Project 7: sinking our teeth into Dead Man's Teeth

Sobolev A., Kupaeva D.

Abstract

Application of metagenomics to the analysis of ancient samples for comparison of microbial communities and investigation of various evolutionary adaptations of Bacteria has recently been of interest. Combination of two main approaches in metagenomics analysis — 16S amplicon sequencing and shotgun sequencing — can shed light on the evolution of diseases and microorganisms that cause them. This project was devoted to the investigation of the bacterial community of the inner tartar layer of dental calculus extracted from fossil specimens. Using qiime2 and MetaPhlAn, we analyzed the taxonomic distribution of this sample using bone sample as a control and compared our results with the data on metagenomes of different tissues obtained from the Human Microbiome Project. Finally, we studied the evolutionary mechanisms of *Tannerella forsythia,* a member of the red complex of periodontal pathogens, by mapping our metagenome assembly to the *T. forsythia* genome.

Introduction

Rapidly growing interest in metagenomics analysis is determined by a wide variety of approaches and biological information one is able to pull from such experiments. Alpha- and beta-diversity of microbial communities provides a lot of information on how microorganisms impact biological processes in host organisms.

There are two main approaches in metagenomics analyses, namely 16S amplicon sequencing and shotgun sequencing. The former has proven to be an efficient and cost-effective strategy for microbiome analysis, albeit vulnerable to biases and limited to taxonomic classification at the genus level. The latter, on the other hand, offers the advantage of species- and strain-level classification of bacteria and allows to examine functional relationships between hosts and bacteria by determining the functional content of samples. However, this approach demands high costs and more advanced bioinformatic requirements[1].

Exploration of different microbiomes can be used to determine effects of different metagenome composition on host health condition in context of diagnosing infections[2], development of neurodegenerative diseases[3], reflecting health of patients[4] and even improve existing assemblies of bacterial genomes based on metagenome-assembled genomes (MAGs) approach[5]. Moreover, metagenomics can be used in terms of analyzing the history of various diseases in host organisms based on the preservation of bacterial species inside ancient samples[6], for instance, samples of the dental calculus.

Calculus is a calcified biofilm that includes the host's biomolecules and the bacterial community of its oral cavity[7]. For medical purposes today, the microbial communities of dental plaque, a biofilm that has not yet been calcified, are most often studied. Dental plaque is responsible for the pathology of the oral cavity - and, accordingly, is the object of close attention of dentists[8]. Tartar - the terminal stage of dental biofilm - is an interesting object for comparison, including because researchers have access to analysis not only of dental calculus (and, accordingly, microbial communities) of modern humans, but also dental calculi of ancient people - from fossil samples[9] .
In the process of plaque development, the dominant taxa change in it, and each stage of maturation is characterized by its own community[10]. However, modern oral hygiene - both personal, such as daily dental cleaning, and with the help of dentists - prevents the formation of mature tartar[11]. On the one hand, this has a good effect on the quality of life of a modern person; at the same time, it complicates the analysis of the evolution of the bacterial community of the human oral cavity[12].
In this study, we compare the bacterial communities of the teeth of modern humans and ancient humans, approaching the problem in a complex way: we analyze the taxonomic composition and isolate elements of the microbiome that are characteristic of fossil dental films, compare them with modern dental films, and also study the genome of an individual representative of the fossil biofilm and analyze the differences at the level of individual genes in the composition of the bacterial genome.

Methods

Analysis of fossil samples with qiime2
In this part, we analyzed the V5 16S RNA amplicon sequencing data of the bacterial community extracted from the inner tartar layer of fossil specimens estimated to be approximately 1000 years old. The samples were sequenced on a Roche GS Junior (454). We use root sequencing data from the same sample as a control for environmental contamination. Sequencing data is available for download.
For the analysis, the qiime2 v.2020.11 pipeline was used. After visualizing the quality of raw reads (*qiime demux summarize* options), we set the parameters for trimming raw reads during preprocessing using the dada2 pipeline component. We removed 32 nucleotides from the left (it is adapter sequences), and left only reads longer than 150 nucleotides, cutting them to this length (*--p-trim-left 32 --p-trunc-len 150*).
After that, we have compiled a list of individual operational taxonomic unit (*qiime feature-table summarize*) and classified them using the sklearn library (*qiime feature-classifier classify-sklearn*) using a ready-made classifier based on the GreenGenes database. Finally, we analyzed the diversity of the community at different taxonomic levels using the qiime taxa barplot (*qiime taxa barplot)*.

Comparison of the metagenome of fossil samples with the metagenomes of different tissues of modern people

In this part, we compared the finished shotgun sequenced metagenome assembly of fossil patient G12 with data from the Human Microbiome Project (SRS014459-Stool.fasta, SRS014464-Anterior_nares.fasta, SRS014470-Tongue_dorsum.fasta, SRS014472-Buccal_mucosa.fasta, SRS014476-Supragingival_plaque.fasta, SRS014494-Posterior_fornix.fasta).

For the analysis, we used the MetaPhlAn software (v 3.0). With its help, we established the taxonomic diversity of our patient's data (*metaphlan* script), data from libraries of modern humans, combined the obtained data into a single table (*merge_metaphlan_tables.py* script) and visualized the results in the form of a heatmap for all samples (*metaphlan_hclust_heatmap.py* script).

Comparison of the ancient genome with the genome of modern bacteria

In this part, we compared the assembly of the ancient human dental calculus metagenome with the *Tannerella forsythia* genome (assembly and annotation). We indexed the *T. forsythia* genome file (*bwa* v 0.7.17-r1188, *index* option), aligned the metagenomic assembly to it (using *bwa mem*), reformatted it to bed format (*samtools view -S -b* options, v.0.1.19-44428cd + *bedtools bamtobed* v2.27.1) . After that, we found regions of the genome of modern bacteria that are absent in the metagenome assembly (*bedtools intersect -v* options).

Results

The amplicon sequencing samples contain single-ended readings grouped into 2 libraries: "bone" and "calculus", 5788 and 5368 readings, respectively. The quality of reads is quite high in the first 180 nucleotides of reading, and drastically decreases after 180 nucleotides (Fig. 1)

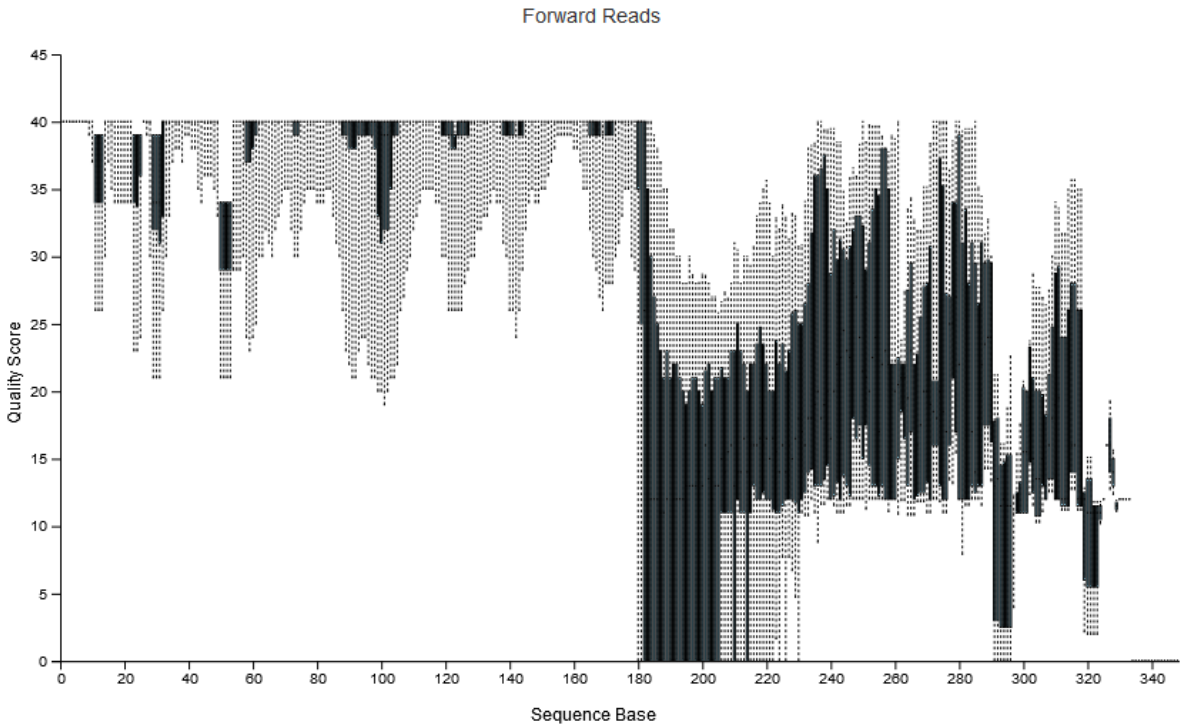

Figure 1. Quality of raw sequences

After cropping and filtering for a length of 150, we were left with 5588 readings for the bone sample and 5179 for the calculus sample. We also tried to crop to a length of 180, but with these settings, too little percentage of the original reads passes through the filter. Some statistical data about them are presented in Table 1.

Table 1. Data statistics after passing the filter

sample-id	input	filtered 180 length	percentage of input passed filter 180 length	filtered 150 length	percentage of input passed filter, 150 length	denoised, 150 length	non-chimeric, 150 length	percentage of input non-chimeric, 150 length
bone	5788	2830	48.89	5588	96.54	5380	5380	92.95
calculus	5362	1446	26.97	5179	96.59	5054	4810	89.71

The taxonomic distribution is shown in Figure 2. As we can see in the graphs, the quantitative representation of the dominant taxa differs significantly in our samples despite the similarity of chemical contents of dental calculus and bone. For instance, in calculus microbiome Firmicutes are more abundant than in bone (65.5% and 10.5%, respectively). On the contrary, Proteobacteria constitute more than a half of bone microbiome and are less abundant in calculus (61.7% and 13.8%, respectively). In general, taxonomic distribution in bone is more diverse than in case of calculus, with aerobic microorganisms galore.

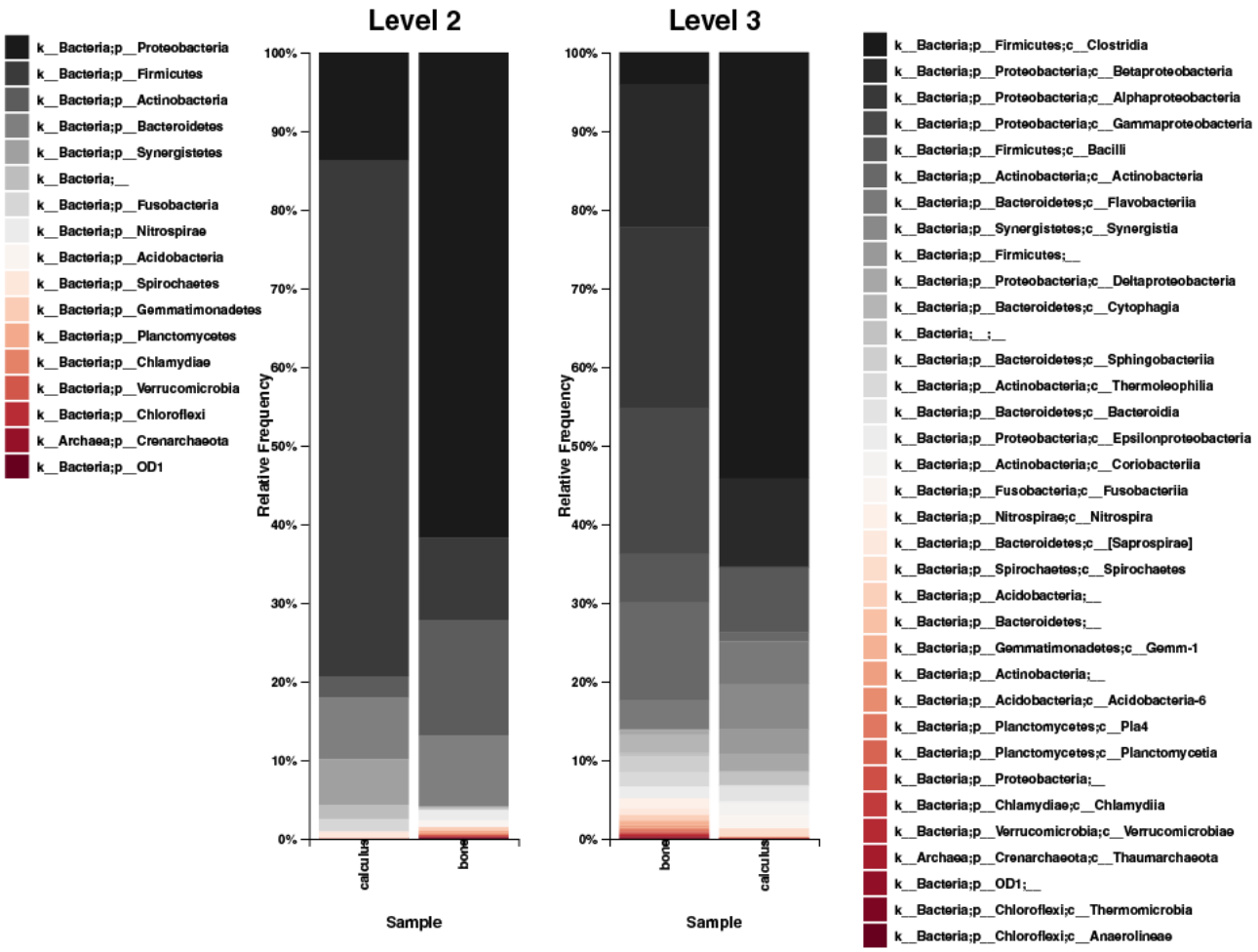

Fig 2. Taxonomic diversity at 2 and 3 taxonomic levels.

At the 7th taxonomic level, we found 36 operational taxonomic units, present in varying amounts in the metagenome, but completely absent in the metagenome of the tooth root. We consider these 36 taxa to be specific for the plaque community. Among these 36 taxa, representatives of the genera Porphyromonas, Tannerella and Treponema are present. These genera are part of the so-called red complex - the community of bacteria responsible for the development of periodontal disease. A complete list of discovered taxa can be found in the supplementary materials 2.

In analyzing the genome-wide plaque shotgun sequencing assembly of fossil patient G12, we also analyzed the taxonomic diversity of the sample. In analyzing the genome-wide plaque shotgun sequencing assembly of fossil patient G12, we also analyzed the taxonomic diversity of the sample. A complete list of encountered taxa is on a separate sheet in Supplementary Materials 2.

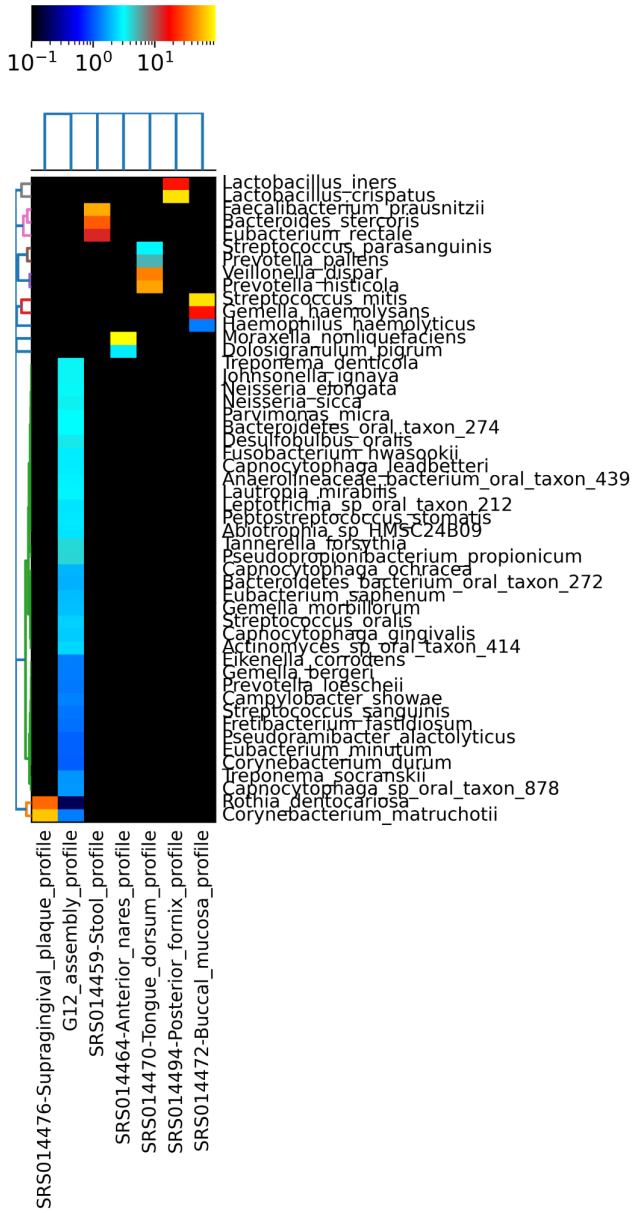

Figure 3. A heatmap comparing the microbial communities of dental calculus in a fossil patient and contemporary specimens.

Comparison with the reference databases of the Human Microbiome Project made it possible to identify the communities to which the metagenome of our samples is closest. The heatmap with comparison is shown in Figure 3. As follows from the heatmap, the only community our data overlaps with is the metagenome supragingival plaque.

Finally, analysis of genes present in the genome of modern Tannerella forsythia, one of the 3 key species of the red complex, and comparison of these genes with genes present in the metagenome of patient G12, revealed regions of the genome that are probably relatively new evolutionary acquisitions.

In total, we found 388 sites that were not represented in the metagenome of ancient humans. After filtering data on the presence of CDS and excluding duplicate positions, 80 genes remain in our list. A complete list of newly acquired protein-coding genes is presented in the attached materials 2, sheet "new genes".

Among interesting and evolutionarily significant new genes, several proteins associated with lanthipeptide metabolism can be distinguished - these are proteins with antimicrobial and antiallodynic activity; antirestriction protein ArdA - possessing antirestriction activity and, accordingly, increasing the likelihood of introducing foreign DNA into the genome; many transposases of different classes - enzymes that embed single-stranded DNA into the genome, as well as other proteins. Another important finding is the tetracycline resistance ribosomal protection protein - this gene is involved in the formation of antibiotic resistance.

Discussion

First, we want to note the importance of choosing a proper threshold to crop the sequences using qiime2. In general, it is assumed that one should crop the sequences near the position where the drop of quality is present. However, after a series of failures, we came up with the fact that it is necessary to experiment and vary this value to ensure that the threshold is chosen properly.

Observed differences in microbiome content in calculus and bone samples showed a high abundance of Clostridiales, Burkholderiales, Lactobacillales, Flavobacteriales and Synergistales in dental calculus. On the other hand, the most abundant families in the bone sample include Burkholderiales, Rhizobiales, Xanthomonadales, Actinomycetales and Bacillales. Most prevalent families in dental calculus are known to be obligately anaerobic[13–15]. These findings are in good agreement with the structure of dental calculus and environmental conditions in it. In addition, microbial diversity in the bone sample is wider than in case of calculus. This can also happen due to the mild environmental conditions in bone since dental calculus is rather difficult to access and populate. This is likely another reason why the project introduction refers to this environment as the “Microbial Pompeii”.

Another of our results is that the calculus community of fossil teeth and the reference Supragingival plaque differ significantly in taxonomic composition. We see this on the heatmap (Figure 3), in which we see a small intersecting fragment, and a high column of the unique bacteria of ancient man's dental calculus. This may be due to the influence of modern oral hygiene. Even within a few days, it has been shown that the community of the oral cavity and plaque changes with brushing[11]. In modern people, mature calculus does not form - that is, the community does not reach the terminal stage, and, accordingly, can be very different from mature calculus.[16] Also, the reason for the differences may be that in modern people we analyze supra-day tartar, and in the ancients - subgingival. These communities differ[11].

If we compare modern and ancient plaque bacteria in terms of metabolism, then first of all I would like to highlight the fact that modern bacteria have acquired a strong adaptation to horizontal gene transfer. This is evidenced by the presence in the genome of many new transposase genes[17], as well as associated genes. We cannot speak with confidence about the reasons for this event without additional analysis of samples intermediate in time. However,

the probable cause of this transformation may be the active use of antibiotics and antimicrobial substances by humans, including for the treatment of periodontal disease, and the associated process of bacteria acquiring antibiotic resistance genes. This is confirmed by the finding of genes directly involved in the formation of antibiotic resistance[18].
In order to confirm or deny the connection between these events, it would be interesting to add an additional point to the study - the metagenome of people who lived in the 20th century, but before the active spread of antibiotic use. So we could understand whether bacteria accumulated these genes uniformly over 1000 years, or all these transposases have moved into the genome over the past 70 years[19].

Supplemental resources

1.Lab notebook

2. Additional data

References

---

1. Rausch, P. *et al.* Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome* **7**, 133 (2019).
2. Rutanga, J. P. *et al.* 16S metagenomics for diagnosis of bloodstream infections: opportunities and pitfalls. *Expert Rev. Mol. Diagn.* **18**, 749–759 (2018).
3. Qian, Y. *et al.* Gut metagenomics-derived genes as potential biomarkers of Parkinson's disease. *Brain* **143**, 2474–2489 (2020).
4. Andrianova, N. V. *et al.* Microbiome-Metabolome Signature of Acute Kidney Injury. *Metabolites* **10**, (2020).
5. Wilkins, L. G. E., Ettinger, C. L., Jospin, G. & Eisen, J. A. Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. *Sci. Rep.* **9**, 3059 (2019).
6. Xu, P. & Gunsolley, J. Application of metagenomics in understanding oral health and disease. *Virulence* **5**, 424–432 (2014).
7. Mann, A. E. *et al.* Differential preservation of endogenous human and microbial DNA in dental

calculus and dentin. *Sci. Rep.* **8**, (2018).

8. Wade, W. G. The oral microbiome in health and disease. *Pharmacol. Res.* **69**, (2013).
9. Velsko, I. M. *et al.* The dental calculus metabolome in modern and historic samples. *Metabolomics* **13**, (2017).
10. Kolenbrander, P. E. *et al.* Communication among Oral Bacteria. *Microbiol. Mol. Biol. Rev.* **66**, 486 (2002).
11. Belstrøm, D. *et al.* Impact of Oral Hygiene Discontinuation on Supragingival and Salivary Microbiomes. *JDR Clinical and Translational Research* **3**, 57 (2018).
12. Velsko, I. M. *et al.* Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. *Microbiome* **7**, (2019).
13. Website. Wells CL, Wilkins TD. Clostridia: Sporeforming Anaerobic Bacilli. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 18. Available from: https://www.ncbi.nlm.nih.gov/books/NBK8219/.
14. Murphy, M. G. & Condon, S. Comparison of aerobic and anaerobic growth of Lactobacillus plantarum in a glucose medium. *Arch. Microbiol.* **138**, 49–53 (1984).
15. Doloman, A., Soboh, Y., Walters, A. J., Sims, R. C. & Miller, C. D. Qualitative Analysis of Microbial Dynamics during Anaerobic Digestion of Microalgal Biomass in a UASB Reactor. *Int. J. Microbiol.* **2017**, 5291283 (2017).
16. Warinner, C., Speller, C. & Collins, M. J. A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, (2015).
17. De Palmenaer, D., Siguier, P. & Mahillon, J. IS4 family goes genomic. *BMC Evol. Biol.* **8**, (2008).
18. Grossman, T. H. Tetracycline Antibiotics and Resistance. *Cold Spring Harb. Perspect. Med.* **6**, (2016).
19. Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* **74**, (2010).