

# Exploring the potential of dental calculus to shed light on past human migrations in Oceania

---

Received: 19 October 2023

---

Accepted: 25 October 2024

---

Published online: 24 November 2024

---

 Check for updates

---

Irina M. Velsko<sup>1</sup>, Zandra Fagnäs<sup>1,2,4</sup>, Monica Tromp<sup>2,3,4</sup>, Stuart Bedford<sup>5,6</sup>, Hallie R. Buckley<sup>4</sup>, Geoffrey Clark<sup>5</sup>, John Dudgeon<sup>7</sup>, James Flexner<sup>8</sup>, Jean-Christophe Galipaud<sup>9</sup>, Rebecca Kinaston<sup>10</sup>, Cecil M. Lewis Jr<sup>11</sup>, Elizabeth Matisoo-Smith<sup>4</sup>, Kathrin Nägele<sup>1</sup>, Andrew T. Ozga<sup>12</sup>, Cosimo Posth<sup>1,13,14</sup>, Adam B. Rohrlach<sup>1,15</sup>, Richard Shing<sup>16</sup>, Truman Simanjuntak<sup>17</sup>, Matthew Spriggs<sup>16,18</sup>, Anatauarii Tamarii<sup>19</sup>, Frédérique Valentin<sup>20</sup>, Edson Willie<sup>16</sup> & Christina Warinner<sup>1,21,22,23</sup> ✉

---

The Pacific islands and Island Southeast Asia have experienced multiple waves of human migrations, providing a case study for exploring the potential of ancient microbiomes to study human migration. We perform a metagenomic study of archaeological dental calculus from 102 individuals, originating from 10 Pacific islands and 1 island in Island Southeast Asia spanning ~3000 years. Oral microbiome DNA preservation in calculus is far higher than that of human DNA in archaeological bone, and comparable to that of calculus from temperate regions. Oral microbial community composition is minimally driven by time period and geography in Pacific and Island Southeast Asia calculus, but is found to be distinctive compared to calculus from Europe, Africa, and Asia. Phylogenies of individual bacterial species in Pacific and Island Southeast Asia calculus reflect geography. Archaeological dental calculus shows good preservation in tropical regions and the potential to yield information about past human migrations, complementing studies of the human genome.

Archaeogenetics studies of ancient human migrations are conventionally conducted by analysing human DNA from skeletal elements, such as teeth and the petrous portion of the temporal bone. Human population movements across the Pacific and Island Southeast Asia (ISEA) have been reconstructed in this manner, revealing successive waves of migration during the Pleistocene and Holocene<sup>1–6</sup>. However, ancient DNA work in the Pacific and ISEA is challenging as the studied populations are closely related and migrations take place over short periods of time. Further, the high temperatures and humidity in the tropics generally increase the rate of DNA decay<sup>7,8</sup>, making human DNA preservation highly variable<sup>1</sup>. Ancient dental calculus, the calcified oral biofilm preserved on teeth, offers a possible alternative approach to studying human migrations in the Pacific and culturally associated sites in ISEA, which has not yet been widely explored<sup>9,10</sup>.

Dental calculus forms when the bacterial biofilm known as dental plaque naturally calcifies on the surfaces of teeth. The dental plaque microbes become encased within the mineral matrix, preserving biomolecules including DNA, proteins, and small molecule metabolites, and this enables studies of oral bacterial communities stretching thousands of years back in time<sup>11</sup>. Because microbes have shorter generation times than humans, they offer the possibility to study the migrations of closely related populations over shorter timescales. Further, DNA within dental calculus is generally better preserved than DNA in skeletal tissues from the same individual<sup>12</sup>, making it a promising substrate to study in areas where skeletal DNA preservation is generally poor, such as the tropics.

The prospect of studying human migration through archaeological dental calculus presents several potential advantages, as it would allow

---

A full list of affiliations appears at the end of the paper. ✉ e-mail: [warinner@fas.harvard.edu](mailto:warinner@fas.harvard.edu)

for a holistic study of an individual's life through just one sample - from migration and diet, to health and disease, and even occupational activities<sup>13</sup>. Dental calculus may prove to be an especially valuable study material in the Pacific, given the speed of settlement of the region and significant cultural changes over a short time. However, no study has to date attempted to investigate past human migrations through the oral microbiome, although the prospect has been discussed<sup>9,10</sup>.

Here, we explore the possibility of using archaeological dental calculus to study past human migrations, using Pacific and ISEA islands as a case study. Shotgun metagenomic sequencing was performed on a total of 102 dental calculus samples from 10 Pacific islands and 1 island in ISEA, spanning a time range of nearly 3000 years (Table 1, Supplementary Data 1, 2). We show that DNA preservation is variable, but that most samples have a well-preserved oral microbiome. This highlights the exceptional preservation of DNA in dental calculus, even in challenging environments. We find that variation in dental calculus microbial community composition does not have a clear geographic or temporal structure, but rather may be influenced by local factors specific to each island. In contrast, phylogenetic analyses of individual oral bacterial taxa exhibit temporal trends, but the ability to detect this signal depends on the selected species' prevalence and abundance. Overall, we find that metagenomic analysis of archaeological dental calculus has the potential to reveal information about past human migration patterns. When combined with human DNA analysis and other approaches, such as paleodietary studies using palaeoproteomics and microremains, it promises to enrich our understanding of the dynamic biological and cultural processes that accompanied past migrations across ISEA and the Pacific.

## Results

### Preservation of endogenous ancient DNA

Obtaining well-preserved ancient DNA is a persistent challenge in archaeogenetic studies of the tropics, as both temperature and humidity contribute to DNA decay<sup>7,8</sup>. Using SourceTracker analysis<sup>14</sup>, we estimated the proportion of microbial taxa originating from endogenous and contaminant sources for each dental calculus sample in this study (Fig. 1A). For 73 of the 102 archaeological dental calculus samples, at least 50% of microbial content is estimated to originate

from an oral microbiome source, indicating good preservation of the dental calculus for these samples, with minimal contamination from exogenous sources. We further assessed the proportion of endogenous oral species using cuperdec<sup>11</sup> (Supplementary Fig. 1) and found high consistency between estimated sample preservation using both methods. A PCA of well-preserved samples shows minimal overlap between dental calculus and the source samples used in SourceTracker, with calculus samples clustering distinctly from all sources (Fig. 1B). All samples that passed the cuperdec threshold for preservation, 73 samples, were carried forward for analysis (Fig. 1C).

The effect of sample age and environmental variables (average rainfall, average temperature, and average evapotranspiration of each island) on the proportion of taxa that could be assigned to oral sources (dental plaque and dental calculus), as well as the proportion of human DNA recovered from calculus, was investigated using beta regression (Fig. 1D). Only sample age was a significant predictor of preservation ( $p=0.046$ ,  $R^2=0.083$ ), yet the low  $R^2$  value indicates that additional, unknown factors more strongly affect preservation. Local environmental factors at the burial site and for each individual grave, such as soil type, humidity, and pH, are plausible candidates, but such data are not available.

Lastly, because the environment of the Pacific is less conducive to DNA preservation than colder, drier climates, we compared preservation patterns between the Pacific/ISEA and northern Europe, and we asked whether the relative preservation of endogenous oral microbiome DNA in dental calculus is greater than that of human DNA in archaeological bone/teeth (Fig. 1E). For the analysis, we considered all oral microbial DNA identified in dental calculus to be endogenous, and all human DNA present in bone/teeth to be endogenous. Overall, we find that endogenous preservation of ancient dental calculus in the Pacific/ISEA is high, in which 72/102 samples (70%) are estimated to derive 80% or more of their composition from an oral microbiome source. This is only somewhat lower than that estimated for dental calculus from northern Europe, where 106/112 samples (95%) are estimated to derive more than 80% of their composition from an oral microbiome source. In contrast, human DNA preservation in skeletal material was much lower for both the Pacific and northern Europe, with 5/183 (2.7%) and 19/464 (4.1%) samples having endogenous DNA content of at least 5%, respectively, for samples in which ancient human DNA was detected. These results suggest that DNA within calculus is less prone to degradation and exogenous contamination than DNA in archaeological skeletal remains and may be a more promising avenue of study for sites in warmer, more humid climates.

### Microbial community composition

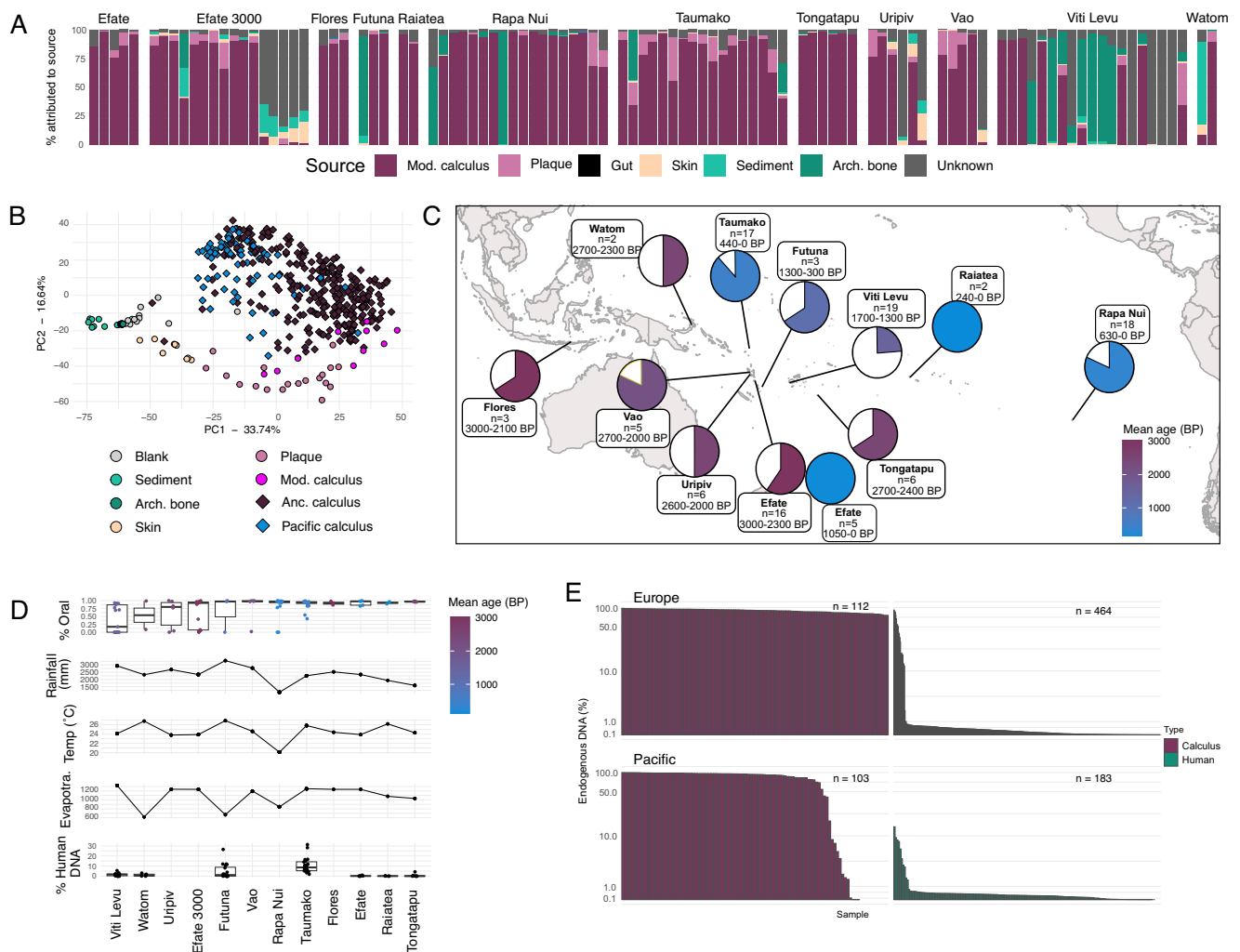
We next asked whether the microbial community of well-preserved Pacific/ISEA calculus samples showed temporal or spatial differences. If present, such patterns might suggest that the calculus microbiome changed with human migration through ISEA and the Pacific, although specific factors affecting such changes would need elucidation. We performed a beta-diversity analysis and visualized the samples using PCA (Fig. 2A) and tested whether multiple clusters were present in the data. Testing the goodness of fit of sample cluster numbers to the data indicated that a single cluster optimally described the data (Supplementary Fig. 2). This corresponded with the visual lack of distinct sample clustering in the PCA plot, in which the samples did not tightly cluster based on island or time period, nor did they plot along a cline that might suggest temporal or geographic change. PERMANOVA determined that the processing lab ( $R^2=0.0275$ ,  $F=2.748$ ,  $p=0.01$ ), island ( $R^2=0.1913$ ,  $F=1.736$ ,  $p=0.001$ ), and average GC content ( $R^2=0.0508$ ,  $F=5.068$ ,  $p=0.002$ ) were the most influential factors in how the samples plotted. When controlling for the lab in which samples were processed, island ( $R^2=0.2045$ ,  $F=1.650$ ,  $p=0.035$ ) and average GC content ( $R^2=0.0761$ ,  $F=7.372$ ,  $p=0.01$ ) were still significant drivers of the sample community composition.

**Table 1 | Archaeological dental calculus samples included in this study**

Island	Island group/Area	# Samples (# passed)	Age BP <sup>a</sup>	Processing lab
Efate	Vanuatu, Pacific	5 (5)	1050–0	Jena
Efate	Vanuatu, Pacific	16 (10)	3000–2300	Otago
Flores	Lesser Sunda Islands, ISEA	3 (3)	3000–2100	Oklahoma
Futuna	Vanuatu, Pacific	3 (2)	1270–290	Jena
Raiatea	Society Islands, Pacific	2 (2)	240–0	Jena
Rapa Nui	Rapa Nui, Pacific	18 (15)	630–0	Oklahoma
Taumako	Duff Islands, Pacific	17 (16)	440–0	Jena
Tongatapu	Tonga, Pacific	6 (5)	2700–2400	Jena
Uripiv	Vanuatu, Pacific	6 (3)	2600–2000	Otago
Vao	Vanuatu, Pacific	5 (4)	2750–2000	Otago
Viti Levu	Fiji, Pacific	6 + 13 (4 + 2) <sup>b</sup>	1700–1300	Oklahoma + Jena
Watom	Bismarck Archipelago, Pacific	2 (1)	2700–2300	Otago

<sup>a</sup>Age BP is the estimated age range for the individuals by site. See Supplementary Data 1 for details.

<sup>b</sup>Samples from Viti Levu were extracted in two labs, either in Oklahoma or in Jena. Of the 6 samples extracted in Oklahoma, 4 were well-preserved, while of the 13 extracted in Jena, only 2 were well-preserved.



**Fig. 1 | Preservation assessment of dental calculus samples.** **A** SourceTracker analysis of species tables. Each bar represents a sample, coloured by the proportion of each contributing source (sources are the same as in (B)). **B** PCA of well-preserved calculus from the Pacific islands (blue) with samples from the same sources used in SourceTracker analysis plus additional ancient calculus. **C** Map of the islands from which samples were collected for this study, with the island name, number of collected samples, and age of the site. Pie charts indicate the fraction of samples from each site that were considered well-preserved, and are coloured by age of the site. Map was produced using the R packages ggmap, plotly, and sf. **D** Comparison of endogenous oral bacterial DNA and human host DNA in samples with environmental conditions on the islands that may affect DNA preservation. No associations between preservation of either oral bacterial DNA or human host DNA

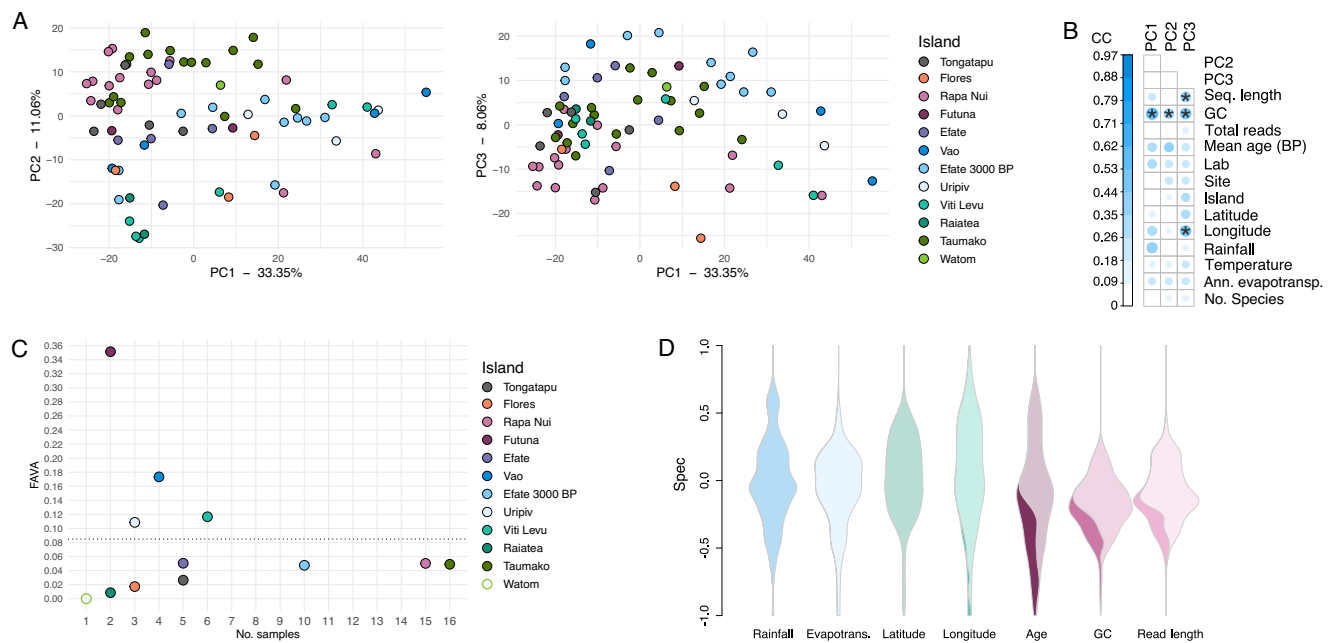
were found with any of the measured environmental factors by ANOVA; values for groups with < 3 samples should be considered unreliable. Boxes show data median, interquartile range (25th–75th percentile) and whiskers indicate minimum and maximum values. **E** Comparison between the Pacific/ISEA and northern Europe (England/the Netherlands) of the percentage of endogenous oral microbiome DNA in ancient dental calculus and human DNA in bones/teeth. Scale is logarithmic. The number of samples in each group is indicated next to the bars. Ancient dental calculus contains high levels of endogenous oral bacterial DNA in the Pacific/ISEA similar to that seen in northern Europe, in contrast to the lower levels of preserved human DNA in bones/teeth from the Pacific islands compared to northern Europe. Arch. bone - archaeological bone; Anc. calculus - ancient calculus; Mod. calculus - modern calculus.

The average GC content of calculus is known to increase with sample age<sup>12,15,16</sup> through the taphonomic loss of AT-rich DNA fragments, and most islands are represented by samples from a single time period, therefore tying island and age/GC content. However, the extent to which loss of AT-rich fragments affects species profiles, perhaps skewing older samples to have higher proportions of high-GC taxa, has not yet been extensively explored. We found that the species with the strongest PC1 loadings (Supplementary Data 3) suggested a taxonomic gradient relating to oxygen tolerance, such as that previously described in calculus samples from the archaeological site of Middenbeemster in the Netherlands<sup>16</sup>, but there was no clear association with GC content.

Low taxonomic assignment rates for the Pacific/ISEA samples and the abundance of non-typical oral taxa that predominantly drive separation along PC1 suggest that taxonomic diversity in these

samples may not be represented in the genomic databases we used for classification. We performed further taxonomic profiling with additional tools and databases (Kraken2 with two customized RefSeq-based databases, MetaPhlan3), but were unable to resolve these issues, highlighting the difficulty of disentangling undescribed diversity from intrinsic biases in ancient calculus datasets, such as very short average DNA read lengths ( $\leq 70$  bp) (Supplementary Figs. 3, 4).

We performed canonical correlation analysis between PC loadings, environmental metadata, and laboratory characteristics, to determine if preservation may be influencing the calculus community composition (Fig. 2B). Sample loadings on PC1 and PC2 were significantly correlated with average GC content, but not with sample metadata. We further tested how much variance exists in the microbial community between samples all together as well as separated by island, to see whether the heterogeneity of microbial composition,



**Fig. 2 | Community species profiles are minimally structured by island.** **A** PCA of well-preserved calculus samples, coloured by the island from which samples were collected; PCs 1, 2, and 3 are shown, accounting for > 50% of the variation in the dataset. **B** Canonical correlation (CC) analysis comparing the positions of calculus samples in the PCA shown in (A) to environmental and laboratory metadata. Pearson two-sided correlation tests were used to determine if the correlations were significant. Metadata with a  $p \leq 0.01$  and CC value  $\geq 0.4$  are marked with an asterisk (\*). Ann. evapotransp. - annual evapotranspiration; No. species - number of species. **C** FAVA values of variance in microbial composition between samples grouped by

island. Dotted line indicates the FAVA value of all samples not separated by island. There is only one sample from Watom, so FAVA could not be calculated for this island, and the dot is left unfilled to indicate that FAVA is NA rather than 0. **D** Specificity of species in calculus samples to island conditions (blue), island location (green), or sequenced library characteristics (red). Dark colours in the violin plots indicate the proportion of species that are significantly associated with that metadata type. No dark colour indicates that no species were specifically associated with that metadata.

rather than the dissimilarity measured by PCA, is high. High variability across samples in a group might indicate a community transitioning from one homeostatic balanced state to another due to a perturbation, potentially indicating a recent community structural shift. Using the R package FAVA<sup>17</sup>, we calculated the FST-based Assessment of Variability across vectors of relative Abundances (FAVA) for each island as well as all islands together (Fig. 2C). FAVA values range from 0, indicating identical variance across samples, to 1, indicating maximal variability with each sample having only a single species. The FAVA values are all close to 0, indicating a similar variance in composition within and across islands. However, we note that the islands with smaller numbers of samples have a wider range of FAVA values, while the FAVA value stabilizes above 10 samples, so larger sample numbers for each island are needed to confirm that variability is similar in islands with fewer samples.

We additionally looked for particular species that were significantly associated with environmental and laboratory metadata using the R package specificity (Fig. 2D). No species were significantly associated with rainfall, evapotranspiration, or latitude, while a few were associated with longitude. In contrast, numerous species were significantly associated with sample age, average GC content, and average read length, which are themselves correlated, with some species significantly associated with more than one of these conditions. These results suggest that the environmental conditions tested here have minimal influence on the reconstructed calculus microbiome community, and instead other unexplored factors may be more influential.

### Comparison with global ancient calculus microbiome profiles

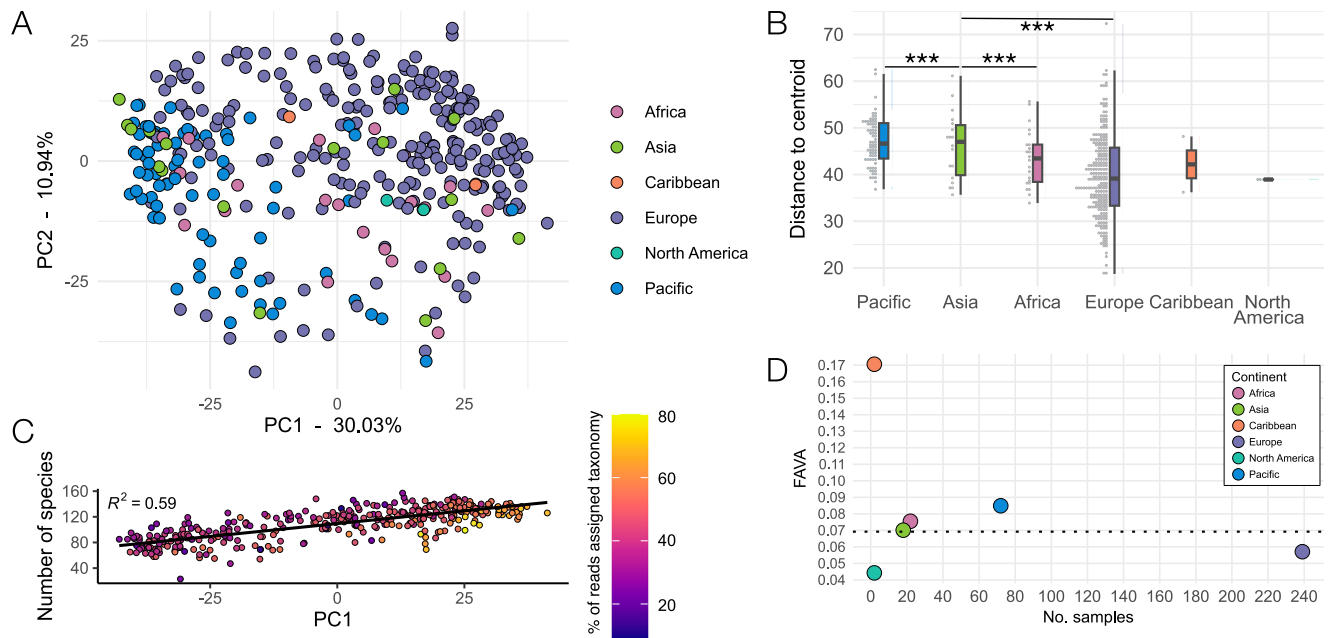
As this is the first large ancient dental calculus dataset published from the Pacific/ISEA, we next wanted to know whether the microbiome communities fall within the known variation of published dental

calculus datasets from across the globe. We compared the species profiles of the Pacific/ISEA calculus samples to those from Europe, Africa, and Asia using PCA (Fig. 3A). The Pacific/ISEA samples and those from Japan<sup>18</sup> cluster at the same end of the plot, suggesting that this region may have a particular species profile characteristic. Sample clustering based on continent/region (hereafter “continent”) was significant by PERMANOVA ( $p < 0.01$ ,  $F = 2.47$ ,  $R^2 = 0.03114$ ), while tests of beta-dispersion between the continents found differences between Asia and the Pacific/ISEA, Africa, and Europe (Fig. 3B,  $p < 0.001$ ). However, the large difference in sample numbers for each continent make these comparisons less reliable.

The distribution of samples across PC1 in the PCA may be largely driven by the number of species in each sample, as there is a moderate correlation between PC1 value and species counts (Fig. 3C). We confirmed that there is limited variance in the microbial composition of samples within each continent and across all continents by calculating the FAVA for each (Fig. 3D). The values are all close to 0, indicating a similar variance in composition within and across continents, which supports that the calculus microbiome community is largely stable across time and geography. The lower average number of species detected in Pacific/ISEA samples compared to other continents may be related to the lower average number of reads in each sample that could be assigned taxonomy (Supplementary Fig. 5), further hinting at unexplored diversity in the Pacific/ISEA calculus.

### Gene content

We next investigated whether microbial gene content distinguishes the Pacific/ISEA calculus samples from those of other continents. Hierarchical clustering of the KEGG orthologs (KOs) detected in samples did not cluster samples by continent or sample age (Fig. 4A), the processing lab, or the study in which they were first presented (Supplementary Figs. 6, 7). Using hierarchical clustering, the samples form



**Fig. 3 | Situating the Pacific/ISEA calculus samples within known ancient dental calculus microbial diversity.** **A** PCA of Pacific/ISEA calculus samples with ancient dental calculus from additional geographic regions. The additional samples are the same as those in Fig. 1B. **B** The distance to the centroid of all samples in the PCA, test by ANOVA \*\*\*  $p = 0.001$ , with Tukey's Honest Significant Differences; values for groups with < 3 samples should be considered unreliable. Boxes show data median,

interquartile range (25th-75th percentile) and whiskers indicate minimum and maximum values. **C** The number of species in each sample, ordered by PC1 loading, and coloured by the percentage of total reads in the sample that were assigned taxonomy. The trend line is fit with a generalized linear model. **D** FAVA values of variance in microbial composition between samples grouped by continent. Dotted line indicates the FAVA value of all samples not separated by continent.

two clusters (Clusters A and B) that correspond to their placement along PC1 in a PCA based on KO abundance (Supplementary Fig. S8A) and based on species abundance (Supplementary Fig. 8B), which is loosely correlated with the number of detectable KOs and species in each sample (Supplementary Fig. S8C). Most of the Pacific/ISEA samples fall in Sample Cluster A, which on average has lower species and KO counts (Supplementary Fig. 8D, E), possibly indicating these samples have more unidentified taxa, with corresponding unannotated genes/metabolic functions.

The KOs formed three clusters (Clusters 1, 2, and 3), and Sample Cluster B was enriched in KOs from Cluster 1 and depleted in KOs from Cluster 3. We grouped the KOs in Clusters 1 and 3 by Pathway and found that one pathway had significantly more KOs in Cluster 1 than 3: Protein families: genetic information processing (Fig. 4B). Given the broadly general cellular processing categories included in this pathway, we assessed the genera that were contributing the orthologs in these pathways to see if we could glean more microbially-relevant information. We found that a high proportion of numerous orthologs were attributed to *Ottowia* (Fig. 4C), specifically *Ottowia* sp. oral taxon 894, a poorly characterized species. The samples in Sample Cluster B have on average higher proportions of *Ottowia* sp. oral taxon 894 than those in Sample Cluster A (Fig. 4C), perhaps indicating a difference in the biofilm environment of these two clusters. However, the presence and abundance of this genus does not appear to drive sample loading in PCA, as the plot structure remains largely unchanged after filtering out *Ottowia* from the input table (Supplementary Fig. 7). Overall, *Ottowia* is the most prevalent genus enriched in KO Cluster 1 that is contributing sample Cluster B (Fig. 4D).

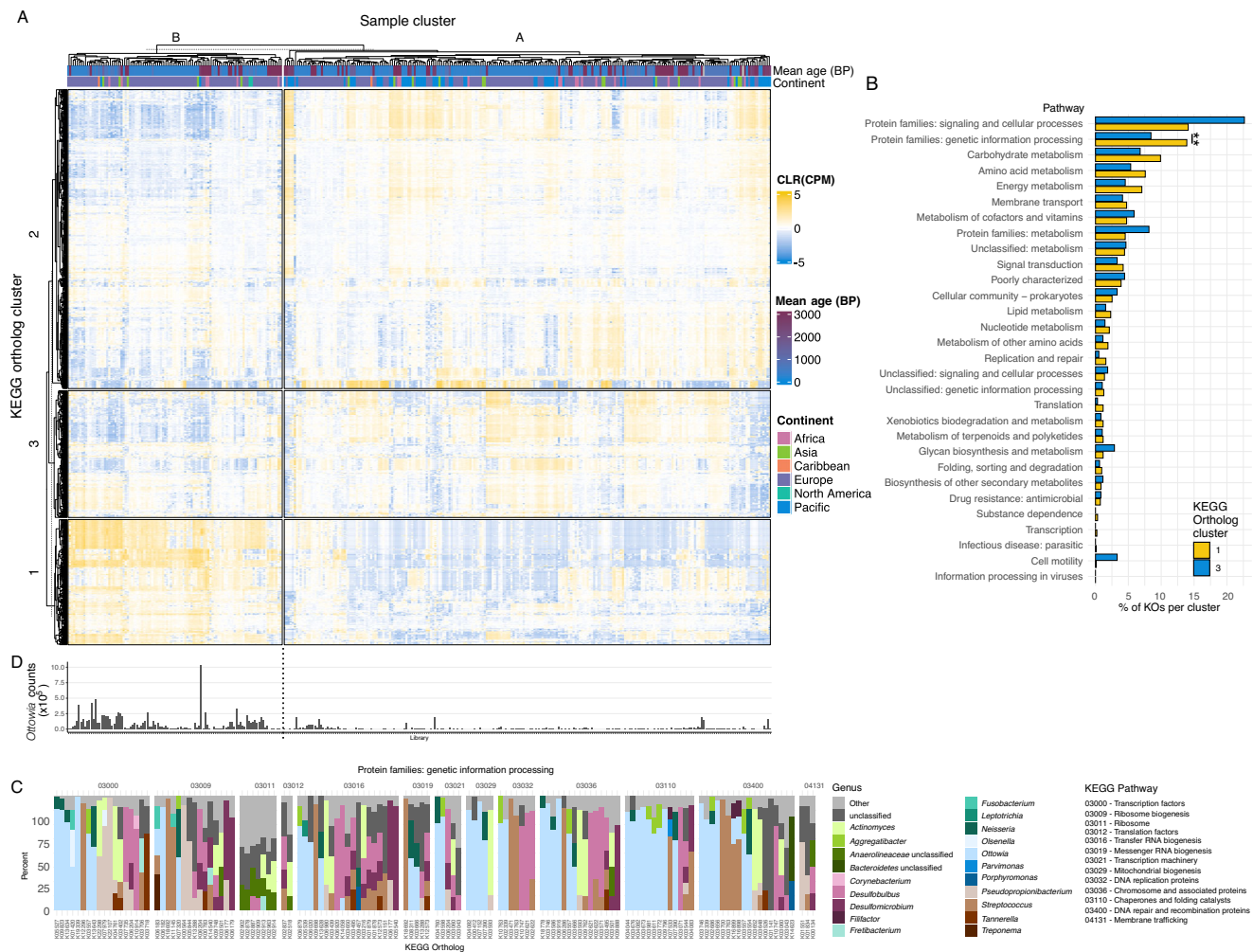
### Phylogenetic analyses

Phylogenetic trees were constructed for *Tannerella forsythia* and *Anaerolineaceae* bacterium oral taxon 439, as both have previously been studied phylogenetically in archaeological dental calculus in relation to human migration<sup>18-21</sup>. In both phylogenetic trees, samples from the same islands generally cluster together (Fig. 5), which was

consistent across multiple tree-building methods (Kendall's coefficient of concordance (W) 0.91-0.96 for *Anaerolineaceae* bacterium oral taxon 439 and 0.83-0.98 for *T. forsythia*) (Supplementary Figs. 9, 10). We found little evidence of recombination in either the *T. forsythia* or *Anaerolineaceae* bacterium oral taxon 439 alignments using Gubbins (Supplementary Data 6, 7, Supplementary Fig. 11), and the branches of a ML tree built from the masked SNP alignment was generally concordant with the branches of a ML tree built from an unmasked SNP alignment (Supplementary Figs. 9, 10), but showed greater differences for *Anaerolineaceae* bacterium oral taxon 439.

The percentage of multiallelic SNPs is generally < 20% for *T. forsythia* (Supplementary Data S4), indicating that the reference strain used for mapping is closely related to the strains present in the samples. For *Anaerolineaceae* bacterium oral taxon 439, however, the number of sites with multiallelic SNPs is much higher (Supplementary Data S5), indicating that the reference genome may be quite distinct from the strains present in the samples, and reads from several strains or species in each sample may be aligning to this reference genome. Within a cluster of samples, there is a tendency for samples with higher levels of heterozygosity to fall basal to other samples (e.g., the cluster of samples from Taumako) in both the NJ and ML trees (Fig. 5, Supplementary Figs. 9, 10). It is likely that the *Anaerolineaceae* bacterium oral taxon 439 phylogeny does not represent that of a single taxon, but rather a collection of closely related species or strains, which adds uncertainty to the tree topology. At present, however, only one reference genome is available for oral bacteria in the family *Anaerolineaceae*, making further within-sample taxonomic disambiguation challenging.

We additionally ran inStrain to test the ability to identify shared strains of *Anaerolineaceae* bacterium oral taxon 439 and *T. forsythia* among our samples. Following testing of inStrain with in silico-generated datasets to understand how ancient DNA damage patterns affect the strain assessments (Supplementary Figs. 12, 13), we found identical strains in samples AMH001 and AMH004 (popANI > 99.999). All other samples had popANI values  $\leq 99.99$ , indicating shared, but not identical, strains (Supplementary Fig. 14). The similarity



**Fig. 4 | KEGG ortholog (KO) enrichment is associated with sample species composition.** **A** Heatmap of clustered samples (clusters **A** and **B**) and KOs (clusters 1, 2, and 3), showing CLR-transformed copies per million (CPM) for each ortholog. **B** Percent of KOs in each KO cluster from **(A)** in KEGG pathways present in all samples. Two-sided Wilcoxon test with FDR correction, \*\*  $p = 0.00272$ . **C** Mean

percent contribution by genera of KOs enriched in KO Cluster 1, which are enriched in sample Cluster B. *Ottovia* is the most prevalent genus contributing to these KOs. **D** Read counts of *Ottovia* in all samples, aligned with panel **(A)**, showing a higher percentage of samples in cluster B have higher *Ottovia* read counts than in cluster A.

of *Anaerolineaceae* bacterium oral taxon 439 strains (popANI > 99.99%) in samples from the same islands, including Viti Levu (SIG), Taumako (NMU), and Rapa Nui (A\*), is reflected in the close phylogenetic clustering of these samples. No identical *T. forsythia* strains were found between any samples (Supplementary Fig. 15), but closely related strains (popANI > 99.91%) were found in samples from Rapa Nui (A\*).

As an additional test for whether there are multiple closely-related species or strains of *Anaerolineaceae* bacterium oral taxon 439 in our samples, we calculated the polymorphic rate over protein-coding genes<sup>22</sup> for the reference genomes of *Anaerolineaceae* bacterium oral taxon 439 and *Tannerella forsythia*. The dN/dS values of samples mapped against the *Anaerolineaceae* bacterium oral taxon 439 genome are on average lower than for *T. forsythia*, and fall below the estimated value for mapping to an incorrect reference genome (Supplementary Figs. 16, 17), indicating multiple strains of this species are likely present, while this is less likely the case for *T. forsythia*.

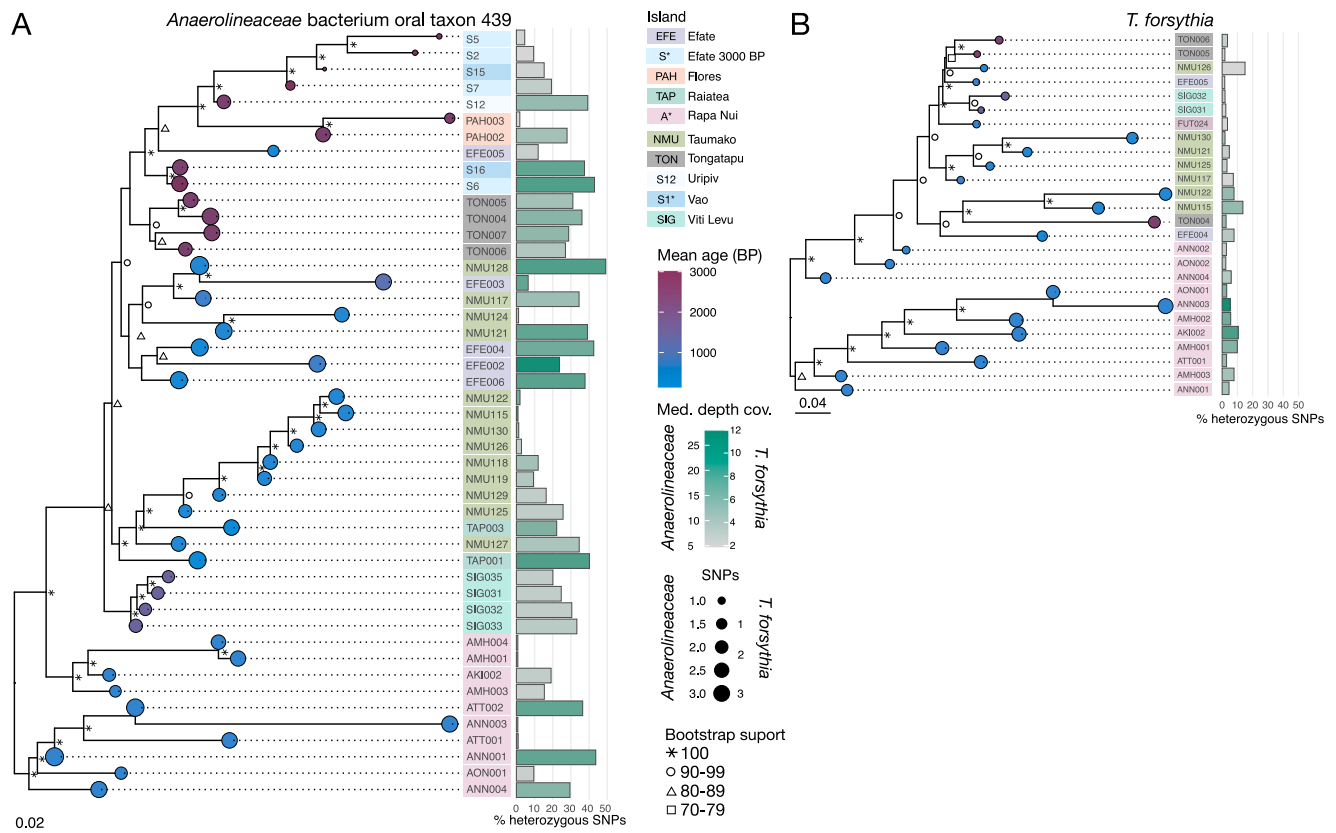
### Dietary DNA

In addition to tracing the microbial changes in dental calculus across the Pacific islands, we sought to examine the potential of recovering eukaryotic, food-derived DNA that may offer insight into dietary patterns across these sites. We were unable to identify any unambiguously

positive evidence for dietary DNA in our samples, which may be due to the low number of non-microbial DNA sequences that were recovered. Alternatively, the sequences may be modern contaminants, or they may be aligning to an inaccurate reference genome<sup>23</sup> (Supplementary Data 8).

### Microparticles

To gain further insights into potential dietary patterns in the Pacific, we examined the microparticle content of the dental calculus samples analysed in this study (Supplementary Figs. 18, 19, 20, 21, 22, and Supplementary Data 9). The microparticle results from Rapa Nui<sup>24,25</sup> and Teouma<sup>26</sup> were previously published elsewhere. Overall, the samples had low microparticle counts compared to other dental calculus examined in the Pacific. Almost all samples contained fungal spores and hyphae, likely from sediment. Some starch granules were observed in the samples, but they could not be distinguished from common manufacturing contaminants associated with gloves and laboratory consumables. Phytoliths and diatoms of likely dietary origin were present, but at low levels. This may indicate a greater reliance on starchy root crops and/or processing of plant foods than at Teouma where there were abundant phytoliths recovered from dental calculus<sup>26</sup>. Compared to the published data from Rapa Nui<sup>24,25</sup>, where numerous diatoms were found, the data presented here suggests



**Fig. 5 | Phylogenetic trees show that bacterial genomes from the same island resemble each other. A** A neighbour-joining tree of *Anaerolineaceae* bacterium oral taxon 439, including only samples with >5X genomic coverage of the taxon and using only homozygous SNPs, with midpoint rooting. **B** A neighbour-joining tree of *Tannerella forsythia* from samples with >2X genomic coverage, using only homozygous SNPs with midpoint rooting. For both trees, the age of the sample (in

years BP) is shown as coloured circles on tree tips, the circle diameter indicates the number of SNPs ( $\times 10,000$ ) in that sample, the island of origin is indicated by a coloured box behind sample IDs, the percentage of heterozygous SNPs is shown as a bar, and the mean coverage of the genome as the colour of the bar. Scale bar indicates the genetic distance.

better freshwater access on these islands than on Rapa Nui. Dietary microparticles were too sparse to draw further conclusions.

## Discussion

Here we show that archaeological dental calculus samples from islands across the Pacific preserve a high proportion of DNA from endogenous oral microbiota, despite a climate that is unfavourable to DNA preservation. This allowed us to assess the diversity of ancient oral microbiomes in an understudied region, to put them in context on a global scale, and to explore the potential of the oral microbiome for tracing human migration through ISEA and the Pacific. While we did not observe temporal or geographic patterns of microbiome species composition in samples from across the Pacific/ISEA or across the globe, we observed a distinct community structure in the Pacific/ISEA calculus compared to samples from Europe spanning a similar time period, suggesting the presence of undescribed microbial diversity in Pacific/ISEA dental calculus oral microbiomes.

Considering the high temperatures and humidity in ISEA and the Pacific, and variable success in human DNA extraction from skeletal elements from the region<sup>1</sup>, the overall high preservation of DNA in dental calculus from the Pacific was an unexpected success. These results provide further support to the exceptional preservation of biomolecules in dental calculus<sup>12</sup> compared to bones and teeth. We found that none of the climatic variables we tested predicted preservation, which suggests that smaller-scale local factors, such as soil biogeochemistry, water exposure, or the microclimate of the burial site, may be more influential. Further studies are needed to investigate local factors that may contribute to preservation, which could help

explain why calculus preservation is high even in regions with conditions unfavourable to DNA preservation in skeletal remains.

Given the high level of DNA preservation in ancient dental calculus in the Pacific, we sought to explore the potential of calculus microbial DNA for tracking human migration or other behavioural or cultural changes associated with island colonization. Investigating correlations of host genetic background with community composition or strain sharing, such as by using estimated ancestry proportions for each individual, was not possible in this study, as we did not have enough calculus samples from individuals with paired human genetic data. This is, however, an exciting avenue to explore in future studies. We had limited success identifying differences in the microbial community composition related to island of origin or time period. This is in line with other studies to date, which have not found substantial microbial differences at the community-wide level related to geography, time period, or oral health<sup>11,16,27–29</sup>, indicating that the species community composition is relatively stable throughout human history. This is supported by studies of modern oral microbiomes, which indicate that perturbations in the community of dental plaque microbiomes are quickly rebalanced<sup>30</sup> and are generally stable across a variety of cultural practices<sup>31–34</sup>, even when gut microbiomes from the same communities under comparison are substantially different.

However, there is growing evidence that the oral species present in ancient dental calculus are not fully represented in current genomic databases, such as NCBI, used for taxonomic profiling. Because of this, community composition analyses may be missing taxa and underestimating diversity, and therefore signals of differences or change may be hidden. The Pacific calculus dataset appears to be particularly

affected by taxonomic database bias, as it has a notably lower taxonomic assignment rate compared to European ancient dental calculus samples for which the distributions of read length and GC content overlap.

Despite the lack of signal at the microbial community level, several studies have demonstrated the utility of phylogenetic reconstruction of abundant species in dental calculus to trace their evolutionary history, including *Anaerolineaceae* bacterium oral taxon 439 and *Tannerella forsythia* in relation to human migration<sup>11,18–21</sup>. Our own phylogenetic reconstructions of these species did not show clear temporal or latitudinal/longitudinal patterning; however, the reconstructed genomes often clustered by island, suggesting that oral strains are more similar within an island than between islands, which was also supported by independent strain identification with inStrain. In phylogenetic trees for both *T. forsythia* and *Anaerolineaceae* bacterium oral taxon 439, the samples from Rapa Nui, the most remote island in this study, are in two clusters that fall on either side of the midpoint root. This pattern is similar to that of *Anaerolineaceae* bacterium taxon 439 and *T. forsythia* presented by Honap, et al.<sup>21</sup>, and suggests that there are 2 distinct lineages of each species present in these samples. Alternatively, these lineages may instead be different species from those of the genomes used as the reference for mapping. However, due to the lack of additional reference genomes for oral *Anaerolineaceae* and *Tannerella*, this possibility is currently difficult to explore.

An outstanding challenge to reconstructing past microbial genomes from metagenomes is distinguishing multiple closely related species and strains within a metagenome. Attempts to study migration patterns through the microbiome, therefore, come with a degree of inherent uncertainty when attributing microbial DNA to particular species and strains. Due to issues such as contaminated (“dirty”) reference genomes<sup>35</sup>, which include sequences not derived from the species of interest, or incomplete databases<sup>36</sup>, it is possible that sequencing reads from multiple species are inappropriately aligned to a reference genome<sup>23,35</sup>, creating noise in the data analysis. Therefore, the reliability of conclusions regarding human migration drawn from trees built using our current approaches is uncertain.

This issue appears to have particularly affected our reconstruction of *Anaerolineaceae* bacterium oral taxon 439, for which we observed high rates of SNP heterozygosity. As there is only a single isolate reference genome of *Anaerolineaceae* bacterium oral taxon 439 sequenced to date, the extent of diversity in this organism both past and present is yet unknown. The high rate of SNP heterozygosity in many samples mapped to this reference genome appeared to affect the branching pattern within clades, despite inclusion of only biallelic SNPs in the alignment used to build our tree. Future sequencing of additional isolates of this species, or reconstruction of metagenome-assembled genomes (MAGs) of this organism from deeply sequenced modern and ancient oral metagenomes<sup>37</sup>, may lead to more accurate strain separation, read alignment, SNP calling, and phylogenetic tree reconstruction. As the study of migratory patterns through ancient host-associated microbiomes is still in its infancy, method development will be fundamental in order to explore the full potential of this field.

Our results indicate the high potential of dental calculus to be well-preserved in geographic and climatic conditions that are otherwise unfavorable to DNA preservation, opening the possibility to explore archaeogenetics data in formerly poorly-accessible locations. Although we did not observe the microbial community composition of the calculus microbiome structuring by island or time period, the low taxonomic assignment rate suggests that there is additional taxonomic diversity in the Pacific calculus samples beyond that currently represented in databases, highlighting the need for studies of dental plaque biodiversity in broad, global contexts. Individual species reconstructions have the potential to reveal evolutionary patterns that mirror the migration patterns of their human hosts, but further work

disentangling closely related species and strains within ancient dental calculus metagenomes, as well as revealing currently unknown species diversity, is needed to allow accurate identification of individual species and to perform reliable phylogenetic reconstructions.

## Methods

### Ethics & inclusion statement

Archaeological research was carried out in close consultation with local communities and in partnership with local cultural councils, museums, and research institutions. Local researchers contributed to the study as co-authors. Export of samples for analysis was approved as part of the permissions processes described below.

**Efate, Vanuatu.** The Vanuatu Cultural Centre (VCC) and the Vanuatu National Cultural Council provided ethical oversight of the study of human remains from Efate. In addition, the leaseholder M.R. Monvoisin and family and the traditional landowners and population of Eratap Village provided support for the study of remains on their ancestral lands. Monica Tromp entered into a research agreement in 2013 with the Vanuatu Cultural Centre to conduct research on human dental calculus from multiple archaeological sites in Vanuatu. The permit number is 10071.

**Flores, Lesser Sunda Islands, Indonesia.** The Pain Haka cemetery on the island of Flores in Indonesia was discovered and excavated by archaeologist Jean-Christophe Galipaud and anthropologists Charles Illouz and Philippe Grangé in 2010 (Institute for Research for Development [IRD] and University of La Rochelle) under a research permit from the Province of Nusa Tenggara Timur. Further excavation and sampling in 2012 were made possible by a joint collaboration between IRD, Puslit Arkenas, Hallie R. Buckley, and University of Otago. Export permits were obtained from Puslit Arkenas and RISTEK (Foreign Research Permit Division, Ministry of Research and Technology/National Research and Innovation Agency). Local communities participated at all stages of the research and gave permission for the removal of samples for analysis. All samples were collected and prepared by Rebecca Kinaston and exported to Otago University for management.

**Futuna, Vanuatu.** The Vanuatu Cultural Centre (VCC) and the Vanuatu National Cultural Council provided ethical oversight of the study of human remains from Futuna. Monica Tromp entered into a research agreement in 2013 with the Vanuatu Cultural Centre to conduct research on human dental calculus from multiple archaeological sites in Vanuatu. The permit number is 10071.

**Raiatea, Society Islands.** Excavations at Taputapuatea on the island of Raiatea were undertaken in 1994 and 1995 by archaeologists of the Centre Polynésien des Sciences Humaines (Tahiti, Polynésie française), and the export permit for subsequent analysis was issued by Service de la Culture et du Patrimoine (Tahiti, Polynésie française) in 2014. More recent communication (March 2020) with Director of Direction de la Culture et du Patrimoine (Tahiti, Polynésie française), granted permission for human dental calculus analyses.

**Rapa Nui, Chile.** Archaeological remains were sampled in 2002 as part of John Dudgeon’s (2008) dissertation research from collections excavated during the National Geographic Easter Island Anthropological Expedition. Led by George Gill of the University of Wyoming, Sergio Rapu, former curator of the Sebastian Englert Museum, and Claudio Cristino of the University of Chile, the expedition collected the skeletal material in several field seasons, ending in 1981. The skeletal remains were curated at the Museo Antropológico Padre Sebastián Englert (MAPSE), Rapa Nui. Approval for the collection of skeletal remains for analysis was approved by the Consejo de Monumentos Nacionales de Chile, and the Museo Antropológico Padre Sebastián



Englert, under then-director Francisco Torres Hochstetter. Skeletal materials from which dental calculus was extracted were repatriated to Rapa Nui in 2009.

**Taumako, Duff Islands, Solomon Islands.** The skeletal remains from Taumako in the Solomon Islands were excavated in the early 1970s by Foss Leach and Janet Davidson. We do not have any record of the original research permit. The skeletal remains were curated at the Anatomy Dept in Otago University firstly by Professor Phil Houghton and then by Hallie Buckley. The skeletal remains were repatriated in 2009 to the Solomon Islands National Museum. At that time permission was given to retain bone and tooth samples for further destructive analyses by the Director Lawrence Kiko.

**Tongatapu, Tonga.** Excavations at Talasiu (Tongatapu) were directed by archaeologist Geoffrey Clark and human and cultural remains export were permitted by Tongan Traditions Committee (Komoti Talafakafonua), Nuku'alofa, Kingdom of Tonga (2016–2025). The excavations of human skeletal remains at Talasiu were conducted in consultation and with the agreement of the Lapaha community.

**Uripiv, Vanuatu.** The Vanuatu Cultural Centre (VCC) and the Vanuatu National Cultural Council provided ethical oversight of the study of human remains from Uripiv. Monica Tromp entered into a research agreement in 2013 with the Vanuatu Cultural Centre to conduct research on human dental calculus from multiple archaeological sites in Vanuatu. The permit number is 10071.

**Vao, Vanuatu.** The Vanuatu Cultural Centre (VCC) and the Vanuatu National Cultural Council provided ethical oversight of the study of human remains from Vao. Monica Tromp entered into a research agreement in 2013 with the Vanuatu Cultural Centre to conduct research on human dental calculus from multiple archaeological sites in Vanuatu. The permit number is 10071.

**Viti Levu, Fiji.** The skeletal remains from the Sigatoka Dunes site on the island of Viti Levu in Fiji were sampled in 2015 from prior excavations curated at the Fiji Museum, Suva, Fiji. The original excavations were conducted under the Sigatoka Salvage Archaeological Project by Simon Best in 1987 and 1988. Sampling of skeletal remains was approved by the Fiji Ministry of Education and the Immigration Department and by the Fiji Museum, and sampling was assisted by museum staff Sepeti Matararaba, Jone Balenaivalu, Elia Nakoro, Sakiusa Kataiwi, and Jotami Naqelelia. Funding for this research was provided by the Nation Science Foundation of the United States, Award # SBS 1216310.

**Watom, Bismarck Archipelago, Papua New Guinea.** Excavations at Watom in 2008 and 2009 were directed by archaeologist Dimitri Anson and Hallie R. Buckley. Research permits for the excavation included permissions for export of human and cultural remains for analysis. In 2008, Hallie Buckley obtained permission from Herman Mandui, Chief Archaeologist of the National Museum and Art Gallery of Papua New Guinea in 2008. We have also had more recent communications (2023) with the Director of the Museum, Alous Kuaso, granting permission for further DNA and other destructive analyses. The excavations of human skeletal remains on Watom Island were conducted in consultation and with the agreement of the Village community.

### Laboratory methods

A total of 102 archaeological dental calculus samples were processed in this study (Table 1, Supplementary Data 1). The dental calculus samples were processed in three groups, hereafter referred to as Jena, Oklahoma and Otago, named after the respective processing lab. Each

sample group was processed using a slightly different extraction and library preparation protocol. Samples from Efate were analysed in two groups - Efate (samples processed in Jena, <1050 BP) and Efate 3000 BP (samples processed in Otago, 3000–2300 BP). Temporal information for the dental calculus samples in this study were obtained either through direct radiocarbon dating of the individual or by cultural association of the burial. Sample extraction and library preparation followed Dabney, et al.,<sup>38</sup> with slight variations between labs.

**Jena.** All sample processing took place in a dedicated cleanroom facility at the Max Planck Institute for Evolutionary Anthropology (MPI-EVA) laboratories (Jena, Germany). Total DNA was extracted from 0.5–7 mg of dental calculus per individual, using a silica column-based extraction protocol optimized for the recovery of short DNA fragments, adapted for dental calculus following the methods described in refs. 12,38,39 and available as downloadable bench protocol through the online protocol repository protocols.io at <https://doi.org/10.17504/protocols.io.bidyka7w><sup>39</sup>. Blanks were processed alongside the samples. The extracts were prepared into double-stranded libraries with partial uracil-DNA-glycosylase (UDG) treatment<sup>40</sup> and dual indexing<sup>41,42</sup> following protocols available through protocols.io at <https://doi.org/10.17504/protocols.io.bmh6k39e> and <https://doi.org/10.17504/protocols.io.4r3l287x3lly/v3><sup>43,44</sup>. The libraries were sequenced to a depth of  $10.5 \pm 2.3$  million reads (mean  $\pm$  standard deviation) on an Illumina Nextseq with 75-bp paired-end sequencing chemistry. Blanks were processed alongside the samples for both extraction and library preparation. Two samples, SIG040 and SIG046, were extracted and sequenced twice, as it was suspected that burial sand may have been unintentionally included during the first processing round.

**Oklahoma.** Total DNA was extracted from 0.8–12.8 mg dental calculus per individual at the ancient DNA facility of the University of Oklahoma Laboratories of Molecular Anthropology and Microbiome Research (LMAMR, Norman, OK, USA), using a protocol very similar to that used in Jena. To remove surface contaminants, dental calculus samples were UV-irradiated for 1 min on each side, and washed with 1 ml 0.5 M EDTA for 15 min. The decontaminated calculus was then resuspended in 1 ml of 0.5 M EDTA solution and incubated overnight at room temperature. A 100  $\mu$ l proteinase K solution (> 600 mAU ml<sup>-1</sup>; Qiagen, Cat. No. RP103B) was added and incubated at 37 °C for 8 h, followed by continued digestion under agitation at room temperature until decalcification was complete. After digestion, the supernatant was added to 14 mL of PB buffer (Qiagen, Cat. No. 19066). This was then centrifuged in a MinElute column (Qiagen, Cat. No. 28006) attached to a Zymo-Spin V column (Zymo Research, Cat. No. C1012) for 4 min at 1500 rcf, rotated, and then centrifuged for an additional 2 min. The column was then dry spun for 1 min at 3400 rcf and washed twice with 700  $\mu$ l PE buffer (Qiagen, Cat. No. 19065) at 9400 rcf. DNA was eluted from the column after a 5 min RT incubation into 30  $\mu$ l of EB buffer (Qiagen, Cat. No. 19086) under centrifugation at 17900 rcf. Blanks were processed alongside the samples. The extracts were thereafter shipped to MPI-EVA, where they were prepared into libraries, as described above, alongside the Jena samples. The samples were sequenced to a depth of  $10.6 \pm 1.5$  million reads on the same NextSeq flow cells as the Jena samples with 75-bp paired-end sequencing chemistry.

**Otago.** Total DNA was extracted from approximately 1–17 mg of dental calculus per individual using a phenol-chloroform aDNA extraction protocol<sup>45</sup>. In brief, dental calculus samples were washed with ultra-pure water and allowed to dry in a laminar flow hood overnight. A second wash was performed using 1 ml of 0.5 M EDTA, with a 30 min incubation time. The supernatant was removed, and the samples were demineralized in 1 ml of 0.5 M EDTA for up to 72 h, until fully demineralized. The supernatant was added to a tube with 750  $\mu$ l of

phenol:chloroform:isoamyl (25:24:1), vortexed, and left on a rotator for 10 min. After centrifuging, the aqueous phase was transferred to 750  $\mu$ l of phenol:chloroform:isoamyl alcohol (25:24:1). The incubation step was repeated, after which the aqueous phase was transferred to 750  $\mu$ l of chloroform:isoamyl alcohol (24:1). After vortexing and mixing by inversion, the mixture was centrifuged and the aqueous phase transferred to 13 ml of 6 M GuSCN and 200  $\mu$ l of silica suspension, and left on a nutator for 30 min. After centrifugation, the supernatant was removed, the silica was resuspended in 1 ml of GuSCN binding buffer, and the supernatant discarded after centrifugation (three times in total). The silica pellet was air dried for 15 min, and DNA eluted twice in 60  $\mu$ l TE (heated to 65–75 °C). Blanks were processed alongside samples through extraction and library preparation. Double-stranded libraries were prepared by blunt-end repairing the DNA strands, and after that ligating and filling in adaptors. The libraries were amplified using KAPA HiFi DNA polymerase (Roche, 07958927001), and no UDG-treatment was performed. The libraries were sequenced using an Illumina MiSeq 75-bp paired-end sequencing chemistry to  $8.3 \pm 7.2$  million reads at the Otago Genomics Facility (Otago, New Zealand).

### General data processing

Data analyses were conducted in R v.4.1.0<sup>46</sup>, unless otherwise stated. General packages used were *tidyverse* v.1.3.1<sup>47</sup>, *readxl* v.1.3.1<sup>48</sup>, *ggpubr* v.0.4.0<sup>49</sup> and *janitor* v.2.1.0<sup>50</sup>. The colour palette for the study is from the R package *microshades* v.0.0.0.9000<sup>51</sup>. Regression models were drawn to data with a generalized linear model with *geom\_smooth* in *ggplot2* as part of *tidyverse*.

### Preprocessing

DNA sequencing data was preprocessed using the *nf-core/eager* v.2.3.3 pipeline<sup>52</sup>. Default options were used unless otherwise stated. Poly-G stretches were removed from the raw data, as they are a common by-product of the two-color chemistry sequencing strategy used by Illumina's NextSeq. Human DNA was removed from the dataset by mapping to the human reference genome GRCh38, and only unmapped reads were retained for downstream microbiome analyses. Taxonomic profiling to produce an OTU table was performed using MALT v.0.4.1<sup>53,54</sup> with a custom database<sup>11</sup>. The database contained all bacterial and archaeal assemblies (scaffold/chromosome/complete levels, up to 10 randomly selected genomes per species) from RefSeq and the human HG19 reference genome via the *nf-core/eager* pipeline. In addition, the dataset was aligned to the NCBI nt database (as of October 2017), to screen for eukaryotic DNA. MEGAN v.6.17.0<sup>55</sup> was used to export OTU tables from the resulting MALT-produced rma6 files, using summarized read counts at both the genus and species level.

Comparative datasets of published microbiome studies were also processed using the same procedures. One comparative dataset was used to assess preservation with the programme SourceTracker<sup>14</sup>, and consisted of 10 non-industrialized gut samples<sup>56,57</sup>, 10 industrialized gut samples<sup>58,59</sup>, 10 skin samples<sup>60</sup>, 10 subgingival and 10 supragingival plaque samples<sup>58</sup>, 10 femur archaeological bone samples<sup>11</sup>, 10 modern dental calculus samples<sup>11</sup> and 10 archaeological sediment samples<sup>61</sup> (Supplementary Data 2). In addition, 10 archaeological petrous bone samples from Taumako and 10 from Viti Levu were included as local environmental controls; this data had been produced during genetic screening of human remains for human population genetic studies at MPI-EVA laboratories in Jena, Germany. Because bones are free of microbes during life, microbes detected in these samples provide a good proxy for local post-mortem colonization<sup>62</sup>.

A second comparative dataset was used to compare calculus species profiles across the globe, and consisted of ancient calculus samples from 6 previously published studies. These included samples from Japan<sup>18</sup>, Europe and Africa<sup>11</sup>, Europe<sup>16,20,27</sup>, and Europe, the Caribbean, North America, and Mongolia<sup>12</sup> (Supplementary Data 2).

### Preservation

Preservation was assessed using SourceTracker v.1.01<sup>14</sup>, PCA, and the R package *cuperdec*<sup>11,63</sup> by comparing the microbial DNA in this study to previously published metagenomic datasets<sup>1,5,16,27,64,65</sup>. For SourceTracker analysis, a species-level OTU table was used as input, and the published reference metagenomic datasets were used as sources. Rarefaction was performed to 10,000 reads, with a training data rarefaction of 5,000 reads. For principal component analysis (PCA), species-level read counts of all dental calculus samples and sources (including 9 modern dental calculus samples) were compared. The R-package *cuperdec*<sup>11,63</sup> was used to identify well-preserved samples (adaptive burn-in method, cut-off 50%), and 73 (out of 102) samples were carried on to further analyses based on this analysis.

Putative environmental and laboratory contaminants in the dental calculus samples were identified using the R package *decontam* v.1.6.0<sup>66</sup>, with the prevalence method. The samples were separated into groups based on the processing lab, and blanks and archaeological bones were used as proxies for contamination sources (cut-off 0.25 for all groups, for both blanks and bones). Contaminants from all labs were combined into a single list of all contaminants, and all contaminants were removed from all samples. The comparative calculus datasets were likewise assessed with *decontam* and all contaminants were combined into a single list, along with those from the Pacific calculus dataset here, and all contaminants were removed from all datasets for the comparative analyses.

To evaluate whether preservation was related to environmental conditions, a dataset consisting of annual average temperature and annual total rainfall was compiled for each island in this study<sup>67</sup>. Missing data for some islands was obtained from alternative sources<sup>68,69</sup>. Annual average evapotranspiration was compiled from<sup>70</sup>.

The proportion of taxa estimated to originate from the oral microbiome (i.e., dental plaque and dental calculus) by SourceTracker was used as a proxy for the preservation of the archaeological dental calculus samples. The effects of environmental variables and/or sample age on preservation were investigated using beta regression with a complementary log-log link function to account for observed heteroscedasticity. Using ANOVA, it was found that the model was not significantly improved by adding the random effects of laboratory and/or island ( $p > 1$ ). Step AIC (using both directions) was used for model selection. To reliably estimate parameters for the model, statistically influential data points were removed, and the model-fitting process was repeated until a stable dataset was reached. Model fit was measured using the  $R^2$  value as suggested by Ferrari and Cribari-Neto<sup>71</sup> for beta regression models.

Preservation of calculus was determined by the percentage of the sample that was determined to be from an oral source (calculus, supragingival plaque, and subgingival plaque) with SourceTracker. Preservation of human DNA was determined by the percentage of endogenous human DNA in shotgun sequenced data, and was taken from published values. Human DNA preservation in the Pacific was taken from Posth, et al. (Supplementary Data 3)<sup>1</sup> and Liu, et al. (Supplementary Data 2)<sup>5</sup>, human preservation in England was from Schiffels, et al. (Supplementary Data 1)<sup>64</sup> and Patterson, et al. (Supplementary Data 1)<sup>65</sup>, and calculus preservation in Europe was taken from SourceTracker values of data from<sup>16,27</sup>.

### Taxonomic profiling

Taxonomic profiles were generated with three tools, MALT, Kraken2, and MetaPhlan3. The MALT table was generated as described above as part of the *nf-core/eager* run, and was used for all community composition analyses. The tools Kraken2 and MetaPhlan3 were used to assess whether altering parameters or using non-standard databases increased the number of taxonomic assignments in samples (Supplementary Fig. 3, Supplementary Fig. 4). Kraken2 was run with default parameters, using two databases (described in ref. 34): a custom RefSeq database and the same custom RefSeq database with the

addition of MAGs from<sup>22</sup>. This allowed us to test whether including additional diversity in the database resulted in a substantial increase in the number of reads assigned to taxonomy. However, we found that it did not, and that this profiler is particularly affected by the read length. MetaPhlan3 was run with two different sets of parameters: default, and custom settings optimized for assignment of ancient DNA (-D 20 -R 3 -N 1 -L 20 -i S,1,0.50 and minimum read length of 35 bp). The use of ancient-optimized parameters substantially increased the number of species identified in each sample (Supplementary Fig. 4A) and we found that the number of species identified by MetaPhlan3 with ancient-optimized parameters and by MALT were correlated (Supplementary Fig. 4B). The optimal number of sample clusters within the Pacific calculus dataset was determined using the Gap statistic with clusGap from the R package cluster<sup>72</sup> with 500 bootstrap replicates, on both the same MALT species table used for compositional analysis, as well as on a MALT genus table that was cleaned and filtered following the same steps as the species table. The optimal number of clusters was found to be one for both the species and genus tables (Supplementary Fig. 2), so no further cluster analysis was performed.

### Community composition

Principal component analysis (PCA, Euclidean distance) was conducted on the decontaminated species table of only the Pacific data, and again on a decontaminated species table of the Pacific data plus the global comparative data, using the R package MixOmics (Supplementary Data 3) with a centred-log ratio transformation. Drivers of variation in the community composition were tested with a PERMANOVA from the R-package vegan v.2.5.7<sup>73</sup>, with euclidean distances and 999 permutations. Homogeneity of multivariate dispersions was tested with the function betadisper from the R package vegan. Canonical correlation was performed using the R package variancePartition<sup>74</sup>. To determine if particular species were associated with environmental conditions on the islands, geographic location, sample age, or library characteristics, we used the R package specificity<sup>75</sup> and the decontaminated species table. We used the R package FAVA<sup>17</sup> to test variability between samples both within and across islands for the Pacific dataset, and within and across continents for the global dataset. Pacific calculus sample clustering was performed following Quagliariello, et al.,<sup>29</sup> with optimal sample cluster number determined using the function clusGap in the R package cluster<sup>72</sup>, however the optimum number of clusters was determined to be 1 and no further cluster analysis was performed (Supplementary Fig. 2). For details see Supplementary Methods.

### Functional analysis

Functional analysis was performed using HUMAnN3<sup>76</sup>. All reads < 50 bp were removed from the fastq files prior to running HUMAnN3, because these are generally too short to be classified after translation. The bowtie2 mapping parameters were adjusted to account for aDNA damage patterns (-D 20 -R 3 -N 1 -L 20 -i S,1,0.50). The standard gene family output table with UniRef90 gene clusters was grouped to KEGG Orthology, and analysis was performed on the KEGG orthologs. Poorly preserved samples as well as those with < 750 orthologs were removed, and orthologs present at < 0.005% abundance were filtered out prior to analysis. The optimal number of sample clusters and KEGG Ortholog clusters was determined using the eclust function of the R package factoextra<sup>77</sup> with kmeans clustering, which uses clusGap of the cluster<sup>72</sup> R package. The gap statistic was visualized with the factoextra function fviz\_gap\_stat, while clustering of sample and KEGG Orthologs was visualized from the output of the eclust function using ggplot. For sample clustering, the optimal cluster number was determined to be 6 (Supplementary Fig. 6A), however, visualization of the clustering demonstrated substantial overlap of the samples of each cluster (Supplementary Fig. 6B). Based on this result, for hierarchical clustering, a sample cluster number of two was chosen to reflect the trajectory of samples in Supplementary Fig. 6B. For KEGG Ortholog

clustering, the algorithm did not converge after 10 iterations, the maximum number of clusters tested. Visualization of the ortholog clustering with 10 clusters (Supplementary Fig. 6C) revealed substantial overlap of the orthologs in different clusters, yet one cluster (cluster 5) included orthologs that clearly separated into two clusters (Supplementary Fig. 6D). Based on visual inspection of Supplementary Fig. 6D, three was selected as the number of ortholog clusters to include in hierarchical analysis.

A PCA was performed using the same steps as with taxonomy. PERMANOVA was performed with the adonis function in the R package vegan<sup>73</sup>. Hierarchical cluster analysis was performed within the Heatmap function of the R package ComplexHeatmap<sup>78</sup>. Both samples and KEGG orthologs were clustered with the “complete” method, and the number of clusters was specified based on the clustering and Gap statistic testing done in the preceding paragraph (2 clusters for samples, and 3 clusters for orthologs).

### Phylogenetic analyses

The nf-core/eager pipeline was used, as described above, to map the non-human reads of the well-preserved samples to the abundant and prevalent oral bacteria *Tannerella forsythia* (strain 92A2, assembly GCA\_00238215.1) and *Anaerolineaceae* bacterium oral taxon 439 (assembly GCA\_001717545.1). Through nf-core/eager, duplicates were removed using Picard MarkDuplicates v.2.22.9, and prior to mapping, damage was clipped off of the reads (two bases for libraries with partial UDG treatment, and seven bases for non-treated libraries). Genotyping was performed with GATK UnifiedGenotyper, allowing for heterozygous calls and using all sites, with the SNP likelihood model. A minimum base coverage of 5 was required. The SNPs were further filtered in order to construct the phylogenies with only homozygous SNPs (defined as the major nucleotide having a frequency greater than 0.9), using MultiVCFanalyzer v.0.0.87<sup>79</sup> (Supplementary Data 4, 5). Only samples with at least 1000 SNPs and a mean genome-wide coverage of at least 2X (for *T. forsythia*) or 5X (*Anaerolineaceae* bacterium oral taxon 439) were included. The coverage requirement was increased for *Anaerolineaceae* bacterium oral taxon 439, because its percentage of heterozygous SNPs was higher.

Recombination was assessed using Gubbins<sup>80</sup> v3.3.5 on a full alignment of *Anaerolineaceae* bacterium oral taxon 439 and *T. forsythia*. Within Gubbins the initial tree was built with FastTree<sup>81</sup> and subsequent trees were built with RaxML<sup>82</sup> v8.2.12, using the best model, GTRGAMMA, as determined by iqtree<sup>83</sup> v2.3.4 and pyjar<sup>84</sup> v1.0, and 100 bootstrap replicates. RaxML was independently run in RaxML-NG<sup>85</sup> v1.2.2 to build trees with 200 bootstrap replicates on the same SNP alignment that was used to build the neighbour-joining tree, as well as on the masked alignment produced by Gubbins. Neighbour-joining trees and maximum likelihood trees were built using distance matrices generated with the TN93 + G4 substitution model, as this was determined by testing with the DECIPHER R package<sup>86,87</sup> to be the best model for each dataset, with 200 bootstrap replicates. The trees were rooted using the midpoint, determined with the R package phangorn<sup>88</sup>. The neighbour-joining phylogenetic trees were constructed and visualized using R packages ape v.5.5<sup>89</sup>, ade4 v.1.7.17<sup>90</sup>, adegenet v.2.1.3<sup>91</sup> and ggtree v.1.99.1<sup>92</sup>, while maximum likelihood trees were built with RaxML-NG<sup>85</sup> v1.2.2 using RaxML v8<sup>82</sup>. Tree concordance by Kendall's coefficient of concordance (W) with the function CADM.global and by a Mantel test with CADM.post in the R package ape<sup>93</sup>. For this, trees were converted to distance matrices with the ape function cophenetic, and distance matrices of the trees being compared were joined as a single file with rbind. The combined distance matrices were used as input for both CADM.global and CADM.post with 99 permutations.

### Strain sharing

Strain sharing across samples was assessed with inStrain<sup>94</sup>. A test dataset was generated in silico to test for the effects of ancient DNA

damage on popANI calculations. For this, two 10 M bp paired-end simulated read datasets were generated to test the effects of aDNA damage patterns and clipping, using the genomes of the 10 species in the ZymoBiomics kit that was used by Olm, et al. for testing, in the proportions described in the kit. The genomes were processed with gargammel<sup>95</sup> in two batches, to create datasets with two different damage profiles based on the damage of the samples sequenced here. Two samples were selected for references, HCLVMBX2-3505-07-00-01\_S8, which is non-UDG treated, and EFE002.B0101, which is UDG-half treated. Both samples were mapped against the *Anaerolineaceae* bacterium oral taxon 439 genome (GCA\_001717545.1) with bwa (-n 0.02 -l 1024)<sup>96</sup>, and the bam files were used as input to mapDamage<sup>97</sup> to generate damage profiles and read length distributions that were used as input to gargammel. The simulated datasets were processed with the nf-core/eager pipeline the same way as samples, and adaptor-trimmed un-collapsed paired reads were pulled out of the eager output with a custom script. The paired end reads were mapped against all 10 reference genomes in the ZymoBiomics kit, combined into a single fasta file with bwa (-n 0.02 -l 1024). The effect of read end masking to remove aDNA damage was tested with a custom python script<sup>98</sup>, with mapped reads from UDG-treated reads masked 1 base on either end, and non-UDG-treated reads masked 9, 11, 13, and 15 bases on either end, based on the C-T transversion ratio along read lengths determined by mapDamage. InStrain profile was run on the unmasked and masked mapped read bam files with insert sizes 12, 24, 36, and 48, and inStrain compare was run on all output of the profile step. Based on the output of these simulated data tests, a mask length of 11 and insert size of 12 were selected for processing real samples, as these returned a maximum popANI with minimal loss of coverage.

The coverage overlap and popANI between all samples was visualized and compared in R (Supplementary Figs 12 and 13).

Strain sharing across Pacific calculus samples was assessed with inStrain for two species that were highly abundant across the Pacific calculus dataset: *Anaerolineaceae* bacterium oral taxon 439 (CP017039.1), and *Tannerella forsythia* (NC\_016610.1). Paired-end reads were mapped against reference genomes for *Actinomyces dentalis* (GCF\_000429225.1), *Anaerolineaceae* bacterium oral taxon 439 (CP017039.1), *Desulfobulbus oralis* (CP021255.1), *Eubacterium minutum* (CP016202.1), *Olsenella* oral taxon 807 (CP012069.2), and *Tannerella forsythia* (NC\_016610.1), combined into a single fasta file. Only *Anaerolineaceae* bacterium oral taxon 439 and *Tannerella forsythia* had sufficient genome breadth and depth of coverage across enough samples for reliable strain comparison with inStrain. Mapped reads in each bam file were masked according to their UDG treatment, with UDG-half-treated samples masked 1 base at both ends, and non-UDG-treated samples masked 11 bases at both ends. inStrain profile was run on the masked bam files with an insert size of 12, and the default value of 5X minimum coverage was required for samples to be included in inStrain analysis. inStrain compare was run on all output files of inStrain profile. We focused only on *Anaerolineaceae* bacterium oral taxon 439 and *Tannerella forsythia*, for which we generated whole genome SNP-based phylogenies. The script polymut.py from cmseq<sup>22</sup> was used to assess whether samples contained multiple strains of the species *Tannerella forsythia* and *Anaerolineaceae* bacterium oral taxon 439 based on the ratio of non-synonymous vs. synonymous sites (dN/dS) in coding regions of the genome. This is a previously published ad-hoc method<sup>22</sup> to assess strain heterogeneity in genomes produced from metagenomes. Testing in that study found that metagenomes with higher polymorphic rates are more likely to contain multiple strains.

We performed a test to determine the expected dN/dS of mapping reads against an incorrect reference genome by using three species of *Fusobacterium*, which were formerly considered subspecies of *F. nucleatum*: *F. nucleatum*, *F. polymorphum*, and *F. vincentii*. These genomes were run through prokka to generate gff files for polymut. The ANI of each genome compared to the other two was

determined with MASH<sup>99</sup> as part of the programme dRep<sup>100</sup> and found to be below the species-cutoff of 95% identity (Supplementary Fig. 16A). We used polymut.py to calculate the dN/dS for short read datasets of each genome mapped to the three reference genomes (Supplementary Fig. 16B–G), and took the average across all genomes, which was 1.94. We took this to be the expected dN/dS value when mapping a species against a closely-related but incorrect reference genome.

Nine simulated short-read datasets were generated from each of the three genomes with three read lengths and three different levels of deamination: long read length (100 bp), medium read length (75 bp), and short read length (50 bp), and no deamination, low deamination, and high deamination. These 9 short-read datasets were each mapped against all three reference genomes using bwa aln and the flags -n 0.01 -l 1024, and the script polymut.py was used to determine the number of synonymous SNPs, the number of non-synonymous SNPs, and the total number of sites compared, for each mapped bam file (Supplementary Fig. 16B–G). We found that the deamination level did not affect the dN/dS values for any of the species mapped to any of the others. Because the Pacific dataset reads have a short read length, we focused on the results of the short read length dataset, and calculated the average dN/dS for short reads across all deamination levels, which is 1.94. We took this to be the expected dN/dS value when mapping a species against a closely-related but incorrect reference genome.

For our samples, collapsed reads were mapped against the *Tannerella forsythia* genome (NC\_016610.1) and the *Anaerolineaceae* bacterium oral taxon 439 genome (CP017039.1) using bwa aln and the flags -n 0.01 -l 1024 to account for ancient DNA damage. Mapped reads were masked according to their UDG treatment as described above (i.e., UDG-half treated reads were masked 1 bp on both ends, and non-UDG treated reads were masked 11 bp on each end). Masked mapped bam files were run through polymut with a minimum coverage requirement of 5X, min quality 30, and dominant frequency threshold of 0.8, and the gff files from NCBI RefSeq for the genome accessions above. We then calculated the ratio of non-synonymous SNPs to synonymous SNPs (dN/dS) for all samples mapped against the *Anaerolineaceae* bacterium oral taxon 439 or *Tannerella forsythia* genome (Supplementary Fig. 17).

### Eukaryotic DNA

In addition to analysing microbial DNA, we also investigated putative eukaryotic DNA within the samples. Identifying eukaryotic taxa within metagenomic datasets is challenging and requires multiple validation steps<sup>23</sup>. Additionally, it should be noted that some of the potential dietary items, such as kava (*Piper methysticum*) and the giant swamp taro (*Cyrtosperma merkusii*), from the Pacific may not be present in reference databases as they do not have sequenced nuclear or organelle genomes. We examined the MALT species table from profiling samples with the NCBI nt database to identify potential dietary hits. Within the well-preserved dental calculus samples, we identified DNA from five non-human eukaryotic species of interest: cattle (*Bos taurus*), dog (*Canis lupus familiaris*), broad fish tapeworm (*Dibothriocephalus latus*), bamboo (*Fargesia denudata*), and wheat (*Triticum aestivum*) (Supplementary Data 8). Other eukaryotic DNA present in the dataset belonged to commonly recognized contaminants, were assigned to genomes with known contamination<sup>23</sup>, or represented highly unlikely taxa; these assignments were excluded from subsequent analysis. Data were next mapped (mapping quality set to 37) to the following genomes using nf-core/eager to produce damage patterns for authentication analysis: cattle (*Bos taurus*), dog (*Canis lupus familiaris*), broad fish tapeworm (*Dibothriocephalus latus*), bamboo (*Fargesia denudata*), and wheat (*Triticum aestivum*). For cattle, mapping was also performed to the zebu genome (*Bos indicus*), as this species may be a closer match for this region<sup>101</sup>. For bamboo, only the chloroplast genome was available. For wheat, the mapping was restricted to the mitochondrial

genome in this initial step, as the full wheat genome is very large (15.4 Gb). Damage patterns were investigated for individuals with at least 200 reads mapping to the specific taxa<sup>23</sup>. For taxa where 200 reads were not reached for any individuals, the damage patterns of the 10 individuals with the highest number of reads were investigated.

### Microparticle analysis

Fifty-four samples of decalcified dental calculus were analysed in this study. Phytoliths and starch granules were identified based on M.T.'s reference collection from several botanic gardens and herbariums in Vanuatu, New Zealand, Hawaii, and Rapa Nui, as well as other published material (for example<sup>102</sup>). Phytoliths were described based on the International Code for Phytolith Nomenclature (ICPN) 2.0<sup>103</sup>. Methods used for Rapa Nui<sup>24,25</sup> and Teouma<sup>26</sup> were previously published. All other samples had been decalcified in 0.5 M EDTA for aDNA extraction. After aDNA extraction, the remaining EDTA aliquot was rinsed in DDI water and a 40 µl drop containing most of the microparticle sample was placed on a glass slide and covered with a glass coverslip. Optical light microscopy was done with a Zeiss Axioscope (located in the Anthropology and Archaeology Department at the University of Otago). Each slide was examined in its entirety in vertical transects using cross-polarized and transmitted light to identify and photograph all microparticles. All microparticles were counted and separated into plant or animal types and then by morphotype. Identification of microparticles was done using published material as well as the Pacific-focused reference collection developed by M.T. The International Code for Phytolith Nomenclature 2.0 was used to name and describe all phytoliths<sup>103</sup>.

**Efate.** The samples from Efate were not very microparticle rich; the most common microparticles were fungal spores and hyphae (commonly found in sediment samples and not possible to refine the ID) (Supplementary Fig. 14A). Sample EFE002.B did contain diatoms, but since it is only one sample and two diatoms, not much can be inferred from this. One starch granule was found in sample EFE003.B; however, it is round and less than 10 µm, which means it could come from just about any starch-containing plant (Supplementary Fig. 14B). A starch granule was also found in sample EFE006.B (Supplementary Fig. 14C). This granule is approximately 10 µm and angular; it is most similar to *Colocasia esculenta* or taro; however, it could also overlap with several other root crops and finding it in isolation makes it difficult to be certain.

**Futuna.** The Futuna samples had the highest number of starch granules of any location examined, and almost all were in one sample, FUT018.B. Of the 27 granules, 18 are ≤ 10 µm and so cannot be confidently identified to family/species. The remaining 9 granules correspond to Types 2 ( $n=3$ ), 3 ( $n=2$ ), 4 ( $n=2$ ), 5 ( $n=1$ ) and 8 ( $n=1$ ) (Supplementary Fig. 15A, B, C, D, E)<sup>45</sup>. Type 2 granules could come from several root and tree crops and corn, which could be contamination, although the possibility that it is a genuine dietary component cannot be ruled out given that Europeans may have traded corn there as early as the 1600 s<sup>104,105</sup>. Type 3 granules could be from several root and tree crops. Type 4 granules are generally from *Tacca leontopetaloides* or Polynesian arrowroot—unfortunately, they also overlap with corn, so contamination cannot be ruled out. The type 5 granule could be from several root and tree crops. The type 8 granule is similar to one found in Lapita-aged samples from Vanuatu but could not be linked to any known reference species. A few different fungal spores and hyphae were also found in this sample. Sample FUT021.B contained one starch granule consistent with *Zea mays* (corn), which is likely contamination. There was also a piece of microcharcoal (Supplementary Fig. 15F) and three probable grass or Cyperaceae phytoliths (Supplementary Fig. 15G, H, I) in this sample.

**Taumako.** Almost all samples from Taumako contained fungal spores and/or hyphae, but not much else. Five starch granules were found in

three samples—all are either too small or damaged to be securely identified except for one, which is consistent with wheat starch and likely contamination (Supplementary Fig. 16A, B, C). Sample NMUI16.A contained unusual microparticles that remain unidentified (Supplementary Fig. 16D); these may be associated with the high marine diet found in stable isotope results from the same population<sup>106</sup>. Finally, a few phytoliths were recovered; most were dentate elongate and likely from grass leaves; some were damaged and could not be adequately described (Supplementary Fig. 16E, F).

**Fiji.** There were quite a few phytoliths found in the Fiji samples, most of which are probably from grass leaves (blocky, elongate, bilobate and rondel types), along with a couple of samples that contained palm phytoliths (Supplementary Fig. 17A–J). These are all quite commonly found within the Pacific. There was one phytolith that could be species-specific that did not match anything in our reference database (Supplementary Fig. 17B); further reference samples would be needed to positively identify it, but it resembles phytoliths from the bark of species with medicinal properties in West Africa<sup>107</sup>. There are also two phytoliths that look like double-peaked glume rice phytoliths (Supplementary Figs. 17A, C), but they are very degraded so the resemblance may be an artefact. Several of the phytoliths appeared black or burnt (Supplementary Fig. 17G, H, I), which may be an indication of fire use. Overall, the Fijian samples were very mineral rich, likely due to insufficient removal of residual sediment for these samples (this issue is also noted in the aDNA methods above, where two samples had to be re-extracted and sequenced due to sediment contamination). Several of the mineral particles are large (probably too large to be inclusions in the dental calculus) and olive green (however, they are unlikely to be obsidian due to the lack of conchoidal fractures) (Supplementary Fig. 17K). Sample SIG044.A contained an unknown fibre probably of plant origin as there are no visible scales or a medulla; it may also be part of an insect (Supplementary Fig. 17L).

**Tongatapu.** Most samples from Tongatapu contained at least one phytolith. There were mostly palm phytoliths (spheroid echinate) (Supplementary Fig. 18B), followed by probable grass phytoliths (Supplementary Fig. 18A) and one instance of a Cyperaceae phytolith (Supplementary Fig. 18C). Two samples also contained sponge spicules (Supplementary Fig. 18D). One sample, TON001.C contained a unique looking fibre that may be a taphonomically damaged feather barbule (Supplementary Fig. 18E).

### Reporting summary

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

### Data availability

For detailed information about the archaeological sites and samples in this study, as well as research consultations and permissions, see Methods. The sequencing data for this study have been deposited in the European Nucleotide Archive under accession PRJEB61887 at <https://www.ebi.ac.uk/ena/browser/view/PRJEB61887>; individual accession numbers are provided in Supplementary Data 1. Published datasets analysed in this study are available in the European Nucleotide Archive; individual accession numbers are provided in Supplementary Data 2. Information about how this data was analysed is provided in Methods and Supplementary Methods. Source data are provided in the “06-publication” folder of [https://github.com/ivelsko/pacific\\_calculus](https://github.com/ivelsko/pacific_calculus), also available through the <https://doi.org/10.5281/zenodo.13903784>.

### Code availability

The scripts used for analysis and figure generation are available at [https://github.com/ivelsko/pacific\\_calculus](https://github.com/ivelsko/pacific_calculus), and through <https://doi.org/10.5281/zenodo.13903784>.

## References

1. Posth, C. et al. Language continuity despite population replacement in remote Oceania. *Nat. Ecol. Evol.* **2**, 731–740 (2018).
2. Skoglund, P. et al. Genomic insights into the peopling of the Southwest Pacific. *Nature* **538**, 510–513 (2016).
3. Pugach, I. et al. Ancient DNA from Guam and the peopling of the Pacific. *Proc. Natl Acad. Sci. USA*. **118**, e2022112118 (2021).
4. Lipson, M. et al. Population turnover in remote Oceania shortly after initial settlement. *Curr. Biol.* **28**, 1157–1165.e7 (2018).
5. Liu, Y.-C. et al. Ancient DNA reveals five streams of migration into Micronesia and matrilocality in early Pacific seafarers. *Science* **377**, 72–79 (2022).
6. Oliveira, S. et al. Ancient genomes from the last three millennia support multiple human dispersals into Wallacea. *Nat. Ecol. Evol.* **6**, 1024–1034 (2022).
7. Dabney, J., Meyer, M. & Pääbo, S. Ancient DNA damage. *Cold Spring Harb. Perspect. Biol.* **5**, a012567 (2013).
8. Smith, C. I., Chamberlain, A. T., Riley, M. S., Stringer, C. & Collins, M. J. The thermal history of human fossils and the likelihood of successful DNA amplification. *J. Hum. Evol.* **45**, 203–217 (2003).
9. Eisenhofer, R., Anderson, A., Dobney, K., Cooper, A. & Weyrich, L. S. Ancient microbial DNA in dental calculus: a new method for studying rapid human migration events. *J. Isl. Coast. Archaeol.* **14**, 149–162 (2019).
10. Tromp, M., Dudgeon, J. V., Buckley, H. R. & Matisoo-Smith, E. A. Dental calculus and plant diet in Oceania. in *The Routledge Handbook of Bioarchaeology in Southeast Asia and the Pacific Islands* (eds. Oxenham, M. & Buckley, H.) 599–622 (Routledge, London and New York, 2016).
11. Fellows Yates, J. A. et al. The evolution and changing ecology of the African hominid oral microbiome. *Proc. Natl Acad. Sci. USA*. **118**, e2021655118 (2021).
12. Mann, A. E. et al. Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. *Sci. Rep.* **8**, 9822 (2018).
13. Velsko, I. M. & Warinner, C. G. Bioarchaeology of the human microbiome. *Bioarchaeology Int.* **1**, 86–99 (2017).
14. Knights, D. et al. Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* **8**, 761–763 (2011).
15. Fagernäs, Z. et al. A unified protocol for simultaneous extraction of DNA and proteins from archaeological dental calculus. *J. Archaeol. Sci.* **118**, 105135 (2020).
16. 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).
17. Morrison, M. L., Xue, K. S. & Rosenberg, N. A. Quantifying compositional variability in microbial communities with FAVA. *bioRxiv* <https://doi.org/10.1101/2024.07.03.601929> (2024).
18. Eisenhofer, R., Kanzawa-Kiriyama, H., Shinoda, K.-I. & Weyrich, L. S. Investigating the demographic history of Japan using ancient oral microbiota. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190578 (2020).
19. Bravo-Lopez, M. et al. Paleogenomic insights into the red complex bacteria *Tannerella forsythia* in pre-hispanic and colonial individuals from Mexico. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190580 (2020).
20. Ottoni, C. et al. Tracking the transition to agriculture in Southern Europe through ancient DNA analysis of dental calculus. *Proc. Natl Acad. Sci. USA*. **118**, e2102116118 (2021).
21. Honap, T. P. et al. Oral metagenomes from Native American Ancestors reveal distinct microbial lineages in the pre-contact era. *Am. J. Biol. Anthropol.* <https://doi.org/10.1002/ajpa.24735> (2023).
22. Pasolli, E. et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176**, 649–662.e20 (2019).
23. 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.* <https://doi.org/10.1016/j.quaint.2020.11.019> (2020).
24. Dudgeon, J. V. & Tromp, M. Diet, geography and drinking water in Polynesia: microfossil research from archaeological human dental calculus, Rapa Nui (Easter Island). *Int. J. Osteoarchaeol.* **24**, 634–648 (2014).
25. Tromp, M. & Dudgeon, J. V. Differentiating dietary and non-dietary microfossils extracted from human dental calculus: the importance of sweet potato to ancient diet on Rapa Nui. *J. Archaeol. Sci.* **54**, 54–63 (2015).
26. Tromp, M. et al. Exploitation and utilization of tropical rainforests indicated in dental calculus of ancient Oceanic Lapita culture colonists. *Nat. Hum. Behav.* **4**, 489–495 (2020).
27. 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).
28. 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).
29. Quagliariello, A. et al. Ancient oral microbiomes support gradual Neolithic dietary shifts towards agriculture. *Nat. Commun.* **13**, 6927 (2022).
30. Zaura, E. et al. Same exposure but two radically different responses to antibiotics: resilience of the salivary microbiome versus long-term microbial shifts in feces. *MBio* **6**, e01693–15 (2015).
31. 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).
32. Clemente, J. C. et al. The microbiome of uncontacted Amerindians. *Sci. Adv.* **1**, e1500183 (2015).
33. McCall, L.-I. et al. Home chemical and microbial transitions across urbanization. *Nat. Microbiol.* **5**, 108–115 (2020).
34. Velsko, I. M., Gallois, S., Stahl, R., Henry, A. G. & Warinner, C. High conservation of the dental plaque microbiome community across populations with differing subsistence strategies and levels of market integration. *bioRxiv* <https://doi.org/10.1101/2022.07.27.501666> (2022).
35. Steinegger, M. & Salzberg, S. L. Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biol.* **21**, 115 (2020).
36. Warinner, C. et al. A robust framework for microbial archaeology. *Annu. Rev. Genomics Hum. Genet.* **18**, 321–356 (2017).
37. Klapper, M. et al. Natural products from reconstructed bacterial genomes of the Middle and Upper Paleolithic. *Science* **380**, 619–624 (2023). eadf5300.
38. Dabney, J. et al. Complete mitochondrial genome sequence of a middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA*. **110**, 15758–15763 (2013).
39. Aron, F. et al. Ancient DNA extraction from dental calculus v1. *protocols.io* ZappyLab, Inc. <https://doi.org/10.17504/protocols.io.bidyka7w> (2020).
40. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130624 (2015).
41. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
42. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448 (2010).

43. Aron, F., Neumann, G. U. & Brandt, G. Half-UDG treated double-stranded ancient DNA library preparation for Illumina sequencing v1. *protocols.io* ZappyLab, Inc. <https://doi.org/10.17504/protocols.io.bmh6k39e> (2020).
44. Stahl, R. et al. Illumina double-stranded DNA dual indexing for ancient DNA v1. *protocols.io* ZappyLab, Inc. <https://doi.org/10.17504/protocols.io.bakticwn> (2019).
45. Tromp, M. Lapita plants, people and pigs. (University of Otago, 2016).
46. R Core Team. R: a language and environment for statistical computing (R Foundation for Statistical Computing, 2021). <https://www.R-project.org/>.
47. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
48. Wickham, H. & Bryan, J. readxl: Read Excel files. <https://CRAN.R-project.org/package=readxl> (2019).
49. Kassambara, A. ggpubr: 'ggplot2' based publication ready plots. <https://CRAN.R-project.org/package=ggpubr> (2018).
50. Firke, S. janitor: Simple tools for examining and cleaning dirty data. <https://CRAN.R-project.org/package=janitor> (2023).
51. Dahl, E., Karstens, L. & Neer, E. microshades: A custom color palette for improving data visualization. <https://karstenslab.github.io/microshades> (2021).
52. Fellows Yates, J. A. et al. Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. *PeerJ* **9**, e10947 (2021).
53. Vågene, Å. J. et al. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2**, 520–528 (2018).
54. Herbig, A. et al. MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. *bioRxiv* <https://doi.org/10.1101/050559> (2016). 050559.
55. Huson, D. H. et al. MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* **12**, e1004957 (2016).
56. Obregon-Tito, A. J. et al. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat. Commun.* **6**, 6505 (2015).
57. Rampelli, S. et al. Metagenome sequencing of the Hadza hunter-gatherer gut microbiota. *Curr. Biol.* **25**, 1682–1693 (2015).
58. Gevers, D. et al. The human microbiome project: a community resource for the healthy human microbiome. *PLoS Biol.* **10**, e1001377 (2012).
59. Sankaranarayanan, K. et al. Gut microbiome diversity among Cheyenne and Arapaho individuals from Western Oklahoma. *Curr. Biol.* **25**, 3161–3169 (2015).
60. Oh, J. et al. Temporal stability of the human skin microbiome. *Cell* **165**, 854–866 (2016).
61. Slon, V. et al. Neandertal and Denisovan DNA from Pleistocene sediments. *Science* **356**, 605–608 (2017).
62. Warinner, C. et al. Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46**, 336–344 (2014).
63. Fellows Yates, J. A. *Cuperdec: Cumulative Percent Decay Curves*. (2020).
64. Schiffels, S. et al. Iron age and Anglo-Saxon genomes from East England reveal British migration history. *Nat. Commun.* **7**, 10408 (2016).
65. Patterson, N. et al. Large-scale migration into Britain during the Middle to Late Bronze Age. *Nature* **601**, 588–594 (2022).
66. Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226 (2018).
67. Zepner, L., Karrasch, P., Wiemann, F. & Bernard, L. ClimateCharts.net – an interactive climate analysis web platform. *Int. J. Digital Earth* **14**, 338–356 (2021).
68. Climates to Travel - world climate guide. <https://www.climatestotravel.com/>.
69. timeanddate.com. <https://www.timeanddate.com/>.
70. ArcGIS Average Annual Evapotranspiration map. <https://www.arcgis.com/home/webmap/viewer.html?layers=ad3f8cc18fc74e6894ee220acd15020a> (2020).
71. Ferrari, S. & Cribari-Neto, F. Beta regression for modelling rates and proportions. *J. Appl. Stat.* **31**, 799–815 (2004).
72. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. cluster: Cluster analysis basics and extensions. <https://CRAN.R-project.org/package=cluster> (2022).
73. Oksanen, J. et al. vegan: Community Ecology Package <https://CRAN.R-project.org/package=vegan> (2022).
74. Hoffman, G. E. & Schadt, E. E. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinform.* **17**, 483 (2016).
75. Darcy, J. L., Amend, A. S., Swift, S. O. I., Sommers, P. S. & Lozupone, C. A. specificity: an R package for analysis of feature specificity to environmental and higher dimensional variables, applied to microbiome species data. *Environ. Microbiome* **17**, 34 (2022).
76. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* **10**, e65088 (2021).
77. Kassambara, A. & Mundt, F. factoextra: Extract and visualize the results of multivariate data analyses. <https://cran.r-project.org/web/packages/factoextra> (2020).
78. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
79. Bos, K. I. et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497 (2014).
80. Croucher, N. J. et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
81. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
82. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
83. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
84. Pupko, T., Pe'er, I., Shamir, R. & Graur, D. A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol. Biol. Evol.* **17**, 890–896 (2000).
85. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
86. Wright, E. Using DECIPHER v2.0 to analyze big biological sequence data in R. *R. J.* **8**, 352 (2016).
87. Wright, E. S. RNAconTest: comparing tools for noncoding RNA multiple sequence alignment based on structural consistency. *RNA* **26**, 531–540 (2020).
88. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
89. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
90. Dray, S. & Dufour, A.-B. The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw., Artic.* **22**, 1–20 (2007).

91. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
92. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
93. Paradis, E. et al. The ape package. *Analyses of phylogenetics and evolution* (2008).
94. Olm, M. R. et al. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat. Biotechnol.* **39**, 727–736 (2021).
95. Renaud, G., Hanghøj, K., Willerslev, E. & Orlando, L. gargammel: a sequence simulator for ancient DNA. *Bioinformatics* **33**, 577–579 (2017).
96. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
97. 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).
98. Meyer, M. et al. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403–406 (2014).
99. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
100. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
101. Wu, D.-D. et al. Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat. Ecol. Evol.* **2**, 1139–1145 (2018).
102. Piperno, D. R. *Phytoliths: A Comprehensive Guide for Archaeologists and Paleoecologists*. (AltaMira Press, Walnut Creek, CA, 2006).
103. International Committee for Phytolith Taxonomy (ICPT). International Code for Phytolith Nomenclature (ICPN) 2.0. *Ann. Bot.* **124**, 189–199 (2019).
104. Flexner, J., Spriggs, M., Bedford, S. & Abong, M. *Beginning Historical Archaeology in Vanuatu: Recent Projects on the Archaeology of Spanish, French, and Anglophone Colonialism*. in *Archaeologies of Early Modern Spanish Colonialism* (eds. Montón-Subías, S., Cruz Berrocal, M. & Ruiz Martínez, A.) 205–227 (Springer International Publishing, Cham, 2016).
105. *Archaeologies of Island Melanesia: Current Approaches to Landscapes, Exchange and Practice*. (ANU Press, Canberra, Australia, 2019).
106. Kinaston, R. L. & Buckley, H. R. Isotopic insights into diet and health at the site of Namu, Taumako Island, Southeast Solomon Islands. *Archaeol. Anthropol. Sci.* **9**, 1405–1420 (2017).
107. Collura, L. V. & Neumann, K. Wood and bark phytoliths of West African woody plants. *Quat. Int.* **434**, 142–159 (2017).

## Acknowledgements

We thank M. Abong of the Vanuatu Cultural Centre (VCC) for assistance during calculus sampling. Sampling of the Teouma dental calculus was done through a research agreement between M.T. and the Vanuatu National Cultural Council. The Teouma Archaeological Project is a joint initiative of the Vanuatu National Museum and the Australian National University (ANU), directed by M.S. and S.B. and at different times R. Regenvanu and M. Abong, both former Directors of the VCC. Funding for the project was provided by the Australian Research Council (DP 0556874 to S.B. and M.S.), the National Geographic Society (SRC 8038–06 to M.S.), the Pacific Biological Foundation (to M.S.), the Department of Archaeology and Natural History and School of Archaeology and Anthropology at the ANU

(S.B. and M.S.), the Snowy Mountains Engineering Corporation Foundation (S.B. and M.S.) and Brian Powell (to S.B. and M.S.). The laboratory research and travel for excavation of the skeletal remains were funded by The Royal Society of New Zealand Marsden Fund (U000407 and 09-U00-106 to H.R.B.) and a University of Otago Research Grant (to H.R.B.). The support of the leaseholder M. R. Monvoisin and family is acknowledged, as is the support and assistance of the traditional landowners and population of Eratap Village. Detailed acknowledgement by the authors of the Vanuatu studies (S.B., H.R.B., J.F., R.S., M.S., E.W., and F.V.) are given in the published site reports for each location, and were supported by the Australian Research Council DP160103578 (to J.F., S.B., and F.V.). The excavation of the Pain Haka site in 2012 was funded by a grant from the Research Institute for Development (UMR Paloc to G.C.) and by additional funding from the French Embassy in Indonesia (to G.C. and T.S.), as well as a University of Otago Research Grant awarded to H.R.B. for the excavation and analysis of the human skeletal remains. Additional funding was provided by an Australian Research Council Discovery Grant (DP170100732 and DP200102872 to G.C.). Genetic data generation and analysis was supported by the Werner Siemens Stiftung (“Paleobiotechnology” to C.W.) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy (EXC 2051, Project-ID 390713860 to C.W.). The Otago research presented here was funded by a University of Otago Doctoral Scholarship (to M.T.), a Royal Society of New Zealand Skinner Fund grant (to M.T.) and an Otago Centre for Electron Microscopy Student Research Award (to M.T.). We thank Alexander Hübner for discussions on strain analyses.

## Author contributions

Z.F., I.M.V., M.T., and C.W. designed and conceived the study. C.W. and I.M.V. supervised the project. A.T., S.B., H.R.B., G.C., J.D., J.F., J.-C.G., R.K., C.M.L., E.M.-S., K.N., A.T.O., C.P., A.B.R., R.S., T.S., M.S., A.T., F.V., and E.W. provided archaeological materials and resources. Z.F., M.T., and A.O. generated the genetic data. Z.F. performed initial data analysis. M.T. performed paleoethnobotanical data generation and data analysis. I.M.V. performed the final genetic data analysis. I.M.V. and C.W. wrote the manuscript; all authors edited the manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## 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/s41467-024-53920-z>.

**Correspondence** and requests for materials should be addressed to Christina Warinner.

**Peer review information** *Nature Communications* thanks Riccardo Nodari and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions 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 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

---

<sup>1</sup>Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. <sup>2</sup>Department of Archaeology, Max Planck Institute for Geoanthropology, Jena, Germany. <sup>3</sup>Southern Pacific Archaeological Research, Archaeology Programme, University of Otago, Dunedin, New Zealand. <sup>4</sup>Department of Anatomy, School of Biomedical Sciences, University of Otago, Dunedin, New Zealand. <sup>5</sup>Department of Archaeology and Natural History, College of Asia and the Pacific, The Australian National University, Canberra, Australia. <sup>6</sup>Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. <sup>7</sup>Department of Anthropology, Idaho State University, Pocatello, ID, USA. <sup>8</sup>Archaeology, School of Humanities, University of Sydney, Sydney, Australia. <sup>9</sup>Research Institute for Development and National Museum of Natural History, Paris, France. <sup>10</sup>BioArch South, Waitati, New Zealand. <sup>11</sup>Department of Anthropology, University of Oklahoma, Norman, OK, USA. <sup>12</sup>Department of Biological Sciences, Halmos College of Arts and Sciences, Nova Southeastern University, Fort Lauderdale, FL, USA. <sup>13</sup>Archaeo- and Palaeogenetics, Institute for Archaeological Sciences, Department of Geosciences, University of Tübingen, Tübingen, Germany. <sup>14</sup>Senckenberg Centre for Human Evolution and Palaeoenvironment at the University of Tübingen, Tübingen, Germany. <sup>15</sup>School of Biological Sciences, The University of Adelaide, Adelaide, Australia. <sup>16</sup>Vanuatu Cultural Centre, Port-Vila, Vanuatu. <sup>17</sup>National Research and Development Centre for Archaeology, Jakarta, Indonesia. <sup>18</sup>School of Archaeology and Anthropology, College of Arts & Social Sciences, The Australian National University, Canberra, Australia. <sup>19</sup>Direction de la Culture et du Patrimoine, Puna'auia, French Polynesia. <sup>20</sup>CNRS UMR 8068, MSH Mondes, Nanterre, France. <sup>21</sup>Faculty of Biological Sciences, Friedrich Schiller University, Jena, Germany. <sup>22</sup>Archaeogenetics Unit, Leibniz Institute for Infection Biology and Natural Products Research Hans Knoll Institute, Jena, Germany. <sup>23</sup>Department of Anthropology, Harvard University, Cambridge, MA, USA. <sup>24</sup>Present address: University of Copenhagen, Globe Institute, Copenhagen, Denmark. ✉ e-mail: [warinner@fas.harvard.edu](mailto:warinner@fas.harvard.edu)