Practice

BIOINFORMATICS INSTITUTE

Ancient dental metagenomes, or Cleaning thousand-year-old teeth and examining dental calculus

Balan Anna[1, 2] and Zolotar Anastasia[1, 3, 4]

[1]Bioinformatics Institute
[2]RSAU – MTAA named after K.A. Timiryazev
[3]National Research Center for Epidemiology and Microbiology named after the honorary academician N. F. Gamaleya, Moscow, Russia
[4]Institute of Higher Nervous Activity and Neurophysiology of RAS, Moscow, Russia

Abstract

Metagenomic analysis of the diverse microbial communities allows us to study ecosystem functioning and human health. Next-generation sequencing (NGS) technique provides almost limitless possibilities for "dissecting" modern microbiomes, and even ancient ones. For this study, we have analyzed samples of dental calculus from several human remains, more than a thousand years old. The metagenome of the discovered and analyzed bacterial community can tell a lot about diseases, nutrition, human life, as well as the evolution of typical bacteria of the human oral cavity. Analysis of 16S rRNA genes as well as shortgun sequencing allows us to reveal taxonomy composition of ancient dental calculus and to indicate that two out of four ancient individuals most probably had severe forms of periodontal disease.

Keywords: Metagenomics, Paleometagenomics, Human microbiome

Introduction

A healthy human body contains over 10^{14} symbiotic microbes that mediate cellular processes, support immunity, and synthesize essential vitamins and amino acids. Imbalance of microbial dynamics, however, has been linked to a variety of human illnesses, including diabetes, cancer, cardiovascular disease, and even neurological disorders (Athanasopoulou *et al.* 2023).

Metagenomic analysis is important because it allows us to study the diverse microbial communities present in different environments such as soil, water, and human gut. These microbial communities play critical roles in ecosystem function and human health. Metagenomic analysis can provide insights into the functional potential of microbial communities, including their ability to decompose organic matter, fix nitrogen, and produce secondary metabolites (Wooley *et al.* 2010).

Two main approaches for metagenomic analysis of microbiomes are 16S rRNA gene sequencing and shotgun sequencing. Sequencing of 16S rRNA gene based on targeting a conserved gene region present in all bacteria and archaea, allowing identification of different taxonomic groups present in a microbial community. Shotgun sequencing, on the other hand, allows to analyze total DNA present in a sample, providing both taxonomic and functional information about the microbial community. Shotgun sequencing can also identify novel genes and metabolic pathways not previously described in known organisms (Shah *et al.* 2011).

In 1990, archaeologists working on a monastic site in Dunheim, Germany dug out several human skulls. They were looking for teeth covered by dental calculus, or tartar, which is a form of hardened dental plaque. It is caused by precipitation of minerals from saliva and gingival crevicular fluid (GCF) in plaque on the teeth. (White 1997). Dental calculus can preserve DNA for 1000s of years, so scientists extracted DNA from the material underneath the dental calculus and studied this microbial community (Warinner *et al.* 2014).

During this study, we have analyzed 16S ribosomal RNA reads as well as shotgun sequencing results of DNA samples extracted from dental calculus of mediaeval people in order to investigate ancient oral microbiomes.

Material and methods

16S rRNA gene sequencing

Researchers sequenced the DNA samples and provided the single-ended sequencing reads of V5 16S ribosomal RNA obtained by an instrument Roche GS Junior (454) (Dudley *et al.* 2012).

All data from the original research are available in the NCBI Sequence Read Archive (SRA) (Lipman *et al.* 2011) under number SRP029257 (BioProject PRJNA216965).

We have used the package QIIME2 v2023.2.0 (“Quantitative Insights Into Microbial Ecology”) (Hall and Beiko 2018), an open-source bioinformatics pipeline for performing microbiome analysis from raw DNA sequencing data. Using this pipeline, amplicon sequencing data was subjected to the following analysis: quality control, trimming, denoising, dereplication, filtering chimeric sequences, clustering into amplicon sequence variants (ASVs). Particulary, DADA2 v1.26.0 (Callahan *et al.* 2016), as a part of QIIME2, was used for clustering into an amplicon sequence variant (ASV). For trimming and filtering parameters —p-trim-left 32 —p-trunc-len 150 were specified according to barcode and primer length and QC distribution.

For taxonomic composition analysis of samples QIIME2 Naive Bayes classifier trained on 16S RNA data (SILVA reference

sequence and taxonomy files) was used. Data for each step of analysis was visualised with QIIME2 as well.

Shortgun sequencing

Dental calculus microbiome DNA sample of individual G12 skeleton (AD 950-1200) sample suffered from periodontal disease was subjected to shotgun sequencing and resulted reads were assembled into contigs. For further comparison with ancient *Tannerella forsythia* genome precomputed assembly for G12 individual was obtained.

Metagenome assembly of G12 sample was aligned to the *Tannerella forsythia* 92A2 with annotation using BWA-mem v0.7.17 (Li 2013). Samtools v1.16.1 (Danecek *et al.* 2021) was used to compress, sort and index the aligned reads and index reference sequence with default parameters. Basic statistics of resulted alignment were obtained using command samtools flagstat.

Further resulted alignment was converted to .bed format by BEDtools v2.27.1 (bedtools bamtobed) ((Quinlan and Hall 2010). Genome regions absent in the ancient strain were subtracted from GFF3 using bedtools intersect (option -v). Resulting annotation file was manually inspected using bash and was visualized in desktop application Integrative Genomics Viewer (IGV) (Thorvaldsdóttir *et al.* 2012).

Results

Amplicon sequencing

Initially, sequencing results contained 5000 reads per analyzed sample, with the median length of 179 nucleotides. As the result of read quality check, we obseved that the read quality score for almost sequences lowers starting at 180 base. Taking into account barcodes and primers sequences length for library preparation and read quality decrease, reads were trimmed.

With DADA2 v1.26.0 pipeline (Callahan *et al.* 2016), we filtered chimeric sequences of sequence V5 16S regions. These regions are great for profiling oral metagenomes (Teng *et al.* 2018). They are constantly present in metadata samples. As a result, more than 90% of reads in each sample passed all the filters and appeared to be non-chimeric. In **Table** 1 basic statistics of trimming and filtering sequencing reads are presented.

One of the main DADA2 step results is a clustering into an amplicon sequence variant (ASV) - a higher-resolution analogue of the traditional OTUs. We've got feature distribution as a result, and it shows that just a few features are present in the majority of the samples, which is perfectly understandable, since we are analyzing conservative sequences.

As the result of microbiome classification (**Figure**1), we found seven dominant phyla in the calculus samples: Firmicutes, Proteobacteria, Bacteroidetes, Synergistetes, Actinobacteria, Fusobacteria, and Spirochaetes. The root samples also contained bacteria from Firmicutes, Proteobacteria, Bacteroidetes, and Actinobacteria but they constituted different proportions and composition.

Shortgun sequencing

Only 1.83% of calculus metagenomic contigs were aligned to the *Tannerella forsythia* 92A2 genome. Analyzing alignment results, We have revealed several regions of zero coverage in the modern strain that were probably gained during evolution. Among them, 196 coding sequences were found. We have distinguished the following groups for acquired genes: restriction modification system, mobile genetic elements, genes related to DNA repair, post-translational modification, antibiotic biosynthesis, drug resistance, and metabolism. Information for some representatives is given in **Table** 2.

Discussion

As we may conclude, QIIME2 (Hall and Beiko 2018) has several advantages over other software packages for metagenomics analysis. It is designed to work with various types of data, including 16S rRNA sequencing and shotgun sequencing data. It also has a user-friendly interface and provides detailed documentation and tutorials to help users navigate the software. That is why we have chosen those tool to work with our data.

454 pyrosequencing is a method for high-throughput sequencing of DNA that uses a series of enzymatic reactions to detect nucleotide incorporation (Harrington *et al.* 2013). This technology has been widely used for metagenomic studies due to its ability to generate long reads and high throughput.

However, one of the limitations of 454 pyrosequencing is the relatively high error rate, which can lead to inaccurate sequence data. To address this issue, quality control measures such as filtering and trimming are often applied to improve the accuracy and reliability of the sequencing data. It explains why the quality of our samples drastically drops down.

As we are working with a V5 region of 16S, RNA we have reads of 105 nucleotides in length, and some features are present in all of the reads, because its a highly conservative region (Garcia-Mazcorro and Barcenas-Walls 2016).

We have analysed the taxonomy of the metagenome. In general, metagenome of roots and calculus can be differentiated. In metagenomes of teeth roots p. Proteobacteria is the most presented bacteria, p. Firmicutes are much less present, p. Synergistota can be found only in little amounts or not present at all, etc.

Calculus metagenome, on the contrary, always has p. Firmicutes as a prevailing species, and p. Synergistota is present much. We can not say if those features are consistent for all the cases, as we are having only six samples in present study.

The red complex is a group of bacteria that are categorized together based on their association with severe forms of periodontal disease (Mohanty *et al.* 2019). We have found *Tannerella forsythia*, *Porphyromonas gingivalis*, and *Treponema denticola*, which are representatives of this microbial group, in ancient dental calculus samples. Only for two of four skulls bacteria the red complex bacteria were found, those are G12 and B61.

It is quite interesting, that none of those genera are present in the root samples. Also all the calculus samples have at least two genera out of three. This result suggests that periodontal disease existed in the medieval age just as it does today, despite major changes in lifestyle, nutrition, and oral hygiene.

As the result of comparing ancient calculus metagenome with modern *Tannerella forsythia* 92A2 strain, we have identified acquired genes, some of them listed in **Table** 2. Respective gene products might improve bacterial survival abilities, increase its virulence and competitiveness among other pathogens.

Horizontal gene transfer (HGT) is well known to be essential for the development, transmission, and maintenance of pathogenicity (Arnold *et al.* 2022). We have revealed several *tra* genes that facilitate HGT F-like plasmids transfer.

Moreover, we have identified several tetracycline antibiotic resistance genes (WP_007366522.1, WP_007366526.1, WP_051322484.1). As previously shown, tetracycline resistance genes may be acquired by large conjugative transposons

Table 1 Basic statistics of initial sequencing reads and their statistics after treir trimming and filtering via QIIME2 v2023.2.0 pipeline (Hall and Beiko 2018)

sample-id	initial reads	filtered reads	passed filter, %	denoised	non-chimeric	input non-chimeric, %
S10-V5-Q-B61-calc	5957	5695	95.6	5575	5559	93.32
S14-V5-P-B17-calc	4491	4282	95.35	4218	4218	93.92
S15-V5-R-B78-calc	4212	4037	95.85	3950	3911	92.85
S16S17-V5-K1-G12-root	5788	5599	96.73	5427	5361	92.62
S16S17-V5-K2-G12-root	5272	5066	96.09	4884	4862	92.22
S18S19-V5-L-B17-root	4955	4790	96.67	4680	4680	94.45
S20S21-V5-M-B61-root	5516	5396	97.82	5240	5182	93.94
S22S23-V5-N-B78-root	4695	4534	96.57	4382	4261	90.76
S8-V5-O-G12-calc	5362	5199	96.96	5108	5092	94.96

Table 2 Examples of several protein coding genes acquired by the *Tannerella forsythia* 92A2 strain

product	protein_id	protein_group
AIPR family protein	WP_014224073.1	restriction modification system
DDE-type integrase/transposase/recombinase	WP_014223573.1	mobile genetic element
IS1 family transposase	WP_099046116.1	mobile genetic element
IS110 family transposase	WP_014225157.1	mobile genetic element
IS1380 family transposase	WP_041590984.1	mobile genetic element
DNA mismatch endonuclease Vsr	WP_041590610.1	DNA repair
GNAT family N-acetyltransferase	WP_007366527.1	post-translational modification
SPASM domain-containing protein	WP_157755359.1	post-translational modification
radical SAM peptide maturase	WP_014224286.1	post-translational modification
class I lanthipeptide	WP_041590503.1	antibiotic biosynthesis
lanthionine synthetase C family protein	WP_014223583.1	antibiotic biosynthesis
lantibiotic dehydratase family protein	WP_041590506.1	antibiotic biosynthesis
thiopeptide-type bacteriocin biosynthesis protein	WP_041590507.1	antibiotic biosynthesis
ParA family protein	WP_007366517.1	HGT of plasmids
conjugative transposon protein TraJ	WP_007366507.1	HGT of plasmids
conjugative transposon protein TraM	WP_007366504.1	HGT of plasmids
conjugative transposon protein TraN	WP_007366503.1	HGT of plasmids
tetracycline resistance ribosomal protection protein	WP_007366526.1	antibiotic resistance
RteC domain-containing protein	WP_007366522.1	antibiotic resistance
TetR/AcrR family transcriptional regulator	WP_051322484.1	antibiotic resistance
dihydrofolate reductase family protein	WP_007366523.1	metabolic enzyme
beta-ketoacyl-ACP synthase III	WP_014226278.1	metabolic enzyme

(Roberts 1996).

Among other acquired genes groups enhancing pathogen viability are those which products are involved in DNA repair, antibiotic biosynthesis, post-translational modification and cellular defence systems, such as restriction-modification systems.

As the conclusion, development of techniques for the detection, identification, and characterization of ancient pathogens has allowed us to better understand the natural history of some infectious diseases and their evolution, unravel and clarify mechanisms of antibiotic resistance as well as provide increasing knowledge on ancient microbiota, ancient human lifestyle and habits, and resolve historical controversies.

Acknowledgments

The template of this pdf document was given by Natalia Khotkina.

Literature cited

Arnold BJ, Huang IT, Hanage WP. 2022. Horizontal gene transfer and adaptive evolution in bacteria. Nat. Rev. Microbiol.. 20:206–218.

Athanasopoulou K, Adamopoulos PG, Scorilas A. 2023. Unveiling the human gastrointestinal tract microbiome: The past, present, and future of metagenomics. Biomedicines. 11.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from illumina amplicon data. Nat. Methods. 13:581–583.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM *et al*. 2021. Twelve years of SAMtools and BCFtools. GigaScience. 10. giab008.

Dudley DM, Chin EN, Bimber BN, Sanabani SS, Tarosso LF, Costa PR, Sauer MM, Kallas EG, O'Connor DH. 2012. Low-cost ultra-wide genotyping using roche/454 pyrosequencing for surveillance of HIV drug resistance. PLoS One. 7:e36494.

Garcia-Mazcorro JF, Barcenas-Walls JR. 2016. Thinking beside the box: Should we care about the non-coding strand of the 16S rRNA gene? FEMS Microbiol. Lett.. 363:fnw171.

Hall M, Beiko RG. 2018. 16S rRNA gene analysis with QIIME2. Methods Mol. Biol.. 1849:113–129.

Harrington CT, Lin EI, Olson MT, Eshleman JR. 2013. Fundamentals of pyrosequencing. Arch. Pathol. Lab. Med.. 137:1296–1303.

Li. 2013. Aligning sequence reads, clone sequences and assembly contigs with bwa-mem. arxiv:1303.3997v2.

Lipman D, Flicek P, Salzberg S, Gerstein M, Knight R. 2011. Closure of the NCBI SRA and implications for the long-term future of genomics data storage. Genome Biol.. 12:402.

Mohanty R, Asopa SJ, Joseph MD, Singh B, Rajguru JP, Saidath K, Sharma U. 2019. Red complex: Polymicrobial conglomerate in oral flora: A review. J. Family Med. Prim. Care. 8:3480–3486.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26:841–842.

Roberts MC. 1996. Tetracycline resistance determinants: mechanisms of action, regulation of expression, genetic mobility, and distribution. FEMS Microbiol. Rev.. 19:1–24.

Shah N, Tang H, Doak TG, Ye Y. 2011. Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. Pac. Symp. Biocomput.. pp. 165–176.

Teng F, Darveekaran Nair SS, Zhu P, Li S, Huang S, Li X, Xu J, Yang F. 2018. Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. Sci. Rep.. 8:16321.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2012. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics. 14:178–192.

Warinner C, Rodrigues JFM, Vyas R, Trachsel C, Shved N, Grossmann J, Radini A, Hancock Y, Tito RY, Fiddyment S *et al*. 2014. Pathogens and host immunity in the ancient human oral cavity. Nat. Genet.. 46:336–344.

White DJ. 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.

Wooley JC, Godzik A, Friedberg I. 2010. A primer on metagenomics. PLoS Comput. Biol.. 6:e1000667.

Supplementary materials

Our lab journals can be found in Google Disk.

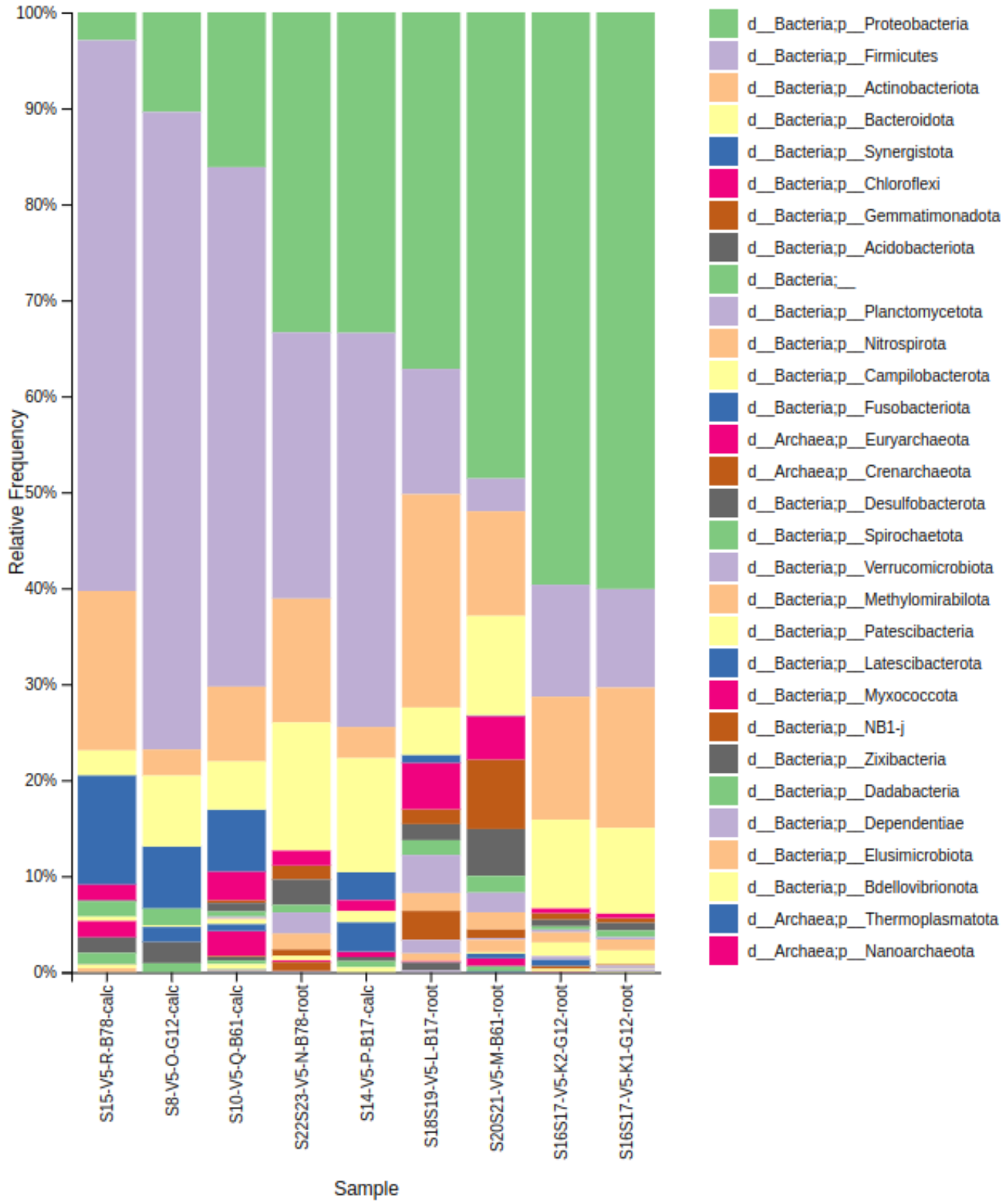

Figure 1 Taxonomic composition of analyzed ancient samples using QIIME2 v2023.2.0 pipeline (Hall and Beiko 2018)