An efficient pipeline for creating metagenomic-assembled genomes from ancient oral microbiomes

Francesca J. Standeven[1], Gwyn Dahlquist-Axe[1], Camilla F. Speller[2], Conor J. Meehan[3], Andrew Tedder[1]

[1]School of Chemistry and Biosciences, University of Bradford

[2]Department of Anthropology, University of British Columbia

[3]Department of Biosciences, Nottingham Trent University

Abstract

Metagenomic-assembled genomes (MAGs) are difficult to recover from ancient DNA (aDNA) due to substantial fragmentation, degradation, and multi-source contamination. These complexities associated with aDNA raise concerns about whether bioinformatic tools intended for interpreting modern DNA are suitable for reconstructing ancient MAGs. Using simulated modern and ancient data, we investigated: 1) how using binning tools designed for modern DNA affects our ability to effectively construct MAGs from ancient genomes; 2) the performance of three different binning tools for aDNA samples; and 3) whether a 'one size fits all' approach is suitable for ancient metagenomics. We established that binning tools for modern DNA performed efficiently on simulated modern and ancient DNA. When applied to 'real' archaeological DNA spanning 5000 years, we retrieve high-confidence MAGs in most cases.

Introduction

Recovering meaningful genomic data at the operational taxonomic unit (OTU) level from ancient microbiome (microbial community) samples requires the use

of metagenomic-assembled genomes (MAGs). This differs from a single ‘species’ approach, in that a MAG represents a microbial genome using a group of sequences from assembled genomes with similar traits (Yang et al. 2021) rather than identifying species based on individual sequences. Reconstructing MAGs, therefore, offers us significant insight into uncultured microbial genomic diversity (Churcheward et al. 2022), can enable the detection of new species and offer evidence into their potential functions in a dynamic ecosystem (Yang et al. 2021), and the integration of ancient MAGs enriches this analysis. For example, a recent study by Klapper et al. (2023) reconstructed ancient Chlorobium MAGs which uncovered unmodified metabolites, allowing for the observation of functional traits over millennia.

Archaeological science has delved into the historical events and lifestyle transitions that had a huge impact on the health of past societies, and the creation of ancient oral MAGs has allowed for an understanding of the origins and evolution of microbial species and strains in the oral cavity. For example, Wagner et al. (2014) reassembled a *Yersinia pestis* draft genome from dental calculus, revealing a novel branch dating back to the Justinian plague. This study showed how their strain differs from strains connected with future pandemics caused by *Y. pestis*, such as the Black Death. Other research has also used *Y. pestis* genomes to evidence its establishment during the Bronze Age (Spyrou et al. 2018). In another study, the reconstruction of *Methanobrevibacter* genomes from historical dental calculus across different regions has revealed previously unknown species that demonstrate a decline of the genus over thousands of years due to changing diets (Granehäll et al. 2021). Metagenomic assembled *Olsenella* sp. Oral taxon 807 genomes showed several missing components related to virulence and biofilm formation compared to modern genomes, implying that the species has recently evolved and adapted to thrive in the oral

microbiome, most likely due to stronger selective pressure exerted by dietary changes (Quagliariello et al. 2022).

Due to the poor quality and quantity of aDNA, combined with high contamination levels (Gaeta 2021), ancient metagenomics is still an extremely challenging field of study; therefore, the efficiency of modern MAG generation techniques when applied to aDNA remains unclear. Depurination, a primary mechanism for miscoding lesions in aDNA, converts cytosine into uracil which is often mistaken for thymine (Dabney et al. 2013). The depurination process can also result in highly fragmented DNA (Kistler et al. 2017). Hydrolytic depurination, followed by beta-elimination, causes single-strand breaks in DNA when exposed to acidic environmental conditions (Dabney et al. 2013). The extent of this damage depends on the location and duration of the DNA's burial; nevertheless, aDNA fragments are generally very short ranging from 40–500 bp (Dabney et al. 2013). The chemical processes that cause damage have the potential to influence the performance of modern MAG assembly methods on aDNA. aDNA is also prone to environmental and modern contamination because it can be situated in the ground for thousands of years, making it susceptible to contamination from environmental microbes in the soil (Kazarina et al. 2019). Likewise, the degraded nature of aDNA means that it is frequently low template, and thus at more risk from modern contamination resulting from post-excavation handling including the laboratory environment (Llamas et al. 2017). Overall, both damage and contamination affect the ability to create ancient MAGs with the same degree of confidence as modern DNA because the effects of these traits on metagenomic assembly are not fully understood.

Although the field of aDNA is quickly progressing, it is still difficult to retrieve high sequencing coverage for historical samples with short read lengths and

various challenges can arise when used with bioinformatic software packages designed for modern long-read DNA. Commonly applied data pre-processing approaches exacerbate these issues by further shortening the read length and likely impacting MAG assembly. For example, necessary decontamination procedures such as BBduk (Joint Genome Institute 2023) have the potential to greatly reduce read length by removing vast amounts of possibly contaminated sequences and can cause short, uneven counts in paired-end reads. Additional tools, such as Trimmomatic (Bolger et al. 2014) or PEAR (Zhang et al. 2014), must sometimes be applied to repair these files by evenly cutting off base pairs that result in even more sequence reduction. Therefore, the short read length in ancient metagenomics emphasises the need for experimentation in genome assembly techniques.

MAGs are valuable resources for archaeological scientists and microbiologists. Therefore, utilising efficient and effective assembly methodologies is paramount for ensuring accurate and authentic ancient genomes. Reconstructing MAGs involves genomic binning where organised contigs (contiguous fragments) are sorted into groups of high similarity, or ‘bins’, whereby each bin corresponds to a MAG (Setubal 2021). A bin is an individual genome reconstruction made from contigs that are similar in composition. Binning methodologies vary, but a common theme involves comparing contigs to reference genomes as part of the decision tree. Several successful binning tools are available for modern DNA, such as Maxbin (Wu et al. 2016), CONCOCT (Alneberg et al. 2014), and MetaBAT (Kang et al. 2015), as well as many MAGs pipelines that have one or more binning tools already implemented (Uritskiy et al. 2018). It has been argued that a combination of binning tools can strengthen binning by the application of many independent variables for contig clustering (Hug 2018). After applying various binning tools, overlaps and redundancy can

be determined. For example, the programme DAS Tool can be used to combine various binning algorithms to rewrite and calculate an optimised, non-redundant selection of bins from a single assembly (Sieber *et al*. 2018). In light of all of the issues with aDNA, it is unclear to what extent existing binning methods and pipelines are effective, and whether further optimisation is required. Despite recent pipeline developments with functionalities geared towards ancient MAG production (Krakau et al. 2022), systematic comparisons of modern MAG binning tools on ancient DNA are lacking.

MAGs give a degree of reliability because they require numerous reads to form their assembly and we can calculate their completeness and contamination; however, the performance of modern DNA binning techniques on ancient sequences is understudied and could result in unreliable MAG outcomes, such as false positives and false negatives. To determine whether standard tools for modern MAG creation are suitable for aDNA, we tested our binning pipeline on simulated 'modern' and 'ancient' data by quantifying MAG success and binning performance. Our method was then applied to a real aDNA dataset obtained from archaeological dental calculus to understand if binning tools developed for modern DNA have aDNA limitations and if factors such as short read length is a barrier to MAG recovery success in aDNA.

Materials

Simulated data

To disentangle the impact of aDNA deamination (damage) on the process of MAG recovery and reconstruction, simulated metagenomic populations were created using InSilicoSeq v1.6.0 (Gourlé et al. 2019). Twenty modern populations were simulated to reflect paired-end (PE) 300 bp MiSeq reads.

‘Populations’ comprised three replicate samples, each of which contained reads taken from 10 microbial genomes (not limited to oral microbial genomes) pulled randomly from NCBI. The frequency of each OTU (per sample) was determined at random (Gourlé et al. 2019). Simulated modern samples then underwent simulated delamination using ‘deamSim’ from gargammel v1.1.4 (Renaud et al. 2017) with the following default parameters -damage 0.03, 0.4, 0.01, 0.7 to damage nucleotide bases. Fragmentation is not explicitly modelled in this dataset. Information on species and abundance for the simulated datasets is provided in **Table S1**.

Archaeological data

The ancient oral microbiome samples used in this research were selected from data reported by (Standeven et al. 2024). The dataset comprises 30 pre-industrial (ca. 2,300 BC–1750), 21 Industrial (ca. 1750–1850), and 54 post-industrial (ca. 1837–1901) samples. Sample ID, European Nucleotide Archive (ENA) accession number, sample type, site location, historical period, and read length of samples are listed in **Table S2.**

Methodology

Our simulated data and real archaeological samples were run through our hybrid-scripted MAGs pipeline (**Fig 1**) described below. The pipeline was first applied to the simulated data to test its performance before real archaeological samples. Then, real archaeological samples were run through the pipeline via period batch (see **materials**) and then per sample to comprehend how ancient characteristics such as low read count affect binning. Damage authentication and decontamination protocols were only applied to real archaeological data with matched laboratory controls.

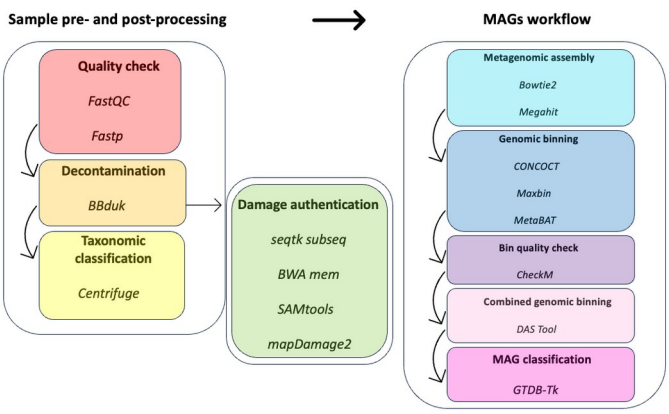


Figure 1. MAGs assembly and QC workflow.

Hybrid-scripted MAGs pipeline

Sample pre-processing

Standeven *et al*. (2024) provide a full overview of quality control, decontamination, and damage authentication protocols for ancient metagenomic sequences. FastQC v0.11.9 (Babraham Bioinformatics 2019) was used with default parameters to quality-check samples, pre- and post-filtering. Fastp v0.23.2 (Chen et al. 2018) was used with default parameters to perform read filtering, trimming, adapter removal, and mismatched bp repair on raw metagenomic sequencing data. To minimise the impact of laboratory and environmental contamination, BBduk (Joint Genome Institute 2023) was used with default parameters to identify and remove homologous sequences found in laboratory and environmental controls - when matching environmental samples (bone and tooth root) were available (Standeven et al. 2024) - from raw trimmed reads. Centrifuge v1.0.3 (Kim et al. 2016) was performed with default parameters to screen for OTUs. Non-human reads were retained from Centrifuge outputs and converted into post-Centrifuge fastq files using seqtk v1.4 ‘subseq’ (Li 2024). To authenticate ancient sequences, we retained only human reads from Centrifuge outputs and converted them into fastq files using seqtk ‘subseq’ which were then mapped to the human genome (hg38) (NCBI

2013) using BWA mem v0.7.17. SAMtools v1.12 (Danecek et al. 2021) was used to prepare the files for mapDamage2 v2.2.2 (Jónsson et al. 2013) which was run with default parameters.

Metagenomic assembly

Metagenomic assembly of post-Centrifuge fastq files was performed using Megahit v1.2.9 (Li et al. 2015), a tool used to assemble large and complex metagenomics *de novo* datasets. The parameters --kmin-1pass --no-mercy were applied for memory efficiency while recovering low-coverage sequences. To ensure file compatibility with later tools, contigs produced by Megahit were reformatted into fasta files using the microbial 'omics platform anvio v7 (Eren et al. 2020). Anvio was used to filter out short contigs <2500bp (-l 2500, --simplify-names) to promote good MAG quality. Studies recommend using contigs no less than 1000 bp for binning (Meziti et al. 2021) and given the ancient nature of our samples, we aimed to be especially cautious and apply stringent parameters to minimize the risk of false positives. Bowtie2 v2.5.0 (Langmead and Salzberg 2012) was used to map raw reads back to the assembly with the parameter -no-unal to suppress unaligned reads.

Genomic binning

Contigs were binned using three individual binning tools: MetaBAT v2.12.1 (Kang *et al*. 2015) with parameters -m 2000 for minimum contig size, and --saveCls to save cluster memberships; CONCOCT v1.1.0 (Alneberg *et al*. 2014) with parameters -c 10000 for maximal number of clusters and a default minimum contig length threshold of 1000; and Maxbin2 v2.2.7 (Wu *et al*. 2016) via anvio with a default minimum contig length threshold of 1000.

To obtain the best quality MAGs, DAS Tool v1.1.6 (Sieber *et al*. 2018) was used with default parameters to merge, de-replicate, and select good-quality bins

225 (≥75% completeness) from each binning tool. After DAS Tool selected bins, bins
226 were then designated for further classification based on a completeness
227 threshold of ≥75%.

229 *Quality checking*

230 We utilised CheckM v1.2.2 (Parks *et al*. 2015) with the automatic lineage
231 workflow (lineage_wf) to assess percentages of completeness and
232 contamination in MAGs. We conformed to recent guidelines (Bowers *et al*. 2017)
233 that group MAGs into either 'low-quality draft' (<50% complete with <10%
234 contamination), 'medium-quality draft' (≥50% complete with <10%
235 contamination), or 'high-quality draft' (>90% complete with <5% contamination)
236 genomes.
237
238 *MAG taxonomic classification*

239 To assign taxonomic labels to the MAGs, we used the Genome Taxonomy
240 Database Toolkit v1.0.2 (GTDB-Tk; Chaumeil *et al*. 2020) using the workflow
241 'classify_wf' to process genomes using both bacterial and archaeal markers.
242

243 **Quantifying MAG success**

244 After running the simulated data through the pipeline shown in **Fig 1**, MAGs
245 success was measured in the following way: Firstly, to determine if MAG
246 classifications corresponded to the correct OTUs, FastANI v1.33 (Jain et al. 2018)
247 was used to calculate the average nucleotide identity (ANI) between MAG files
248 and original OTUs (see **Table S3** for FastANI outputs detailing query and
249 reference genome, ANI, fragment mappings and total query fragments). Then,
250 a Python script was run to query GenBank via Entrez (the NCBI retrieval system)
251 for tax ID numbers for each initial OTU and each MAG in FastANI files, then the
252 tax IDs were matched to calculate true and false positives (see **Table S4** for

information on tax ID and tax name corresponding with OTUs). A paired t-test was performed on the number of true and false positive MAGs between modern and ancient sample sets to measure overall MAG retrieval success.

It is important to understand factors that may influence failure to retrieve MAGs for each OTU in a given sample. To disentangle factors linked to MAG retrieval success, such as OTU abundance in the initial sample, OTU genome size, and OTU read count, we used FastANI v1.33 (Jain et al. 2018) to attribute MAGs from both modern and ancient simulated samples to an OTU. We used an ANI of 96% (Ciufo et al. 2018) to determine MAG identity. MAG retrieval success (MAG produced or not) was then compared to OTU abundance in the initial sample population pool and OTU genome size, as well as OTU abundance and read count.

Quantifying binning tool performance

To determine whether binning using a single binning tool performs adequately for aDNA, or whether a multi-binning tool approach is required, we compared the frequency in which all three binning tools produce MAGs for the same OTU above our acceptance threshold of 75% completeness, before the DAS tool step. This analysis was performed on simulated ancient data.

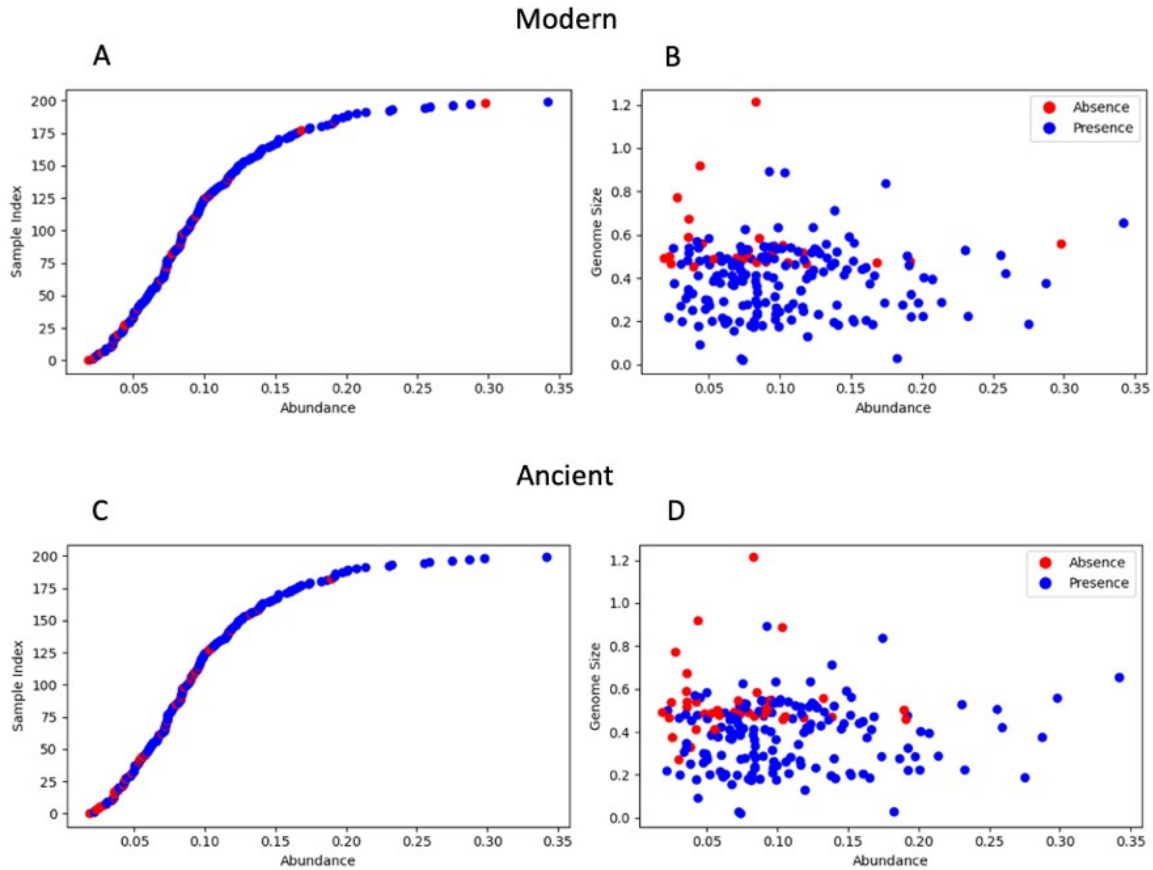
Applying modern genome assembly methods to real archaeological data

After running the pipeline on simulated data, the MAG pipeline described above (**Fig 1**) was applied to real archaeological dental calculus to understand if binning is effective on single samples in comparison to groups of samples separated by historical period.

Results

MAG success on simulated data

283 85.5% (171 out of 200) of simulated modern OTUs and 79.5% (159 out of 200) of
284 simulated ancient OTUs were represented by MAGs. The paired t-test indicated
285 no significant difference (p-value = 0.94) in MAG success between modern and
286 ancient simulated groups. Neither simulated group had evidence of false
287 positive MAGs. While MAG retrieval success was not significantly different
288 between simulated modern and ancient samples, the MAG pipeline failed to
289 retrieve MAGs (false negatives) in simulated ancient samples with a higher
290 frequency (20.5% in ancient vs. 14.5% in modern). As expected, low OTU
291 abundance, shown with varying read count (sample index), in both simulated



292 sets results in failure to retrieve MAGs (**Fig 2a & c**), although we also see OTUs
293 with high abundance (>10%) fail in both sequence types. This appears to
294 correlate with OTU genome size (**Fig 2b & d**), with larger genomes increasing
295 the likelihood of failure to retrieve MAGs even with moderate to high
296 abundance.
297

Figure 2. MAGs presence (blue) and absence (red) based on sorted abundance and read count (a & c) and abundance and genome size (b & d) for both simulated modern and ancient groups.

Binning tool performance on simulated data

Binning tools all achieve overlapping bins in only 44% of cases. In the remaining 56% of cases where one or more of the binning tools retrieved a unique bin not retrieved by the other tools, each binning tool (CONCOCT = 35%; Maxbin = 40%; and MetaBAT = 25%) retrieved a proportionately similar number unique bins, although CONCOCT and Maxbin performed slightly better.

Archaeological MAGs

Following the simulated data experiments, we ran the pipeline on our real archaeological metagenomic dataset to examine binning performance on samples via period and per sample.

With one exception, none of the binning tools were able to retrieve MAGs from single samples. A medieval tooth sample (LC26T; 125 bp) was the only instance in which MAGs (with varying quality) were identified. While some of these MAGs corresponded with those found in the pooled samples from the pre-industrial period, there was not a complete overlap. MAG retrieval was much more successful when samples were grouped into meaningful period categories. CheckM completeness and contamination levels for all bins (before DAS Tool selection) across all periods are summarised below and presented in **Table 2** and **Fig 3**. A more detailed summary of information on the final MAGs, including completeness and contamination, GTDB classification, and genome quality can be found in **Table S5**. In total, 13 pre-industrial bins (CONCOCT=6, Maxbin=2, MetaBAT=5), 7 Industrial bins (CONCOCT=4, Maxbin=0, MetaBAT=3), and 18 Post-Industrial bins (CONCOCT=7, Maxbin=2, MetaBAT=9) from the DAS Tool bin

list output met the ≥75% completeness threshold. CheckM values indicate that the majority of bins (34 out of 38 bins) range from medium (≥50% complete with <10% contamination) to high (≥90% complete with <5% contamination) quality.

Table 2. Table of Check M completeness and contamination values for bins, including bin period, binning tool, and number of bins.

Date	Binning tool	Number of bins	Check M completeness (%)	Check M contamination (%)
pre-Industrial	CONCOCT	38	Mean 49.81 Median 55.59 IQR 66.18 SD 33.71 SE 5.46	Mean 3.73 Median 1.5 IQR 3.55 SD 8.19 SE 1.32
	Maxbin	25	Mean 66.27 Median 69.20 IQR 39.21 SD 25.39 SE 5.09	Mean 15.25 Median 13.32 IQR 24.25 SD 16.38 SE 3.27
	MetaBAT	26	Mean 62.60 Median 66.53 IQR 53.29 SD 28.60 SE 5.61	Mean 6.38 Median 1.52 IQR 4.19 SD 21.62 SE 4.24
Industrial	CONCOCT	35	Mean 64.79 Median 63.65 IQR 47.52 SD 27.75 SE 4.69	Mean 2.10 Median 0.70 IQR 1.74 SD 3.97 SE 0.75
	Maxbin	18	Mean 71.40 Median 71.49 IQR 18.81 SD 17.05 SE 4.01	Mean 15.33 Median 9.12 IQR 20.19 SD 17.46 SE 4.11
	MetaBAT	22	Mean 53.39 Median 57.11 IQR 36.26 SD 24.19 SE 5.15	Mean 1.12 Median 1.12 IQR 1.82 SD 1.05 SE 0.22
post-Industrial	CONCOCT	64	Mean 43.28 Median 38.27 IQR 59.02 SD 30.61 SE 3.82	Mean 1.92 Median 0.81 IQR 3.40 SD 2.76 SE 0.34
	Maxbin	31	Mean 70.75 Median 77.54 IQR 55.44 SD 22.77 SE 4.09	Mean 19.91 Median 6.62 IQR 36.95 SD 20.79 SE 3.73

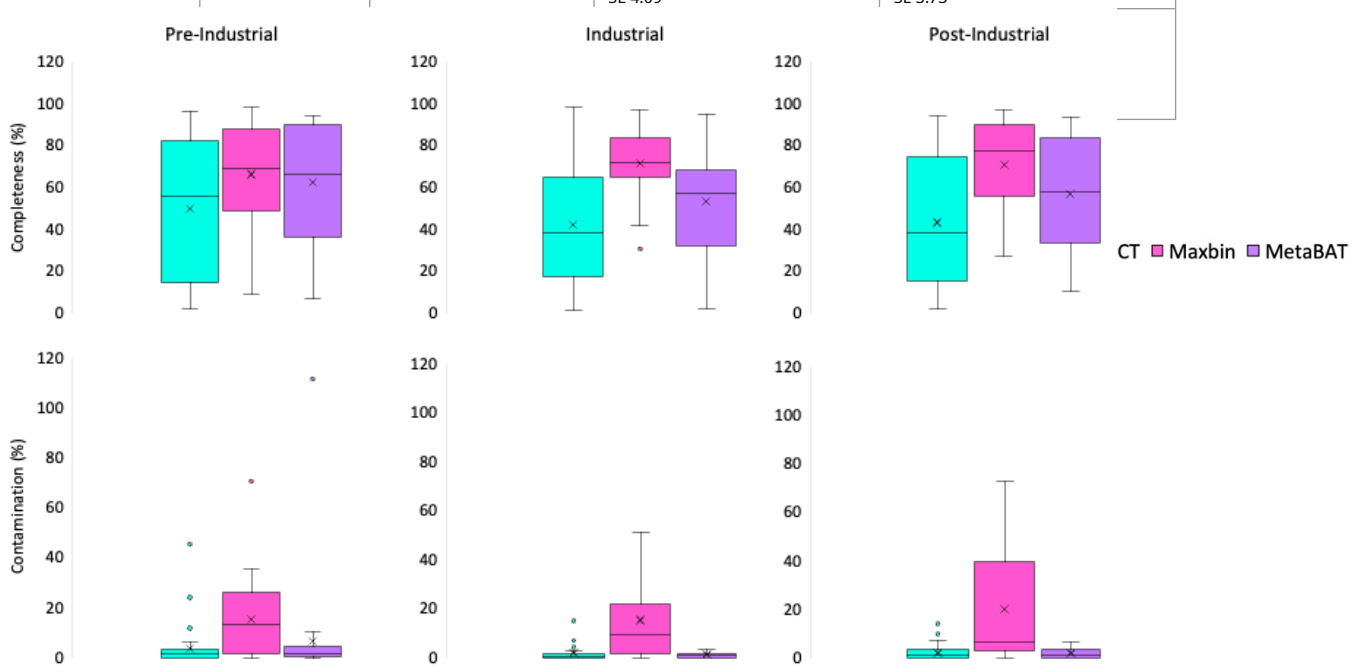


Figure 3. Varying contamination and completeness statistics (CheckM) for CONCOCT, Maxbin, and MetaBAT bins (before DAS Tool selection) across the Pre-Industrial, Industrial, and Post-Industrial periods.

Discussion

Modern MAG binning tools work on aDNA

To determine if MAG retrieval tools, as they are commonly integrated into pipelines for recovery of MAGs from modern microbiome samples, are also effective on aDNA, we first applied a standard pipeline (**Fig 1**) containing three binning tools (CONCOCT, MetaBAT, Maxbin) to 20 simulated modern and ancient datasets. MAG recovery success in modern simulated samples (85.5%) was greater than in ancient samples (which included the same reads, with simulated damage) (79.5%), but this difference was not statistically significant (p-value = 0.94). The pipeline did not produce false positives (absent MAGs) for either group. This suggests that under controlled conditions, i.e. with simulated data, modern binning tools can be applied to aDNA with some degree of success.

Ancient and modern metagenomics lack studies on links between OTU abundance and genome size and MAG recovery; however, this research shows that OTU abundance in a sample is important and that its combination with OTU genome size plays a key role in MAG recovery. OTUs with low sample abundance and large genome sizes have an increased chance of MAG recovery failure. While this is an issue for both groups, ancient samples have a larger risk, due to sequence degradation impacting sequence similarity. Read count and abundance did not seem to affect MAG success in any group.

Factors that impact MAG success such as abundance and genome size are likely to be complicated in ‘real’ aDNA samples due in part to different rates of DNA degradation linked to factors such as time, soil pH, and temperature, which may decrease the accuracy of MAG retrieval. Additionally, sequencing strategy may also impact MAG retrieval success in ancient samples. In our simulated data set, the read length was 300 bp. While this isn’t unreasonable for modern data with contemporary sequencing technology, DNA fragmentation linked to the factors mentioned above may make recovering fragments of this length from aDNA samples difficult. Certainly, most aDNA samples sequenced pre-industrial will not meet this read length standard. This means that one might anticipate that aDNA samples with varying degrees of fragmentation and degradation linked to time since burial and burial conditions, will likely have reduced MAG retrieval success. The older a sample (or the worse its preservation), the higher the likelihood of false negative OTUs. This may be compounded further by the sequencing read length strategy used, with short read lengths in combination with high DNA degradation likely to have a high impact on MAG retrieval success.

In addition to read lengths and damage, another potential factor is the impact of sequencing depth on MAGs retrieval. A MAGs simulation study (Steinke et al. 2024) found that even slight changes, as minimal as 1%, in coverage depth can affect the recovery of a MAG. Shallow sequencing depth may thus hinder the representation of low-abundance OTUs in MAGs. Recent pipelines combining third-generation and ultra-deep second-generation sequencing have improved sequencing depth for low-abundance subpopulations in human microbiomes and enhanced binning performance (Jin et al. 2022). Therefore, if low genomic coverage is significantly affecting the retrieval of modern MAGs, the field must consider developing better extraction and sequencing techniques for aDNA.

Single vs. combined binning tools on aDNA

For the real archaeological metagenomes, each binning tool produced a different number of bins with different levels of contamination and completeness (see **Fig. 3**), with Maxbin producing bins showing contamination levels as high as 19.91% while the other two produced bins below 7%, despite having used the same data. The differences in binning results raise questions about which tool is performing the best and whether a multi-binning strategy is necessary at all.

Using simulated data, we established that combining all three binning tools is likely to outperform a pipeline that uses only one of the binning tools, as the three tools found completely overlapping bin sets in only 44% of cases. This likely means that when applied to degraded ancient data, using a pipeline which employs only one binning tool is likely to lead to a significant underestimation of OTUs in any given sample. Again, given that this is simulated ancient data, with only moderate degradation and fragmentation on 300 bp paired-end data, this may be an underestimation of the number of OTU MAGs that may be missed with 'real' aDNA which will have variable degradation and potentially shorter read lengths.

This type of research regarding binning combination to look at OTU representation in MAGs is lacking in the field of ancient and modern metagenomics so there are not many studies to compare this to. But we do know that other research yielded the most high-quality bins when combining multiple bins on modern metagenomic data compared to using just one binning tool (Uritskiy et al. 2018; Yue et al. 2020; Rühlemann et al. 2022). More research like this using simulated data with artificial damage would be

advantageous in the field of ancient metagenomics because then we would be able to see the effect that damage has on MAG success.

Optimal amount of samples for MAG assembly

To understand factors affecting MAG retrieval success in real archaeological data, we applied our multi-binning pipeline approach to real aDNA from dental calculus samples across multiple periods using different read lengths. With a 'single sample' approach, in which we applied the pipeline to each sample individually, the results were relatively striking, namely, retrieving MAGs is frequently not possible. This may not be surprising, given that aDNA samples will be fragmented, often poor quality, and samples may well be sequenced from relatively low quantities of DNA. The abundance of any given OTU in the sample may also play a key role. While samples with longer read lengths had slightly higher retrieval success, it is clear that a single-sample approach is not meaningful with aDNA, and that grouping samples, based on meaningful parameters (sample age, location etc) is a more appropriate strategy for MAG retrieval. Some of these samples are likely part responsible for GTDB-Tk not always being able to assign family, genus, or species of bacteria to MAGs, and it would be useful if more studies were to investigate the binning of different periods to further investigate how read count/damage ('ancientness') of certain samples contributes to poorly defined MAGs (taxonomically undeterminable past family level).

Conclusion

This research has revealed that modern MAGs binning tools can retrieve medium to high quality MAGs that reflect original OTUs from highly damaged

reads. Our findings have also revealed that a combined binning approach on multiple samples is more beneficial than binning single samples and using one tool to bin samples. We emphasise the importance of testing the suitability of modern tools on simulated data in ancient metagenomics as aDNA has varying characteristics that may affect MAG success. Moreover, we urge for more research to investigate the impact of damaged aDNA with low-coverage short-read length on MAG creation.

Acknowledgments

The authors acknowledge the use of the University of Bradford High Performance Computing Service in the completion of this work.

Author Contributions

Conceptualization: A.T., C.J.M. Supervision: A.T., C.J.M., and C.S. Funding acquisition: C.S., A.T., C.J.M. Investigation: F.J.S, G D-A., and A.T. Formal Analysis: F.J.S, G. D-A., A.T. and C.J.M. Data Curation, Visualization and Writing - Original Draft: F.J.S. Writing - Review & Editing: all authors.

Bibliography

Achtman, M. and Zhou, Z. (2020) Metagenomics of the modern and historical human oral microbiome with phylogenetic studies on Streptococcus mutans and Streptococcus sobrinus. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1812), 20190573.

Alneberg, J., Smári Bjarnason, B., de Bruijn, I., Schirmer, M., Quick, J., Umer, I., Lahti, L., Loman, N., Andersson, A. and Quince, C. (2014) Binning metagenomic contigs by coverage and composition. *Nature Methods* 11, 1144–1146.

Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., Manghi, P., Dubois, L., Huang, K. D., Thomas, A. M., Nickols, W. A., Piccinno, G., Piperni, E., Punčochář, M., Valles-Colomer, M., Tett, A., Giordano, F., Davies, R., Wolf, J., Berry, S. E., Spector, T. D., Franzosa, E. A., Pasolli, E., Asnicar, F., Huttenhower, C. and Segata, N. (2023) Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology*.

Borry, M., Hübner, A., Rohrlach, A. B. and Warinner, C. (2021) PyDamage: automated ancient damage identification and estimation for contigs in ancient DNA de novo assembly. *PeerJ* 9, e11845.

Bos, K. I., Harkins, K. M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S. A., Bryant, J. M., Harris, S. R., Schuenemann, V. J., Campbell, T. J., Majander, K., Wilbur, A. K., Guichon, R. A., Wolfe Steadman, D. L., Cook, D. C., Niemann, S., Behr, M. A., Zumarraga, M., Bastida, R., Huson, D., Nieselt, K., Young, D., Parkhill, J., Buikstra, J. E., Gagneux, S., Stone, A. C. and Krause, J. (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 514 (7523), 494-497.

Bos, K. I., Schuenemann, V. J., Golding, G. B., Burbano, H. A., Waglechner, N., Coombes, B. K., McPhee, J. B., DeWitte, S. N., Meyer, M., Schmedes, S., Wood, J., Earn, D. J. D., Herring, D. A., Bauer, P., Poinar, H. N. and Krause, J. (2011) A draft genome of Yersinia pestis from victims of the Black Death. *Nature* 478 (7370), 506-510.

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., Schulz, F., Jarett, J., Rivers, A. R., Eloe-Fadrosh, E. A., Tringe, S. G., Ivanova, N. N., Copeland, A., Clum, A., Becraft, E. D., Malmstrom, R. R., Birren, B., Podar, M., Bork, P., Weinstock, G. M., Garrity, G. M., Dodsworth, J. A., Yooseph, S., Sutton, G., Glöckner, F. O., Gilbert, J. A., Nelson, W. C., Hallam, S. J., Jungbluth, S. P., Ettema, T. J. G., Tighe, S., Konstantinidis, K. T., Liu, W.-T., Baker, B. J., Rattei, T., Eisen, J. A., Hedlund, B., McMahon, K. D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G. W., Rinke, C., Lapidus, A., Meyer, F., Yilmaz, P., Parks, D. H., Murat Eren, A., Schriml, L., Banfield, J. F., Hugenholtz, P. and Woyke,

T. (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology* 35 (8), 725-731.

Brodmann, T., Endo, A., Gueimonde, M., Vinderola, G., Kneifel, W., de Vos, W. M., Salminen, S. and Gómez-Gallego, C. (2017) Safety of novel microbes for human consumption : Practical examples of assessment in the European Union. *Frontiers in microbiology* 8, 1725-1725.

Churcheward, B., Millet, M., Bihouée, A., Fertin, G. and Chaffron, S. (2022) MAGNETO: An Automated Workflow for Genome-Resolved Metagenomics. *mSystems* 7 (4), e0043222-e0043222.

Dabney, J., Meyer, M. and Paabo, S. (2013) Ancient DNA Damage. *Cold Spring Harbor Perspectives in Biology* 5 (7), a012567-a012567.

Eisenhofer, R. and Weyrich, L. (2018) Proper Authentication of Ancient DNA Is Still Essential. *Genes* 9 (3), 122.

Gourlé, H., Karlsson-Lindsjö, O., Hayer, J. and Bongcam-Rudloff, E. (2019) Simulating Illumina metagenomic data with InSilicoSeq. *Bioinformatics* 35 (3), 521-522.

Granehäll, L., Huang, K. D., Tett, A., Manghi, P., Paladin, A., O'Sullivan, N., Rota-Stabelli, O., Segata, N., Zink, A. and Maixner, F. (2021) Metagenomic analysis of ancient dental calculus reveals unexplored diversity of oral archaeal Methanobrevibacter. *Microbiome* 9 (1).

Hug, L. A. (2018) Subsampled Assemblies and Hybrid Nucleotide Composition/Differential Coverage Binning for Genome-Resolved Metagenomics. Springer New York. 215-225. https://dx.doi.org/10.1007/978-1-4939-8728-3_14

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

Klapper, M., Hübner, A., Ibrahim, A., Wasmuth, I., Borry, M., Haensch, V. G., Zhang, S., Al-Jammal, W. K., Suma, H., Fellows Yates, J. A., Frangenberg, J., Velsko, I. M., Chowdhury, S., Herbst, R., Bratovanov, E. V., Dahse, H.-M., Horch, T., Hertweck, C., González Morales, M. R., Straus, L. G., Vilotijevic, I., Warinner, C. and Stallforth, P. (2023) Natural products from reconstructed bacterial genomes of the Middle and Upper Paleolithic. *Science (American Association for the Advancement of Science)* 380 (6645), 619-624.

Li, D., Liu, C.-M., Luo, R., Sadakane, K. and Lam, T.-W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31 (10), 1674-1676.

Ma, L., Xia, Y., Li, B., Yang, Y., Li, L.-G., Tiedje, J. M. and Zhang, T. (2016) Metagenomic Assembly Reveals Hosts of Antibiotic Resistance Genes and the Shared Resistome in Pig, Chicken, and Human Feces. *Environmental Science & Technology* 50 (1), 420-427.

Nurk, S., Meleshko, D., Korobeynikov, A. and Pevzner, P. A. (2017) metaSPAdes: a new versatile metagenomic assembler. *Genome Research* 27 (5), 824-834.

Quagliariello, A., Modi, A., Innocenti, G., Zaro, V., Conati Barbaro, C., Ronchitelli, A., Boschin, F., Cavazzuti, C., Dellù, E., Radina, F., Sperduti, A., Bondioli, L., Ricci, S., Lognoli, M., Belcastro, M. G., Mariotti, V., Caramelli, D., Mariotti Lippi, M., Cristiani, E., Martino, M. E., Muntoni, I. M. and Lari, M. (2022) Ancient oral microbiomes support gradual Neolithic dietary shifts towards agriculture. *Nature Communications* 13 (1).

Renaud, G., Hanghøj, K., Willerslev, E. and Orlando, L. (2017) gargammel: a sequence simulator for ancient DNA. *Bioinformatics* 33 (4), 577-579.

Santiago-Rodriguez, T., Fornaciari, G., Luciani, S., Toranzos, G., Marota, I., Giuffra, V. and Cano, R. (2017) Gut Microbiome and Putative Resistome of Inca and Italian Nobility Mummies. *Genes* 8 (11), 310.

Setubal, J. C. (2021) Metagenome-assembled genomes: concepts, analogies, and challenges. *Biophysical Reviews* 13 (6), 905-909.

Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G. and Banfield, J. F. (2018) Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology* 3 (7), 836-843.

Spyrou, M. A., Tukhbatova, R. I., Wang, C.-C., Valtueña, A. A., Lankapalli, A. K., Kondrashin, V. V., Tsybin, V. A., Khokhlov, A., Kühnert, D., Herbig, A., Bos, K. I. and Krause, J. (2018) Analysis of 3800-year-old Yersinia pestis genomes suggests Bronze Age origin for bubonic plague. *Nature Communications* 9 (1).

Veech, J. A., Summerville, K. S., Crist, T. O. and Gering, J. C. (2002) The additive partitioning of species diversity: recent revival of an old idea. *Oikos* 99 (1), 3-9.

Velsko, I. M., Frantz, L. A. F., Herbig, A., Larson, G. and Warinner, C. (2018) Selection of appropriate metagenome taxonomic classifiers for ancient microbiome research. *mSystems* 3 (4), e00080-18.

Wagner, D. M. P., Klunk, J. B. S., Harbeck, M. P., Devault, A. M. A., Waglechner, N. M., Sahl, J. W. P., Enk, J. M., Birdsell, D. N. P., Kuch, M. M., Lumibao, C. M., Poinar, D. M. A., Pearson, T. P., Fourment, M. P., Golding, B. P., Riehm, J. M. P., Earn, D. J. D. P., DeWitte, S. P., Rouillard, J.-M. P., Grupe, G. P., Wiechmann, I. P., Bliska, J. B. P., Keim, P. S. P., Scholz, H. C. P., Holmes, E. C. P. and Poinar, H. D. (2014) Yersinia pestis and the Plague of Justinian 541–543 AD: a genomic analysis. *The Lancet infectious diseases* 14 (4), 319-326.

Wu, Y.-W., Simmons, B. A., Singer, S. W. and Lawrence Berkeley National Lab, B. C. A. (2016) MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32 (4), 605-607.

Yang, C., Chowdhury, D., Zhang, Z., Cheung, W. K., Lu, A., Bian, Z. and Zhang, L. (2021) A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. *Computational and structural biotechnology journal* 19, 6301-6314.

Babraham Bioinformatics (2019) *FastQC.* Babraham Institute. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ Accessed 4th January.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114-2120.

Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34 (17), i884-i890.

Churcheward, B., Millet, M., Bihouée, A., Fertin, G. and Chaffron, S. (2022) MAGNETO: An Automated Workflow for Genome-Resolved Metagenomics. *mSystems* 7 (4), e0043222-e0043222.

Ciufo, S., Kannan, S., Sharma, S., Badretdin, A., Clark, K., Turner, S., Brover, S., Schoch, C. L., Kimchi, A. and DiCuccio, M. (2018) Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *International journal of systematic and evolutionary microbiology* 68 (7), 2386-2392.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M. and Li, H. (2021) Twelve years of SAMtools and BCFtools. *GigaScience* 10 (2).

Duru, I. C., Lecomte, A., Shishido, T. K., Laine, P., Suppula, J., Paulin, L., Scheperjans, F., Pereira, P. A. B. and Auvinen, P. (2023) Metagenome-assembled microbial genomes from Parkinson's disease fecal samples. Cold Spring Harbor Laboratory.

Eisenhofer, R. and Weyrich, L. (2018) Proper Authentication of Ancient DNA Is Still Essential. *Genes* 9 (3), 122.

Eren, A. M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S. E., Schechter, M. S., Fink, I., Pan, J. N., Yousef, M., Fogarty, E. C., Trigodet, F., Watson, A. R., Esen, Ö. C., Moore, R. M., Clayssen, Q., Lee, M. D., Kivenson, V., Graham, E. D., Merrill, B. D., Karkman, A., Blankenberg, D., Eppley, J. M., Sjödin, A., Scott, J. J., Vázquez-Campos, X., McKay, L. J., McDaniel, E. A., Stevens, S. L. R., Anderson, R. E., Fuessel, J., Fernandez-Guerra, A., Maignien, L., Delmont, T. O. and Willis, A. D. (2020) Community-led, integrated, reproducible multi-omics with anvi'o. *Nature Microbiology* 6 (1), 3-6.

Farrer, A. G., Wright, S. L., Skelly, E., Eisenhofer, R., Dobney, K. and Weyrich, L. S. (2021) Effectiveness of decontamination protocols when analyzing ancient DNA preserved in dental calculus. *Scientific Reports* 11 (1).

Fellows Yates, J. A., Lamnidis, T. C., Borry, M., Valtueña, A. A., Fagernäs, Z., Clayton, S., Garcia, M. U., Neukamm, J. and Peltzer, A. (2021) Reproducible, portable, and efficient ancient genome reconstruction with nfcore/ eager. *PeerJ (San Francisco, CA)* 9.

González, C. D., Rangavittal, S., Vicedomini, R., Chikhi, R. and Richard, H. (2023) aKmerBroom: Ancient oral DNA decontamination using Bloom filters on k-mer sets. *iScience* 26 (11), 108057-108057.

Gourlé, H., Karlsson-Lindsjö, O., Hayer, J. and Bongcam-Rudloff, E. (2019) Simulating Illumina metagenomic data with InSilicoSeq. *Bioinformatics* 35 (3), 521-522.

Joint Genome Institute (2023) *BBduk Guide.* JGI. Accessed 27 February.

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. and Aluru, S. (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* 9 (1).

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. and Orlando, L. (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29 (13), 1682-1684.

Ke, S., Weiss, S. T. and Liu, Y.-Y. (2022) Dissecting the role of the human microbiome in COVID-19 via metagenome-assembled genomes. *Nature Communications* 13 (1).

Kim, D., Song, L., Breitwieser, F. P. and Salzberg, S. L. (2016) Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Research* 26 (12), 1721-1729.

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., Bushman, F. D., Knight, R. and Kelley, S. T. (2011) Bayesian community-wide culture-independent microbial source tracking. *Nature Methods* 8 (9), 761-763.

Krakau, S., Straub, D., Gourlé, H., Gabernet, G. and Nahnsen, S. (2022) nf-core/mag: a best-practice pipeline for metagenome hybrid assembly and binning. *NAR genomics and bioinformatics* 4 (1), lqac007-lqac007.

Langmead, B. and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9 (4), 357-359.

Li, D., Liu, C.-M., Luo, R., Sadakane, K. and Lam, T.-W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31 (10), 1674-1676.

Link, V., Kousathanas, A., Veeramah, K., Sell, C., Scheu, A. and Wegmann, D. (2017) ATLAS: Analysis Tools for Low-depth and Ancient Samples. Cold Spring Harbor Laboratory.

Martí, J. M. (2019) Recentrifuge: Robust comparative analysis and contamination removal for metagenomics. *PLOS Computational Biology* 15 (4), e1006967.

Musich, R., Cadle-Davidson, L. and Osier, M. V. (2021) Comparison of Short-Read Sequence Aligners Indicates Strengths and Weaknesses for Biologists to Consider. *Frontiers in Plant Science* 12.

Nayfach, S., Shi, Z. J., Seshadri, R., Pollard, K. S. and Kyrpides, N. C. (2019) New insights from uncultivated genomes of the global human gut microbiome. *Nature* 568 (7753), 505-510.

Nie, S., Jing, Z., Wang, J., Deng, Y., Zhang, Y., Ye, Z. and Ge, Y. (2023) The link between increased Desulfovibrio and disease severity in Parkinson's disease. *Applied Microbiology and Biotechnology* 107 (9), 3033-3045.

Nissen, J. N., Johansen, J., Allesøe, R. L., Sønderby, C. K., Armenteros, J. J. A., Grønbech, C. H., Jensen, L. J., Nielsen, H. B., Petersen, T. N., Winther, O. and Rasmussen, S. (2021) Improved metagenome binning and assembly using deep variational autoencoders. *Nature Biotechnology* 39 (5), 555-560.

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J. and Nieselt, K. (2016) EAGER: efficient ancient genome reconstruction. *Genome Biology* 17 (1).

Pessi, I. S., Popin, R. V., Durieu, B., Lara, Y., Tytgat, B., Savaglia, V., Roncero-Ramos, B., Hultman, J., Verleyen, E., Vyverman, W. and Wilmotte, A. (2023) Novel diversity of polar Cyanobacteria revealed by genome-resolved metagenomics. *Microbial genomics* 9 (7).

Pochon, Z., Bergfeldt, N., Kırdök, E., Vicente, M., Naidoo, T., Van Der Valk, T., Altınışık, N. E., Krzewińska, M., Dalen, L., Götherström, A., Mirabello, C., Unneberg, P. and Oskolkov, N. (2022) aMeta: an accurate and memory-efficient ancient Metagenomic profiling workflow. Cold Spring Harbor Laboratory.

Sanctis, B., Money, D., Pedersen, M. W. and Durbin, R. (2022) A theoretical analysis of taxonomic binning accuracy. *Molecular ecology resources* 22 (6), 2208-2219.

Schmieder, R., Edwards, R. and Argonne National Laboratory, A. I. L. (2011) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PloS one* 6 (3), e17288-e17288.

Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M., Willerslev, E. and Orlando, L. (2014) Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature protocols* 9 (5), 1056-1082.

Seppey, M., Manni, M. and Zdobnov, E. M. (2019) BUSCO: Assessing Genome Assembly and Annotation Completeness. Springer New York. 227-245. https://dx.doi.org/10.1007/978-1-4939-9173-0_14

Singh, N. K., Wood, J. M., Patane, J., Moura, L. M. S., Lombardino, J., Setubal, J. C. and Venkateswaran, K. (2023) Characterization of metagenome-assembled genomes from the International Space Station. *Microbiome* 11 (1).

Uritskiy, G. V., Diruggiero, J. and Taylor, J. (2018) MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6 (1).

Zhang, J., Kobert, K., Flouri, T. and Stamatakis, A. (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30 (5), 614-620.

Zhu, J., Tian, L., Chen, P., Han, M., Song, L., Tong, X., Sun, X., Yang, F., Lin, Z., Liu, X., Liu, C., Wang, X., Lin, Y., Cai, K., Hou, Y., Xu, X., Yang, H., Wang, J., Kristiansen, K., Xiao, L., Zhang, T., Jia, H. and Jie, Z. (2022) Over 50,000 Metagenomically Assembled Draft Genomes for the Human Oral Microbiome Reveal New Taxa. *Genomics, proteomics & bioinformatics* 20 (2), 246-259.

Dabney, J., Meyer, M. and Paabo, S. (2013) Ancient DNA Damage. *Cold Spring Harbor Perspectives in Biology* 5 (7), a012567-a012567.

Jin, H., You, L., Zhao, F., Li, S., Ma, T., Kwok, L.-Y., Xu, H. and Sun, Z. (2022) Hybrid, ultra-deep metagenomic sequencing enables genomic and functional characterization of low-abundance species in the human gut microbiome. *Gut Microbes* 14 (1).

Kazarina, A., Gerhards, G., Petersone-Gordina, E., Kimsis, J., Pole, I., Zole, E., Leonova, V. and Ranka, R. (2019) Analysis of the bacterial communities in ancient human bones and burial soil samples: Tracing the impact of environmental bacteria. *Journal of archaeological science* 109, 104989.

Kistler, L., Ware, R., Smith, O., Collins, M. and Allaby, R. G. (2017) A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic acids research* 45 (11), 6310-6320.

Li, H. (2024) *seqtk.* GitHub: GitHub, Inc. https://github.com/lh3/seqtk Accessed 26 August.

Llamas, B., Valverde, G., Fehren-Schmitz, L., Weyrich, L. S., Cooper, A. and 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), 1-14.

Meziti, A., Rodriguez-R, L. M., Hatt, J. K., Peña-Gonzalez, A., Levy, K. and Konstantinidis, K. T. (2021) The Reliability of Metagenome-Assembled Genomes (MAGs) in Representing Natural Populations: Insights from Comparing MAGs against Isolate Genomes Derived from the Same Fecal Sample. *Applied and environmental microbiology* 87 (6).

NCBI (2013) *Genome assembly GRCh38.* https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.26/ Accessed 31 July.

Quagliariello, A., Modi, A., Innocenti, G., Zaro, V., Conati Barbaro, C., Ronchitelli, A., Boschin, F., Cavazzuti, C., Dellù, E., Radina, F., Sperduti, A., Bondioli, L., Ricci, S., Lognoli, M., Belcastro, M. G., Mariotti, V., Caramelli, D., Mariotti Lippi, M., Cristiani, E., Martino, M. E., Muntoni, I. M. and Lari, M. (2022) Ancient oral microbiomes support gradual Neolithic dietary shifts towards agriculture. *Nature Communications* 13 (1).

Rühlemann, M. C., Wacker, E. M., Ellinghaus, D. and Franke, A. (2022) <tt>MAGScoT</tt>: a fast, lightweight and accurate bin-refinement tool. *Bioinformatics* 38 (24), 5430-5433.

Steinke, K., Pamp, S. J. and Munk, P. (2024) MAGICIAN: MAG simulation for investigating criteria for bioinformatic analysis. *BMC genomics* 25 (1), 55-55.

Uritskiy, G. V., Diruggiero, J. and Taylor, J. (2018) MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6 (1).

Yue, Y., Huang, H., Qi, Z., Dou, H.-M., Liu, X.-Y., Han, T.-F., Chen, Y., Song, X.-J., Zhang, Y.-H. and Tu, J. (2020) Evaluating metagenomics tools for genome binning with real metagenomic datasets and CAMI datasets. *BMC bioinformatics* 21 (1), 1-334.

Standeven, F. J., Dahlquist-Axe, G., Hendy, J., Fiddyment, S., Holst, M., McGrath, K., Collins, M., Mundorff, A., Meehan, C. J., Tedder, A. and Speller, C. F. (2024) An extensive archaeological dental calculus dataset spanning 5000 years for ancient human oral microbiome research. *bioRxiv*, 2024.09.17.613443.