

A roadmap to mammalian oral microbiome evolution with dental calculus

Jaelle C. Brealey¹, Henrique G. Leitão¹, Tom van der Valk¹, Wenbo Xu¹, Katia Bougiouri¹, Love Dalén²,
Katerina Guschanski¹

¹Department of Ecology and Genetics, Animal Ecology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36, Uppsala, Sweden

²Department of Bioinformatics and Genetics, Swedish Museum of Natural History, 104 05, Stockholm, Sweden

Correspondence:

Jaelle Brealey, Department of Ecology and Genetics, Animal Ecology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36, Uppsala, Sweden. Email: jaelle.brealey@ebc.uu.se

Katerina Guschanski, Department of Ecology and Genetics, Animal Ecology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36, Uppsala, Sweden. Email: katerina.guschanski@ebc.uu.se

Abstract

Animals and their associated microbiomes share a long evolutionary history, influenced by a complex interplay between extrinsic environmental and intrinsic host factors. However, we know little about microbiome responses to long-lasting environmental and host-centred processes, which require studying microbiome changes through time. Here, we apply a temporal metagenomics approach to dental calculus, the calcified oral microbial biofilm. We establish dental calculus as a valuable tool for the study of host microbiome evolution by characterising the taxonomic and functional composition of the oral microbiome in a variety of wild mammals. We detect oral pathogens in individuals with evidence of oral disease, assemble near-complete bacterial genomes from historical specimens, characterise antibiotic resistance genes even before the advent of industrial antibiotic production, reconstruct components of the host diet and recover host genetic profiles. Our work demonstrates how dental calculus can be used in the future to study the evolution of oral microbiomes and pathogens, and the impact of anthropogenic changes on wildlife and the environment.

Main

Hosts and their associated microbiomes share a long evolutionary history, influenced by a complex interplay between host genetics and the environment¹⁻⁴. On short temporal scales, experimental work in model organisms has provided important insights into a multitude of factors shaping the host-associated microbial community, including the effect of diet, reproductive status and rapid environmental change⁵⁻⁷. However, there is limited understanding of how the host-associated microbiome is influenced by population-level extrinsic and intrinsic processes that span multiple host generations, such as long-term environmental and host demographic changes. Similar to fossilised plant and animal remains, microbial fossils can offer a unique opportunity to study microbiome evolution and quantify how complex microbial communities and their individual members change through time. Unfortunately, host-associated microbiomes rarely persist after the

host's death. An accessible and well-preserved exception is dental calculus, the calcified form of the microbial biofilm that forms on mammalian teeth⁸. Besides representing the oral microbial community, dental calculus also captures invading pathogens, host cells and dietary biomolecules, thus providing the opportunity for integrative investigations of microbial, dietary and host genetic factors from the same source material through sequencing of preserved DNA^{9–12}. Dental calculus undergoes periodic mineralisation during an individual's life, which reduces external contamination and facilitates DNA preservation through time⁸.

To date, dental calculus has been studied in humans, where DNA sequencing has revealed shifts in oral microbiome composition associated with cultural transitions and allowed tracking of host-pathogen co-evolution through time^{9–11}. Many other mammals produce dental calculus, offering the opportunity to study the link between evolutionary, ecological and demographic factors, and microbiome evolution. Yet, this rich source of information remains virtually unexplored. Dental calculus is readily available from museum-preserved and archaeological specimens. Natural history collections have been extensively used to study the effects of human-driven changes in animal populations over the last few hundred years, such as population declines, range shifts and introduction of pollutants^{13–16}. It may also be possible to infer the effects of these processes on host-associated microbiomes and the potential of microbiomes to confer adaptations to changing environments¹⁷. To establish dental calculus as a standard research tool for the study of host-associated microbiome evolution in diverse mammalian species, we used DNA sequencing to characterise the historical dental calculus microbiome of three evolutionarily distant mammalian species with distinct ecology, diet and physiology: European reindeer (*Rangifer tarandus*), Scandinavian brown bear (*Ursus arctos*), and eastern gorilla (*Gorilla beringei*). Reindeer are group-living ruminant herbivores with a multigastric digestive system and specialised hypsodont molars adapted to an abrasive, fibrous diet. Brown bears are solitary omnivores with brachydont molars more adapted for a partially carnivorous diet. Gorillas are group-living folivores and specialised

hindgut fermenters. The close evolutionary relationship between gorillas and humans (the major source of the microbial reference databases used for microbiome taxonomic characterisation) and the previous successful reconstruction of the chimpanzee oral microbiome from dental calculus¹¹ prompted us to include gorillas to aid characterisation of previously unexplored microbiomes from the other two host species. We outline strategies to overcome the challenges of working with historical microbial DNA from non-model host species, including contamination issues and reference database biases, and demonstrate that a wealth of evolutionary, ecological and conservation-relevant information can be obtained from historical dental calculus samples of diverse host species.

Results

Oral microbiome signature can be successfully recovered from dental calculus of non-human mammals

DNA extraction, Illumina shotgun sequencing and metagenomics analyses were carried out on dental calculus collected from five reindeer (including forest, mountain and high arctic Svalbard ecotypes), six brown bears (from western and eastern Europe) and two eastern gorillas (one Grauer's and one mountain gorilla) dating from 1861 to 1961 (Supplementary Table 1). Contamination of samples with modern DNA is a major problem faced by all historical genomic and metagenomic studies¹⁸. We used a combination of laboratory and bioinformatics procedures to quantify and reduce contamination. In the laboratory, we tested three surface decontamination treatments on the two gorilla dental calculus samples: UV exposure, a wash in EDTA-containing buffer, and no surface decontamination as a control. Following microbial taxonomic assignment with Kraken2¹⁹, we determined that samples decontaminated with the EDTA wash contained the highest proportion of oral taxa (Supplementary Figure 1), and they were therefore retained for downstream analyses. We employed a multi-step bioinformatics approach, which relied on the availability of negative controls and the hallmarks of ancient DNA to identify and filter out contaminant taxa. More specifically, we flagged taxa as contaminants if they were present in the negative controls from the extraction and library

preparation stages²⁰, had higher relative abundance in low biomass samples²¹, had predominantly long DNA fragments (Supplementary Figure 2) and lacked the typical post-mortem deamination patterns¹⁸ (Supplementary Figure 3).

To determine the proportion of microbial communities in our samples that could be assigned to either well-characterised oral microbiome taxa or putative contaminating sources, we used the Bayesian classification tool SourceTracker²². While there was substantial variation between samples, an oral microbiome signature was present in dental calculus samples from all three host species (Figure 1a). Generally, the oral microbiome signature was highest in the two gorilla individuals (Supplementary Figure 4), likely the result of improved taxonomic classification due to the close phylogenetic relationship between gorillas and the source of the oral microbiome database (humans). Human skin and laboratory reagents were the most common contamination sources, and were particularly abundant in specimens that did not undergo surface decontamination before DNA extraction (Supplementary Figure 5). One such bear sample (Ua6) contained high levels of contaminants (>70%) and no detectable oral microbiome signature, and was therefore excluded from all microbial analyses.

Host species had distinct oral microbiome composition (Figure 1b; for abundance-based comparison see Supplementary Figure 6 and Supplementary Table 2). Approximately 30% of the variation between samples could be explained by host species ($p=0.006$) and < 10% each by surface decontamination, sequencing depth and abundance of human reads as measure of human DNA contamination (Supplementary Table 2). Microbial taxa that showed the greatest host species specificity (i.e. were only present in a single host species) were generally associated with either the mammalian oral microbiome or other mammalian body sites (Supplementary Table 3). To independently verify the host species-specific differences between reindeer and bear (gorillas were excluded due to the low sample size of 2), we used a random forests classifier on presence/absence

data. The correct host species could be assigned in 90.0% of cases, and the most important taxa for determining the host species generally included oral taxa, such as *Streptococcus* species, which were highly abundant in the bears but almost entirely absent in reindeer (Figure 1c-d). Reindeer individuals also had lower microbial diversity than bears and gorillas (Supplementary Figure 7, Shannon diversity index median $\pm$ interquartile range: 1.777 ± 0.288 (reindeer) vs. 3.794 ± 0.014 (gorilla) vs. 3.836 ± 0.602 (bear), $p=0.003$). This low alpha diversity of the reindeer samples may reflect poor representation of reindeer oral taxa in the reference database, rather than true reduced diversity.

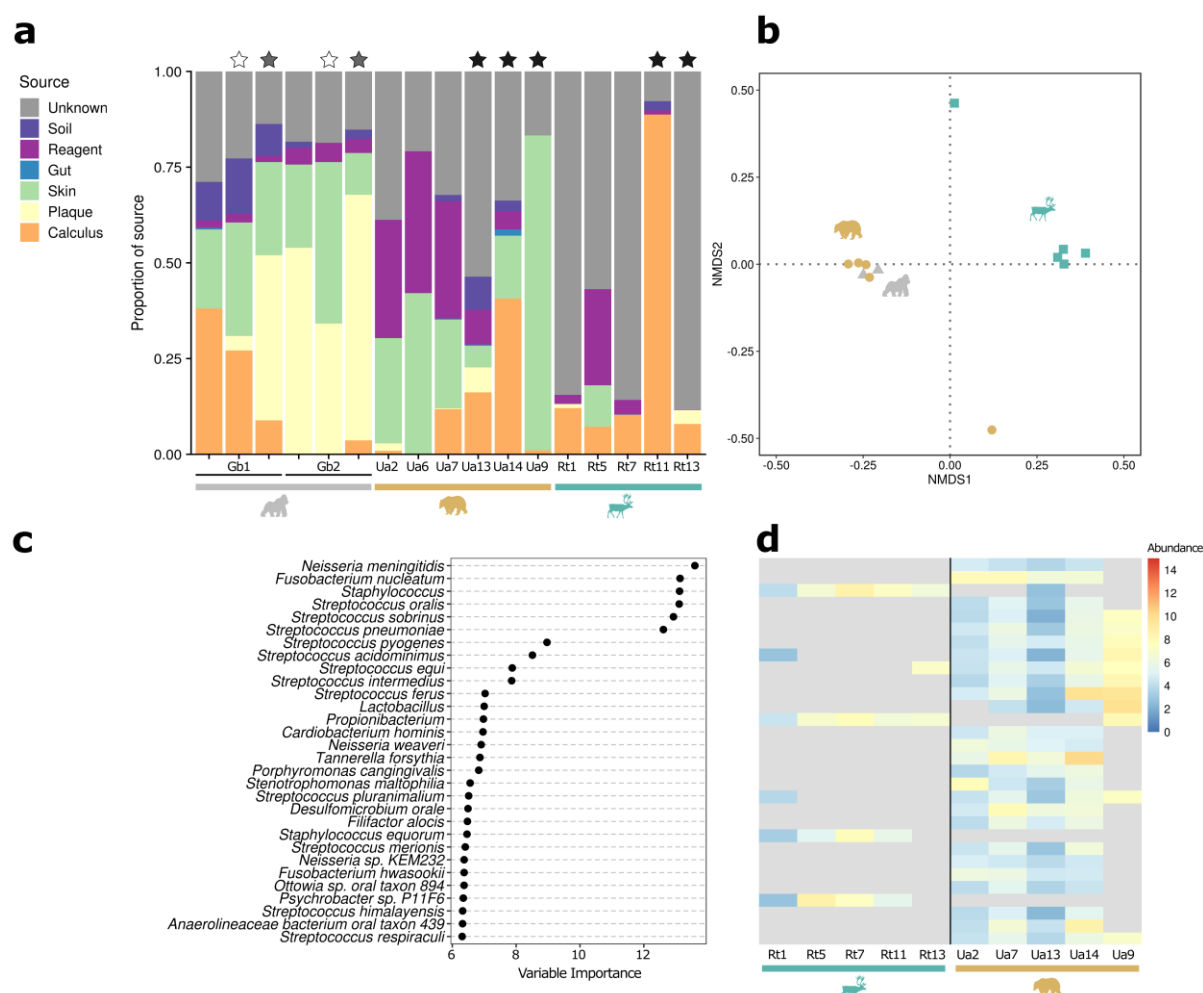


Figure 1. Dental calculus of non-human mammals shows an oral microbiome signature and contains host-specific taxa. **a)** Proportions of source contributions to the microbial communities (identified taxonomically at the species and genus level) contained in the dental calculus samples estimated by

SourceTracker. Stars above bars indicate samples in which surface decontamination was performed before DNA extraction (white: UV only, grey: EDTA wash only, black: UV followed by EDTA wash). Note that each gorilla sample was processed in three different ways to determine the most efficient decontamination strategy (see Methods). **b)** NMDS ordination on Jaccard presence/absence distance matrix of microbial taxa per sample, coloured by host species. **c)** Random forest variable importance plot of the 30 most discriminatory taxa comparing bear and reindeer samples, based on presence/absence data after contamination filtering. **d)** CLR normalised abundance of the top 30 taxa in **(c)** in the bear and reindeer samples. Taxa that were not detected in a sample are coloured grey.

Oral pathogens in wild animals and functional repertoire of the mammalian oral microbiome

While oral diseases are often thought of as an affliction of human dietary practices²³, oral pathologies have been observed in animals^{24,25}, including wild North American black bears²⁶ and captive brown bears²⁷. One of the brown bear specimens in our study (Ua9) was sampled from a caries lesion (Figure 2a). The most abundant microbial taxa in Ua9 included members of the *Lactobacillus casei* group (*L. casei*, *L. paracasei*, *L. rhamnosus* and *L. zae*) and mutans streptococci (such as the closely related *Streptococcus mutans* and *S. ratt*²⁸), species that have been associated with caries lesions in humans^{29,30} (Figure 2b). Species in the *L. casei* group in particular have been associated with active caries lesions in humans^{29,30}. One additional bear individual had signs of caries (Ua7), but was sampled from a healthy tooth, rather than a caries lesion, and its microbiome taxonomically appeared more similar to the other bear samples without macroscopic signs of oral disease (Figure 2b). The biogeography of the oral microbiome in humans has been found to be site-specific³¹, suggesting that even within a diseased individual, the oral microbiome could differ between healthy and carious sites. Cariogenic bacteria including mutans streptococci and *S. mitis* group species were identified in most bear samples without signs of oral disease, as were bacterial

species associated with the periodontal 'red complex' pathogens (*Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia*)³². However, these samples lacked members of the *L. casei* group. Human studies suggest that dental caries is a complex disease that depends more on the environment of the oral cavity and the functions performed by the microorganisms, than on the specific taxonomic composition³³. The same potentially pathogenic taxa were also found in gorilla and some reindeer samples, suggesting that they represent members of the healthy mammalian oral cavity.

By performing shotgun sequencing, we were also able to gain first insights into the functional potential of the microbial communities captured in dental calculus. We used the HumanN2 pipeline³⁴ to assign KEGG orthologues³⁵ to our filtered microbial reads, and cluster these functions into MetaCyc metabolic pathways³⁶. In contrast to the taxonomic results, we observed no clear functional differences between host species (Supplementary Figure 8 and Supplementary Table 2). Instead, we identified core functional pathways shared across individuals and host species that were generally involved in essential metabolic processes, such as energy production and biomolecule synthesis (Supplementary Figure 9), functions common to most living organisms³⁷. Conserved functional properties despite differing community compositions is a common phenomenon in ecosystems, including the human microbiome^{37,38}. However, this inference could also be driven by limitations of the functional reference databases, which are mostly centred on global mechanisms³⁹ and do not allow for more fine-grained functional characterisation. We did identify pathways that have been shown to be enriched in human oral sites, such as those involved in biosynthesis of components of lipopolysaccharide (e.g. ADP-D-glycero- β -D-manno-heptose and lipid A)³⁷. Among the basic functions that can be expected from an oral microbial community, we identified a number of metabolic pathways that were particularly abundant in the carious bear Ua9. They were generally involved in carbohydrate fermentation and acid production (Figure 2c), functions commonly performed by bacteria that colonise the oral cavity and are also associated with the emergence and

progression of dental caries³³. Correspondingly, in Ua9, enzymes encoded by *L. paracasei* and mutans streptococci substantially contributed to one of these pathways (P461-PWY, Figure 2c).

The presence of the cariogenic bacterium *L. paracasei* in Ua9 was further confirmed through *de novo* assembly of a high-quality metagenome-assembled genome (MAG) (Supplementary Table 4). The presence of typical DNA damage patterns adds support that the MAG is likely endogenous (Figure 2d). We additionally recovered five other high-quality MAGs from two bear and two reindeer specimens, which had typical DNA damage patterns and were taxonomically classified as strains related to potentially oral bacteria (*Streptococcus* and *Haemophilus*) (Supplementary Table 4). One *Streptococcus* MAG assembled from Ua9 was identified as closely related to *S. rattus*, which likely reflects the high abundance of mutans streptococci identified by Kraken (Figure 2b). Two other high-quality MAGs were recovered from one bear specimen but could not be assigned known taxonomy and thus possibly represent novel bacteria specific to the bear oral cavity.

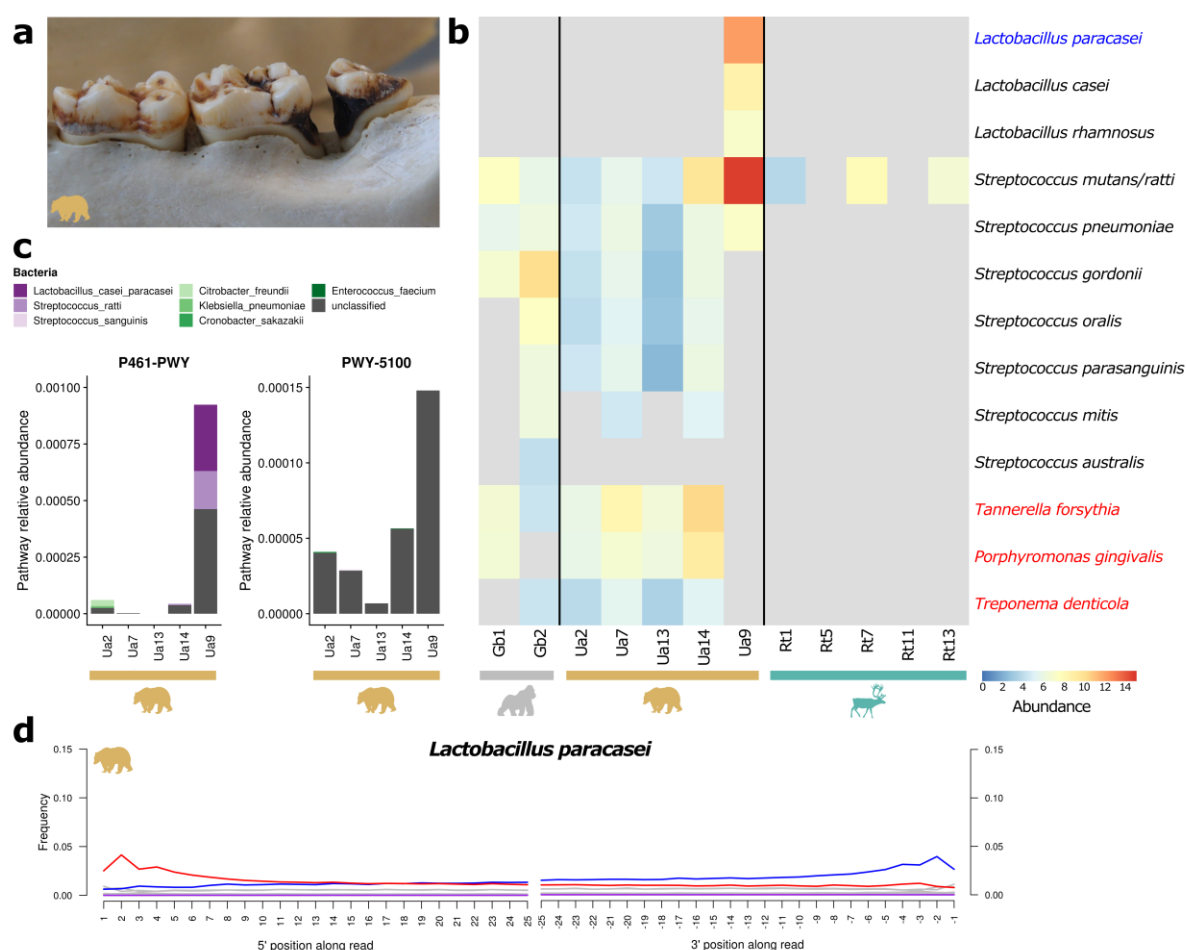


Figure 2. Identification of oral pathogens in a specimen with evidence of oral disease. **a)** Sampling site from a caries lesion from the bear specimen Ua9. **b)** Kraken2 CLR normalised abundance of cariogenic bacteria (*Streptococcus* and *Lactobacillus* species) and periodontal pathogens (*Treponema denticola*, *Porphyromonas gingivalis* and *Tannerella forsythia*, highlighted in red text). Taxa that were not detected in a sample are coloured grey. **c)** Relative abundance of MetaCyc metabolic pathways involved in sugar fermentation to acids in brown bear samples (P461-PWY: hexitol fermentation to lactate, formate, ethanol and acetate; PWY-5100: pyruvate fermentation to acetate and lactate). Relative contribution of bacterial species to each pathway in each sample is shown where known. Note the differences in y-axis scale for each pathway. **d)** Deamination plot of Ua9 reads mapping to the metagenome-assembled genome *L. paracasei* (highlighted in blue in **a**) showing the frequency of C-to-T (red), G-to-A (blue) and all other (grey) substitutions along the 5' and 3' ends of the reads. The drop in C-to-T and G-to-A substitution frequencies at position 1 is

explained by reduced ligation efficiency to barcoded adapters with terminal C or G bases⁴⁰
(Supplementary Tables 9-10).

Antimicrobial resistance (AMR) genes are present in wild animal microbiomes

The presence of bacteria carrying AMR genes has been documented in the human oral microbiome^{10,41}. We therefore investigated whether oral microbial communities of wild animals contain AMR genes and whether their abundance differs across host species. To this end, we investigated both global AMR potential in the entire microbial community and the diversity of AMR genes chromosomally encoded by an oral bacterial pathogen with reported multi-drug resistance, *Acinetobacter baumannii*⁴². To survey global AMR potential, we blasted the contamination-filtered reads against the Comprehensive Antibiotic Resistance Database (CARD)⁴³ and assigned the top match for each read to its respective gene family under the Antibiotic Resistance Ontology (ARO) (Figure 3a). To identify AMR genes encoded by *A. baumannii*, we extracted reads aligned by MALT⁴⁴ to the species using MEGAN⁴⁵. We used the EAGER pipeline⁴⁶ to identify reads that mapped to the plasmid-free chromosome and subsequently processed these reads in CARD, as above (Figure 3b). We also confirmed that the identified *A. baumannii* chromosomal reads exhibited characteristic deamination patterns consistent with post-mortem DNA damage (Figure 3c). Both globally and in the *A. baumannii*-specific analysis, gorillas and bears tended to have greater AMR potential compared to reindeer, in terms of both relative abundance (proportion of reads) and diversity (number) of AMR gene families identified. Similar trends were observed in the global dataset when a more conservative, marker-based method of AMR gene family identification was used (Supplementary Figure 10). With our limited sample numbers, we observed no obvious differences in AMR potential in samples collected before and after the start of industrial-scale antibiotics production in the 1940s⁴⁷. However, our analyses are in line with previous studies on historical human and permafrost microbiomes that demonstrate that many of the underlying molecular

mechanisms conferring resistance to (modern) antibiotics have existed in the environment long before mass antibiotics production^{10,48}. For example, we identified the presence of PER beta-lactamases in *A. baumannii* reads in both Ua7 (dating from 1922) and Ua9 (dating from 1942) (Figure 3b). PER-1 is an extended-spectrum beta-lactamase conferring resistance to third generation cephalosporins and was first observed in *Acinetobacter* species in the late 1990s^{49,50}. This investigation opens doors for the use of dental calculus as a tool to study AMR evolution through time in wild animals from diverse geographic locations, and for determining the potential of wildlife to serve as reservoirs for clinically-relevant AMR factors.

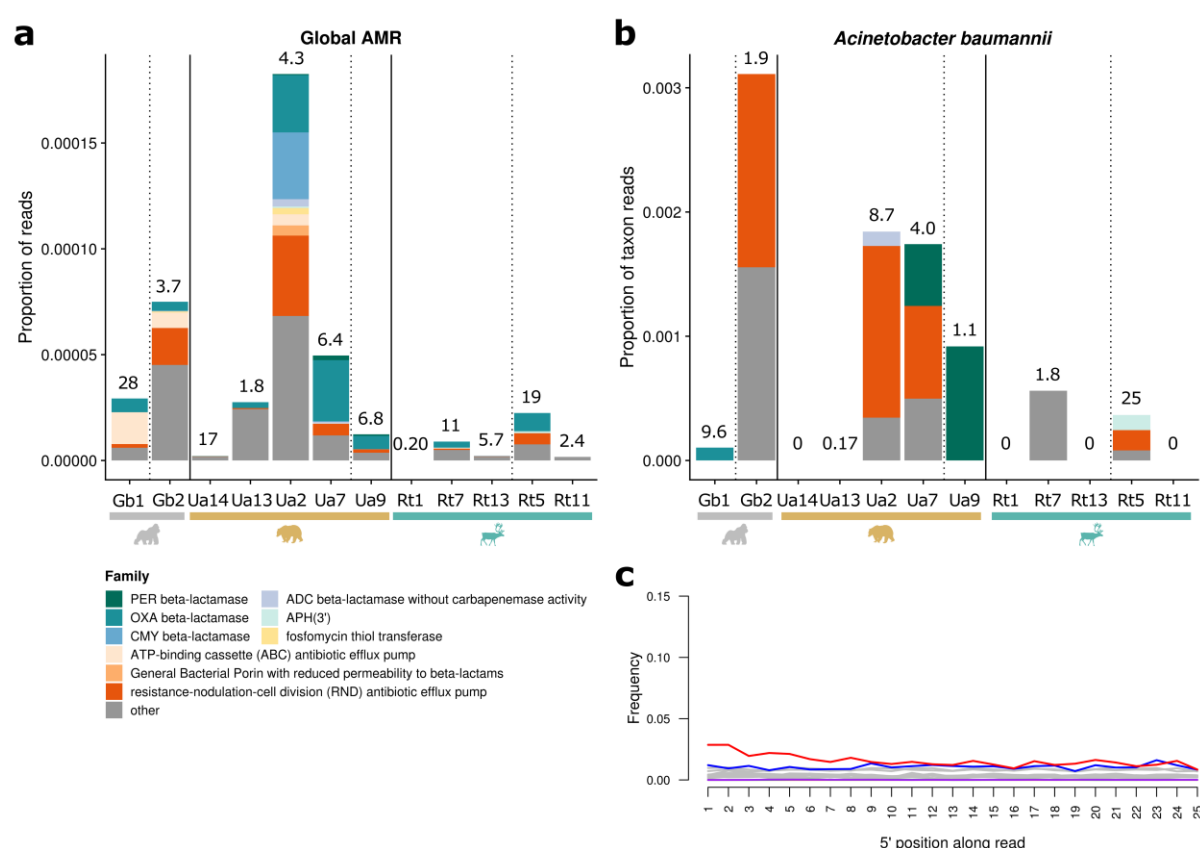


Figure 3. Antimicrobial resistance (AMR) genes can be recovered from historical dental calculus from specimens collected prior to the advent of industrial-scale antibiotics production in the 1940s. **a)** Proportion of global AMR genes identified in all contamination-filtered reads in each sample. **b)** Proportion of *Acinetobacter baumannii* chromosomally-encoded AMR genes identified in each

sample. Samples are grouped by host species and ordered by year with pre-1940 samples separated from post-1940 samples by dashed vertical lines. The top nine most abundant AMR gene families are shown, with the remainder grouped into ‘other’. Number of contamination-filtered reads (in millions) is shown above the bars for each sample in (a), and number of reads (in thousands) mapping to the chromosome of *A. baumannii* is shown in (b). **c)** *A. baumannii* is an endogenous member of the mammalian oral microbiome, as evidenced by the presence of typical DNA post-mortem damage patterns exemplified in reindeer Rt5. Frequency of C-to-T (red), G-to-A (blue) and all other (grey) substitutions are shown along the 5’ end of the reads.

Dental calculus as source of dietary information

To explore the potential of dental calculus to provide insights into the dietary composition of each host species, we taxonomically profiled all eukaryotic reads (excluding those mapping to host and human during data preprocessing) from our samples using MALT⁴⁴ and MEGAN⁴⁵ (Figure 4a). We observed a number of likely spurious mappings and/or contamination, such as to the bovine genus *Bos*, which showed mapping of reads from all three host species and the negative controls (Figure 4b). In bears, only few plant-based dietary components were identified and we found no clear patterns for mammalian or invertebrate putative dietary items, as similar taxa were also present in some reindeer and/or gorilla samples, which are not expected to consume mammals. We were able to infer population-specific dietary characteristics in gorilla and reindeer samples, although in many cases the taxa identified in our analyses are likely close relatives to the consumed species, which are not well represented in the reference databases. For example, among reads that mapped to the Poaceae family in the mountain gorilla sample, approximately 28% mapped to *Phyllostachys*, a genus of Asian giant timber bamboo (Figure 4b). Mountain gorillas are known to consume a related *Arundinaria alpina* bamboo⁵¹, the genome of which is not currently available. We also identified *Galium* vines in the mountain gorilla (Figure 4b), consistent with the known dietary preferences of Virunga mountain gorillas for which these plants are among the six most frequently consumed

taxa^{51,52}. *Salicaceae* plants, e.g. *Salix* (willows) were identified in all reindeer (Figure 4b), consistent with the known browsing behaviour of these animals⁵³. A number of Arctic plants were identified in the Svalbard reindeer (Rt1 and Rt7) that are known or likely components of the high Arctic reindeer diet, including *Saxifraga* spp. and *Oxyria* spp⁵⁴ (Figure 4b). Furthermore, reads assigned to rumen ciliates from the Ophryoscolecidae family (*Entodinium caudatum* and *Epidinium ecaudatum*) were identified in one of the reindeer samples (Rt11). These protozoa are important facilitators of digestive processes in ruminants⁵⁵. Similarly, in the microbial analyses, several microbial taxa specific to the reindeer were identified as *Methanobrevibacter* species associated with the rumen of domesticated bovines and ovines⁵⁶. Ruminants regurgitate large amounts of rumen material into the oral cavity when chewing cud, and a study in sheep has found that oral swabs contain a proportion of rumen-associated microbes⁵⁷. It is therefore plausible that dental calculus of ruminants also captures a subset of the rumen microbiome.

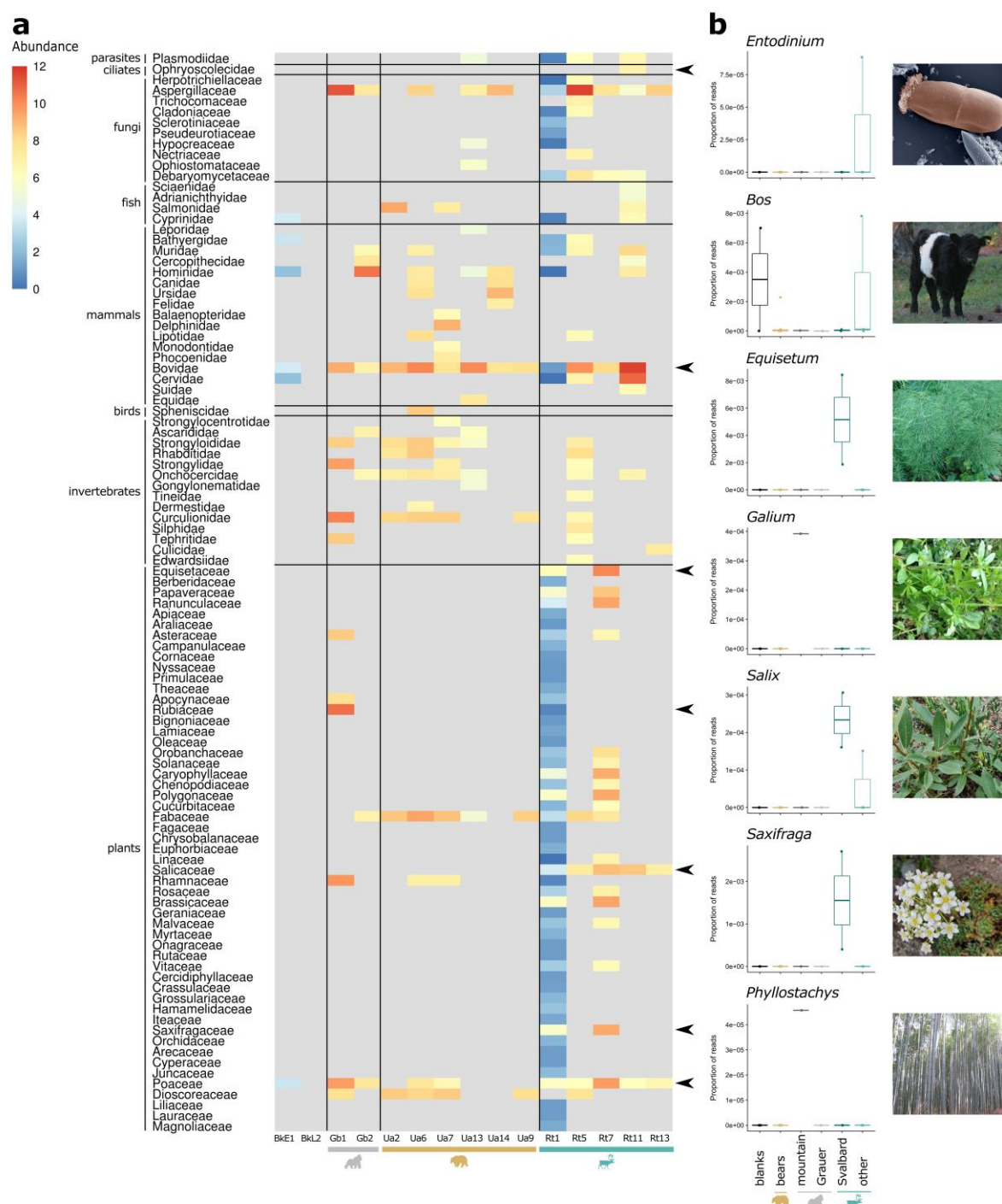


Figure 4. Host diet can be inferred from dental calculus. **a)** MALT/MEGAN CLR normalised abundance of eukaryotic reads at the family taxonomic level. Taxa that were not detected in a sample are coloured grey. Broad groups of eukaryotes are designated by horizontal black lines. Black arrows indicate selected families for which genus-specific relative abundances were plotted in (b). **b)** Proportion of reads mapping to specific eukaryotic genera in different host species, including blank

controls (visualised as Tukey boxplots). Gorilla samples are divided into the two subspecies (mountain and Grauer's gorillas) and reindeer are divided into the Svalbard ecotype and "other" ecotypes (mountain and forest) to illustrate population-specific differences in dietary components. The Bovidae genus *Bos* is included as an example of spurious mappings, due to the presence of reads in samples from multiple host species and blank controls. Image credit: *Entodinium caudatum* photo by Sharon Franklin and colourisation by Stephen Ausmus (<https://www.ars.usda.gov/oc/images/photos/feb06/d383-2/>), other images by Katerina Guschanski and Jaelle Brealey.

Recovery of host genomic profiles

In addition to microbial remains, dental calculus also entraps host cells and thus can serve as a source for host DNA^{12,58}. We therefore identified host DNA preserved in dental calculus samples (including bear Ua6 excluded from the microbial analyses) by mapping the reads to reference genomes of the host species' closest phylogenetic relatives. For mitochondria, 2.6-99.7% (median 91.2%) of the genome was covered by at least one read in each study sample with the coverage depth ranging from 0.02 to 174x (Supplementary Table 5). Indeed, the high abundance of host reads in some samples allowed us to reconstruct complete mitochondrial genomes from five specimens (Supplementary Table 5). For nuclear genomes, 0.004-21.3% (median 0.346%) were covered by at least one read, with maximum genome-wide coverage of 0.3x. We compared the recovered genomic profiles to published genomes from the same species. Mitochondrial haplotypes could be placed within species-specific mitochondrial networks (Figure 5a and Supplementary Figures 11-12, note that only a single reference genome is currently available for reindeer, which limited our inferences for this species). In bears, the mitochondrial haplotype network reflected known differences in colonization history of Scandinavia from west and east⁵⁹: one of the two Swedish bear samples (Ua14 from central Sweden) was placed together with central European bears and the other (Ua9

from northern Sweden) with Russian/Alaskan individuals (Figure 5a). Projection of low-coverage host nuclear genomes onto the PCA space pre-calculated from high-quality published genomes clearly assigned all study samples to their correct species (reindeer), subspecies (gorilla), and even to broader geographic populations of origin (brown bear) (Figure 5b-d).

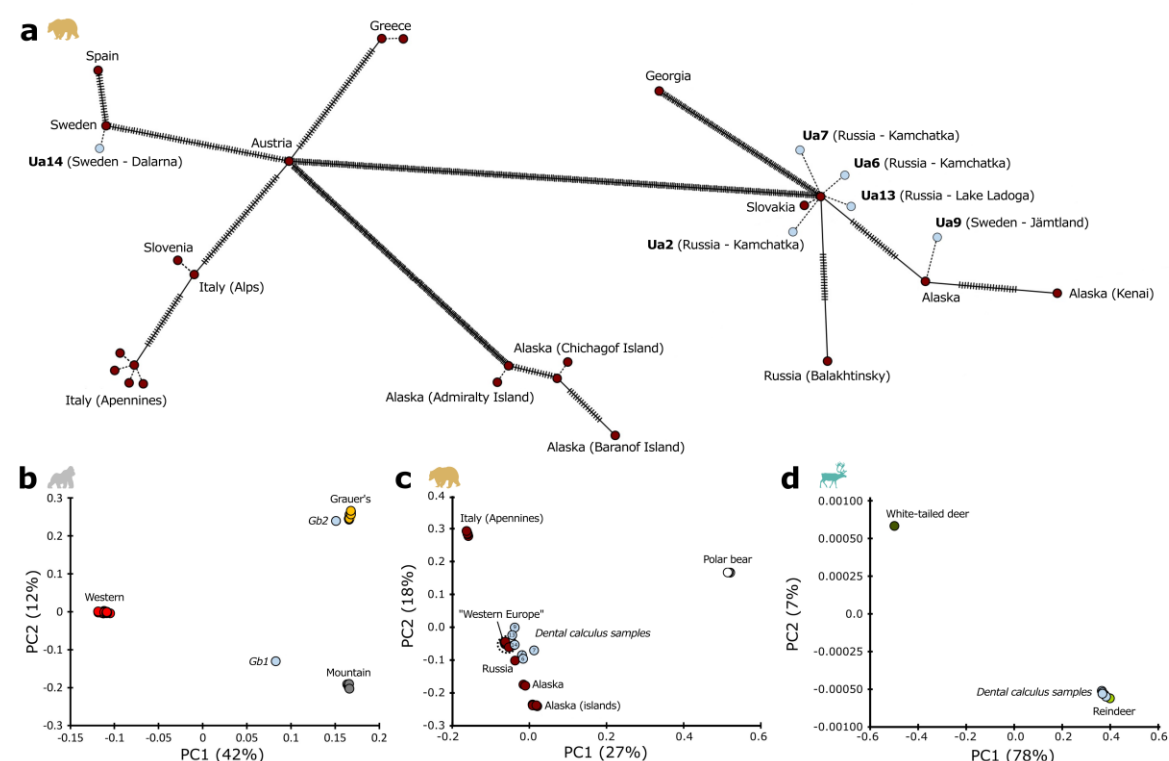


Figure 5. Host population genetic structure can be reconstructed from dental calculus. **a)** mtDNA haplotype network for brown bears. Each circle represents a sample, with ticks on the connecting lines showing the number of base pair substitutions between the haplotypes. Dotted lines represent identical haplotypes or in the case of dental calculus samples (shown in light blue), the predicted most closely related haplotype. Sample labels include country and locality of specimen collection. **b-d)** PCA of modern high-coverage genomes and low-coverage dental calculus samples projected onto the modern genomes for gorilla (b), brown bear (c) and reindeer (d). Samples cluster together with their respective species. Dental calculus gorilla samples cluster most closely to their subspecies of origin. Dental calculus brown bear samples from western Europe and

Russia cluster with the modern genomes from the “Western Europe” clade (consisting of Spain, Greece, Slovenia, Italy (Alps), Sweden and Slovakia) and the Russian clade, respectively. Dental calculus reindeer samples cluster with the single available reindeer reference genome.

Dealing with current limitations: Contamination and database biases

Dental calculus has been little explored outside of human research, yet it can be a treasure trove of information about evolutionary and ecological processes of the host and its oral microbiome. However, research on microbiomes from the past, including from dental calculus, is hindered by a number of challenges. We have established rigorous methodology for overcoming the problem of contamination, which affects all historical genomic and metagenomic studies¹⁸. Through a combination of laboratory and computational procedures, we reduced the proportion of contaminants and systematically increased the proportion of the bacterial communities attributed to the oral microbiome (Supplementary Figure 1). However, the SourceTracker results indicated that some proportion of human skin taxa remained in our samples. Several bacterial species colonise multiple niches within the host^{37,38}, which can obscure distinction of a genuine signal from a likely contaminant. For example, *S. mitis*, *Staphylococcus epidermidis* and *Corynebacterium matruchotii* are found in the human mouth, nostrils and skin³⁸. This limitation is not specific to our dataset, but poses greater problems for studies based on historical samples that are expected to be subject to contamination. The other common limitation faced by our and many other studies is the reliance on microbial reference databases. These databases are heavily biased towards microbial species with medical or agricultural significance⁶⁰, restricting read-based analyses of metagenomics data from non-human hosts. A large proportion of microbial taxa identified in our samples remained unassigned to any source microbiome by SourceTracker. Although some of these taxa may be members of other microbial communities not included as a source in our analysis, we expect that by studying a novel environment (the non-human oral microbiome) we will encounter unique microbial taxa. In the absence of a dedicated reference database from the study species, an alternative

approach is *de novo* MAG assembly^{60,61}. Given the fragmented and damaged nature of ancient DNA, this technique poses great challenges for historical microbiome studies. However, our study demonstrates that with deeper sequencing MAG recovery may be able to complement read-based analyses of historical microbiome samples. This is particularly important in non-model species, where reference database bias is a problem.

Discussion

With the development of high-throughput sequencing techniques and methodological advances in metagenomic analyses of ancient samples, the time is ripe to venture into the study of the diversity of environmental and host-associated microbial communities from the past. The temporal perspective provided by historical and ancient samples allows us to study many fundamental evolutionary processes, including those with direct relevance to human and ecosystem health. Our study describes a rigorous roadmap for the analysis of historical microbiomes and illuminates a magnitude of biological questions that can benefit from the study of dental calculus remains. We demonstrate that a single sample source can be used to link the host microbial community to host genetics, diet and even disease. While larger sample sizes are needed to substantiate the biological inferences of our preliminary findings, our study establishes dental calculus as a tool for the exploration of a broad variety of research topics. Questions of interest include the evolution of host-associated microbiome through multi-generational changes in the host environment, such as climate change, alterations in host population demography and genetic diversity, the invasion of new habitats, or changes in competitive regimes. Temporal sampling of dental calculus also provides insights into oral disease emergence and the progression of antimicrobial resistance in host-associated microbiomes. These processes can be of interest to both evolutionary biologists and the public health sector, since wild animal populations can act as sources and reservoirs for emerging zoonotic pathogens^{62,63} and contribute to the spread of antimicrobial resistance^{64,65}. In addition, our dietary results indicate that dental calculus can be used to infer population-specific dietary

characterises, particularly if complemented with microfossil analysis from the same material and stable isotope analysis of teeth or bones^{66–68}, which can be extended to extinct species (e.g. ⁶⁹). We envision that in the future, our integrative approach will be applied to many different questions in ecology and evolution.

Materials and methods

Sample collection

Dental calculus was collected from two eastern gorilla (*Gorilla beringei*) specimens from the Royal Museum for Central Africa (Brussels, Belgium), as well as five reindeer (*Rangifer tarandus*) and six brown bear (*Ursus arctos*) specimens from the Swedish Natural History Museum (Stockholm, Sweden). Skulls were macroscopically examined for dental calculus deposits and evidence of oral diseases. Calculus was removed from the surfaces of the teeth with disposable sterile scalpel blades and deposited in sterile microcentrifuge tubes.

Sample processing and DNA extraction

All laboratory protocols were performed in a dedicated ancient DNA laboratory following stringent procedures to minimise contamination¹⁸. Initially, calculus samples were processed without surface decontamination (Supplementary Table 1). We then tested the effect of surface contamination on the two gorilla samples using three treatments: exposing the calculus to UV light (245 nm) for 10 min, washing the calculus in 500 µl of 0.5 ethylenediaminetetraacetate (EDTA) for 30 seconds, and no surface decontamination as a control. Based on real-time PCR of libraries prepared from these samples (see below), we determined that neither decontamination treatment substantially reduced library quantity and thus continued with a surface decontamination procedure consisting of the UV light exposure followed by the EDTA wash for all subsequent calculus samples (Supplementary Table 1). The pellet was taken forward for DNA extraction. Sample weight ranged from < 5 mg to 20 mg. DNA extractions were performed following a silica-based method⁷⁰. Briefly, samples were incubated

overnight at 37°C in extraction buffer (0.45M EDTA, 0.25 mg/ml Proteinase K). The DNA from the supernatant was combined with binding buffer (3M sodium acetate, 5M guanidine-hydrochloride, 40% (v/v) isopropanol, 0.05% (v/v) Tween-20) and processed through the spin columns from High Pure Viral Nucleic Acid Large Volume kits (Roche, Switzerland). Purified DNA was eluted in either 45 µl of EB buffer (10 mM tris-hydrochloride (pH 8.0) (Qiagen, The Netherlands) or 45 µl of TE buffer (10 mM tris-hydrochloride (pH 8.0), 1 mM EDTA), both supplemented with 0.05% (v/v) Tween-20 (Supplementary Table 1). Ten blanks were included during five extraction batches and carried through library preparation.

Library preparation and sequencing

Double-stranded Illumina libraries were prepared following ⁷¹ and we included a double-barcoding double-indexing strategy to guard against index hopping and retain absolute certainty about sample of origin^{40,72}. Briefly, blunt-end repair reactions were performed using 20 µl of each extract and purified using MinElute columns with elutions in 22 µl of EB buffer (Qiagen, The Netherlands). Adapters containing inline 7 bp barcodes (Supplementary Tables 1 and 6) were ligated to both ends of the blunt-ended DNA, which was subsequently purified with MinElute columns and eluted in 22 µl of EB buffer. After the adapter fill-in reaction, *Bst* 2.0 polymerase (New England BioLabs, USA) was inactivated with a 15 min incubation at 80°C. Seven library blanks were included in four library preparation batches. The adapter-ligated libraries were quantified using a real-time PCR assay with preHyb primers⁴⁰ (Supplementary Table 6) and the estimated fragment number was used to approximate the number of indexing PCR cycles needed for sequencing. All extraction and library blanks were consistently lower in DNA content than samples, as measured by real-time PCR, thus one extraction blank and one library blank were randomly selected for subsequent indexing and sequencing. Libraries were double-indexed with unique P5 and P7 indices so that each sample had a unique barcode-index combination (Supplementary Table 1). Indexing PCR reactions were performed with 18 µl of adapter-ligated library in 50 µl reactions, with 1 µl PfuTurbo C_x hotstart

polymerase (2.5 U/μl, Agilent Technologies, USA), 5 μl 10X PfuTurbo C_x reaction buffer, 0.5 μl dNTP mix (25 mM) and 1 μl of each indexing primer (10 μM). After an initial incubation for 2 min at 95°C, 12 cycles of 30 sec at 95°C, 30 sec at 59°C and 1 min at 72°C were performed, followed by a final step of 10 min at 72°C. Reactions were purified with MinElute columns and eluted in 10 μl of EBT buffer. The indexed libraries were quantified using a real-time PCR assay with i7 and i5 indexing primers⁴⁰ (Supplementary Table 6) and library DNA fragment length distribution was determined by the 2200 TapeStation system. The mean fragment length after library preparation and excluding the 148 bp adapter sequences was 70 bp, similar to what has been observed in previous historical sequencing libraries^{12,73}. Five microliters of each sample library were pooled along with a randomly selected extraction blank and library blank. Size selection was performed on the pooled library with AMPure XP beads (Beckman Coulter, USA), selecting for fragments approximately 100-500 bp in length, and the purified library eluted in 36 μl of EB buffer. The final pooled library was quantified using a Qubit High Sensitivity fluorometer and on the 2200 TapeStation system. The pooled library was first sequenced by SciLifeLab Uppsala on 2 lanes of the Illumina HiSeq 2500 using paired-end 125 bp read length v4 chemistry, followed by an additional 2 lanes on the Illumina HiSeq 2500 in rapid mode using paired-end 100 bp read length v2 chemistry.

Data processing

Sequencing data were demultiplexed and assigned to each sample with an in-house python script based on the unique combination of barcodes and indices. Overlapping paired-end reads were merged and adapters and low quality terminal bases (phred scores ≤ 30) were removed with AdapterRemoval v2.2.2⁷⁴. Barcode sequences were removed from the 5' and 3' ends of merged reads with an in-house python script. Forward reads from the unmerged read pairs (i.e. pairs that did not contain overlapping regions of at least 11 bp between the forward and reverse reads) were also retained for analyses. The 5' barcode was removed with an in-house python script and the 3' barcode with any remaining adapter sequence removed with AdapterRemoval. Reads from the two

lanes within the same sequencing run were concatenated into a single file per sample. Merged reads from the two separate runs were also concatenated into a single file per sample. Reads with a length < 30 bp were filtered out with AdapterRemoval and reads with mean base quality < 30 were filtered out with PrinSeq-Lite v0.20.4⁷⁵. Duplicate reads were removed by randomly keeping one read among those reads having an identical sequence. The Illumina sequencing control phage PhiX was spiked into our sequencing runs and has been reported to have been erroneously integrated into many microbial reference genomes^{76,77}. Reads were therefore mapped to PhiX (accession: GCA_000819615.1) with bwa mem v0.7.17^{78,79} and the unmapped reads retained with SAMTools v1.9⁸⁰ and BEDTools v2.21.0⁸¹. To remove reads originating from the host organism and from human contamination, we mapped all reads in a sample to a combined reference consisting of the human genome⁸² (RefSeq accession: GCF_000001405.38) and the respective host genome (GCF_000151905.2 (*Gorilla gorilla gorilla*)⁸³, GCF_003584765.1 (*U. arctos horribilis*)⁸⁴ and GCA_004026565.1 (*Rangifer tarandus*)) with bwa mem. The unmapped reads were retained with SAMTools for downstream microbial taxonomic analyses.

Taxonomic assignment

Merged and unmerged unmapped reads were assigned taxonomy using the *k*-mer based classifier Kraken2 v2.0.7¹⁹ with the standard Kraken2 database (all archaea, bacteria, viruses and the human genome in RefSeq; built 2019-03-01). We used Kraken-biom (<https://github.com/smdabdoub/kraken-biom>) to extract the summarised number of reads assigned at the genus and species levels. These assignments were taken into R (<https://www.R-project.org/>) for further processing and analysis.

We also taxonomically binned the contamination-filtered reads (see below) against a wider database by alignment with MALT v0.4.0⁴⁴ (parameters as per ¹²: semi-global alignment, 85% minimum identity threshold, 0.01% minimum support threshold and a top percent value of 1.0) against an

index built from the entire NCBI nucleotide database
(<ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nt.gz>, modified 31-05-2018) using a step size of 4 to
reduce index size. The results were viewed in MEGAN Community Edition v6.10.5⁴⁵. Due to relatively
low specificity leading to spurious alignments that required manual curation, we did not use the
MALT/MEGAN results for microbiome analyses, but took them forward for specific bacterial
alignments as well as for eukaryotic taxonomic assignments (see below).

Identifying contamination

We used several complementary approaches to identify and remove contaminating bacterial taxa
from the Kraken taxonomy assignments.

SourceTracker analysis: Potential contribution of source microbiomes to samples was estimated with
SourceTracker v1.0²² in R, using the Kraken2 genus- and species-level assignments. Source
sequencing reads were processed through the same pipeline as sample reads, and included soil⁸⁵,
human skin⁸⁶, human gut^{37,38}, human supragingival plaque^{37,38}, human medieval dental calculus¹² and
laboratory reagent²⁰ microbiomes (Supplementary Table 7).

Fragment length: Given our library fragment length distribution peak at 70 bp, consistent with
historical degraded DNA, and our paired-end sequencing approach of 100 bp, we expected that the
majority of reads stemming from authentic historical microorganisms would be successfully merged
by AdapterRemoval. In contrast, modern contaminating taxa with longer fragment lengths should be
predominantly found among the unmerged reads. We therefore compared the raw read counts of
each taxon between the merged and unmerged reads on a per sample basis (Supplementary Figure
2). The difference in read number between the unmerged and merged reads for each taxon in each
sample was calculated and the distribution of all differences > 0 (i.e. more reads observed in the
unmerged reads) was investigated (Supplementary Figure 2). Taxa identified as outliers (defined as a

difference > 1.5 interquartile ranges above the third quartile) in at least one sample were filtered out as putative contaminants. While it is possible that certain microbial structural features, e.g. the thick cell wall in Gram-positive bacteria, might bias preservation towards specific taxa and thereby affect our fragment length analysis¹⁸, a recent study of microbial DNA perseverance in human dental calculus found no association between cellular structures and microbial DNA fragmentation¹². Nonetheless, we confirmed that the majority of taxa in our putative contaminant list were non-host-associated environmental microorganisms through a literature search.

Statistical identification of contaminants: Contamination derived from environmental sources (including the laboratory) is expected to be present in samples at approximately similar absolute amounts and will therefore be disproportionally more abundant in samples with low numbers of total sequenced reads^{20,87}. This concept has been implemented in the R package decontam²¹, which was used to identify and remove all taxa that showed an inverse relationship between taxon abundance and total number of sequences included in the sequencing pool per sample, as estimated by real-time PCR.

Presence in blanks: SourceTracker analysis demonstrated that the two blank samples contained taxa associated with soil and human skin microbiomes (Supplementary Figure 1). However, low levels of sample cross-contamination are common during laboratory processing. Conservatively filtering out all taxa observed in the blanks might remove genuine signals. We therefore screened all taxa that were present in the blanks against the Human Oral Microbiome Database (HOMD)⁸⁸. Taxa that were not present in HOMD were classified as contaminants and removed from further analysis.

Removal of contaminants from reads: For all downstream analyses, reads mapping to all species-level taxa (bacteria, archaea and viruses) identified as putative contaminants were removed from the fastq files. To this end, one reference genome for each taxon was downloaded from GenBank

(Supplementary Table 8; accessed 2019-04-11) and reads were mapped to the reference genomes with bwa mem v0.7.17, relaxing the mismatch parameter (-B) to 3 in order to map reads from closely related strains. Only unmapped reads were passed onto further analyses.

Microbial analyses

Genome size normalisation: Taxa with larger genomes will generally contribute more sequencing reads to a library, biasing read-based abundance estimates⁸⁹. To account for this bias, the number of reads per taxon within a sample was normalised by dividing by the estimated average genome size of each respective taxon. Estimated average prokaryotic and viral genome sizes were calculated using publicly available genome sizes from the RefSeq database (accessed 2019-02-15)⁹⁰. In cases where no genome size data were available for a given species, the average genome size of taxa in that genus was used.

Abundance filtering: Taxa present at < 0.03% relative abundance (normalised read count divided by sum of normalised read counts in a sample) were removed, as filtering of low-abundance species reduces false-positive taxonomic assignments⁹¹. The filtering threshold was selected by testing a series of thresholds commonly applied in metagenomics studies, ranging from 0.01 – 0.1%⁹²(Supplementary Figure 13). From this analysis, we identified the threshold (0.03%) that yielded a microbial community with a complexity which was most similar to what has been observed in other dental calculus¹² and oral microbiome studies^{93–96} (i.e. approximately 100-300 taxa).

Abundance normalization: In high-throughput microbiome sequencing datasets, the total number of reads obtained is an arbitrary value set by the sequencing instrument and absolute abundance of each taxon is unknown⁹⁷. Therefore, to account for the compositional nature of the data, we applied the centred log-ratio (CLR) transformation⁹⁸. Because log transformation is only possible for positive values, we dealt with 0 count values by adding a pseudo-count to the normalised read count for

every taxon in every sample. Due to the genome size normalisation, we set the pseudo-count to the equivalent of one read divided by the average genome size for all taxa.

Statistical analyses: We used the R package *vegan*⁹⁹ for diversity estimates. The Shannon index¹⁰⁰ was used to estimate alpha diversity, via the diversity function on the raw read count data (i.e. prior to genome size normalisation and CLR transformation), because the calculation of the metric required positive integers. Differences between host species were investigated with an AIC-based stepwise regression to determine the best-fit general linear model, with surface decontamination, sequencing depth (number of unmapped reads per sample) and proportion of human reads per sample included as covariates.

Beta diversity (a measure of inter-individual variation) was investigated based on the presence/absence of microbial taxa. Jaccard distances, calculated using the *vegdist* function, were used for ordination with non-metrical multidimensional scaling (NMDS), and permutational multivariate analysis of variance (PERMANOVA) with the *adonis* function in *vegan*. Host species, surface decontamination, number of unmapped reads per sample, and proportion of human reads per sample were included as covariates in the *adonis* model. To determine whether differences in within-group variation between host species was biasing inferences of a host-specific oral microbiome signature, a distance-based test for homogeneity of multivariate dispersions was performed with the *vegan* function *betadisper*. No such differences were detected, adding confidence to the PERMANOVA results. The same set of beta analyses were repeated using Euclidean distances calculated on the CLR normalised abundance data.

To identify taxa which discriminated between host species, we carried out a random forest classification based on presence/absence data using the R package *randomForest*¹⁰¹ with 10000 trees. This approach reports the out-of-bag estimated error (how often an individual was incorrectly

assigned to a host species) and the variable importance (mean decrease in accuracy) of each taxon, which reflects the importance of the given taxon in determining the correct host species. For this analysis, the gorilla samples were excluded due to low sample size (n=2). We also investigated a subset of taxa unique to each host species and shared by > 50% of samples from this species to determine whether they were likely oral taxa. As the two gorilla samples did not share any taxa, we investigated all of the taxa unique to each of the two samples. We then compared these sets of taxa to the HOMD⁸⁸ and classified those present as 'oral'. Taxa not present in the HOMD were manually classified as 'oral' (based on the presence in the oral microbiome of non-human mammals), 'host-associated' (present in non-oral mammalian microbiomes) or 'not host-associated' through a literature search.

Functional analysis

Classification: The functional genic content of the microbial community in dental calculus was characterised by running the contamination-filtered reads through the HUMAnN2 pipeline³⁴, which identifies species-specific genes with the taxonomic profiler MetaPhlAn2¹⁰² and a built-in microbial pangenome database representing all known nonredundant protein-coding potential for each species identified by MetaPhlAn2, and more general functional characterisation by alignment with DIAMOND¹⁰³ against the UniRef90¹⁰⁴ database. The mappings are weighted by quality and sequence length to estimate species-specific and total community gene family abundance. Metabolic pathways are also reconstructed based on genes annotated to metabolic enzymes in MetaCyc³⁶, and the pathway abundance and coverage are reported.

CLR normalisation, NMDS ordination and PERMANOVA were carried out for the pathway abundances as for the microbial analyses outlined above. Core pathways were defined as those containing > 50% of the required enzymes in a sample (i.e. per sample pathway coverage > 0.5) for

the total community (i.e. not stratified by microbial species). Relative (proportional) abundance for specific pathways stratified by microbial species was also calculated by HUMAnN2 for each sample.

Antimicrobial resistance profiles

Global: Contamination-filtered reads were aligned to the Comprehensive Antibiotic Resistance Database (CARD) v3.0.1 (modified 2019-02-19)⁴³, a curated collection of resistance determinant sequences, with blast v2.7.1+^{105,106} using default parameters. The Antibiotic Resistance Ontology (ARO) accession number associated with each CARD sequence was used to obtain the AMR gene family of each sequence. Where reads matched multiple sequences in the CARD, the best hit was identified based on highest bit score. Where multiple hits had the same bit score, we compared the ARO terms and if all hits shared the same ARO information, we randomly chose one hit to carry forward. When ARO information was not identical, we manually identified a common higher level term of the hits: for example, if all hits were to different beta-lactamases, we referred to the gene family as “unclassified beta-lactamase”. We then calculated the number of reads per ARO accession per sample and normalised it by sample sequencing depth (number of reads post-contamination filtering). The abundance of ARO gene families for each sample was calculated by summing across the ARO accession abundances associated with each gene family.

Targeted: All complete genomes for *A. baumannii* (n=133) available at NCBI Genome (accessed 2019-03-28) were downloaded and the fasta files concatenated into one file representing the ‘pangenome’. The plasmid sequences were removed to exclude analysis of reads that map to plasmids, since plasmids can be exchanged between bacterial species through horizontal gene transfer, which can lead to erroneous taxonomic identification. Reads assigned by MALT to *A. baumannii* on at least the species level were identified in each sample by MEGAN. These reads were then mapped to the plasmid-free pangenome with bwa mem. All mapped reads were extracted with SAMTools (samtools fastq -F4) and aligned to CARD as per the global AMR analysis outlined above.

The abundance of *A. baumannii* ARO gene families for each sample was calculated by summing across ARO accession abundances normalised by number of reads mapping to each bacteria.

Marker-based: Contamination-filtered reads were also compared to a set of AMR gene family marker sequences built via ShortBRED¹⁰⁷ from CARD v3.0.1 (modified 2019-02-19) using UniProt Reference Cluster 90 (UniRef90, modified 2019-05-08)¹⁰⁸ as the background. Normalised counts (reads per kilobase of reference sequence per million sample reads) for each family in each sample were subsequently analysed as above.

Metagenome-assembled genome (MAG) recovery

We attempted to recover MAGs from our samples following a similar strategy as described in Zhou *et al.* 2018¹⁰⁹ and Parks *et al.* 2017⁶¹. Reads were assembled into contigs with MEGAHIT¹¹⁰. Sample depth profiles were generated by mapping reads back to the contigs with BBMap v38.08 (Bushnell B., *BBMap*, <https://sourceforge.net/projects/bbmap/>). Contigs > 1500 bp in length were grouped into genome bins based on co-abundance and sequence composition probabilities using MetaBAT2¹¹¹, where each bin theoretically represents the genome of a single strain. The quality of each bin was assessed using the lineage-specific workflow in CheckM¹¹². High quality bins were classified as those with an estimated > 90% genome completeness and < 5% strain contamination⁶¹. The taxonomic lineage of each high quality bin, as defined in the Genome Taxonomy Database (GTDB)¹¹³, was determined with GTDB-Tk (<https://github.com/ecogenomics/gtdbtk>) using the classify workflow.

mapDamage profiles

Deamination rates and plots were generated using mapDamage¹¹⁴ via the EAGER pipeline⁴⁶. During this investigation, we observed unusual damage patterns in several of our samples – rather than the expected incremental rise in deamination towards read ends, several samples showed a drop in the

frequency of C-to-T substitutions from the penultimate base to the terminal base (e.g. Figure 2d [*L. paracasei* from Ua9] compared to Supplementary Figure 3 [*Streptococcus* sp. Z15 from Rt7]). It has been suggested previously that barcodes with a terminal G or C have reduced ligation efficiency to damaged cytosines at 5' ends of reads⁴⁰. We therefore examined damage profiles for 14 most abundant bacterial taxa in the post-filtering dataset (Supplementary Table 9), only considering cases with more than 10000 reads mapping to a taxon reference genome in a given sample. We found that samples containing barcodes with a terminal G or C were more likely to show the unusual pattern ($\chi^2 = 31.11$, $df = 1$, $p\text{-value} < 0.0001$) (Supplementary Table 10).

Dietary component recovery

Counts of reads assigned by MALT to eukaryotic taxa were exported from MEGAN at the family and genus level and imported into R. Taxa with <10 assigned reads across all samples were excluded from further analysis. At the family level, assigned read counts were normalised across samples by adding a pseudo-count of 0.1 and applying the CLR transformation. The proportion of specific genera per sample was calculated by taking genus level read counts and normalising by sample sequencing depth.

Host genome recovery

We collected published whole genome sequence data for 1 reindeer (GCA_004026565.1), 1 white-tailed deer (GCA_002102435.1), 47 gorillas^{115–117}, 24 brown bears and 3 polar bears^{118–121}. Reads were mapped either merged (ancient samples), or paired-end (modern samples) to an outgroup reference genome assembly (white-tailed deer: GCA_002102435.1, human: GCA_000151905.3 and polar bear: GCA_000687225.1) using bwa mem on default settings. Next, we excluded reads with a mapping quality score < 30 and removed duplicate reads with SAMTools v0.1.19. Additionally, for each study host species, we mapped all reads to the mitochondrial references of white-tailed

deer (NC_015247.1), polar bear (NC_003428.1) and western gorilla (NC_011120.1), respectively, following the same pipeline.

To investigate host genomic variation, we generated pseudo-haploid sequences for each individual by randomly selecting a single high-quality base call (BaseQuality ≥ 30 , MapQuality ≥ 30) at each site covered by at least 1 read, excluding sites within repetitive regions (as identified from the repeatmask tracts) and for modern genomes sites with >2 times genome-wide coverage to minimize false bases from spurious mappings. A reference set of high quality polymorphisms was made from all bi-allelic autosomal sites in the modern genomes and we then projected the low-coverage dental calculus samples onto the precalculated PCA space using the `lsqproject` function in `smartpca`¹²².

Mitochondrial genome variation was investigated following the same pipeline but calling the majority allele at each covered site. We then used `popart`¹²³ to create haplotype networks for all near-complete mitochondrial genomes ($>90\%$ complete). For the incomplete mitochondrial genomes, we calculated pairwise-divergence to each of the modern genomes to obtain the most likely closest related haplotype.

In addition we also obtained *de novo* mitochondrial genomes using `MITObim v1.9`¹²⁴, a mitochondrial baiting and iterative mapping method, which reconstructs complete mitochondrial genomes from next generation sequencing data. The mitochondrial genomes of the western lowland gorilla (*Gorilla gorilla gorilla*, NC_011120.1), the brown bear (*Ursus arctos*, NC_003427.1) and the reindeer (*Rangifer tarandus*, KM506758.1) were used as bait sequences for the mitochondrial genome assembly of the eastern gorilla, brown bear and reindeer samples, respectively. The merged host reads were used as input files for each sample. All mitochondrial genomes were annotated with the `MITOS WebServer`¹²⁵ using default parameters.

732

733 *Data availability*

734 All sequencing data are archived at the European Nucleotide Archive under the accession number

735 PRJEBxxxxx. Sample metadata is provided in Supplementary Table 1.

736

737 **Acknowledgements**

738 We thank Daniela Kalthoff and Emmanuel Gilissen for providing access to the mammalian museum

739 specimens, Peter Niehoff and Riana Minocher for assisting with sampling, and James Fellows Yates,

740 Franziska Aron, Johanna von Seth and Tatiana Feuerborn for laboratory and analysis advice.

741 Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of

742 the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ

743 Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg

744 Foundation. We also acknowledge the National Bioinformatics Infrastructure (NBIS) for providing

745 computational resources to this project. The project was supported by the Formas grant (2016-

746 00835) and the National Projects grant (NP00039) to KG.

747

748 **References**

749 1. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences.

750 *Proc. Natl. Acad. Sci.* **110**, 3229–3236 (2013).

751 2. Youngblut, N. D. *et al.* Host diet and evolutionary history explain different aspects of gut

752 microbiome diversity among vertebrate clades. *Nat. Commun.* **10**, 2200 (2019).

753 3. Foster, K. R., Schluter, J., Coyte, K. Z. & Rakoff-Nahoum, S. The evolution of the host

754 microbiome as an ecosystem on a leash. *Nature* **548**, 43–51 (2017).

755 4. Gomez, A. *et al.* Host genetic control of the oral microbiome in health and disease. *Cell Host*

756 *Microbe* **22**, 269–278 (2017).

757 5. Chevalier, C. *et al.* Gut Microbiota Orchestrates Energy Homeostasis during Cold. *Cell* **163**,

- 758 1360–1374 (2015).
- 759 6. Carmody, R. N. *et al.* Diet dominates host genotype in shaping the murine gut microbiota. *Cell*
760 *Host Microbe* **17**, 72–84 (2015).
- 761 7. Koren, O. *et al.* Host remodeling of the gut microbiome and metabolic changes during
762 pregnancy. *Cell* **150**, 470–480 (2013).
- 763 8. Warinner, C., Speller, C., Collins, M. J. & Lewis, C. M. Ancient human microbiomes. *J. Hum.*
764 *Evol.* **79**, 125–136 (2015).
- 765 9. Adler, C. J. *et al.* Sequencing ancient calcified dental plaque shows changes in oral microbiota
766 with dietary shifts of the Neolithic and Industrial revolutions. *Nat. Genet.* **45**, 450–455 (2013).
- 767 10. Warinner, C. *et al.* Pathogens and host immunity in the ancient human oral cavity. *Nat.*
768 *Genet.* **46**, 336–344 (2014).
- 769 11. Weyrich, L. S. *et al.* Neanderthal behaviour, diet, and disease inferred from ancient DNA in
770 dental calculus. *Nature* **544**, 357–361 (2017).
- 771 12. Mann, A. E. *et al.* Differential preservation of endogenous human and microbial DNA in
772 dental calculus and dentin. *Sci. Rep.* **8**, 9822 (2018).
- 773 13. Holmes, M. W. *et al.* Natural history collections as windows on evolutionary processes. *Mol.*
774 *Ecol.* **25**, 864–881 (2016).
- 775 14. Van Der Valk, T. *et al.* Significant loss of mitochondrial diversity within the last century due to
776 extinction of peripheral populations in eastern gorillas. *Sci. Rep.* **8**, 1–10 (2018).
- 777 15. van der Valk, T., Díez-del-Molino, D., Marques-Bonet, T., Guschanski, K. & Dalén, L. Historical
778 genomes reveal the genomic consequences of recent population decline in eastern gorillas.
779 *Curr. Biol.* **29**, 165–170 (2019).
- 780 16. van't Hof, A. E., Edmonds, N., Dalíková, M., Marec, F. & Saccheri, I. J. Industrial Melanism in
781 British Peppered Moths Has a Singular and Recent Mutational Origin. *Science (80-.).* **332**,
782 958–60 (2011).
- 783 17. Alberdi, A., Aizpurua, O., Bohmann, K., Zepeda-Mendoza, M. L. & Gilbert, M. T. P. Do

- 784 vertebrate gut metagenomes confer rapid ecological adaptation? *Trends Ecol. Evol.* **31**, 689–
785 699 (2016).
- 786 18. Key, F. M., Posth, C., Krause, J., Herbig, A. & Bos, K. I. Mining Metagenomic Data Sets for
787 Ancient DNA: Recommended Protocols for Authentication. *Trends Genet.* **33**, 508–520 (2017).
- 788 19. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using
789 exact alignments. *Genome Biol.* **15**, R46 (2014).
- 790 20. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based
791 microbiome analyses. *BMC Biol.* **12**, 87 (2014).
- 792 21. Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical
793 identification and removal of contaminant sequences in marker-gene and metagenomics
794 data. *Microbiome* **6**, 226 (2018).
- 795 22. Knights, D. *et al.* Bayesian community-wide culture-independent microbial source tracking.
796 *Nat. Methods* **8**, 761–763 (2011).
- 797 23. Lassalle, F. *et al.* Oral microbiomes from hunter-gatherers and traditional farmers reveal
798 shifts in commensal balance and pathogen load linked to diet. *Mol. Ecol.* **27**, 182–195 (2018).
- 799 24. Kennedy, R. *et al.* The microbiome associated with equine periodontitis and oral health. *Vet.*
800 *Res.* **47**, 49 (2016).
- 801 25. Gorrel, C. Periodontal disease and diet in domestic pets. *J. Nutr.* **128**, 2712S–2714S (1998).
- 802 26. Manville, Albert M., I. Variability of dental diseases in two populations of Great Lakes black
803 bears. *Bears Their Biol. Manag.* **8**, 129–134 (1990).
- 804 27. Wenker, C. J. *et al.* Dental health status and endodontic treatment of captive brown bears
805 (*Ursus Arctos* ssp.) living in the Bernese bear pit. *J. Vet. Dent.* **15**, 27–34 (1998).
- 806 28. Gao, X. Y., Zhi, X. Y., Li, H. W., Klenk, H. P. & Li, W. J. Comparative genomics of the bacterial
807 genus *Streptococcus* illuminates evolutionary implications of species groups. *PLoS One* **9**,
808 e101229 (2014).
- 809 29. Neves, B. G., Stipp, R. N., da Silva Bezerra, D., de Figueiredo Guedes, S. F. & Rodrigues, L. K. A.

810 Molecular detection of bacteria associated to caries activity in dentinal lesions. *Clin. Oral*
811 *Investig.* **21**, 2053–2061 (2017).

812 30. Tanzer, J. M., Livingston, J. & Thompson, A. M. The microbiology of primary dental caries in
813 humans. *J. Dent. Educ.* **65**, 1028–37 (2001).

814 31. Lazarevic, V., Whiteson, K., Hernandez, D., François, P. & Schrenzel, J. Study of inter- and
815 intra-individual variations in the salivary microbiota. *BMC Genomics* **11**, 523 (2010).

816 32. Socransky, S., Haffajee, A., Cugini, M., Smith, C. & Kent, R. J. Microbial complexes in
817 subgingival plaque. *J Clin Periodontol* **25**, 134–44 (1998).

818 33. Takahashi, N. & Nyvad, B. The role of bacteria in the caries process: Ecological perspectives. *J.*
819 *Dent. Res.* **90**, 294–303 (2011).

820 34. Franzosa, E. A. *et al.* Species-level functional profiling of metagenomes and
821 metatranscriptomes. *Nat. Methods* **15**, 962–968 (2018).

822 35. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopaedia of Genes and Genomes. *Nucl. Acids Res.*
823 **28**, 27–30 (2000).

824 36. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res.*
825 **46**, D633–D639 (2018).

826 37. Lloyd-price, J. *et al.* Strains, functions, and dynamics in the expanded Human Microbiome
827 Project. *Nature* **550**, 61–66 (2018).

828 38. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome.
829 *Nature* **486**, 207–214 (2012).

830 39. Prakash, T. & Taylor, T. D. Functional assignment of metagenomic data: Challenges and
831 applications. *Brief. Bioinform.* **13**, 711–727 (2012).

832 40. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase
833 treatment for screening of ancient DNA. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20130624
834 (2015).

835 41. Xie, G. *et al.* Community and gene composition of a human dental plaque microbiota

- obtained by metagenomic sequencing. *Mol. Oral Microbiol.* **25**, 391–405 (2010).
42. Tommasi, R., Brown, D. G., Walkup, G. K., Manchester, J. I. & Miller, A. A. ESKAPEing the labyrinth of antibacterial discovery. *Nature Reviews Drug Discovery* **14**, 529–542 (2015).
43. Jia, B. *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).
44. 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).
45. Huson, D. H. *et al.* MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* **12**, e1004957 (2016).
46. Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
47. Landecker, H. Antibiotic resistance and the biology of history. *Body Soc.* **22**, 19–52 (2016).
48. Dcosta, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011).
49. Vahaboglu, H. *et al.* Widespread detection of per-1-type extended-spectrum β -lactamases among nosocomial Acinetobacter and Pseudomonas aeruginosa isolates in Turkey: A nationwide multicenter study. *Antimicrob. Agents Chemother.* **41**, 2265–2269 (1997).
50. Manchanda, V., Sanchaita, S. & Singh, N. Multidrug Resistant Acinetobacter. *J. Glob. Infect. Dis.* **2**, 291–304 (2010).
51. Grueter, C. C. *et al.* Long-Term Temporal and Spatial Dynamics of Food Availability for Endangered Mountain Gorillas in Volcanoes National Park, Rwanda. *Am. J. Primatol.* **75**, 267–280 (2013).
52. Galbany, J. *et al.* Tooth wear and feeding ecology in mountain gorillas from Volcanoes National Park, Rwanda. *Am. J. Phys. Anthropol.* **159**, 457–465 (2016).
53. Skogland, T. Comparative summer feeding strategies of Arctic and Alpine Rangifer. *J. Anim. Ecol.* **49**, 81–98 (1980).
54. Staaland, H. *et al.* The introduction of reindeer to Brøggerhalvøya, Svalbard: grazing preference and effect on vegetation. *Rangifer* **13**, 15–19 (1993).

55. Zeitz, J. O. *et al.* Effect of the rumen ciliates *Entodinium caudatum*, *Epidinium ecaudatum* and *Eudiplodinium maggii*, and combinations thereof, on ruminal fermentation and total tract digestion in sheep. *Arch. Anim. Nutr.* **66**, 180–199 (2012).
56. Janssen, P. H. & Kirs, M. Structure of the archaeal community of the rumen. *Appl. Environ. Microbiol.* **74**, 3619–3625 (2008).
57. Kittelmann, S., Kirk, M. R., Jonker, A., McCulloch, A. & Janssen, P. H. Buccal swabbing as a noninvasive method to determine bacterial, archaeal, and eukaryotic microbial community structures in the rumen. *Appl. Environ. Microbiol.* **81**, 7470–7483 (2015).
58. Ozga, A. T. *et al.* Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus. *Am. J. Phys. Anthropol.* **160**, 220–228 (2016).
59. Xenikoudakis, G. *et al.* Consequences of a demographic bottleneck on genetic structure and variation in the Scandinavian brown bear. *Mol. Ecol.* **24**, 3441–3454 (2015).
60. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
61. Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542 (2017).
62. Gong, S. & Bao, L. The battle against SARS and MERS coronaviruses: Reservoirs and Animal Models. *Anim. Model. Exp. Med.* **1**, 125–133 (2018).
63. Reperant, L. A., Moesker, F. M. & Osterhaus, A. D. M. E. Influenza: from zoonosis to pandemic. *ERJ Open Res.* **2**, 00013-2016 (2016).
64. Vittecoq, M. *et al.* REVIEW: Antimicrobial resistance in wildlife. *J. Appl. Ecol.* **53**, 519–529 (2016).
65. Arnold, K. E., Williams, N. J. & Bennett, M. ‘Disperse abroad in the land’: The role of wildlife in the dissemination of antimicrobial resistance. *Biology Letters* **12**, 20160137 (2016).
66. Power, R. C., Salazar-García, D. C., Wittig, R. M., Freiberg, M. & Henry, A. G. Dental calculus evidence of Taï Forest Chimpanzee plant consumption and life history transitions. *Sci. Rep.* **5**,

- 1–13 (2015).
67. Mowat, G. & Heard, D. C. Major components of grizzly bear diet across North America. *Can. J. Zool.* **84**, 473–489 (2006).
68. Cristiani, E. *et al.* Dental calculus and isotopes provide direct evidence of fish and plant consumption in Mesolithic Mediterranean. *Sci. Rep.* **8**, 8147 (2018).
69. Henry, A. G., Brooks, A. S. & Piperno, D. R. Microfossils in calculus demonstrate consumption of plants and cooked foods in Neanderthal diets (Shanidar III, Iraq; Spy I and II, Belgium). *Proc. Natl. Acad. Sci.* **108**, 486–491 (2011).
70. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci.* **110**, 15758–15763 (2013).
71. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5448 (2010).
72. Valk, T. van der, Vezzi, F., Ormestad, M., Dalen, L. & Guschanski, K. Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. *Mol. Ecol. Resour.* (2019). doi:10.1111/1755-0998.13009
73. van der Valk, T., Lona Durazo, F., Dalén, L. & Guschanski, K. Whole mitochondrial genome capture from faecal samples and museum-preserved specimens. *Mol. Ecol. Resour.* **17**, e111–e121 (2017).
74. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
75. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
76. Mukherjee, S., Huntemann, M., Ivanova, N., Kyrpides, N. C. & Pati, A. Large-scale contamination of microbial isolate genomes by illumina Phix control. *Stand. Genomic Sci.* **10**, 18 (2015).

914 77. Hooper, R. *et al.* Host-derived population genomics data provides insights into bacterial and
915 diatom composition of the killer whale skin. *Mol. Ecol.* **28**, 484–502 (2019).

916 78. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.
917 *Bioinformatics* **25**, 1754–1760 (2009).

918 79. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*
919 *Prepr.* arXiv:1303.3997 (2013). doi:10.1186/s13756-018-0352-y

920 80. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–
921 2079 (2009).

922 81. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic
923 features. *Bioinformatics* **26**, 841–842 (2010).

924 82. Schneider, V. A. *et al.* Evaluation of GRCh38 and de novo haploid genome assemblies
925 demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 849–864
926 (2017).

927 83. Scally, A. *et al.* Insights into hominid evolution from the gorilla genome sequence. *Nature*
928 **483**, 169–175 (2012).

929 84. Taylor, G. A. *et al.* The genome of the North American brown bear or grizzly: *Ursus arctos* ssp.
930 *horribilis*. *Genes (Basel)*. **9**, 598 (2018).

931 85. Johnston, E. R. *et al.* Metagenomics reveals pervasive bacterial populations and reduced
932 community diversity across the Alaska tundra ecosystem. *Front. Microbiol.* **7**, 1–16 (2016).

933 86. Oh, J. *et al.* Biogeography and individuality shape function in the human skin metagenome.
934 *Nature* **514**, 59–64 (2014).

935 87. Lusk, R. W. Diverse and widespread contamination evident in the unmapped depths of high
936 throughput sequencing data. *PLoS One* **9**, e110808 (2014).

937 88. Chen, T. *et al.* The Human Oral Microbiome Database: a web accessible resource for
938 investigating oral microbe taxonomic and genomic information. *Database (Oxford)*. (2010).
939 doi:10.1093/database/baq013

- 940 89. Frank, J. A. & Sørensen, S. J. Quantitative Metagenomic Analyses Based on Average Genome
941 Size Normalization. *Appl. Environ. Microbiol.* **77**, 2513–2521 (2011).
- 942 90. O’Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic
943 expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
- 944 91. Velsko, I. M., Frantz, L. A. F., Herbig, A., Larson, G. & Warinner, C. Selection of appropriate
945 metagenome taxonomic classifiers for ancient microbiome research. *mSystems* **3**, e00080-18
946 (2018).
- 947 92. Peabody, M. A., Van Rossum, T., Lo, R. & Brinkman, F. S. L. Evaluation of shotgun
948 metagenomics sequence classification methods using in silico and in vitro simulated
949 communities. *BMC Bioinformatics* **16**, 363 (2015).
- 950 93. Dewhirst, F. E. *et al.* The human oral microbiome. *J. Bacteriol.* **192**, 5002–5017 (2010).
- 951 94. Kilian, M. *et al.* The oral microbiome - an update for oral healthcare professionals. *Br. Dent. J.*
952 **221**, 657–666 (2016).
- 953 95. Dewhirst, F. E. *et al.* The canine oral microbiome. *PLoS One* **7**, 1–12 (2012).
- 954 96. Dewhirst, F. E. *et al.* The feline oral microbiome: A provisional 16S rRNA gene based
955 taxonomy with full-length reference sequences. *Vet. Microbiol.* **175**, 294–303 (2015).
- 956 97. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome datasets are
957 compositional: And this is not optional. *Front. Microbiol.* **8**, 1–6 (2017).
- 958 98. Aitchison, J. *The Statistical Analysis of Compositional Data*. (Chapman & Hall, 1986).
- 959 99. Oksanen, A. J. *et al.* Vegan: Community ecology package. R package version 2.5-3.
960 <https://CRAN.R-project.org/package=vegan>. (2018). doi:10.4135/9781412971874.n145
- 961 100. Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* (1948).
962 doi:10.1002/j.1538-7305.1948.tb00917.x
- 963 101. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R news* **2**, 18–22
964 (2002).
- 965 102. Truong, D. T. *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat.*

966 *Methods* **12**, 902–903 (2015).

967 103. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat.*
968 *Methods* (2015). doi:10.1038/nmeth.3176

969 104. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: Comprehensive and
970 non-redundant UniProt reference clusters. *Bioinformatics* (2007).
971 doi:10.1093/bioinformatics/btm098

972 105. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search
973 tool. *J. Mol. Biol.* **215**, 403–410 (1990).

974 106. Madden, T. L. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421
975 (2009).

976 107. Kaminski, J. *et al.* High-Specificity Targeted Functional Profiling in Microbial Communities with
977 ShortBRED. *PLoS Biol.* **11**, e1004557 (2015).

978 108. Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B. & Wu, C. H. UniRef clusters: A
979 comprehensive and scalable alternative for improving sequence similarity searches.
980 *Bioinformatics* **31**, 926–932 (2015).

981 109. Zhou, Z. *et al.* Pan-genome Analysis of Ancient and Modern *Salmonella enterica*
982 Demonstrates Genomic Stability of the Invasive Para C Lineage for Millennia. *Curr. Biol.* **28**,
983 2420–2428 (2018).

984 110. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: An ultra-fast single-node
985 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
986 *Bioinformatics* **31**, 1674–1676 (2015).

987 111. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately
988 reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).

989 112. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing
990 the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
991 *Genome Res.* **25**, 1043–55 (2015).

992 113. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny
993 substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).

994 114. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. MapDamage2.0: Fast
995 approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**,
996 1682–1684 (2013).

997 115. Prado-Martinez, J. *et al.* Great ape genetic diversity and population history. *Nature* **499**, 471–
998 475 (2013).

999 116. Xue, Y. *et al.* Mountain gorilla genomes reveal the impact of long-term population decline
1000 and inbreeding. *Science (80-.)*. **348**, 242–245 (2015).

1001 117. Besenbacher, S., Hvilsom, C., Marques-Bonet, T., Mailund, T. & Schierup, M. H. Direct
1002 estimation of mutations in great apes reconciles phylogenetic dating. *Nat. Ecol. Evol.* **3**, 286–
1003 292 (2019).

1004 118. Barlow, A. *et al.* Partial genomic survival of cave bears in living brown bears. *Nat. Ecol. Evol.*
1005 *2018 210* **2**, 1563 (2018).

1006 119. Cahill, J. A. *et al.* Genomic Evidence for Island Population Conversion Resolves Conflicting
1007 Theories of Polar Bear Evolution. *PLoS Genet.* **9**, e1003345 (2013).

1008 120. Benazzo, A. *et al.* Survival and divergence in a small group: The extraordinary genomic history
1009 of the endangered Apennine brown bear stragglers. *Proc. Natl. Acad. Sci.* **114**, E9589–E9597
1010 (2017).

1011 121. Cahill, J. A. *et al.* Genomic evidence of geographically widespread effect of gene flow from
1012 polar bears into brown bears. *Mol. Ecol.* **24**, 1205–17 (2015).

1013 122. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genet.* **2**,
1014 e190 (2006).

1015 123. Leigh, J. W. & Bryant, D. popart : full-feature software for haplotype network construction.
1016 *Methods Ecol. Evol.* **6**, 1110–1116 (2015).

1017 124. Hahn, C., Bachmann, L. & Chevreux, B. Reconstructing mitochondrial genomes directly from

- 1018 genomic next-generation sequencing reads - A baiting and iterative mapping approach.
- 1019 *Nucleic Acids Res.* **41**, e129 (2013).
- 1020 125. Bernt, M. *et al.* MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol.*
- 1021 *Phylogenet. Evol.* **69**, 313–319 (2013).
- 1022