Quaternary International xxx (xxxx) xxx

Contents lists available at ScienceDirect

Quaternary International

journal homepage: www.elsevier.com/locate/quaint

Do I have something in my teeth? The trouble with genetic analyses of diet from archaeological dental calculus

Allison E. Mann [a,*], James A. Fellows Yates [b,c], Zandra Fagernäs [b], Rita M. Austin [d,e,f,i], Elizabeth A. Nelson [b,g,h], Courtney A. Hofman [e,f]

[a] Department of Biological Sciences, Clemson University, Clemson, SC, USA
[b] Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany
[c] Institut für Vor- und Frühgeschichtliche Archäologie und Provinzialrömische Archäologie, Ludwig Maximilian University, Munich, Germany
[d] Department of Anthropology, Smithsonian National Museum of Natural History, Washington, DC, USA
[e] Department of Anthropology, University of Oklahoma, Norman, OK, USA
[f] Laboratories of Molecular Anthropology and Microbiome Research, University of Oklahoma, Norman, OK, USA
[g] Department of Anthropology, University of Texas at Arlington, Arlington, TX, USA
[h] Institut für Naturwissenschaftliche Archäologie, Eberhard Karls Universität Tübingen, 72074, Tübingen, Germany
[i] Natural History Museum Oslo, University of Oslo, Oslo, Norway

ARTICLE INFO

Keywords:
Diet
Dental calculus
Metagenomics
Ancient DNA

ABSTRACT

Dental calculus and other preserved microbiome substrates are an attractive target for dietary reconstruction in past populations through a variety of physical, chemical, and molecular means. Recently, studies have attempted to reconstruct diet from archaeological dental calculus using archaeogenetic techniques. While dental calculus may provide a relatively stable environment for DNA preservation, the detection of plants and animals possibly consumed by an individual through DNA analysis is primarily hindered by microbial richness and incomplete reference databases. Moreover, high genomic similarity within eukaryotic groups - such as mammals - can obfuscate precise taxonomic identification. In the current study we demonstrate the challenges associated with accurate taxonomic identification and authentication of dietary taxa in ancient DNA data using both synthetic and ancient dental calculus datasets. We highlight common errors and sources of contamination across ancient DNA datasets, provide recommendations for dietary DNA validation, and call for caution in the interpretation of diet from dental calculus and other archaeological microbiome substrates.

1. Introduction

Diet is a fundamental component of human culture, biology, and evolution. Shifts in food procurement, production, and processing are inherently linked to shifts in human society and major evolutionary events (Goodman and Redclift, 2002; Larsen, 2003; Bocquet-Appel and Bar-Yosef, 2008; Ma et al., 2016; Andrews and Johnson, 2020). What people choose to eat (or not to eat) provides insight into cultural values and beliefs (Messer, 1984). Archaeological study of the interrelationships between people and foods, such as plants and animals, has revealed complex cultural practices and socio-political structures (Atalay and Hastorf, 2006; Tung et al., 2016; Morehart and Morell-Hart, 2015) organized around food consumption and distribution in which diet is directly related to social status (Cuéllar, 2013), ceremonial events, control of food supply, and the establishment of trade (Williams, 2010; Tung and Knudson, 2018). Likewise, the effect of environmental factors such as climate shifts and geologic processes on dietary resources can be investigated through the study of ancient diet (Messner and Stinchcomb, 2014; Nelson et al., 2016). How food was produced and procured throughout human history also provides important historical context for understanding human health in the modern era. Given its importance in understanding the human condition, archaeologists work to reconstruct past diets using a variety of techniques, including analyses of faunal assemblages (e.g., Blasco et al., 2013), paleobotanical remains (e.g., Pearsall, 2018), coprolite analysis (e.g., Reinhard and Bryant, 1992), and dental analyses of wear, development, and disease (e.g., Forshaw, 2014; Molnar et al., 1972; White and Folkens, 2005). The introduction of bulk biomolecular methods, such as stable isotope

* Corresponding author. Clemson University Department of Biological Sciences, 111 Jordan Hall, Clemson, SC, 29634.
E-mail address: amann3@g.clemson.edu (A.E. Mann).

https://doi.org/10.1016/j.quaint.2020.11.019
Received 3 August 2020; Received in revised form 2 November 2020; Accepted 12 November 2020
Available online 19 November 2020
1040-6182/

analysis from ancient teeth, bone, paleo-residues, and artefacts, has allowed for further resolution of paleodiets (e.g., Evershed, 2008; Makarewicz and Sealy, 2015).

More recently, advances in techniques that target ancient biomolecules such as proteins (Warinner et al., 2014a; Hendy et al., 2018; Geber et al., 2019) and the application of next generation DNA sequencing (Adler et al., 2013; Warinner et al., 2014b; Weyrich et al., 2017; Maixner et al., 2018; Ottoni et al., 2019; Ozga et al., 2019) have been explored as an alternative to more traditional methods. Using DNA to study the diet of past humans is an attractive approach, as in principle it enables identification of specific food items, instead of a general picture of bulk dietary intake (as with e.g., isotope studies), or only (generally) higher taxonomic-level identifications (e.g., with botanical remains) (Schlumbaum et al., 2008). Most earlier genetic analyses of diet from the archaeological record have focused on paleofeces or other preserved tissues, (Poinar et al., 1998; Hofreiter et al., 2003; Bon et al., 2012; Maixner et al., 2018) but the preservation of DNA is highly inconsistent and many early studies were plagued with authentication problems, which restricted further development of this area of research (Cooper and Poinar, 2000). A slightly more unexpected archaeological material - birch pitch mastics (ancient 'chewing gum') - has recently been shown to be a source of both human and dietary DNA (Jensen et al., 2019; Kashuba et al., 2019). However, these types of materials are relatively rare in the archaeological record and are only preserved in exceptional circumstances. In contrast, dental calculus, a mineralized microbial biofilm that is a long term source of host, microbial, and ambient biomolecules, is an increasingly attractive target for dietary reconstruction. This is due to dental calculus being more commonly found associated with well-preserved archaeological skeletons than other potential sources of dietary DNA (although this varies across collections, age of the individual, oral hygiene practices, and archaeological contexts) (White, 1997; Austin et al., 2019).

Dietary DNA analysis has been carried out using two major genetic techniques - amplicon sequencing and metagenomics. Amplicon sequencing involves the amplification and massive parallel sequencing of a specific region of DNA that targets one or more species of interest (e.g., Poinar et al., 1998; Sawafuji et al., 2020). While less expensive than metagenomics, amplicon sequencing has been shown to introduce biases due to the fragmented nature of ancient DNA (aDNA) (Ziesemer et al., 2015) and choosing a region of a genome suitable for species identification given the restrictions of aDNA (i.e., short DNA fragments), can be difficult. Metagenomic sequencing, on the other hand, involves untargeted DNA extraction and sequencing, and theoretically allows every DNA molecule to be detected using high throughput sequencing technology (also known as a 'metagenome' when multi-organismal DNA is present in a sample).

Dietary reconstructions from metagenomic aDNA datasets generated from human (Adler et al., 2013; Warinner et al., 2014b; Weyrich et al., 2017) and primate (Ottoni et al., 2019; Ozga et al., 2019) calculus have been previously attempted, but generally very few sequencing reads (sequenced DNA molecules) from potential dietary sources were recovered. For example, Warinner et al. (2014b) identified a total of 487 eukaryotic reads (0.0003% of the entire sequencing dataset), of which only three could be assigned to the species level for three likely dietary organisms, and Weyrich et al. (2017) identified an average of 571 putative dietary molecules per sample. With so little data, authenticating the antiquity of these DNA molecules is difficult, and often studies have instead sought to validate their finds by other methods, such as microfossil analysis (Henry and Piperno, 2008; Hardy et al., 2012; Tromp and Dudgeon, 2015) and paleoproteomics (Warinner et al., 2014a; Hendy et al., 2018; Geber et al., 2019). These studies raise an important question: why are putative dietary organisms difficult to detect in metagenomic datasets generated from dental calculus, which otherwise serves as a rich source of aDNA?

Accurate detection of animal and plant taxa that may represent foodstuffs in metagenomic datasets generated from dental calculus or other archaeological sources is hindered by (1) the fragmented and damaged nature of the DNA molecules, (2) the relatively low number of dietary biomolecules as compared to those derived from the endogenous and environmental microbial community, (3) differences in genome size and similarity across different species, and (4) reference database errors or lack of representation.

A primary challenge for all ancient metagenomic analyses is DNA sequence authentication. Post-mortem taphonomic processes, including intracellular nucleases, microbial activity, and natural chemical processes (Eglinton et al., 1991; Lindahl, 1993; Allentoft et al., 2012), continually act on ancient remains, resulting in molecular damage, fragmentation, and contamination from modern and environmental sources (Dabney et al., 2013). However, the types of aDNA damage that occur are generally consistent, with characteristic nucleotide misincorporation damage patterns and fragmentation (Sawyer et al., 2012). Depending on the tissue type, age, and general preservation, DNA fragments recovered with current laboratory methods are generally expected to be less than 100 base pairs (bp) and as small as 25bp (Briggs et al., 2007; Dabney et al., 2013; Glocke and Meyer, 2017). Fragmentation occurs rapidly (Kistler et al., 2017) due to DNA depurination, whereby the bonds between certain nucleotides are cleaved, leaving exposed abasic sites at the end of aDNA molecules (Hofreiter et al., 2001; Dabney et al., 2013). Next, single stranded overhangs (i.e., unequal ends of the two strands of DNA) that can occur after depurination, expose the single-stranded nucleotides to a greater chance of chemical processes - most commonly deamination - converting cytosines to uracils (a nucleotide normally found in RNA), through the loss of an amino group. When repairing these ends for next-generation sequencing, deaminated cytosines (C) are then 'mistakenly' read as thymines (T) and later as guanine (G) to adenine (A) mutations on the complementary strand of molecules, respectively (Hofreiter et al., 2001; Briggs et al., 2007; Green et al., 2009). This 'misreading' of the original nucleotide is often called 'miscoding' or 'misincorporation' lesions. Programs like mapDamage2.0 (Jónsson et al., 2013), PMDtools (Skoglund et al., 2014), and DamageProfiler (Neukamm et al., 2020) evaluate the ancient origin of DNA sequences, where the ends of ancient and fragmented molecules are expected to have increased frequencies of C to T misincorporations. These misincorporations are then plotted to assess whether DNA from that particular organism is modern contamination (i.e., no increased frequency) or may represent likely aDNA (Fig. 1a and b).

In addition to fragmented and damaged DNA, dental calculus is primarily a calcified microbial 'biofilm' (oral microbiome) that during the lifetime of the individual adheres to the enamel surface unless mechanically removed. As the vast majority of biomass in dental calculus is composed of bacterial cells (approximately 70%) (Marsh and Bradshaw, 1995), retrievable DNA informative of diet is therefore easily eclipsed by microbial DNA. Dietary DNA, in contrast, mostly will reside in saliva or loosely bound to oral surfaces in a transient manner and is primarily removed from the oral cavity through swallowing. Moreover, humans often rely on starch-rich organs of plants which do not contain as much retrievable DNA as other tissues. For example, leaves of the potato plant contain over 10 times and roots 20 times more DNA per gram of tissue than tubers (Scott et al., 1984). Cooking and other types of food processing may also degrade the quality of DNA before being consumed (Aslan et al., 2009; Gryson, 2010).

The remaining dietary fragments that will get trapped within the oral biofilm must be separated from the DNA of the microbial community during analysis. As aDNA fragments are generally short, any one aDNA fragment is much less likely to be unique or specific to a particular organism due to the limited number of combinations that four nucleotides can make. Therefore, DNA fragments originally from a dietary source can be misassigned to different unrelated taxa, particularly when it is underrepresented in the database. Next, although microbes and eukaryotes are quite divergent, genomes within eukaryotes can share large sequences of DNA even across divergent species. For example, while the human and chimpanzee genomes are approximately 96% similar

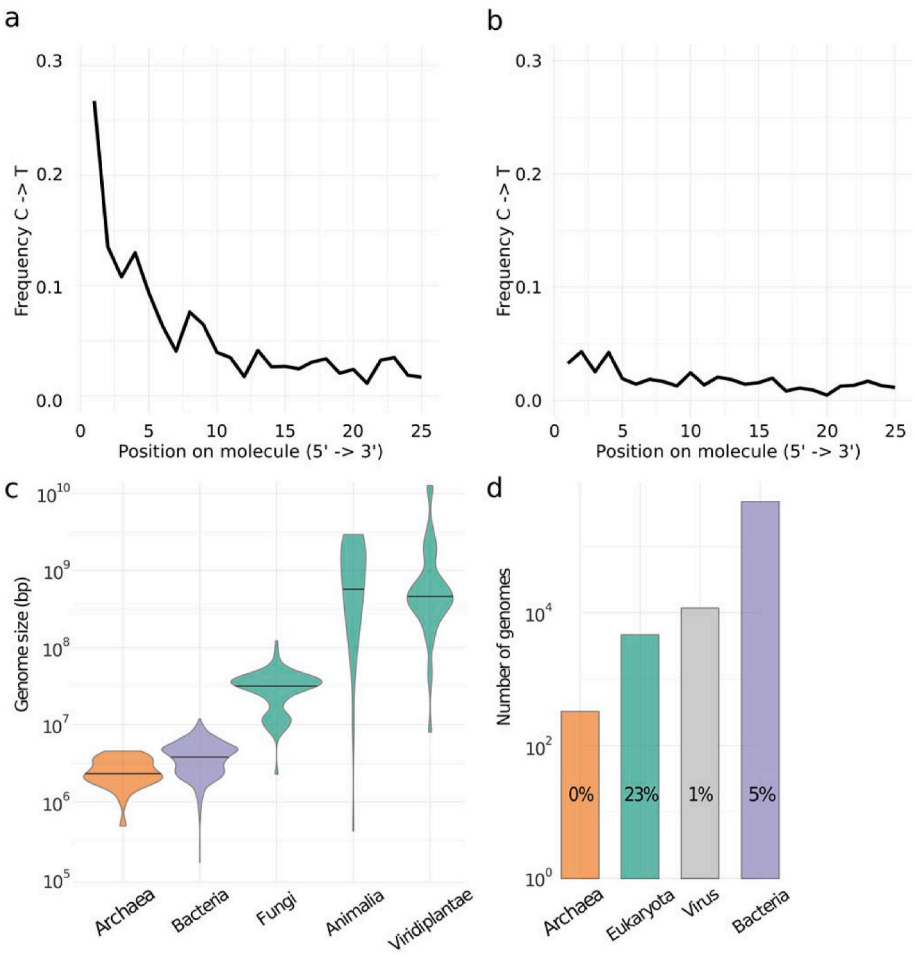

Fig. 1. Typical damage patterns indicative of ancient and modern samples and distribution of genomes in NCBI. (a) Authentic ancient samples are characterized by an increase in cytosine to thymine changes at the end of reads due to chemical damage on single stranded overhangs typical of short fragmented ancient DNA reads (ERR2900752). (b) Modern DNA exhibits no pattern of cytosine to thymine changes (ERR3307054). Plots generated using mapDamage2 (Jónsson et al., 2013). (c) Genome sizes per species (with horizontal line showing median). (d) Total number of sequenced genomes, with number on bars showing percentage of genomes from species generally considered to be model organisms.

(Waterson et al., 2005), the much more distantly related zebrafish (*Danio rerio*) also shares approximately 70% of its genome with humans (Howe et al., 2013). Meanwhile, strains of bacteria that belong to the same species may have up to 20% genome sequence divergence (Lan et al., 2000). Many metagenomic profiling tools rely on the concept of 'lowest common ancestor' (LCA) to assess the specificity of taxonomic assignment. LCA tools assess whether a sequencing read can map to different genomes equally well. If so, the tool will 'look up' a phylogenetic tree and assign the read to the most recent common ancestor of the two reference genomes. For example, a read aligning equally well to both a chimpanzee and human sequence will be assigned to the Hominini subfamily, and therefore, in the context of aDNA, makes informative identification more difficult (such as reported by Ottoni et al., 2019). The larger size of eukaryotic genomes can also pose a problem (Fig. 1c), as these often have much longer stretches of highly-repetitive non-coding regions. Reads deriving from these regions often have 'low sequence complexity' or low specificity, that again can map to many different organisms as they are not specific only to closely related taxa (e.g., at species or genus level).

Finally, identifying dietary eukaryotes from metagenomic datasets is complicated by poor representation of eukaryotes in publicly available reference databases. Genome sequencing projects have historically focused on economically or biomedically relevant organisms and despite recent pushes to improve the diversity of genome databases (Genome 10K, 2009), database biodiversity still remains low. As of March 2020, 4637 eukaryotic genomes were deposited in the NCBI reference database, 23% of which are from organisms that are generally used as experimental models in biomedical or other biological research (Fig. 1d). Lacking proper genetic references, the detection of many dietary eukaryotes is therefore limited. Furthermore, the quality of these modern reference genomes can vary - with many reconstructed genomes representing chimeras of the target taxon and contaminating species (both microbial, eukaryotic, and synthetic DNA from sequencing) (Merchant et al., 2014; Steinegger and Salzberg, 2020).

The purpose of the current study is to illustrate some of the specific challenges associated with dietary analysis of archaeological materials using metagenomic approaches, and to provide both researchers and peer reviewers guidelines to evaluate the confidence of dietary interpretations from metagenomic datasets. To this end we perform analysis to show the effects of low read numbers on the reliability of aDNA damage pattern detection. We generate a synthetic dataset to represent a typical oral microbial community from dental calculus with spiked-in genetic material from ten dietary eukaryotes to test the accuracy and precision of genetic analyses of diet in a mixed microbial community. To demonstrate this with real data, we also analyze 13 previously published archaeological and a modern dental calculus sample to identify common sources of reference database contamination or misidentification. We find that accurate identification of potential dietary organisms with a commonly used metagenomic profiling technique is weak and most reads are unable to be classified. Moreover, a subset of reads across all organisms, both eukaryotic and microbial, are classified to the wrong genus or species. We also report eukaryotic organisms detected across metagenomic datasets generated from dental calculus that likely represent reference database contamination or annotation errors.

This study highlights the caution that must be taken when analyzing potential dietary information from metagenomic datasets and provides suggestions for future directions into the robust validation of ancient

dietary DNA from dental calculus and other microbiome substrates.

2. Materials and methods

To demonstrate the effect of low numbers of identified dietary reads on damage profiles - one of the most typical authentication criteria of aDNA - we performed a downsampling experiment on single genomes of well-preserved aDNA samples. To assess whether these effects are the same across different types of organisms, we downloaded metagenomic aDNA sequencing data of a range of contexts: a human petrous bone as a typical target of aDNA studies (Gamba et al., 2014), an Atlantic cod (*Gadus morhua*) sample as a possible dietary species (Star et al., 2017), archaeological calcified nodules containing *Gardnerella vaginalis* to represent a clonal pathogenic microbe (Devault et al., 2017), and an ancient human dental calculus sample containing *Tannerella forsythia,* a common microbiome symbiont (Mann et al., 2018). We chose to analyze bacteria in addition to eukaryotes due to the increasing interest into food preparation processes such as fermentation, and therefore bacteria may increasingly become target organisms for genetic analysis of diet (Gibbons and Rinker, 2015; Sibbesson, 2019). We used nf-core/eager (v2.1.0) (Fellows Yates et al., 2020) to simulate a typical screening procedure, removing adapters with AdapterRemoval (v2) (Schubert et al., 2016) and mapping the sequencing reads with bwa aln (v0.7.17) (Li and Durbin, 2009) against each respective reference genome, using a slightly relaxed edit distance of 0.03 to account for DNA damage (Schubert et al., 2012). The resulting BAM file of mapped-only reads was then processed using samtools (v1.10) (Li et al., 2009) and the GNU 'shuf' tool to downsample each BAM file to 25, 50, 100, 200, 500, 1,000, 5000 and 10000 reads (with 100 sampling-with-replacement replicates for each). Damage profiles were then generated using DamageProfiler (v0.4.9) (Neukamm et al., 2020), and visualized in R (v3.6.3) (R Core Team, 2017) with the data.table (Dowle et al., 2020) and tidyverse set of packages (v1.3.0) (Wickham et al., 2019).

To explore the challenges of identification and specificity of dietary DNA sequences during metagenomic taxon identification, we generated synthetic dental calculus metagenomes containing possible dietary taxa. We chose five plant and five animal species to represent a wide range of potential dietary sources. Representative genomes were retrieved from NCBI and include: peanut (*Arachis hypogaea*: GCA_003086295.2), tomato (*Solanum lycopersicum*: GCA_000188115.3), peach (*Prunus persica*: GCA_000346465.2), wheat (*Triticum aestivum*: GCA_900519105.1), corn (*Zea mays*: GCA_000005005.6), red deer (*Cervus elaphus*: GCA_002197005.1), rabbit (*Oryctolagus cuniculus*: GCA_000003625.1), pig (*Sus scrofa*: GCA_000003025.6), chicken (*Gallus gallus*: GCA_000002315.5), and Atlantic salmon (*Salmo salar*: GCA_000233375.4). During selection of these sequences, accessions were compared to a list of genome sequences that have been previously flagged as being contaminated with microbial or host DNA, to ensure genome sequences are reliable representations of each given species (Steinegger and Salzberg, 2020). Any suspected contaminated sequences were removed from downstream analysis. We generated a mock oral microbial community using representative genomes of 19 bacterial species as well as a single archaeal species for a total of 20 taxa. Species were chosen to represent microbes commonly found in the human oral microbiome. Information for all microbes in the mock oral and dietary datasets can be found in Supplementary Table 1. From these genome sequences, we generated simulated ancient metagenomic sequencing libraries using Gargammel (v1.1.2) (Renaud et al., 2017) with a minimum read length of 25 bp, a maximum read length of 125 bp, and damage patterns based on Briggs et al. (2007). Because the host should be a predominant source of eukaryotic DNA in dental calculus, given regular contact of dental calculus with human oral tissues and saliva, we also generated a simulated dataset from the human reference genome (GCF_000001405.39) in an identical manner to compare to the mock diet results. After identifying adapter sequences for each dataset using –identify-adapters, we quality filtered, trimmed, and merged each simulated dataset with AdapterRemoval (v2.3.0) (Schubert et al., 2016) with a minimum quality phred score of 25. Only those reads that were successfully merged, or were unmerged forward reads that passed quality checks, were used in downstream analyses. We removed exact sequence duplicates from each synthetic dataset using VSEARCH (v1.1.3) (Rognes et al., 2016). We then generated mock oral and dietary/host datasets by spiking in a randomly sampled 50, 500, or 5000 reads of each dietary/host source into the mock oral microbiome community so that three datasets for each dietary/host organism was generated each equaling five million reads total. A taxonomic assignment for each read in all datasets was made using KRAKEN2 (v2.0.8) (Wood et al., 2019) and the NCBI nucleotide (nt) database (January 2020) as reference. An alluvial plot illustrating the proportion of microbial reads that were classified correctly at each taxonomic level was generated with RawGraphs (Mauri et al., 2017). Lollipop plots of mock dietary results were generated with ggplot2 (Wickham, 2016) in R version 3.6.1 (R Core Team, 2017). To demonstrate that the challenges documented in this manuscript are not unique to the KRAKEN2 pipeline, taxonomic assignment of a single mock diet sample (*S. lycopersicum* 5000 spike-in) was also performed with DIAMOND (v.0.9.14) (Buchfink et al., 2015).

Finally, we analyzed 14 published dental calculus datasets representing a wide geographic, temporal, and preservation range to identify commonly misidentified eukaryotes in metagenomic datasets. We chose 13 archaeological dental calculus samples including two individuals each from Camino del Molino, Iberia (CMOL53, CMOL214: 3290-4870 BP); Arbulag Soum, Khoövsgoöl, Mongolia (H10, H24: 2880-3600 BP); Samdzong, Nepal (S40, S41b: 1550-1300 BP); Anse à la Gourde, Guadeloupe (F1948, F349A: 975-555 BP); Norris Farms, Illinois, USA (NF217, NF47: 650 BP); and Middenbeemster, the Netherlands (S454, S108: 339-84 BP) (Mann et al., 2018; Ziesemer et al., 2019) and a single Neanderthal individual (ELSIDRON1L7: 48400 ± 3200 BP) (Weyrich et al., 2017; Wood et al., 2013). To provide a modern context and to evaluate differences between modern and ancient samples, a dental calculus sample from a living individual from Spain was also included (Velsko et al., 2019). Reads for each dataset were quality filtered, merged, and taxonomically assigned in an identical manner to the synthetic datasets. These samples were chosen to represent a wide geographic and temporal span, but also because they were processed in three different laboratories, to account for methodological differences.

Figures (Fig. 1c and d) illustrating the distribution of published genomes were generated using an assembly report downloaded from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/ASSEMBLY_REPORTS/species_genome_size.txt.gz, as of March 2020). The genomes were binned into taxonomic groups after which the number of available genomes and genome size was calculated. Viruses were omitted from the genome size calculations, and synthetic sequences from all illustrations.

Conda environments for analytical reproducibility, as well as processing and figure generation scripts can be found https://github.com/aemann01/diet_calculus and are archived in Zenodo (https://doi.org/10.5281/zenodo.4265311).

3. Results

3.1. Effects of low read counts on damage patterns

First, we assessed the ability to observe damage patterns when downscaling well-preserved ancient DNA datasets of single eukaryotic genomes to numbers of reads typical of putative dietary taxa found in dental calculus. Across all species types analyzed we observed that the fewer the mapped reads, the weaker the damage signals (Fig. 2a). All four samples show clear damage patterns when using all reads (Fig. 2b). For the two eukaryotic genomes - *G. morhua* - (cod) and *Homo sapiens* (human) - at least 500 reads were required to consistently visualize sufficient damage pattern signals across all downsampling replicates. In contrast, the two bacterial genomes required greater numbers of reads to visualize a signal, with around 5000 for a clean clonal genome and 10000 for a 'commensal' taxon in a microbially diverse sample, to reliably display clear signs of damage across replicates. An example

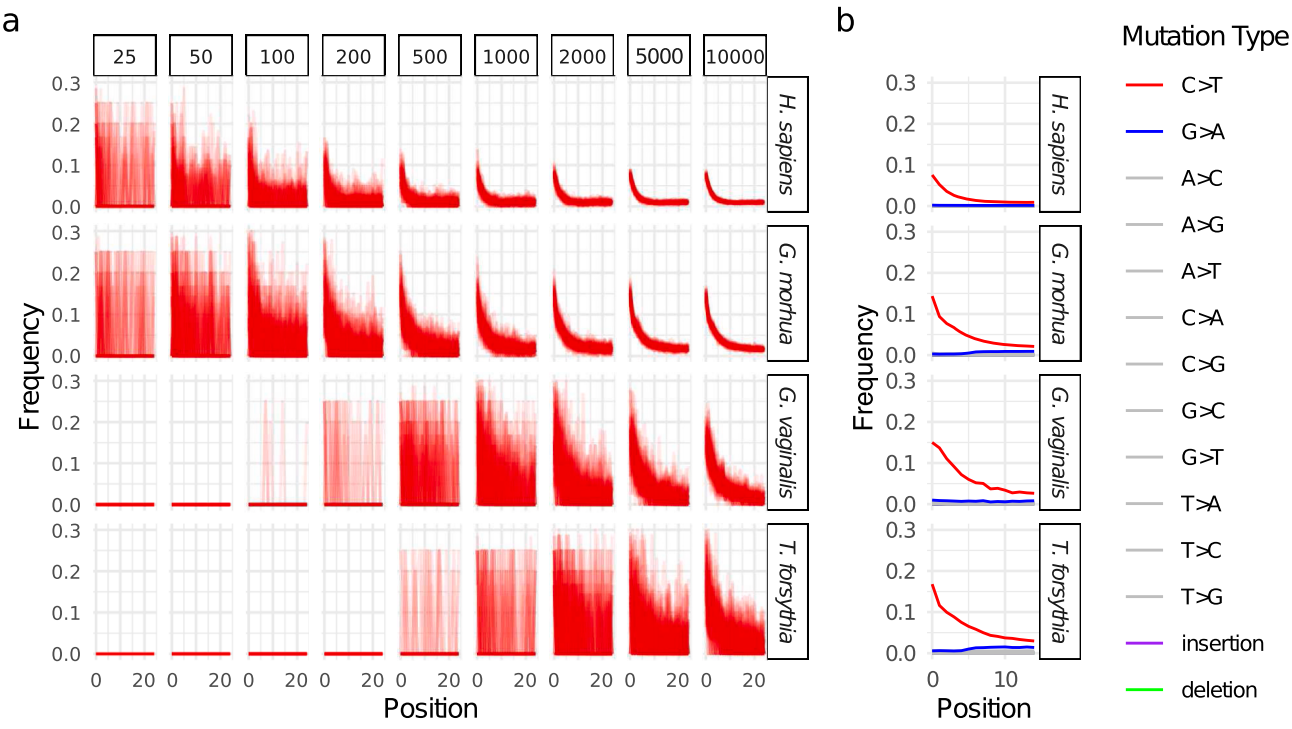

Fig. 2. Effects of low read counts on damage profiles. Four different ancient DNA sample sources were mapped against their respective reference genomes. Each mapped file was then downsampled to different levels of mapped reads, representing different levels of typical dietary hits. (a) represents damage profiles of 100 downsamplings to each set of number of reads. (b) represents the original damage profile of each reference genome with all mapped reads. The number of reads informing each species in (b) are: *H. sapiens* - 7943448, *G. morhua* - 5963543, *G. vaginalis* - 36721, *T. forsythia* - 134372. Note that for both *G. vaginalis* and *T. forsythia*, 100 downsamplings were not always reached for 25-100 read levels, due to insufficient coverage.

damage profile of a single downsampling replicate can be seen in Supplementary Figure 1.

3.2. Identification of dietary reads from synthetic datasets

Overall, accurate identification of dietary reads from synthetic dental calculus samples was poor, with an average of 67.51% (±17.26) having no taxonomic classification and 55.45% (±29.07) classified at either the wrong genus or a taxonomic classification level higher than genus across all mock dietary datasets. Accurate species or genus level assignment was variable between dietary sources but consistent across different spike-in levels (Fig. 3). This is in sharp contrast to the synthetic oral microbes where 56% of reads were assigned to the correct genus, 49% to the correct species, and only 39% unassigned (Fig. 4). Among plants, tomato had the highest proportion of reads correctly assigned to the species level (51.34% ± 4.16) followed by peanut (28.81% ± 2.83), corn (19.97% ± 14.77), wheat (14.11% ± 0.85), and peach (5.67% ± 5.67). Among vertebrates, human had the highest proportion of correctly identified reads at the species level (39.36% ± 4.04) followed by pig (12.67% ± 5.86), chicken (9.86% ± 3.42), salmon (6.17% ± 6.78), and rabbit (5.04% ± 1.18) (Supplementary Table 2). Reads generated from the red deer were never correctly classified at either the genus or species level. Instead, most reads that could be assigned were designated as wild yak (*Bos mutus*) followed by the Bovidae family which notably, does not include deer (Cervidae). Misidentification of the source of dietary reads was common across all synthetic datasets, both at the species and genus level, occasionally at relatively high read counts (Table 1). For example, 239 reads originating from the red deer were misidentified as wild yak while 161 were assigned to bighorn sheep (Table 1). Importantly, correct identification of reads to the species level is strongly correlated with the number of available genome assemblies in the reference database for that genus (Spearman correlation: 5000: $R^2 = 0.87$; 500: $R^2 = 0.91$; 50: $R^2 = 0.88$) (Supplementary Figure 2). Full KRAKEN2 results for all spike-in levels can be found in Supplementary Table 3. To test the impact of another popular bioinformatic pipeline, we analyzed the best performing dietary species from our KRAKEN2 results, *S. lycopersicum*, with DIAMOND. We found that DIAMOND performed substantially poorer, with only 0.24% of reads assigned to *S. lycopersicum*, and 97.58% were unable to be assigned to any taxonomic level (Supplementary Table 4).

Synthetic bacterial reads were occasionally misidentified as eukaryotic organisms (0.02% of all reads) including reads assigned to the common carp (*Cyprinus carpio*), mouse (*Mus musculus*), golden eagle (*Aquila chrysaetos*), human (*Homo sapiens*), and blunt-snouted clingfish (*Gouania willdenowi*) among others. These misidentifications highlight that many eukaryotic entries in the database are contaminated with microbial DNA (Laurence et al., 2014; Merchant et al., 2014) or are otherwise problematic. Full results from the microbial dataset can be found in Supplementary Table 5.

3.3. Identification of dietary reads from real datasets

Next we examined published datasets generated from archaeological and modern calculus to identify eukaryotic organisms that are commonly misidentified in metagenomic data. As the dental calculus samples evaluated here represent a variety of sampling contexts, ages, cultures, subsistence strategies, environmental contexts, and were processed in different laboratories, we expect that non-host eukaryotes detected across a large proportion of samples are common sources of error in metagenomic datasets. We found a wide range of species that were found in all datasets at relatively high average read counts (Table 2), 58.6% of which are fish including the model organisms *Oryzias latipes* and *Danio rerio*. Notably the common carp was found in all samples at unexpectedly higher counts than any other non-host organism ($\bar{x} = 204286$ across all samples) reflecting a high degree of contamination with real or synthetic sequences used during sequencing of this reference genome (a known problem with the *Cyprinus carpio* reference genome: http://grahametherington.blogspot.com/2

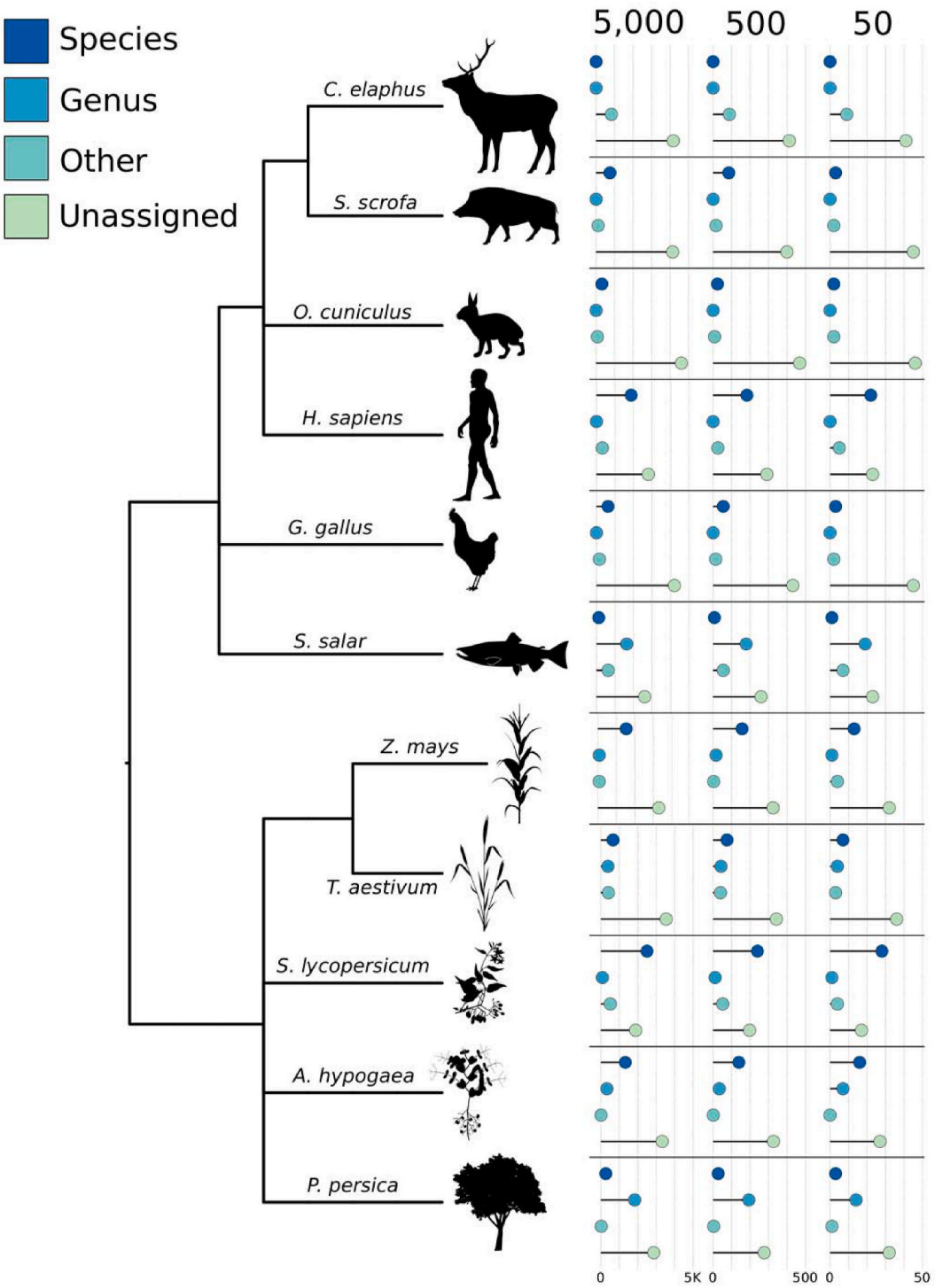

Fig. 3. Most synthetic ancient reads generated from dietary genomes cannot be classified to the species level. Lollipop plots indicate the proportion of reads at each spike-in level (5000, 500, 50) that were correctly identified at the species level, only to the genus level or the correct genus but wrong species, the incorrect genus or higher taxonomic group ("other"), or could not be assigned to any taxonomic level. Species images downloaded from phylopic.org.

014/09/why-you-should-qc-your-reads-and-your.html). Variation in standard deviation for each organism reflects differences in sequencing depth across samples after post processing ($\bar{x} = 16449051 \pm 18627823$) for most organisms (Pearson correlation: $R^2 > 0.7$). Deviations from this pattern include *Cyprinus carpio* ($R^2 = 0.4$), *Erinaceus europaeus* ($R^2 = 0.4$), *Sparus aurata* ($R^2 = -0.07$), and *Camelus dromedarius* ($R^2 = -0.2$). The Neanderthal calculus sample largely drives both the *Cyprinus carpio* and *Erinaceus europaeus* read counts, which contributes 99.28% and 69.60% to the total number of reads matching to these organisms across all samples, respectively. Negative correlation between the sequencing depth and *Sparus aurata* and *Camelus dromedarius* is driven by a disproportionate number of reads derived from a single or two samples (*S. aurata*: S41, *C. dromedarius*: S454 & NF47). Importantly, many of these same species were identified as sources of misidentification in the synthetic bacterial and dietary datasets which lends support to them being contaminated or misannotated genomes in the NCBI nt database (seen in Table 2).

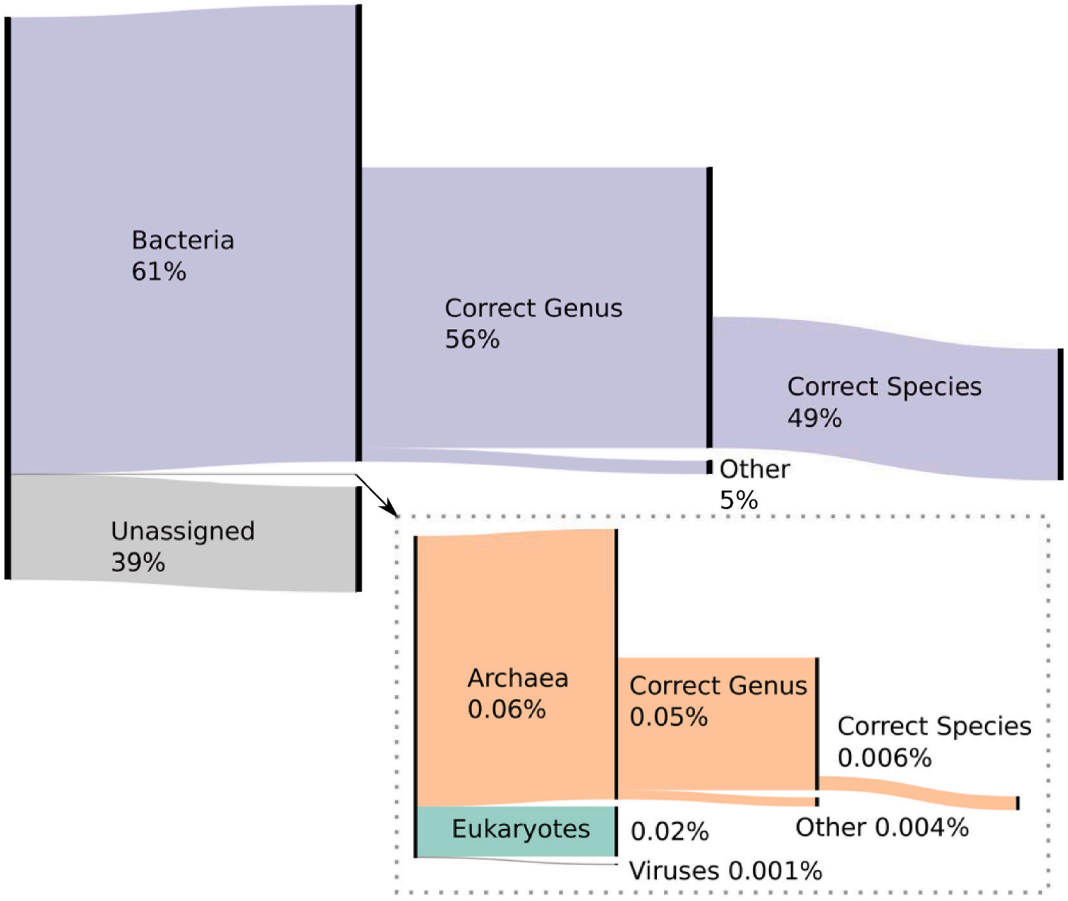

Fig. 4. Alluvial diagram of synthetic microbial DNA. Most reads generated from microbial genomes representing a mock oral microbiome were assigned to bacteria. Of the total dataset 56% of bacterial species were assigned to the correct genus and 49% to the correct species. 0.02% of bacterial or archaeal reads were misassigned as eukaryotic. Box is zoom in of small branch indicated by arrow.

Table 1
Species mismatches in the 5000 spike-in synthetic dataset with identifications of five or more reads. Species misidentification is common across datasets. Query indicates the true origin of reads, subject is the assigned taxonomic identification by KRAKEN2. Count indicates the number of reads assigned to the misidentified subject.

Query	Subject	Count	Common name of misidentified taxa
C. elaphus	*Bos mutus*	239	Wild yak
C. elaphus	*Ovis canadensis*	161	Bighorn sheep
S. salar	*Coregonus* sp. *'balchen'*	79	Blue whitefish
C. elaphus	*Odocoileus virginianus texanus*	34	White tailed deer
T. aestivum	*Aegilops tauschii*	43	Tausch's goatgrass
G. gallus	*Aquila chrysaetos*	19	Golden eagle
C. elaphus	*Bos taurus*	14	Domesticated cattle
S. scrofa	*Bos mutus*	11	Wild yak
G. gallus	*Apteryx mantelli*	11	North Island brown kiwi
S. scrofa	*Ovis canadensis*	9	Bighorn sheep
G. gallus	*Streptopelia turtur*	9	European turtle dove
C. elaphus	*Sus scrofa*	8	Wild boar
T. aestivum	*Hordeum vulgare*	7	Barley
C. elaphus	*Homo sapiens*	7	Human
S. scrofa	*Homo sapiens*	6	Human
G. gallus	*Meleagris gallopavo*	6	Wild turkey
S. scrofa	*Lutra lutra*	5	Eurasian otter
O. cuniculus	*Homo sapiens*	5	Human
G. gallus	*Phasianus colchicus*	5	Ring necked pheasant

4. Discussion and conclusions

Results from this study highlight the challenges associated with proper identification and validation of DNA from dietary sources in archaeological dental calculus. While dietary reconstruction using genomic techniques is increasingly popular, identification of eukaryotes that may represent foodstuffs in mixed microbial communities is difficult, and a lack of validation criteria makes most claims of dietary information from metagenomic samples suspect. Read count and/or relative abundance are not sufficient to prove the existence of dietary DNA in metagenomic data and instead further validation criteria are necessary for robust interpretations of dietary reconstructions of the past.

We have demonstrated here that typical aDNA authentication criteria, such as damage patterns, can often be difficult to assess when dealing with possible dietary genetic material retrieved from dental calculus. Clear signals of damage required at least ~500 reads for eukaryotic authentication, and thousands for microbes. The low yield of putative dietary DNA molecules from the substrate is therefore often insufficient to generate authentication profiles that allow for the assessment of the presence of damage (as rightly reported on dietary hits from baboon dental calculus by Ottoni et al., 2019). This is especially important for highly diverse species, such as seen with, but not limited to microbes, where many more reads were required to reliably detect damage patterns (Fig. 2a). A possible cause for this difference is the presence of DNA from a greater biodiversity of closely related microbial taxa that are often present in burial environments and degraded archaeological material, compared to DNA extracted from a single bone of a single species. The large abundance of many different species in a

Table 2
The usual suspects. Non-host (*Homo sapiens*) eukaryotes at the species level found in all metagenomic samples with an average of 100 reads or higher.

Species	Common name	NCBI taxonomy ID	Average read count (±SD)	Average relative abundance (±SD)
Cyprinus carpio	Common carp	7962	204286 ± 785091	0.5166 ± 1.8482
Erinaceus europaeus	European hedgehog	9365	721 ± 1818	0.0038 ± 0.0040
Mus musculus	House mouse	10090	390 ± 457	0.0024 ± 0.0008
Oryzias latipes	Japanese rice fish	8090	226 ± 264	0.0014 ± 0.0004
Aquila chrysaetos	Golden eagle	223781	223 ± 250	0.0014 ± 0.003
Apteryx mantelli	North Island brown kiwi	202946	210 ± 253	0.0013 ± 0.004
Coregonus* sp. *'balchen'	Whitefish	861768	203 ± 222	0.0013 ± 0.003
Sphaeramia orbicularis	Orbiculate cardinalfish	375764	188 ± 207	0.0012 ± 0.0003
Salmo trutta	Brown trout	8032	185 ± 205	0.0012 ± 0.0004
Danio rerio	Zebrafish	7955	171 ± 187	0.0011 ± 0.003
Myripristis murdjan	Soldierfish	586833	170 ± 207	0.0010 ± 0.0003
Scleropages formosus	Asian arowana	113540	161 ± 193	0.0010 ± 0.003
Spirometra erinaceieuropaei	Tapeworm	99802	155 ± 193	0.0009 ± 0.003
Sparus aurata	Gilt-head bream	8175	154 ± 194	0.009 ± 0.003
Ovis canadensis	Bighorn sheep	112262	148 ± 156	0.0010 ± 0.009
Salarias fasciatus	Lawnmower blenny	181472	147 ± 163	0.0009 ± 0.0002
Bos mutus	Domestic yak	72004	146 ± 155	0.0009 ± 0.0005
Neostethus bicornis	Southeast Asian fish	300306	137 ± 160	0.0008 ± 0.003
Chanos chanos	Milkfish	29144	136 ± 159	0.0008 ± 0.0003
Pecten maximus	Great scallop	6579	133 ± 152	0.008 ± 0.002
Thalassophryne amazonica	Prehistoric monster fish	390379	131 ± 162	0.0007 ± 0.0002
Gouania willdenowi	Blunt-snouted clingfish	441366	131 ± 145	0.0009 ± 0.0003
Lutra lutra	Eurasian otter	9657	129 ± 145	0.0008 ± 0.0002
Lateolabrax maculatus	Asian seabass	315492	122 ± 137	0.0007 ± 0.0003
Gadus morhua	Atlantic cod	8049	119 ± 135	0.0007 ± 0.0002
Caligus rogercresseyi	Sea louse	217165	118 ± 153	0.0007 ± 0.0003
Haemonchus contortus	Barber's pole worm	6289	115 ± 149	0.0007 ± 0.0003
Streptopelia turtur	European turtle dove	177155	113 ± 133	0.0007 ± 0.0005
Echeneis naucrates	Live sharksucker	173247	109 ± 128	0.0006 ± 0.0002
Camelus dromedarius	Dromedary	9838	107 ± 238	0.0018 ± 0.0046
Betta splendens	Betta fish	158456	105 ± 134	0.0006 ± 0.0003
Rhinatrema bivittatum	Two lined caecilian	194408	105 ± 122	0.0006 ± 0.0002
Denticeps clupeoides	Denticle herring	299321	104 ± 129	0.0006 ± 0.0001

sample means that there is a greater chance of DNA sequences being able to misalign to the reference genome of a particular target organism (Warinner et al., 2017). This results in 'false-positive' mutations compared to the reference genome being detected and thus adds noise to damage pattern profiles. DNA originating from putative dietary sources, on the other hand, require fewer species-level assigned reads to authenticate using this method if there are fewer competing closely-related species within the substrate, greatly reducing noise at low read counts as compared to microbes. Authors and reviewers should therefore be aware of unusual or unreported damage patterns and consider the noise that can occur when dealing with low numbers of reads before making assertions of authenticity. For example, some laboratory protocols used for high quality aDNA genome reconstruction will remove damage (e.g., Briggs et al., 2010), to improve detection of true biological mutations in a genome. However, these protocols are only suitable for samples that yield hundreds of thousands to millions of aDNA reads of an organism and should only be used after prior authentication of sequences. This highlights that alternative strategies are needed to assist in the assessment of the authenticity of any reported dietary taxa. Enrichment techniques, such as those employed in other metagenomic contexts (Ozga et al., 2016; Slon et al., 2017; Maixner et al., 2018) or more stringent statistical analysis to assess confidence of damage signals may need to be further developed (Weiβ et al., 2015).

The majority of reads generated in the synthetic datasets across all putative dietary and host organisms could not be assigned to any taxonomic level, illustrating that many regions across eukaryotic genomes can map equally well to a variety of disparate sources. This process is exacerbated in the context of fragmented and damaged aDNA, even when a high number of reads are retained in the dataset (i.e., 5000). Moreover, reads that were assigned to a taxon were often too unspecific (i.e., assigned at the family level) to provide any meaningful dietary analysis or were assigned to the wrong organism altogether. While these mismatches are often quite obvious, limiting discussion of dietary information to those organisms that 'make sense' in terms of the archaeological context or time period belies the complexity and often perplexing nature of eukaryotic results from metagenomic data. For example, while reads matching common grape vine (*Vitis vinifera*) reference sequences - a species that has long been an important viticulture crop in Europe (Ramos-Madrigal et al., 2019) - were found in both archaeological samples from individuals who lived in the Netherlands (n = 21 & 17, respectively), both also had a comparable number of reads matching the Tasmanian devil (*Sarcophilus harrissii* n = 21 & 25, respectively). Analogously, while 37 reads from the Neanderthal individual matched the genus *Pinus* which includes species that produce edible nuts that are thought to be a component of Neanderthal diet (Weyrich et al., 2017), a comparable number of reads in the same dataset matched sequences of the Arizona bark scorpion (*Centruroides sculpturatus*, n = 37), cheetah (*Acinonyx jubatus*, n = 36), and the kakapo (*Strigops habroptila*, n = 35), which are unlikely to be true dietary results. The highest read count for non-host eukaryotes (excluding carp) found in archaeological samples include European hedgehog (*Erinaceus europaeus,* F349, n = 403; F1948, n = 271; CMOL214, n = 275; NF217, n = 135; S108, n = 298; S37, n = 534; ELSIDRON1L7, n = 7021), mouse (*Mus musculus*, H24b, n = 269; H10b, n = 521; CMOL53, n = 393), dromedary (*Camelus dromedarius*, NF47, n = 563; S454, n = 759), and bighorn sheep (*Ovis canadensis*, S41, n = 370). Importantly, this identification of unlikely dietary taxa is not limited to ancient samples. We find many eukaryotes in the modern human dental calculus sample that may be a signal of diet, but also those that are highly unlikely to be so, both at relatively high read counts. For example, while 479 reads matched brown trout (*Salmo trutta*), an important resource in northern Spain (Vera et al., 2018), 582 matched the North Island brown kiwi (*Apteryx mantelli*) and 528 to the Asian arowana (*Scleropages formosus*) (0.001% of the total dataset).

While dental calculus can provide a long-term resource of ancient biomolecules, including DNA, the reconstruction of diet from

archaeological substrates using metagenomic approaches remains a challenge. These challenges may be alleviated by expanding coverage of eukaryotic taxa in genomic reference databases and increased sampling depth. Particularly when aligning short aDNA reads with few unique species-informative substitutions to a collection of eukaryotes with large portions of shared genome sequences, alignment to expanded databases will lower the frequency of species misidentification by pushing uninformative reads up the tree (using LCA based methods), as well as increase the resolution of taxonomically informative reads by providing adequate comparative references (Fig. 5). In the meantime, dietary reconstruction from metagenomic datasets requires vigorous authentication and cautious interpretation. To this end, we provide a suggested list of criteria for the evaluation of dietary analyses from metagenomic studies of ancient dental calculus. It should be noted that it is rare that all points can be completely addressed by researchers, particularly when working with museum samples that do not have detailed sample history records or are utilizing previously published data. However, it is important that studies should at least show that they are aware of the following points to ensure transparency and assist reviewers in assessing the strength of claims of identification of dietary DNA.

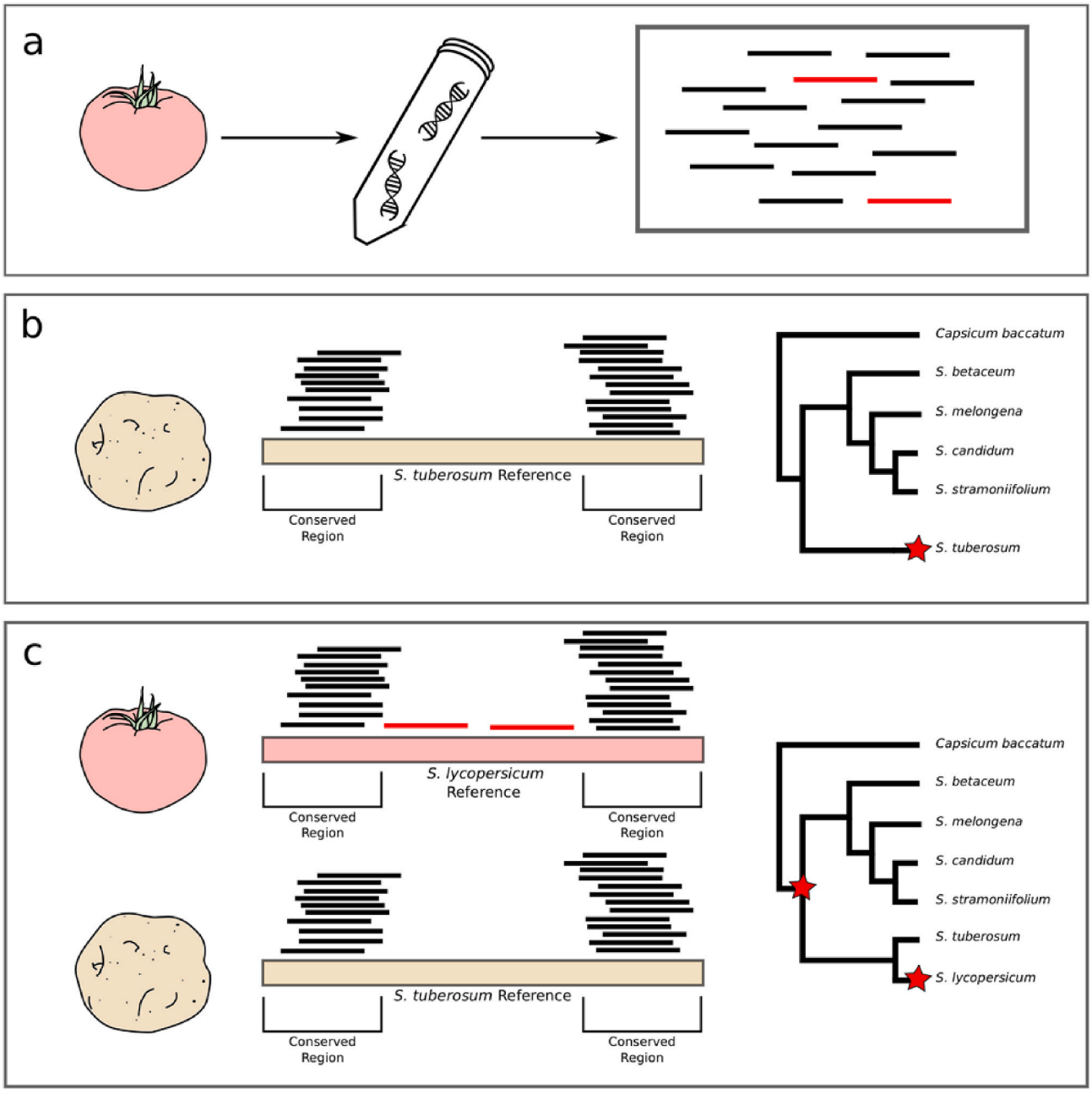

Fig. 5. The importance of reference databases. (a) DNA extracted from a tomato (*S. lycopersicum*) resulting in a library of short DNA molecules of which ~92% are conserved between tomato and potato (*S. tuberosum*) (Sato et al., 2012) (black reads) and 8% have high taxonomic resolution for tomato (red reads) (b) If reads from tomato are compared to a database that only contains the potato reference (and no other members of the genus or other closely related taxa), conserved reads will be mapped to conserved regions and misidentified as potato and species specific reads will be discarded (c) If both tomato and potato are represented, conserved reads will be classified at the genus level as they match both species equally well and species specific reads will be retained, illustrating the importance of ensuring a wide diversity organisms are present in metagenomic databases. *Solanum* tree based on Bohs and Olmstead (1997). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4.1. Checklist for authors and reviewers

4.1.1. Field/storage contamination

DNA contamination can be introduced to a sample from a variety of sources before any molecular analysis is initiated, necessitating close documentation of sample context, excavation practices, and sample collection procedures for the interpretation of potential dietary DNA. Environmental contaminants from soils, percolating/leaching water sources, etc. are assumed to be present in most ancient DNA samples. However, less commonly documented are the conditions of the excavation itself: What precautions were taken against contaminating excavation units and the subsequent biological specimens that were collected (i.e., washing hands, wearing gloves, cleaning tools between excavation units, etc.)? Was the site and by extension the biological remains from the site disturbed by looters and/or animals prior to or during excavation? What measures were taken to ensure the tools for sampling (i.e., dental extractors, dental picks, etc.) would not introduce modern contaminants or cross-contamination between samples? Additionally, were samples washed (to be avoided), stored in old food containers, or treated with glues or consolidants made of plant/animal products (Nicholson et al., 2002)? Lastly, practices to protect specimens from pests or preserve them in museum collections can introduce contaminants. Careful identification of these factors, sample histories and contexts can provide critical information for interpreting the presence of unexpected taxa or authenticating dietary components.

4.1.2. Is the laboratory methodology suitable for aDNA authentication?

Are laboratory conditions introducing dietary DNA? Laboratories that are not adapted for aDNA processing, such as facilities that have shared workbenches or air flow with modern DNA laboratories, may introduce eukaryotic DNA fragments into the samples from the laboratory environment. Some reagents used for sample processing may also contain traces of eukaryotic DNA (e.g., bovine or rat serum albumin) (Leonard et al., 2007). Are short molecules selected for? Extraction methods that recover short reads are more likely to capture authentic ancient eukaryotic DNA, compared to kits for modern research that require and select for longer DNA sequences. Does the methodology remove DNA damage? For example, full UDG treatment of DNA libraries will remove all damage from the DNA fragments, whereas partial UDG treatment will leave damage on the first bases (Rohland et al., 2015) which can then be used for damage verification. Moreover, read 'carry-over' between next-generation sequencing runs on the same machine (Nelson et al., 2014) and indicies 'bleeding' during sample preparation and sequencing (Warinner et al., 2017) can result in false positives (where indices of one library are occasionally transferred to DNA molecules of another). Reads should therefore have unique dual indices (Kircher et al., 2012), or even additional 'inline' barcodes (Valk et al., 2020), to correct for cross-run contamination or chimeric reads that are assigned to the wrong sample. The summary by Llamas and colleagues is a useful reference resource as to the setup typically required for all types of aDNA studies (Llamas et al., 2017).

4.1.3. Is the database and pipeline suitable?

Many databases are biased towards certain classes of organisms (e.g., those that are clinically or economically important) or types of genomes (such as noted by Ozga et al., 2019), which may not be suitable nor appropriate for the analysis of diet in metagenomic datasets. More balanced databases should therefore be constructed to reduce misidentification due to the attraction of reads of over-represented genomes (such as the approach taken by Zhou et al., 2017). Furthermore, if dietary reconstruction is the goal, the database should include representative genomes for known or suspected dietary eukaryotes for the site or region of study as well as closely related species whenever possible. Moreover, as annotation errors or contamination is a common problem for many reference databases, including NCBI nt, suspected signals of diet should be independently verified (e.g., through analyses using a separate database or additional phylogenetic analyses). While stringently curated databases can reduce off-target taxonomic assignments they can also limit the breadth of organisms that may be detected. For example, while BLASTX approaches (translating from nucleotide to amino-acid sequences) (e.g., with DIAMOND) has been used in the past due to improved quality of protein databases (e.g., Weyrich et al., 2017), this suffers from a range of drawbacks, particularly when applied to ancient DNA. Firstly, using only amino-acid sequences reduces database searches against protein-coding regions of the genome only. Secondly, short aDNA molecules translate into even shorter amino-acid sequences that may match many different proteins. Thirdly, typically only high-quality genomes are placed in protein databases, which reduces the number of possible taxa that may be available in databases. Therefore, BLASTX-like approaches are generally more suitable as a confirmatory analysis, depending on the nature of the data as well as taxa of interest.

4.1.4. Is the taxonomic resolution analyzed sufficient for reliable interpretation?

Generally, taxonomic levels above genus should not be relied on for dietary reconstruction due to non-genetically informed definitions of taxonomic ranks, thus varying naming criteria. Moreover, within different genera the strength of interpretations can vary. For example, there are four recognized species within the genus *Gallus* (including the domesticated chicken) commonly known as junglefowl. A strong signal of *Gallus* could therefore be reliably interpreted as the consumption of junglefowl (though species level differentiation may not be possible). On the other hand, the genus *Solanum* has more than 1000 named species, both edible and inedible, including potato, tomato, and eggplant, and a strong *Solanum* signal alone cannot reliably be interpreted as an edible species.

4.1.5. Have unexpected taxa been reported and evaluated?

Organisms that are unexpected or may represent mismatches should be reported and properly evaluated, including those that are incompatible with the geographic location, or are model organisms along with those that are suspected dietary organisms.

4.1.6. Have laboratory controls been evaluated alongside samples?

Parallel processing of negative controls (e.g., molecular laboratory grade water) alongside true samples is necessary for the proper authentication of DNA extracted from archaeological materials, like dental calculus, as well as the tracking of sources of contamination. These should be reported at both extraction and library construction steps, particularly if the latter is performed at non-aDNA specialist commercial or centralized sequencing facilities. If low frequency dietary reads are found in true samples, it should be verified that laboratory controls do not also contain suspicious signals of these putative dietary taxa. In addition to negative controls, positive controls collected from the archaeological site (e.g., the surrounding soil or deposits and concretions originating from the teeth themselves) sequenced along true samples may be useful for characterizing the impact of environmental taxa on data interpretation. However, note that this may only be possible when archaeological excavations are recent, this will likely be impossible for samples from museum collections.

4.1.7. Has evidence of aDNA authenticity been provided?

If sufficient numbers of reads are available, characteristics of authentic aDNA, including short fragment lengths and higher frequencies of cytosine to thymine shifts at the ends of reads, should be reported (Fig. 2).

4.1.8. Is there other evidence supporting that dietary organisms were consumed?

If sufficient reads are not available to verify the presence and authenticity of dietary organisms, archaeological or other proxy evidence for the consumption of that particular food should be provided.

Finally, if strong signals of eukaryotic taxa are found, care must be taken to make practical and reasonable interpretations of taxa. Reporting of unexpected putative dietary items should include literature reviews to ensure a taxon 'makes sense' as a dietary hit. For example, in one of the earliest metagenomic dietary DNA studies, Weyrich et al. (2017) reported moss (*Physcomitrella patens*) as being possibly consumed by a Neanderthal from El Sidrón cave. However, it was later pointed out that moss is both inedible (at least to modern humans) and devoid of nutrition (Dickson et al., 2017), thus making further speculation of any reason to consume moss - utilitarian or otherwise – tenuous. In contrast, Ottoni et al. (2019) reported reservations about hits to cucumber in Egyptian baboon calculus due to geographic implausibility. Furthermore, the inherent limits of the utility of DNA for interpretation must be acknowledged. For example, raw DNA sequences alone cannot indicate which tissue the DNA was derived from, and thus potentially consumed (such as the 'pine nut' interpretation from the identification of *Pinus* DNA by Weyrich et al., 2017); something palaeoproteomics is more applicable to. Therefore, when unusual results are found, further interdisciplinary analysis such as paleoproteomics (Hendy et al., 2018), pollen analysis, evidence of butchery, or the improved design of genetic primers or capture methods that target specific dietary organisms (Maixner et al., 2018; Sawafuji et al., 2020) may be ways to strengthen interpretation of diet in the past from dental calculus metagenomics.

In summary, while dental calculus may be an important source of genetic evidence for dietary organisms, low DNA yields, current metagenomic methods, and limited references may confound many dietary reconstruction attempts. With these limitations in mind, further development of tailored methods for the analysis of eukaryotic results from metagenomic data, and the use of multiple lines of evidence beyond genetics alone, will greatly enhance dietary reconstructions from archaeological dental calculus.

Data availability

Public dataset EBI ENA (https://www.ebi.ac.uk/ena/) accession numbers (archaeological and modern dental calculus samples): ERR3307054, ERR2900732, ERR2900736, ERR2900740, ERR2900744, ERR2900748, ERR2900752, ERR2900756, ERR2900760, ERR2900764, ERR2900768, ERR2900772, ERR2900776. Data for ELSIDRON1L7 downloaded from https://www.oagr.org. For read downscaling on damage experiment: ERR1943572, and from NCBI SRA (https://www.ncbi.nlm.nih.gov/sra) SRR4885939, SRR1187682 and SRR6877313. Conda environment and scripts for all analyses can be found at https://github.com/aemann01/diet_calculus (DOI: 10.5281/zenodo.4265311).

Author contributions

Allison E. Mann: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - original draft, Writing - review & editing. James A. Fellows Yates: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - original draft, Writing - review & editing. Zandra Fagernäs: Conceptualization, Formal analysis, Visualization, Writing - original draft, Writing - review & editing. Rita M. Austin: Conceptualization, Writing - original draft, Writing - review & editing. Elizabeth A. Nelson: Conceptualization, Writing - original draft, Writing - review & editing. Courtney Hofman: Conceptualization, Formal analysis, Writing - review & editing

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

J.A.F.Y. was partially funded by the European Research Council (ERC) under the European Union's Horizon 2020 research innovation programme (ERC-2015-StG 678901-FoodTransforms to Philipp W. Stockhammer, Ludwig Maximilian University Munich, Germany). Z.F. was supported by the Werner Siemens Stiftung through Dr. Christina Warinner. This research was supported in part through computational resources provided by the Department of Archaeogenetics at the Max Planck Institute for the Science of Human History (J.A.F.Y). We thank Stephanie Schnorr for comments on an early version of this manuscript and two anonymous reviewers who improved this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.quaint.2020.11.019.

References

Adler, C.J., Dobney, K., Weyrich, L.S., Kaidonis, J., Walker, A.W., Haak, W., Bradshaw, C. J.A., Townsend, G., Sołtysiak, A., Alt, K.W., Parkhill, J., Cooper, A., 2013. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. Nat. Genet. 45, 450–455. https://doi.org/10.1038/ng.2536.

Allentoft, M.E., Collins, M., Harker, D., Haile, J., Oskam, C.L., Hale, M.L., Campos, P.F., Samaniego, J.A., Gilbert, M.T.P., Willerslev, E., Zhang, G., Scofield, R.P., Holdaway, R.N., Bunce, M., 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proc. Biol. Sci. 279, 4724–4733. https://doi.org/10.1098/rspb.2012.1745.

Andrews, P., Johnson, R.J., 2020. Evolutionary basis for the human diet: consequences for human health. J. Intern. Med. 287, 226–237. https://doi.org/10.1111/joim.13011.

Aslan, Ö., Hamill, R.M., Sweeney, T., Reardon, W., Mullen, A.M., 2009. Integrity of nuclear genomic deoxyribonucleic acid in cooked meat: implications for food traceability. J. Anim. Sci. 87, 57–61. https://doi.org/10.2527/jas.2008-0995.

Atalay, S., Hastorf, C.A., 2006. Food, meals, and daily activities: food habitus at neolithic çatalhöyük. Am. Antiq. 71, 283–319. https://doi.org/10.2307/40035906.

Austin, R.M., Sholts, S.B., Williams, L., Kistler, L., Hofman, C.A., 2019. Opinion: to curate the molecular past, museums need a carefully considered set of best practices. Proc. Natl. Acad. Sci. Unit. States Am. 116, 1471–1474. https://doi.org/10.1073/pnas.1822038116.

Blasco, R., Rosell, J., Fernández Peris, J., Arsuaga, J.L., Bermúdez de Castro, J.M., Carbonell, E., 2013. Environmental availability, behavioural diversity and diet: a zooarchaeological approach from the TD10-1 sublevel of gran dolina (sierra de Atapuerca, burgos, Spain) and bolomor cave (valencia, Spain). Quat. Sci. Rev. 70, 124–144. https://doi.org/10.1016/j.quascirev.2013.03.008.

Bocquet-Appel, J., Bar-Yosef, O., 2008. The Neolithic Demographic Transition and its Consequences. Springer Science & Business Media.

Bohs, L., Olmstead, R.G., 1997. Phylogenetic relationships in Solanum (solanaceae) based on ndhF sequences. Syst. Bot. 22, 5–17. https://doi.org/10.2307/2419674.

Bon, C., Berthonaud, V., Maksud, F., Labadie, K., Poulain, J., Artiguenave, F., Wincker, P., Aury, J.-M., Elalouf, J.-M., 2012. Coprolites as a source of information on the genome and diet of the cave hyena. Proc. Biol. Sci. 279, 2825–2830. https://doi.org/10.1098/rspb.2012.0358.

Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., Pääbo, S., 2007. Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. Unit. States Am. 104, 14616–14621. https://doi.org/10.1073/pnas.0704665104.

Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., Pääbo, S., 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res. 38, e87. https://doi.org/10.1093/nar/gkp1163.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60. https://doi.org/10.1038/nmeth.3176.

Cooper, A., Poinar, H.N., 2000. Ancient DNA: do it right or not at all. Science (New York, N.Y.) 289, 1139. https://doi.org/10.1126/science.289.5482.1139b.

Cuéllar, A.M., 2013. The archaeology of food and social inequality in the andes. J. Archaeol. Res. 21, 123–174. https://doi.org/10.1007/s10814-012-9061-x.

Dabney, J., Meyer, M., Pääbo, S., 2013. Ancient DNA damage. Cold Spring Harbor Perspectives in Biology 5. https://doi.org/10.1101/cshperspect.a012567.

Devault, A.M., Mortimer, T.D., Kitchen, A., Kiesewetter, H., Enk, J.M., Golding, G.B., Southon, J., Kuch, M., Duggan, A.T., Aylward, W., Gardner, S.N., Allen, J.E., King, A. M., Wright, G., Kuroda, M., Kato, K., Briggs, D.E., Fornaciari, G., Holmes, E.C., Poinar, H.N., Pepperell, C.S., 2017. A molecular portrait of maternal sepsis from Byzantine Troy. eLife 6, e20983. https://doi.org/10.7554/eLife.20983.

Dickson, J.H., Oeggl, K., Stanton, D., 2017. 'Forest Moss': no part of the European Neanderthal diet. Antiquity 91. https://doi.org/10.15184/aqy.2017.165.

Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Short, T., Lianoglou, S., Antonyan, E., Bonsch, M., Parsonage, H., Ritchie, S., Ren, K., Tan, X., Saporta, R., Seiskari, O., Dong, X., Lang, M., Iwasaki, W., Wenchel, S., Broman, K., Schmidt, T.,

Arenburg, D., Smith, E., Cocquemas, F., Gomez, M., Chataignon, P., Blaser, N., Selivanov, D., Riabushenko, A., Lee, C., Groves, D., Possenriede, D., Parages, F., Toth, D., Yaramaz-David, M., Perumal, A., Sams, J., Morgan, M., Quinn, M., Storey, R., Saraswat, M., Jacob, M., Schubmehl, M., Vaughan, D., Hocking, T., Silvestri, L., Barrett, T., Hester, J., Damico, A., Freundt, S., Simons, D., Andrade, E.S. de, Miller, C., Meldgaard, J.P., Tlapak, V., Ushey, K., Eddelbuettel, D., 2020. Data.table: extension of "data.frame.

Eglinton, G., Logan, G.A., Ambler, R.P., Boon, J.J., Perizonius, W.R.K., Eglinton, G., Curry, G.B., 1991. Molecular preservation. Phil. Trans. Roy. Soc. Lond. B Biol. Sci. 333, 315–328. https://doi.org/10.1098/rstb.1991.0081.

Evershed, R.P., 2008. Organic residue analysis in archaeology: the archaeological biomarker revolution*. Archaeometry 50, 895–924. https://doi.org/10.1111/j.1475-4754.2008.00446.x.

Forshaw, R., 2014. Dental indicators of ancient dietary patterns: dental analysis in archaeology. Br. Dent. J. 216, 529–535. https://doi.org/10.1038/sj.bdj.2014.353.

Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., Whittle, A., Dani, J., Raczky, P., Higham, T.F.G., Hofreiter, M., Bradley, D.G., Pinhasi, R., 2014. Genome flux and stasis in a five millennium transect of European prehistory. Nat. Commun. 5 https://doi.org/10.1038/ncomms6257.

Geber, J., Tromp, M., Scott, A., Bouwman, A., Nanni, P., Grossmann, J., Hendy, J., Warinner, C., 2019. Relief food subsistence revealed by microparticle and proteomic analyses of dental calculus from victims of the Great Irish Famine. Proc. Natl. Acad. Sci. U.S.A. 116, 19380–19385. https://doi.org/10.1073/pnas.1908839116.

Genome 10K, 2009. A proposal to obtain whole-genome sequence for 10 000 vertebrate species. J. Hered. 100, 659–674. https://doi.org/10.1093/jhered/esp086.

Gibbons, J.G., Rinker, D.C., 2015. The genomics of microbial domestication in the fermented food environment. Curr. Opin. Genet. Dev. 35, 1–8. https://doi.org/10.1016/j.gde.2015.07.003.

Glocke, I., Meyer, M., 2017. Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. Genome Res. 27, 1230–1237. https://doi.org/10.1101/gr.219675.116.

Goodman, D., Redclift, M., 2002. Refashioning Nature: Food, Ecology and Culture. Routledge.

Green, R.E., Briggs, A.W., Krause, J., Prüfer, K., Burbano, H.A., Siebauer, M., Lachmann, M., Pääbo, S., 2009. The Neandertal genome and ancient DNA authenticity. EMBO J. 28, 2494–2502. https://doi.org/10.1038/emboj.2009.222.

Gryson, N., 2010. Effect of food processing on plant DNA degradation and PCR-based GMO analysis: a review. Anal. Bioanal. Chem. 396, 2003–2022. https://doi.org/10.1007/s00216-009-3343-2.

Hardy, K., Buckley, S., Collins, M.J., Estalrrich, A., Brothwell, D., Copeland, L., García-Tabernero, A., García-Vargas, S., Rasilla, M. de la, Lalueza-Fox, C., Huguet, R., Bastir, M., Santamaría, D., Madella, M., Wilson, J., Cortés, Á.F., Rosas, A., 2012. Neanderthal medics? Evidence for food, cooking, and medicinal plants entrapped in dental calculus. Naturwissenschaften 99, 617–626. https://doi.org/10.1007/s00114-012-0942-0.

Hendy, J., Warinner, C., Bouwman, A., Collins, M.J., Fiddyment, S., Fischer, R., Hagan, R., Hofman, C.A., Holst, M., Chaves, E., Klaus, L., Larson, G., Mackie, M., McGrath, K., Mundorff, A.Z., Radini, A., Rao, H., Trachsel, C., Velsko, I.M., Speller, C.F., 2018. Proteomic evidence of dietary sources in ancient dental calculus. Proceedings. Biological Sciences 285. https://doi.org/10.1098/rspb.2018.0977.

Henry, A.G., Piperno, D.R., 2008. Using plant microfossils from dental calculus to recover human diet: a case study from Tell al-Raqā'i, Syria. J. Archaeol. Sci. 35, 1943–1950. https://doi.org/10.1016/j.jas.2007.12.005.

Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A. von, Pääbo, S., 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acids Res. 29, 4793–4799.

Hofreiter, M., Betancourt, J.L., Sbriller, A.P., Markgraf, V., McDonald, H.G., 2003. Phylogeny, diet, and habitat of an extinct ground sloth from Cuchillo Curá, Neuquén Province, southwest Argentina. Quat. Res. 59, 364–378. https://doi.org/10.1016/S0033-5894(03)00030-9.

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J.C., Koch, R., Rauch, G.-J., White, S., Chow, W., Kilian, B., Quintais, L.T., Guerra-Assunção, J.A., Zhou, Y., Gu, Y., Yen, J., Vogel, J.-H., Eyre, T., Redmond, S., Banerjee, R., Chi, J., Fu, B., Langley, E., Maguire, S.F., Laird, G.K., Lloyd, D., Kenyon, E., Donaldson, S., Sehra, H., Almeida-King, J., Loveland, J., Trevanion, S., Jones, M., Quail, M., Willey, D., Hunt, A., Burton, J., Sims, S., McLay, K., Plumb, B., Davis, J., Clee, C., Oliver, K., Clark, R., Riddle, C., Elliott, D., Threadgold, G., Harden, G., Ware, D., Begum, S., Mortimore, B., Kerry, G., Heath, P., Phillimore, B., Tracey, A., Corby, N., Dunn, M., Johnson, C., Wood, J., Clark, S., Pelan, S., Griffiths, G., Smith, M., Glithero, R., Howden, P., Barker, N., Lloyd, C., Stevens, C., Harley, J., Holt, K., Panagiotidis, G., Lovell, J., Beasley, H., Henderson, C., Gordon, D., Auger, K., Wright, D., Collins, J., Raisen, C., Dyer, L., Leung, K., Robertson, L., Ambridge, K., Leongamornlert, D., McGuire, S., Gilderthorp, R., Griffiths, C., Manthravadi, D., Nichol, S., Barker, G., Whitehead, S., Kay, M., Brown, J., Murnane, C., Gray, E., Humphries, M., Sycamore, N., Barker, D., Saunders, D., Wallis, J., Babbage, A., Hammond, S., Mashreghi-Mohammadi, M., Barr, L., Martin, S., Wray, P., Ellington, A., Matthews, N., Ellwood, M., Woodmansey, R., Clark, G., Cooper, J.D., Tromans, A., Grafham, D., Skuce, C., Pandian, R., Andrews, R., Harrison, E., Kimberley, A., Garnett, J., Fosker, N., Hall, R., Garner, P., Kelly, D., Bird, C., Palmer, S., Gehring, I., Berger, A., Dooley, C.M., Ersan-Ürün, Z., Eser, C., Geiger, H., Geisler, M., Karotki, L., Kirn, A., Konantz, J., Konantz, M., Oberländer, M., Rudolph-Geiger, S., Teucke, M., Lanz, C., Raddatz, G., Osoegawa, K., Zhu, B., Rapp, A., Widaa, S., Langford, C., Yang, F., Schuster, S.C., Carter, N.P., Harrow, J., Ning, Z., Herrero, J., Searle, S.M.J., Enright, A., Geisler, R., Plasterk, R.H.A., Lee, C., Westerfield, M., Jong, P.J. de, Zon, L.I., Postlethwait, J.H., Nüsslein-Volhard, C., Hubbard, T.J.P., Crollius, H.R., Rogers, J., Stemple, D.L., 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature 496, 498–503. https://doi.org/10.1038/nature12111.

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

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., Orlando, L., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29, 1682–1684. https://doi.org/10.1093/bioinformatics/btt193.

Kashuba, N., Kırdök, E., Damlien, H., Manninen, M.A., Nordqvist, B., Persson, P., Götherström, A., 2019. Ancient DNA from mastics solidifies connection between material culture and genetics of mesolithic hunter–gatherers in Scandinavia. Communications Biology 2, 1–10. https://doi.org/10.1038/s42003-019-0399-1.

Kircher, M., Sawyer, S., Meyer, M., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 40, e3. https://doi.org/10.1093/nar/gkr771.

Kistler, L., Ware, R., Smith, O., Collins, M., Allaby, R.G., 2017. A new model for ancient DNA decay based on paleogenomic meta-analysis. Nucleic Acids Res. 45, 6310–6320. https://doi.org/10.1093/nar/gkx361.

Lan, R., Reeves, P.R., Lan, R., Reeves, P.R., 2000. Intraspecies variation in bacterial genomes: the need for a species genome concept. Trends Microbiol. 8, 396–401. https://doi.org/10.1016/S0966-842X(00)01791-1.

Larsen, C.S., 2003. Animal source foods and human health during evolution. J. Nutr. 133, 3893S–3897S. https://doi.org/10.1093/jn/133.11.3893S.

Laurence, M., Hatzis, C., Brash, D.E., 2014. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. PLoS One 9, e97876. https://doi.org/10.1371/journal.pone.0097876.

Leonard, J.A., Shanks, O., Hofreiter, M., Kreuz, E., Hodges, L., Ream, W., Wayne, R.K., Fleischer, R.C., 2007. Animal DNA in PCR reagents plagues ancient DNA research. J. Archaeol. Sci. 34, 1361–1366. https://doi.org/10.1016/j.jas.2006.10.023.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Subgroup, 1000 Genome Project Data Processing, 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Lindahl, T., 1993. Instability and decay of the primary structure of DNA. Nature 362, 709–715. https://doi.org/10.1038/362709a0.

Llamas, B., Valverde, G., Fehren-Schmitz, L., Weyrich, L.S., Cooper, A., Haak, W., 2017. From the field to the laboratory: controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. Star: Science & Technology of Archaeological Research 3, 1–14. https://doi.org/10.1080/20548923.2016.1258824.

Ma, M., Dong, G., Jia, X., Wang, H., Cui, Y., Chen, F., 2016. Dietary shift after 3600 cal yr BP and its influencing factors in northwestern China: evidence from stable isotopes. Quat. Sci. Rev. 145, 57–70. https://doi.org/10.1016/j.quascirev.2016.05.041.

Maixner, F., Turaev, D., Cazenave-Gassiot, A., Janko, M., Krause-Kyora, B., Hoopmann, M.R., Kusebauch, U., Sartain, M., Guerriero, G., O'Sullivan, N., Teasdale, M., Cipollini, G., Paladin, A., Mattiangeli, V., Samadelli, M., Tecchiati, U., Putzer, A., Palazoglu, M., Meissen, J., Lösch, S., Rausch, P., Baines, J.F., Kim, B.J., An, H.-J., Gostner, P., Egarter-Vigl, E., Malfertheiner, P., Keller, A., Stark, R.W., Wenk, M., Bishop, D., Bradley, D.G., Fiehn, O., Engstrand, L., Moritz, R.L., Doble, P., Franke, A., Nebel, A., Oeggl, K., Rattei, T., Grimm, R., Zink, A., 2018. The iceman's last meal consisted of fat, wild meat, and cereals. Curr. Biol. 28, 2348–2355. https://doi.org/10.1016/j.cub.2018.05.067 e9.

Makarewicz, C.A., Sealy, J., 2015. Dietary reconstruction, mobility, and the analysis of ancient skeletal tissues: expanding the prospects of stable isotope research in archaeology. Journal of Archaeological Science, Scoping the Future of Archaeological Science: Papers in Honour of Richard Klein 56, 146–158. https://doi.org/10.1016/j.jas.2015.02.035.

Mann, A.E., Sabin, S., Ziesemer, K., Vågene, Å.J., Schroeder, H., Ozga, A.T., Sankaranarayanan, K., Hofman, C.A., Fellows Yates, J.A., Salazar-García, D.C., Frohlich, B., Aldenderfer, M., Hoogland, M., Read, C., Milner, G.R., Stone, A.C., Lewis, C.M., Krause, J., Hofman, C., Bos, K.I., Warinner, C., 2018. Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. Sci. Rep. 8, 9822. https://doi.org/10.1038/s41598-018-28091-9.

Marsh, P.D., Bradshaw, D.J., 1995. Dental plaque as a biofilm. J. Ind. Microbiol. 15, 169–175. https://doi.org/10.1007/BF01569822.

Mauri, M., Elli, T., Caviglia, G., Uboldi, G., Azzi, M., 2017. RAWGraphs: a visualisation platform to create open outputs. In: Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter, CHItaly '17. Association for Computing Machinery, New York, NY, USA, pp. 1–5. https://doi.org/10.1145/3125571.3125585.

Merchant, S., Wood, D.E., Salzberg, S.L., 2014. Unexpected cross-species contamination in genome sequencing projects. PeerJ 2, e675. https://doi.org/10.7717/peerj.675.

Messer, E., 1984. Anthropological perspectives on diet. Annu. Rev. Anthropol. 13, 205–249. https://doi.org/10.1146/annurev.an.13.100184.001225.

Messner, T.C., Stinchcomb, G.E., 2014. Peopling the environment: interdisciplinary inquiries into socioecological systems incorporating paleoclimatology and

geoarchaeology. Method and theory in paleoethnobotany. University Press of Colorado, pp. 257–274.

Molnar, S., Barrett, M.J., Brian, L., Brace, C.L., Brose, D.S., Dewey, J.R., Frisch, J.E., Ganguly, P., Gejvall, N.-G., Greene, D.L., Poirier, F.E., Pourchet, M.J., Rhine, S., Turner, C.G., Van Valen, L., Von Koenigswald, G.H.R., Wilkinson, R.G., Wolpoff, M. H., Wright, G.A., 1972. Tooth wear and culture: a survey of tooth functions among some prehistoric populations [and comments and reply]. Curr. Anthropol. 13, 511–526. https://doi.org/10.1086/201284.

Morehart, C., Morell-Hart, S., 2015. Beyond the ecofact: toward a social paleoethnobotany in mesoamerica. J. Archaeol. Method Theor 22, 483–511. https://doi.org/10.1007/s10816-013-9183-6.

Nelson, M.C., Morrison, H.G., Benjamino, J., Grim, S.L., Graf, J., 2014. Analysis, optimization and verification of illumina-generated 16S rRNA gene amplicon surveys. PloS One 9. https://doi.org/10.1371/journal.pone.0094249.

Nelson, M.C., Ingram, S.E., Dugmore, A.J., Streeter, R., Peeples, M.A., McGovern, T.H., Hegmon, M., Arneborg, J., Kintigh, K.W., Brewington, S., Spielmann, K.A., Simpson, I.A., Strawhacker, C., Comeau, L.E.L., Torvinen, A., Madsen, C.K., Hambrecht, G., Smiarowski, K., 2016. Climate challenges, vulnerabilities, and food security. Proc. Natl. Acad. Sci. Unit. States Am. 113, 298–303. https://doi.org/10.1073/pnas.1506494113.

Neukamm, Judith, Alexander, Peltzer, Nieselt, Kay, 2020. DamageProfiler: Fast damage pattern calculation for ancient DNA. BioRxiv. https://doi.org/10.1101/2020.10.01.322206.

Ottoni, C., Guellil, M., Ozga, A.T., Stone, A.C., Kersten, O., Bramanti, B., Porcier, S., Van Neer, W., 2019. Metagenomic analysis of dental calculus in ancient Egyptian baboons. Sci. Rep. 9, 19637. https://doi.org/10.1038/s41598-019-56074-x.

Ozga, A.T., Nieves-Colón, M.A., Honap, T.P., Sankaranarayanan, K., Hofman, C.A., Milner, G.R., Lewis, C.M., Stone, A.C., Warinner, C., 2016. Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus. Am. J. Phys. Anthropol. 160, 220–228. https://doi.org/10.1002/ajpa.22960.

Ozga, A.T., Gilby, I., Nockerts, R.S., Wilson, M.L., Pusey, A., Stone, A.C., 2019. Oral microbiome diversity in chimpanzees from gombe national park. Sci. Rep. 9, 17354. https://doi.org/10.1038/s41598-019-53802-1.

Pearsall, D., 2018. Case Studies in Paleoethnobotany: Understanding Ancient Lifeways through the Study of Phytoliths, Starch, Macroremains, and Pollen. Routledge, New York, NY.

Poinar, H.N., Hofreiter, M., Spaulding, W.G., Martin, P.S., Stankiewicz, B.A., Bland, H., Evershed, R.P., Possnert, G., Pääbo, S., 1998. Molecular coproscopy: dung and diet of the extinct ground sloth nothrotheriops shastensis. Science 281, 402–406. https://doi.org/10.1126/science.281.5375.402.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramos-Madrigal, J., Runge, A.K.W., Bouby, L., Lacombe, T., Samaniego Castruita, J.A., Adam-Blondon, A.-F., Figueiral, I., Hallavant, C., Martínez-Zapater, J.M., Schaal, C., Töpfer, R., Petersen, B., Sicheritz-Pontén, T., This, P., Bacilieri, R., Gilbert, M.T.P., Wales, N., 2019. Palaeogenomic insights into the origins of French grapevine diversity. Nature Plants 5, 595–603. https://doi.org/10.1038/s41477-019-0437-5.

Reinhard, K., Bryant Jr., V, 1992. Coprolite analysis: a biological perspective on archaeology. Archaeol. Method Theor. 245–288.

Renaud, G., Hanghøj, K., Willerslev, E., Orlando, L., 2017. gargammel: a sequence simulator for ancient DNA. Bioinformatics 33, 577–579. https://doi.org/10.1093/bioinformatics/btw670.

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584. https://doi.org/10.7717/peerj.2584.

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., Reich, D., 2015. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. Phil. Trans. Biol. Sci. 370, 20130624. https://doi.org/10.1098/rstb.2013.0624.

Sato, S, Tabata, S, Hirakawa, H, Klein Lankhorst, R.M., de Jong, H, van Ham, R.C.H.J., Datema, E, Smit, S, Schijlen, E.G.W.M., van Haarst, J.C., Peters, S.A., Henkens, M.H. C., van Staveren, M.J., Mooijman, P.J.W., Hesselink, T, van de Belt, J, Szinay, D, Bai, Y, Visser, R.G.F., 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485, 635–641. https://doi.org/10.1038/nature11119.

Sawafuji, R., Saso, A., Suda, W., Hattori, M., Ueda, S., 2020. Ancient DNA analysis of food remains in human dental calculus from the Edo period, Japan. PloS One 15, e0226654. https://doi.org/10.1371/journal.pone.0226654.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., Pääbo, S., 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS One 7, e34131. https://doi.org/10.1371/journal.pone.0034131.

Schlumbaum, A., Tensen, M., Jaenicke-Després, V., 2008. Ancient plant DNA in archaeobotany. Veg. Hist. Archaeobotany 17, 233–244. https://doi.org/10.1007/s00334-007-0125-7.

Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J.F., AL-Rasheid, K.A., Willerslev, E., Krogh, A., Orlando, L., 2012. Improving ancient DNA read mapping against modern reference genomes. BMC Genom. 13, 178. https://doi.org/10.1186/1471-2164-13-178.

Schubert, M., Lindgreen, S., Orlando, L., 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes 9, 88. https://doi.org/10.1186/s13104-016-1900-2.

Scott, N.S., Tymms, M.J., Possingham, J.V., 1984. Plastid-DNA levels in the different tissues of potato. Planta 161, 12–19. https://doi.org/10.1007/BF00951454.

Sibbesson, E., 2019. Reclaiming the rotten: understanding food fermentation in the neolithic and beyond. Environ. Archaeol. 1–12. https://doi.org/10.1080/14614103.2018.1563374, 0.

Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., Jakobsson, M., 2014. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc. Natl. Acad. Sci. U.S.A. 111, 2229–2234. https://doi.org/10.1073/pnas.1318934111.

Slon, V., Hopfe, C., Weiß, C.L., Mafessoni, F., Rasilla, M. de la, Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M.V., Miller, R., Stewart, J.R., Derevianko, A.P., Jacobs, Z., Li, B., Roberts, R.G., Shunkov, M.V., Lumley, H. de, Perrenoud, C., Gušić, I., Kućan, Ž., Rudan, P., Aximu-Petri, A., Essel, E., Nagel, S., Nickel, B., Schmidt, A., Prüfer, K., Kelso, J., Burbano, H.A., Pääbo, S., Meyer, M., 2017. Neandertal and denisovan DNA from pleistocene sediments. Science 356, 605–608. https://doi.org/10.1126/science.aam9695.

Star, B., Boessenkool, S., Gondek, A.T., Nikulina, E.A., Hufthammer, A.K., Pampoulie, C., Knutsen, H., André, C., Nistelberger, H.M., Dierking, J., Petereit, C., Heinrich, D., Jakobsen, K.S., Stenseth, N.Chr, Jentoft, S., Barrett, J.H., 2017. Ancient DNA reveals the arctic origin of viking age cod from haithabu, Germany. Proc. Natl. Acad. Sci. U. S.A. 114, 9152–9157. https://doi.org/10.1073/pnas.1710186114.

Steinegger, M., Salzberg, S.L., 2020. Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. bioRxiv 2020. https://doi.org/10.1101/2020.01.26.920173, 1.26.920173.

Tromp, M., Dudgeon, J.V., 2015. 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. https://doi.org/10.1016/j.jas.2014.11.024.

Tung, T.A., Knudson, K.J., 2018. Stable isotope analysis of a pre-Hispanic Andean community: reconstructing pre-Wari and Wari era diets in the hinterland of the Wari empire, Peru. Am. J. Phys. Anthropol. 165, 149–172. https://doi.org/10.1002/ajpa.23339.

Tung, T.A., Miller, M, DeSantis, L, Sharp, E.A., Kelly, J, 2016. Patterns of violence and diet among children during a time of imperial decline and climate change in the ancient Peruvian Andes. The archaeology of food and warfare. Springer, pp. 193–228.

Valk, T. van der, Vezzi, F., Ormestad, M., Dalén, L., Guschanski, K., 2020. Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. Molecular Ecology Resources 20, 1171–1181. https://doi.org/10.1111/1755-0998.13009.

Velsko, I.M., Fellows Yates, J.A., Aron, F., Hagan, R.W., Frantz, L.A.F., Loe, L., Martinez, J.B.R., Chaves, E., Gosden, C., Larson, G., Warinner, C., 2019. Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. Microbiome 7, 102. https://doi.org/10.1186/s40168-019-0717-3.

Vera, M., Martinez, P., Bouza, C., 2018. Stocking impact, population structure and conservation of wild brown trout populations in inner Galicia (NW Spain), an unstable hydrologic region. Aquat. Conserv. Mar. Freshw. Ecosyst. 28, 435–443. https://doi.org/10.1002/aqc.2856.

Warinner, C., Hendy, J., Speller, C., Cappellini, E., Fischer, R., Trachsel, C., Arneborg, J., Lynnerup, N., Craig, O.E., Swallow, D.M., Fotakis, A., Christensen, R.J., Olsen, J.V., Liebert, A., Montalva, N., Fiddyment, S., Charlton, S., Mackie, M., Canci, A., Bouwman, A., Rühli, F., Gilbert, M.T.P., Collins, M.J., 2014a. Direct evidence of milk consumption from ancient human dental calculus. Sci. Rep. 4, 7104. https://doi.org/10.1038/srep07104.

Warinner, C., Rodrigues, J.F.M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R.Y., Fiddyment, S., Speller, C., Hendy, J., Charlton, S., Luder, H.U., Salazar-García, D.C., Eppler, E., Seiler, R., Hansen, L.H., Castruita, J.A. S., Barkow-Oesterreicher, S., Teoh, K.Y., Kelstrup, C.D., Olsen, J.V., Nanni, P., Kawai, T., Willerslev, E., Mering, C. von, Lewis, C.M., Collins, M.J., Gilbert, M.T.P., Rühli, F., Cappellini, E., 2014b. Pathogens and host immunity in the ancient human oral cavity. Nat. Genet. 46, 336–344. https://doi.org/10.1038/ng.2906.

Warinner, C., Herbig, A., Mann, A., Fellows Yates, J.A., Weiß, C.L., Burbano, H.A., Orlando, L., Krause, J., 2017. A robust framework for microbial archaeology. Annu. Rev. Genom. Hum. Genet. 18, 321–356. https://doi.org/10.1146/annurev-genom-091416-035526.

Waterson, R.H., Lander, E.S., Wilson, R.K., The Chimpanzee Sequencing and Analysis Consortium, 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437, 69–87. https://doi.org/10.1038/nature04072.

Weiß, C.L., Dannemann, M., Prüfer, K., Burbano, H.A., 2015. Contesting the presence of wheat in the British Isles 8,000 years ago by assessing ancient DNA authenticity from low-coverage data. eLife 4, e10005. https://doi.org/10.7554/eLife.10005.

Weyrich, L.S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A.G., Alt, K.W., Caramelli, D., Dresely, V., Farrell, M., Farrer, A.G., Francken, M., Gully, N., Haak, W., Hardy, K., Harvati, K., Held, P., Holmes, E.C., Kaidonis, J., Lalueza-Fox, C., Rasilla, M. de la, Rosas, A., Semal, P., Soltysiak, A., Townsend, G., Usai, D., Wahl, J., Huson, D.H., Dobney, K., Cooper, A., 2017. Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. Nature 544, 357–361. https://doi.org/10.1038/nature21674.

White, D.J., 1997. Dental calculus: recent insights into occurrence, formation, prevention, removal and oral health effects of supragingival and subgingival deposits. Eur. J. Oral Sci. 105, 508–522. https://doi.org/10.1111/j.1600-0722.1997.tb00238.x.

White, T., Folkens, P., 2005. Human Bone Manual. Academic Press.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Sprinter-Verlag, New York, NY.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. Journal of Open Source Software 4, 1686. https://doi.org/10.21105/joss.01686.

Williams, E., 2010. Salt production and trade in ancient mesoamerica. In: Staller, J., Carrasco, M. (Eds.), Pre-Columbian Foodways: Interdisciplinary Approaches to Food,

Culture, and Markets in Ancient Mesoamerica. Springer, New York, NY, pp. 175–190. https://doi.org/10.1007/978-1-4419-0471-3_7.

Wood, R.E., Higham, T.F.G., Torres, T.D., Tisnérat-Laborde, N., Valladas, H., Ortiz, J.E., Lalueza-Fox, C., Sánchez-Moral, S., Cañaveras, J.C., Rosas, A., Santamaría, D., Rasilla, M.D.L., 2013. A new date for the neanderthals from El Sidrón cave (asturias, northern Spain)*. Archaeometry 55, 148–158. https://doi.org/10.1111/j.1475-4754.2012.00671.x.

Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. Genome Biol. 20, 257. https://doi.org/10.1186/s13059-019-1891-0.

Yates, J.A.F., Lamnidis, T.C., Borry, M., Valtueña, A.A., Fagernäs, Z., Clayton, S., Garcia, M.U., Neukamm, J., Peltzer, A., 2020. Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. bioRxiv 2020. https://doi.org/10.1101/2020.06.11.145615, 6.11.145615.

Zhou, Z., Luhmann, N., Alikhan, N.-F., Quince, C., Achtman, M., 2017. Accurate reconstruction of microbial strains from metagenomic sequencing using representative reference genomes. bioRxiv 215707. https://doi.org/10.1101/215707.

Ziesemer, K.A., Mann, A.E., Sankaranarayanan, K., Schroeder, H., Ozga, A.T., Brandt, B. W., Zaura, E., Waters-Rist, A., Hoogland, M., Salazar-García, D.C., Aldenderfer, M., Speller, C., Hendy, J., Weston, D.A., MacDonald, S.J., Thomas, G.H., Collins, M.J., Lewis, C.M., Hofman, C., Warinner, C., 2015. Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. Sci. Rep. 5, 16498. https://doi.org/10.1038/srep16498.

Ziesemer, K.A., Ramos-Madrigal, J., Mann, A.E., Brandt, B.W., Sankaranarayanan, K., Ozga, A.T., Hoogland, M., Hofman, C.A., Salazar-García, D.C., Frohlich, B., Milner, G.R., Stone, A.C., Aldenderfer, M., Lewis, C.M., Hofman, C.L., Warinner, C., Schroeder, H., 2019. The efficacy of whole human genome capture on ancient dental calculus and dentin. Am. J. Phys. Anthropol. 168, 496–509. https://doi.org/10.1002/ajpa.23763.