><i>Eggerthia_catenaformis</i></b>	<b>0.89 ± 0.5</b>	<b>0.75 ± 0.3</b>	-	<b>3.89 ± 0.3</b>
	<i>Neisseria_sicca</i>	-1.82 ± 0.9	2.13 ± 0.8	-	-
Caries	<b><i>Streptococcus_sobrinus</i></b>	<b>1.94 ± 0.9</b>	-	-	<b>-1.88 ± 0.4</b>
	<i>Prevotella_shahii</i>	-1.88 ± 0.9	-1.67 ± 0.7	-2.27 ± 0.7	-
	<i>Prevotella_aurantiaca</i>	-2.17 ± 0.9	-2.13 ± 0.9	-	-2.64 ± 0.9
	<i>Porphyromonas_gingivalis</i>	-4.33 ± 1.8	-	-	-

**Table 2.** Differential abundance of the marker species in BLE samples using the beta-binomial method. The taxa with increased differential abundance in the marker set is shown in bold. Markers: the taxa that we found differentially abundant in dysbiosis conditions.

values. Eelgrass (*Zostera marina*) in ble004 and mallard (*Anas platyrhynchos*) in ble007 were present in four thresholds. Wolf, European robin (*Erithacus rubecula*), European crab apple (*Malus sylvestris*) in ble004 and European turtle dove (*Streptopelia turtur*) were present in three thresholds. Moreover, we authenticated DNA reads from mistletoe (*Viscum album*), apple (*Malus domestica*), tufted duck (*Aythya fuligula*), and limpet (*Patella vulgata*) in other Kraken2 thresholds.



**Figure 3.** The class prediction probability values from at least six different best models were combined and the predictions are shown for Periodontitis models.

Sample	Taxid	Latin name	English name	Reads after Processing	Mean Read length
ble004	13,451	<i>Corylus avellana</i>	Hazelnut	30,038	54.98
ble004	37,610	<i>Erithacus rubecula</i>	European robin	4209	50.48
ble004	9860	<i>Cervus elaphus</i>	Red deer	1942	56.41
ble004	296,552	<i>Zoster marina</i>	Eelgrass	913	68.44
ble004	8032	<i>Salmo trutta</i>	Brown trout	825	71.28
ble007	8839	<i>Anas platyrhynchos</i>	Mallard	172,172	65.23
ble007	1,771,552	<i>Streptopelia turtur</i>	European turtle dove	3348	66.34
ble007	8032	<i>Salmo trutta</i>	Brown trout	3164	69.40
ble008	9612	<i>Canis lupus</i>	Wolf	1,294,009	76.48
ble008	13,451	<i>Corylus avellana</i>	Hazelnut	177,612	79.29
ble008	9627	<i>Vulpes vulpes</i>	Red fox	5633	69.33
ble008	494,514	<i>Vulpes lagopus</i>	Arctic fox	4318	67.65

**Table 3.** Authenticated eukaryotes using Kraken2 confidence interval threshold of 0.5.

These species passed the visual confirmation criteria: edit distance, length distribution, and deamination patterns.

In particular, wolf and hazelnut in ble008, mallard in ble007, and hazelnut in ble004 samples had more than 10,000 reads aligned to the respective reference sequences. Moreover, we found 1059 reads mapped to the mitochondrial genome of red fox in the ble008 sample.



## Discussion

To broaden our understanding of human health and everyday life in Mesolithic Scandinavia, we investigated the metagenomic content of three pieces of ancient pitch material from one site. Using this data, we explored oral microbiome composition, diet related DNA reads, and resource utilisation.

The non-metric multidimensional scaling on Bray-Curtis<sup>32</sup> distances grouped chewed pitch materials close to human oral microbiome and ancient dental calculus samples. Also, *SourceTracker*<sup>43</sup> analysis showed a contribution of around 50% of modern oral microbiome and ancient dental calculus to ancient pitch samples. Moreover, there is a high relative abundance of species that are known to be commensal oral bacteria. This suggests that the DNA from chewed ancient pitch pieces is not specific to the pitch, nor to a specific tissue, but to most of the oral area in which the pitch was chewed.

To validate the presence of these commensal oral bacterial species we used two approaches. First, we compiled a large database curated from full length genomic sequences of Archaea, Bacteria, Viruses and Fungi, and used *malp*<sup>38</sup> (an alignment-based metagenomic classification method) to confirm the presence of known oral commensal bacteria. Also, we de novo assembled the DNA reads and created draft ancient bacterial genomes. Therefore, our analysis proves that the metagenomic content in ancient pitch pieces is closer to oral microbiome profile and contains commensal oral bacterial species.

A general problem with ancient microbial paleopathology is that there can be bacterial presence without physical symptoms. Generally, the presence of pathogenic species correlates with dysbiotic states but not necessarily in a causal way. Thus, our results suggest dysbiosis but we cannot show disease symptoms.

Subsequently we examined the microbial species in pitch pieces that are over abundant. To do this, we used a modern salivary microbiome dataset that contains healthy, periodontitis, and dental caries conditions and ancient dental calculus samples that were identified as periodontitis.

Recently, Nearing et al.<sup>44</sup> tested several differential abundance tools and showed a large variability between methods. Thus, they suggest using a consensus, that is, i. e. on results where several tools (based on different methods) agree.

In our study we applied two methods to identify bacterial species that correlate with dysbiosis conditions: a statistical method based on the beta-binomial distribution, and machine learning using Random Forests. The logic is, as suggested by Nearing et al.<sup>44</sup>, that the agreement between several techniques is a stronger argument for the detection of a given condition. Along the same line of reasoning, the reliability for dysbiosis detection increases when it is detected by several methods.

Using the modern salivary microbiome datasets, we identified specific bacterial species that are correlated with dysbiosis conditions and we search for the presence of these species in ancient chewed pitch pieces.

When we compare healthy and periodontitis modern samples, we observe increased abundance of *Streptococcus*, *Actinomyces*, *Eubacterium*, *Dialister*, and *Treponema* species along with a reduction of abundance of *Neisseria*, (Fig. 2) which is coherent with the findings of Li et al.<sup>45</sup>

*Streptococcus mutans* has been the most studied marker for dental caries, even though it is not considered to be the only cause of dental caries. For example, even in the absence of *Streptococcus mutans*, several acidophilic bacteria (e.g. *Lactobacillus* species) have been shown to co-occur with dental caries<sup>46,47</sup>.

Previous studies have shown that *Treponema denticola*, *Slackia exigua*<sup>48</sup>, *Eubacterium brachy*, and *Streptococcus anginosus* are likely associated with periodontitis. The predictor bacterial species from the best random forest models in this study identified most marker species and added additional potential markers. Moreover, in the caries random forest models, we found the marker *Streptococcus mutans*, which has been extensively studied in connection with dental caries cases. This is in line with the results of the beta-binomial test.

We confirmed the presence of the marker species in the ancient pitch material. We observed an increased abundance of periodontitis characteristic bacteria in all samples. Moreover, random forests models predict high probability for periodontitis in the Mesolithic individuals who chewed the pitch pieces. This is especially clear in ble004, where we had the highest number of differentially abundant markers.

The list of bacteria from the Syltholm chewed pitch material, which is a type of material similar to ours, indicates the presence of periodontitis specific bacterial taxa from the red-complex species. This sample dates to the Neolithic transition period but the Mesolithic human genetic component is compatible with our material<sup>8</sup>.

When contextualised, the results are compatible with what is generally known about the Mesolithic Scandinavian society. The Mesolithic population density was low<sup>49</sup>, with limited possibilities for pandemic-causing microbes to spread between humans, but not restricting the presence of bacteria from other sources than humans, like for example those causing systemic diseases including infective endocarditis. The wider use of the teeth, as tools, likely increased the risk for collecting periodontitis causing oral microbes.

The genetic traces we found complements the osteological and palaeobotanical results found from Huseby Klev and the species list in Table 3 reflects the palaeoenvironment of a settlement site located in the transitional zone between terrestrial and marine ecosystems. In particular, the remains of wolves, red fox, hazelnut, red deer and mallard were found in the several layers of Huseby Klev. In this perspective, the genetic data is consistent with the archaeological remains on site and reflects a Mesolithic environment<sup>21,50</sup>.

Moreover, other eukaryotic remains also fit the description of other Mesolithic sites in Europe. For example, hazelnut shells are frequently found at contemporary sites<sup>1,8,51</sup>. Also, reports show that limpets could be a dietary component in the Mesolithic period to complement meat from ungulates<sup>52</sup>. Bird bones were commonly used to produce tools in the Mesolithic. DNA sequences from Tufted duck, European robin, and mallard could therefore derive from birds used as food and/or raw material.

Even though remains of salmonids were not present, the presence of other species like cod (*Gadus morhua*) indicates that fishing was part of the subsistence strategy<sup>21</sup>. It has been suggested that salmonids are usually not

detectable due to poor fish bone preservation<sup>53</sup>. Moreover, reference genomes for fishes are mostly not complete. Therefore, it is possible that the DNA reads assigned to trout because of scanty reference genomes.

Mistletoe is a hemiparasitic plant that infects hard and softwood trees. It is frequently found in Europe and often discovered by macroscopic analysis at Mesolithic sites in central Europe. Reports suggest that the plant could be used to produce poison for arrowheads, in addition to being utilised for medicinal purposes<sup>54</sup>. Crab apple pieces, cores and wood from apple trees are also found at the site<sup>41</sup> as in several Mesolithic sites, and it is native to Europe.

In sample ble008, red fox, arctic fox and wolf were identified. These all three are members of the *Canidae* family, and they share a common ancestor. Thus separating these three species can be problematic. The presence of red fox is consistent with Mesolithic Scandinavian fauna, and their presence in the pitch samples is perhaps surprising but still consistent with osteological material from Huseby Klev. Furs of *Canidae* were probably used for clothing and the meat may have been consumed. Even though recent reports have demonstrated hunting of arctic foxes in Southern Poland 25,000 years ago<sup>55</sup>, the climate around Huseby Klev was too mild (annual mean temperature was 8°C) for this species to live<sup>56</sup>. We therefore used a competitive mapping approach and confirmed the presence of the mitochondrial genome of the red fox with 31% breadth of coverage.

Boethius et al.<sup>21</sup> analysed strontium content of faunal and floral remains from Huseby Klev to trace evidence of mobility. The authors suggest that the residents of Huseby Klev were mobile, and used waterways to travel distances up to 20 km for hunting. The presence of terrestrial animal DNA in the pitch material indicates hunting of those species which is in agreement with the osteo-archaeological material from the Huseby Klev site.

Recently Mann et al.<sup>18</sup> demonstrated the challenges in identifying eukaryotic reads in ancient samples. In essence, the relatively low number of aDNA sequences, and shared evolutionary ancestry in eukaryotic species create problems to authenticate species identifications. Moreover, they suggest that thousands of DNA reads would be needed for proper ancient DNA authentication. This is why we relied on stringent authentication criteria to identify the eukaryotic DNA reads.

In conclusion, we found that the chewed pitch material is an excellent source of ancient DNA, which can be used to understand diet, raw material use, and oral health conditions in prehistoric and early modern populations. We reconstructed several ancient bacterial genomes and found notable amounts of oral pathogens. Furthermore, we compared the microbial profile in chewed pitch mastic particles to modern human salivary microbiomes. Our results show that the bacterial profile is similar to modern periodontitis cases.

The samples also contained eukaryotic species like hazelnut, trout, mallard, red fox, limpet, and red deer. These species provide sources for reconstructing Mesolithic diet and paleoenvironment, while the presence of many of the species identified in this study is supported by archaeological finds. These DNA data may provide clues to specific dietary items but do not represent the whole variety of food sources used during the Scandinavian Mesolithic. They do, however, indicate what members of this particular group had been orally processing before or during their interaction with the pitch.

Our results portray a group of Mesolithic hunter-gatherers in south-western Scandinavia, where individuals likely suffered from dental diseases and were utilising resources from different domains through oral manipulation.

## Materials and methods

### The specific environmental/history/collection context

The Huseby Klev materials were unearthed and collected by archaeologists (including two of the co-authors of this article) during the excavation of this coastal hunter-fisher-gatherer site in the 90s<sup>50</sup>.

The material assemblage was rich and well preserved: human bones, animal bones, plant remains and pieces of masticated pitch were found. The exceptional preservation was presumably caused by a layer of clay which sealed the site into a geological time capsule. Pitch or “tar” pieces are a rather common find at Stone Age sites in Eurasia, and particularly in Scandinavia, where well preserved pieces of masticated pitch are found dating well into the Iron Age<sup>57</sup>. The pitch pieces were produced from birch bark tar, thus making birch a valuable resource during the entire Stone Age. There are about 90 masticated pitch pieces from the site, while tooth imprints in six of them have been cast and osteologically analysed. All the analysed pieces had imprints suggesting individuals younger than 20 years: three aged between 5 and 11 and three teenagers<sup>50</sup>. The pieces we used for the study were directly sampled from the material collection, which is under the guardianship of Bengt Nordqvist.

### Experimental procedure

The DNA extraction and library building was performed at Stockholm University, AFL, in ancient DNA dedicated facilities, all designed to have positive pressure, separate airlock, HEPA filters for the incoming air, UV lights at the ceiling, and constant sodium 3% hypochlorite disinfection (NaOCl). Sampling of the masticated pitch pieces was performed in a “drill laboratory”, a separate room to the “main laboratory” where exclusively sampling of the materials was undertaken to minimise contamination. The material was irradiated in a UV oven at about 6 J/cm<sup>2</sup> at 254 nm. When the necessary amount of the samples (between 50 and 100 mg) were collected into 2 mL screw cap tubes, the tubes were taken to the main laboratory for further extraction and library building. A blank was added at this stage. We checked the DNA concentration after extraction using Qubit 3.0 high sensitive double strand DNA fluorometric assay (Thermo Fisher Scientific). We did not observe any trace of contaminating DNA in the blank extractions.

In the main lab extraction was performed using two methods: the Yang extraction method<sup>58</sup> which is commonly used for extraction of bone tissue (incubation at 55 °C in 1000 µL of buffer per sample containing 0.5 M EDTA pH8, 1 M Urea, and 10 µL of Proteinase K 10 mg/mL, followed by concentration of the supernatant using Amicon membrane filters and purification with MinElute spin columns to obtain 110 µL of final product) and the



QIAamp PowerFecal DNA Kit (Qiagen) with which we obtained 100  $\mu\text{L}$  of final product, following the instructions of the kit with minor adjustments.

The libraries were also built in the main laboratory, following a modified protocol by Meyer and Kircher<sup>59</sup> for double-stranded blunt-end-repair libraries and damage-repaired libraries as in<sup>56</sup> pre-treating the extracts with USER enzyme. Libraries were amplified with 10  $\mu\text{M}$  index primers, later double-indexed, using AmpliTaq Gold 1000 Units 5 U/ $\mu\text{L}$  (Applied Biosystems) for blunt-end libraries, and AccuPrime™ Pfx DNA Polymerase (2.5 U/ $\mu\text{L}$ ) (Invitrogen) polymerase for damage repair libraries (damage repair applied to only ble004 sample). We determined the number of cycles using quantitative PCR, with reagents from Thermo Scientific and Biomers. Libraries were sequenced on the Illumina HiSeq X platform at the SciLife centre in Stockholm and SciLife centre in Uppsala. The damage repaired libraries of the ble004 sample are shown in Table S12.

### DNA libraries and external datasets

After receiving the sequences, *AdapterRemoval*<sup>60</sup> was used to trim adapters, nucleotides less than 20 base quality, and filtered shorter than 10 nucleotides and subsequently subjected to human demography pipeline.

In this study we focused on DNA sequences that do not map to the human reference genome. To do this we first filtered human DNA reads from fastq files and we used *cutadapt*<sup>61</sup> to trim remaining adapters, and nucleotides less than 20 base quality, and 10 nucleotides were also removed. Final sequencing depth of the fastq files is shown in Table S12.

In addition, we used published ancient and modern datasets to understand the microbial composition of ancient mastics. First, we used pre-calculated *MetaPhlan3* relative abundances of modern human microbiome samples using the *curatedMetagenomicData* package<sup>25–30,62,63</sup>. Next, we downloaded published DNA data from ancient dental calculus and chewed pitch studies<sup>8–11,14–17,19,22–24</sup>. At last, we used published studies that focus on sediments to identify environmental microbiome contribution to pitch samples<sup>28,33,34,64</sup> (Table S13).

For differential abundance analysis, we used a dataset containing salivary microbiome data of 12 healthy, 12 caries, and 10 periodontitis patients<sup>40,41</sup>.

We downloaded the fastq files from European Nucleotide archive and processed paired-end DNA sequences using *AdapterRemoval* tool<sup>61</sup> tool by trimming bases that have less than 20 phred quality score and collapsed forward and reverse reads if they had overlapping regions with 11 or more nucleotides.

### Metagenomic profiling of the ancient mastics

We used *MetaPhlan3* with default options to profile the fastq files of the ancient samples<sup>62</sup>. This tool uses a marker database (that contains approximately 13,500 bacterial and archaeal, 3,500 viral, and 110 eukaryotic marker DNA fragments from ~ 17,000 reference genomes) to profile a DNA library and produces a microbial relative abundance table. This describes the proportion of DNA fragments that belong to a particular taxon, and a sequence alignment map (sam) file that contains the alignment information. After profiling each DNA library, we removed optical PCR duplicates from sam files with *picard-tools*<sup>65</sup>, sam files that correspond to each library were merged for each sample, and relative abundance tables were calculated with *MetaPhlan3* for each sample.

To further analyse the dataset, we prepared a set of reference and representative genomes of bacterial, archaeal, fungal, and viral species from the NCBI RefSeq database<sup>66</sup> ( $N = 6133$ ) and indexed them using the *malt* alignment tool<sup>38</sup> and we aligned each fastq file to this indexed reference collection with the same tool. We used the *dust-masker*<sup>67</sup> tool to mask repetitive regions that could create false-positive classifications. Then, we aligned the sequenced libraries to the database using *malt*<sup>38</sup> with default options, and we obtained binary *rma6* files. These files contain the alignment and taxonomy information for each sequence. Afterwards, we used *MEGAN*<sup>68</sup> to extract the species absolute abundance values for each fastq file.

### Comparing microbial profiles from ancient samples with modern human microbiome

To understand the metagenomic composition in ancient mastic material, we compared the relative microbial abundances from the ancient mastic samples against a dataset consists of modern human oral cavity ( $N = 725$ ), modern human nasal cavity ( $N = 93$ ), modern human skin ( $N = 42$ ), modern human stool ( $N = 500$ ), modern human vagina ( $N = 86$ ), ancient human dental calculus ( $N = 395$ ), and chewed pitch ( $N = 1$ ). The site composition of the oral microbiome samples are anterior nares ( $N = 93$ ), buccal mucosa ( $N = 119$ ), hard palate ( $N = 1$ ), keratinized gingiva ( $N = 6$ ), palatine tonsils ( $N = 6$ ), saliva ( $N = 5$ ), subgingival plaque ( $N = 72$ ), supragingival plaque ( $N = 127$ ), and tongue dorsum ( $N = 198$ ). Some of the oral microbiome samples do not have any site information and therefore were not included in the list.

We calculated pairwise Bray–Curtis distances between each compositional vector using the *vegdist* function in the *vegan* package of the R statistical platform<sup>32,69,70</sup>. The distance matrix was visualised using a non-metric multidimensional scaling (NMDS) implemented in *metaMDS* in the *vegan* library with a K value of 7. The first three axes were selected for visualisation.

At last, we used *SourceTracker*<sup>43</sup> to quantify the oral microbiome content in chewed pitch materials<sup>30</sup>. To do this, we selected species taxonomic nodes from the *MetaPhlan3* abundance matrix, and we marked the comparison dataset from modern human oral microbiome samples and ancient dental calculus samples as source materials. Additionally, we included sediment ( $N = 23$ ), and shallow marine sediment ( $N = 21$ ) as environmental sources. Then we used the *SourceTracker* tool with the default options, to quantify the source contribution to the ancient samples (Table S14 – S22).

### Generating full genome ancient bacterial sequences

We collected all reads that were assigned to species' taxonomy nodes and combined them into one fastq file for each species, with all reads that align to at least one reference genome of that species, and nothing else.

Based on the taxonomic assignment results, we selected the corresponding reference genomes ( $n = 281$ ). And we aligned each fastq file to the respective reference genome, separately.

To produce full genome sequences, we extracted aDNA reads for each bacteria using *malt-extract*<sup>38</sup>.

Thereafter, DNA reads for each sample were merged and aligned to their respective bacterial full genome sequences using *bwa*<sup>71</sup>.

Sequence alignment files were filtered using *SAMtools*<sup>72</sup> with alignment quality (q30) parameter to remove bad alignments, and DNA sequences with length less than 30 nucleotides were discarded from the alignment file. Optical PCR duplicates were removed using *Picard-tools* with the *mark-duplicates* option. Genome coverage was calculated using *SAMtools*<sup>72</sup>. Deamination patterns were obtained by the *mapDamage* tool<sup>73</sup> with default options. User enzyme treated (damage repaired, *dr*) ble004 libraries were excluded from the analysis to calculate reliable deamination patterns. Results from these libraries are denoted with the suffix *nondr*. Finally, we generated genome authentication plots that contain edit distance distribution, length distribution, deamination patterns, and genome coverage plots. Damage repaired ble004 libraries were added in the analysis after we obtained reliable deamination patterns in non-damage repaired libraries.

Also, we co-assembled DNA reads using the *megahit* tool with the k-list parameter 21,29,39,59,79,99,119<sup>74</sup>. In essence, we de novo assembled all DNA reads that belong to one particular sample and we classified the contigs using *krakenuniq* tool with the full NT database that contains Eukaryotic full length genomes<sup>75,76</sup>. To calculate the authenticity of the contigs, we used *pyDamage* tool<sup>77</sup>. To assess the assembly quality, we used the *Quast* tool<sup>78</sup> and specified reference genomes of each oral microbe that were used in the reference-based genome assembly step. With this approach we aimed to estimate the assembled proportion of each microbe (Table S2).

### Beta-binomial modelling to find overabundant microbes in ancient pitch materials

A beta-binomial statistical model was used to detect which commensal oral microbes are overabundant in ancient mastics. We first used modern salivary microbiome datasets that contain healthy and dysbiosis cases (periodontitis and dental caries), restricted to the species that are commonly found in human oral cavities, as described in the Human Oral Microbiome Database. Then we used the beta-binomial method implemented in the *corncob*<sup>79,80</sup> R package to build differential abundance models for each bacterial species. The cases where the differences between healthy and dysbiosis are statistically significant (multiple tests corrected using Benjamini–Hochberg method with a cut-off value of 0.05) suggests that these microbes could be markers of dysbiosis.

After selecting the marker species for dysbiosis in the modern salivary microbiome, we applied the same methodology to ancient mastics. First, the sequencing depth of ancient mastic libraries was scaled so the average depth of each library matched the mean value of the healthy salivary datasets. This way all distributions have the same order of magnitude and are therefore easier to compare.

Then, a beta-binomial test was applied to each sample, and significant taxa and models were extracted. Afterwards, we searched for marker species in the list of significant species, and we restricted the species list to the species in the Human Oral Microbiome Database.

Moreover, the same methodology was used to identify the differences between pitch materials and ancient dental calculus samples that had been identified with periodontitis<sup>17</sup>.

### Machine learning to find predictors of dysbiosis

To estimate the overall health of the people who chewed the pitch material, we used a machine-learning classification method known as random forests<sup>42</sup>. Random forests aggregate predictions from different classification trees to reduce variance and to improve robustness<sup>75</sup>. In essence, the method creates several simple classification trees, each with a randomly chosen subset of features, and combines their predictions through voting.

In our case, the features correspond to the different species considered, and the data is the relative abundance of these species found in each sample. The labels used for training are the dysbiosis conditions: caries, periodontitis, or healthy. In this part, we calculated the relative abundance restricted to oral specific microbial species.

Each random forest is trained using only part of the data (i.e., bagging), and uses the remaining data to evaluate an out-of-bag error rate, which is numerically close to the N-fold classification error rate<sup>76</sup>. In other words, a low out-of-bagging error indicates that the random tree will generalise well for new data.

By evaluating the mean decrease in Gini impurity (MDG), random forests can rank each feature (i.e., species) sorting out which ones are more relevant to characterise the health condition. We used the MDG ranking to iteratively remove the least important feature, until five features remain, following Darst et al.<sup>81</sup>. In each iteration we created ten independent random forests, we averaged the MDG for each feature, and we discarded the feature with the lowest average MDG. This procedure was repeated until five features remained. With this approach, we kept only the features that contribute to the overall accuracy of the model.

This whole feature selection process was repeated three times. Finally, we sorted all the trees built in the feature selection process by their out-of-bagging error rate, and kept the best two models in each major iteration to make a model pool.

To have a simpler interpretation, we handled the periodontitis and caries cases separately, building two independent model pools.

We used the *randomForest* function of the R statistical package with default parameters to build random forests models<sup>70,82</sup>.

Multiple models trained with the same feature set were combined using the *combine* function in the *randomForest* library. This model pool was used to estimate the probability of each health condition for the ancient material, using the *predict* function over the relative abundance tables from the Mesolithic pitch materials. These probability values were then plotted using the *ggplot2* and *ggpubr* libraries on the R statistical package<sup>83,83</sup>.

## Eukaryotic taxonomic classification

To study the presence of eukaryotic species, we first removed the bacterial reads that we identified from the previous step. Then, we used *Kraken2*<sup>84</sup> to classify these reads using the full NCBI NT database that contains the all available reference sequences at that time.

We also used several confidence thresholds to filter the possible false positive classification (0.1–0.5). We manually restricted the resulting abundance table to keep only eukaryotic taxa with at least 10,000 DNA reads.

To authenticate these eukaryotic DNA reads, we extracted reference sequences for each eukaryote from the *Kraken2* database. Then we used *bwa* with custom parameters ( $L = 16,500$ ,  $N = 0.01$ ,  $O = 2$ ) to align the reads assigned to each taxa against their corresponding genome<sup>71</sup>.

After alignment, we removed nucleotides shorter than 30 bp, and validated the antiquity of the eukaryotic DNA reads using the same aDNA authentication protocol that we used for bacteria. We plotted the edit distance distribution, length distribution, and deamination plots for each set of DNA reads aligned to a eukaryotic species. Each plot was manually scored between 0 and 1 (1 for ideal and 0 for not ideal) based on the visual confirmation criteria: edit distribution should peak at low values and decline rapidly, there should be deamination patterns close to the reads' extremes, and the distribution of DNA length should be concentrated in small values. Then, the entries were sorted according to the scores and to the number of DNA reads aligned to the reference genome (Table S7–S11).

## Data and code availability

Next generation sequencing data of ancient mastic materials are available on NCBI Sequence Read Archive via PRJNA994900 Bioproject accession number. The original codes, metagenomic abundance data and final random forests models are published with <https://doi.org/10.5281/zenodo.10252965>.

Received: 23 June 2023; Accepted: 30 November 2023

Published online: 18 January 2024

## References

- Larsson, L. The mesolithic of southern scandinavia. *J. World Prehistory* **4**, 257–309 (1990).
- Berg-Hansen, I. Chapter 4. Continuity and change in Late- and Post-Glacial social networks: knowledge transmission and blade production methods in Ahrensburgian and Early Mesolithic North West Europe (2018). in (2018). doi:<https://doi.org/10.1558/equinox.30719>.
- Breivik, H. M. Palaeo-oceanographic development and human adaptive strategies in the Pleistocene-Holocene transition: a study from the Norwegian coast. *The Holocene* **24**, 1478–1490 (2014).
- Riede, F. The resettlement of northern Europe. in *Oxford Handbook of the Archaeology and Anthropology of Hunter-Gatherers* (eds. Cummings, V., Jordan, P. & Zvelebil, M.) 556–581 (Oxford University Press, 2014). doi:<https://doi.org/10.1093/oxfordhb/9780199551224.013.059>.
- Kashuba, N. *et al.* Ancient DNA from mastics solidifies connection between material culture and genetics of mesolithic hunter-gatherers in Scandinavia. *Commun. Biol.* **2**, 1–10 (2019).
- Manninen, M. A. *et al.* First encounters in the north: cultural diversity and gene flow in early mesolithic scandinavia. *Antiquity* **95**, 310–328 (2021).
- Warinner, C., Speller, C. & Collins, M. J. A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Philos. Trans. R Soc. B Biol. Sci.* **370**, 20130376 (2015).
- Jensen, T. Z. T. *et al.* A 5700 year-old human genome and oral microbiome from chewed birch pitch. *Nat. Commun.* **10**, 5520 (2019).
- Velsko, I. M. *et al.* Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. *Microbiome* **7**, 102 (2019).
- Velsko, I. M. *et al.* Ancient dental calculus preserves signatures of biofilm succession and interindividual variation independent of dental pathology. *PNAS Nexus* **1**, pgac148 (2022).
- Fellows Yates, J. A. *et al.* The evolution and changing ecology of the African hominid oral microbiome. *Proc. Natl. Acad. Sci.* **118**, e2021655118 (2021).
- Pezo, L. & Eggers, S. Caries Through Time: An Anthropological Overview. in *Contemporary Approach to Dental Caries* (ed. Li, M.-Y.) (InTech, 2012). doi:<https://doi.org/10.5772/38059>.
- 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).
- Ottoni, C. *et al.* Tracking the transition to agriculture in Southern Europe through ancient DNA analysis of dental calculus. *Proc. Natl. Acad. Sci.* **118**, e2102116118 (2021).
- Scorrano, G. *et al.* Genomic ancestry, diet and microbiomes of upper Palaeolithic hunter-gatherers from San Teodoro cave. *Commun. Biol.* **5**, 1–13 (2022).
- Weyrich, L. S. *et al.* Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* **544**, 357–361 (2017).
- Honap, T. P. *et al.* Oral metagenomes from Native American Ancestors reveal distinct microbial lineages in the pre-contact era. *Am. J. Biol. Anthropol.* **1**, (2023).
- Mann, A. E. *et al.* Do I have something in my teeth? The trouble with genetic analyses of diet from archaeological dental calculus. *Quat. Int.* doi:<https://doi.org/10.1016/j.quaint.2020.11.019> (2020).
- Modi, A. *et al.* Combined methodologies for gaining much information from ancient dental calculus: testing experimental strategies for simultaneously analysing DNA and food residues. *Archaeol. Anthropol. Sci.* **12**, 10 (2020).
- Modi, A. *et al.* Combined metagenomic and archaeobotanical analyses on human dental calculus: a cross-section of lifestyle conditions in a copper age population of central Italy. *Quat. Int.* **653–654**, 69–81 (2023).
- Boethius, A. Huseby Klev and the Quest for Pioneer Subsistence Strategies: Diversification of a Maritime Lifestyle. in *The Ecology of Early Settlement in Northern Europe - Conditions for Subsistence and Survival* 99–128 (2018). doi:<https://doi.org/10.1558/equinox.30915>.
- Granehall, L. *et al.* Metagenomic analysis of ancient dental calculus reveals unexplored diversity of oral archaeal Methanobrevibacter. *Microbiome* **9**, 197 (2021).

23. Kazarina, A. *et al.* The postmedieval Latvian oral microbiome in the context of modern dental calculus and modern dental plaque microbial profiles. *Genes* **12**, 309 (2021).
24. Mann, A. E. *et al.* Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. *Sci. Rep.* **8**, 9822 (2018).
25. Brito, I. L. *et al.* Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**, 435–439 (2016).
26. Chu, D. M. *et al.* Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat. Med.* **23**, 314–326 (2017).
27. Ghensi, P. *et al.* Strong oral plaque microbiome signatures for dental implant diseases identified by strain-resolution metagenomics. *NPJ Biofilms Microbiomes* **6**, 47 (2020).
28. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
29. Lassalle, F. *et al.* Oral microbiomes from hunter-gatherers and traditional farmers reveal shifts in commensal balance and pathogen load linked to diet. *Mol. Ecol.* **27**, 182–195 (2018).
30. David, L. A. *et al.* Gut microbial succession follows acute secretory diarrhea in humans. *mBio* **6**, 00381–00315 (2015).
31. Ferretti, P. *et al.* Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. *Cell Host Microbe* **24**, 133–145.e5 (2018).
32. Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
33. Slon, V. *et al.* Neanderthal and denisovan DNA from pleistocene sediments. *Science* **356**, 605–608 (2017).
34. Gaffney, V. *et al.* Multi-proxy characterisation of the storegga tsunami and its impact on the early holocene landscapes of the Southern North sea. *Geosciences* **10**, 270 (2020).
35. Ardelean, C. F. *et al.* Evidence of human occupation in Mexico around the Last Glacial Maximum. *Nature* **584**, 87–92 (2020).
36. Smith, O. *et al.* Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. *Science* **347**, 998–1001 (2015).
37. Key, F. M., Posth, C., Krause, J., Herbig, A. & Bos, K. I. Mining Metagenomic data sets for ancient DNA: recommended protocols for authentication. *Trends Genet.* **33**, 508–520 (2017).
38. Hübner, R. *et al.* HOPS: automated detection and authentication of pathogen DNA in archaeological remains. *Genome Biol.* **20**, 280 (2019).
39. Wessel, S. W. *et al.* Quantification and qualification of bacteria trapped in chewed gum. *PLOS ONE* **10**, e0117191 (2015).
40. Yang, F. *et al.* Characterization of saliva microbiota's functional feature based on metagenomic sequencing. *SpringerPlus* **5**, 2098 (2016).
41. Belström, D. *et al.* Metagenomic and metatranscriptomic analysis of saliva reveals disease-associated microbiota in patients with periodontitis and dental caries. *NPJ Biofilms Microbiomes* **3**, 23 (2017).
42. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
43. Knights, D. *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* **8**, 761–763 (2011).
44. Nearing, J. T. *et al.* Microbiome differential abundance methods produce different results across 38 datasets. *Nat. Commun.* **13**, 342 (2022).
45. He, J., Li, Y., Cao, Y., Xue, J. & Zhou, X. The oral microbiome diversity and its relation to human diseases. *Folia Microbiol. (Praha)* **60**, 69–80 (2015).
46. Lemos, J. *et al.* The biology of *Streptococcus mutans*. *Microbiol. Spectr.* <https://doi.org/10.1128/microbiolspec.GPP3-0051-2018> (2019).
47. Banas, J. A. & Drake, D. R. Are the mutans streptococci still considered relevant to understanding the microbial etiology of dental caries?. *BMC Oral Health* **18**, 129 (2018).
48. Hiranmayi, K. V., Sirisha, K., Ramoji Rao, M. V. & Sudhakar, P. Novel pathogens in periodontal microbiology. *J. Pharm. Bioallied Sci.* **9**, 155–163 (2017).
49. Lundström, V., Peters, R. & Riede, F. Demographic estimates from the Palaeolithic Mesolithic boundary in Scandinavia comparative benchmarks and novel insights. *Philos. Trans. R. Soc. B. Biol. Sci.* **376**, 20200037 (2021).
50. Nordqvist, B. Huseby klev. En kustboplats med bevarat organiskt material från äldsta mesolitikum till järnålder. 153.
51. Larsson, L., Liljegren, R., Magnell, O. & Ekstrom, J. Archaeo-faunal aspects of bog finds from Hassleberga, southern Scania, Sweden. in *39, pp 61–74 (2002)* vol. 39 61–74 (Jutland Archaeological Society, 2002).
52. Garcia-Escárcaga, A. & Gutiérrez-Zugasti, I. The role of shellfish in human subsistence during the Mesolithic of Atlantic Europe: an approach from meat yield estimations. *Quat. Int.* **584**, 9–19 (2021).
53. Butler, D. H., Koivisto, S., Brumfeld, V. & Shahack-Gross, R. Early Evidence for Northern Salmonid fisheries discovered using novel mineral proxies. *Sci. Rep.* **9**, 147 (2019).
54. Jacomet, S. & Vanderpe, P. The search for a needle in a haystack – New studies on plant use during the Mesolithic in southwest Central Europe. *J. Archaeol. Sci. Rep.* **41**, 103308 (2022).
55. Pryor, A. J. E. *et al.* Mobility and season of death of the Arctic foxes killed by Gravettian hunters at Kraków Spadzista. *Poland. J. Archaeol. Sci. Rep.* **33**, 102520 (2020).
56. Antonsson, K. Holocene temperatures in Bohuslän, southwest Sweden: a quantitative reconstruction from fossil pollen data. *Boreas* **36**, 400–410 (2008).
57. Aveling, E. M. & Heron, C. Chewing tar in the early Holocene: an archaeological and ethnographic evaluation. *Antiquity* **73**, 579–584 (1999).
58. Yang, D. Y., Eng, B., Wayne, J. S., Dudar, J. C. & Saunders, S. R. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol.* **105**, 539–543 (1998).
59. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
60. Lindgreen, S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes* **5**, 337 (2012).
61. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
62. Beghini, F. *et al.* Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery. *eLife* **10**, e65088 (2021).
63. Pasolli, E. *et al.* Accessible, curated metagenomic data through ExperimentHub. *Nat. Methods* **14**, 1023–1024 (2017).
64. Smith, W. C., van Brakel, W., Gillis, T., Saunderson, P. & Richardus, J. H. The missing millions: a threat to the elimination of leprosy. *PLoS Negl. Trop. Dis.* **9**, e0003658 (2015).
65. Picard toolkit. *Broad Institute, GitHub repository* (2019).
66. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–745 (2016).
67. Morgulis, A., Gertz, E. M., Schäffer, A. A. & Agarwala, R. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **13**, 1028–1040 (2006).
68. Huson, D. H. *et al.* MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLOS Comput. Biol.* **12**, e1004957 (2016).
69. Oksanen, J. *et al.* vegan: Community Ecology Package. (2022).
70. R Core Team. *R: A Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, 2021).



71. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
72. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).
73. 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).
74. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
75. Breitwieser, F. P., Baker, D. N. & Salzberg, S. L. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. *Genome Biol.* **19**, 198 (2018).
76. Pochon, Z. *et al.* aMeta: an accurate and memory-efficient ancient metagenomic profiling workflow. *Genome Biol.* **24**, 242 (2023).
77. Borry, M., Hübner, A., Rohrlach, A. B. & Warinner, C. PyDamage: automated ancient damage identification and estimation for contigs in ancient DNA de novo assembly. *PeerJ* **9**, e11845 (2021).
78. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinforma. Oxf. Engl.* **29**, 1072–1075 (2013).
79. Martin, B. D., Witten, D. & Willis, A. D. MODELING MICROBIAL ABUNDANCES AND DYSBIOSIS WITH BETA-BINOMIAL REGRESSION. *Ann. Appl. Stat.* **14**, 94–115 (2020).
80. Escapa, I. F. *et al.* New insights into human nostril microbiome from the expanded human oral microbiome database eHOMD a resource for the microbiome of the human aerodigestive tract. *mSystems* **3**, 00187–18 (2018).
81. Darst, B. F., Malecki, K. C. & Engelman, C. D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genet.* **19**, 65 (2018).
82. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R news* **2**, 5 (2002).
83. Kassambara, A. ggpubr: 'ggplot2' Based Publication Ready Plots. (2022).
84. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genom. Biol.* **15**, R46 (2014).

## Acknowledgements

The data was generated within the ATLAS project, supported by Riksbankens Jubileumsfond and the Pioneers of NW Europe project, The Research Council of Norway project no. 231305. Bioinformatics analysis was carried out using the Uppsala Multidisciplinary Centre for Advanced Computational Science cluster system project snic-2018/150-9 and Mersin University BAP project 2019-3-AP3-3729. Dr. Emrah Kırdök was supported by the The Scientific and Technological Research Council of Turkey BİDEB-2219 post-doctoral research grant at Stockholm University. Prof. Anders Götherström was supported by The Swedish Research Council project no. 2019-00849. Authors would like to thank Maja Krzewinska, and Petter Larsson for their assistance and A. Murat Eren for useful comments.

## Author contributions

Conceptualisation A.G., P.P., E.K., N.K.; Methodology, E.K. A.A., A.G.; Validation E.K., A.A.; Formal Analysis E.K., A.A.; Investigation N.K.; Resources A.G., P. P., B. N.; Writing—Original Draft E.K., A.G., B.A., N.K.; Writing—Review & Editing E.K., A.G., B.A., A.A, P.P., N.K., M.M., A.K., H.D., A.M.L, B.N., J.S, M.J.; Visualisation E.K. A.A.; Supervision A.G., P.P.; Project Administration E.K. A.G.; Funding Acquisition A.G., P.P., M.J., J.S.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-48762-6>.

**Correspondence** and requests for materials should be addressed to E.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024