**The history of human oral diseases through metagenomics analysis***Gainova Kristina, Ivanova Eugenia (Strebulaeva)*

**Abstract**

Microbial oral diseases, such as periodontitis, affect people throughout the history of mankind as can be seen from archeological evidence. Dental calculus from dead man’s teeth is a great source of well-preserved DNA of bacterial communities. In this study we examined dental samples of the dead man from Dunheim, Germany, who had suffered from periodontitis and was buried more than 2000 years ago (ca. AD 950-1200). Using results of the sequencing of the V5 region 16s DNA and GreenGenes database we found the pathogenic bacteria associated with periodontitis nowadays, and this finding specifies the same periodontitis cause for ancient and modern humans. Also, the metagenome assembly from the calculus sample was used for exploration of ancient *Tannerella forsythia* genome and its difference from modern one. There was strong evidence that transposable elements contribute to the acquisition of novel genes. Obtained results are very important in studying pathogenic bacteria evolution during human history and its adaptation to changes in human diet.

**Introduction**

To date, the majority of research aims at studying microbial community composition. There is gaining evidence suggesting that the development of many human diseases such as obesity, inflammatory bowel diseases, periodontitis, HIV, and so on can be linked to microbes that naturally reside in the body. The first step towards microbiome wide association studies is the characterization of the composition of human microbiome. The main difficulty researchers face is that a lot of microorganisms cannot be cultured by standard techniques, and the uncultured fraction includes diverse organisms that are only distantly related to the cultured ones [1]. Thus, to understand genetic diversity of microbial community culture-independent methods are essential among which is metagenomics.

There are two widely used approaches in comparative metagenomics - sequencing of the conserved 16S ribosomal RNA (rRNA) gene and shotgun whole genome sequencing (WGS) [2]. They allow us to identify microorganisms and evaluate their diversity and abundance in various environments.

16S rRNA gene sequencing utilizes PCR method to amplify parts of the hypervariable regions (V1-V9) of the bacterial 16S rRNA gene. Then molecular barcodes are added to amplicons from separate samples. Next, obtained molecules are mixed together and sequenced. Raw data of sequencing is trimmed, corrected and compared with the 16S reference database and finally a taxonomy profile is generated.

Unlike the 16S sequencing method, shotgun whole genome sequencing is an approach which sequences all genomic DNA existing in a sample. The library preparation step is similar to whole genome sequencing. The next step is a quality trimming and comparison to a reference database to generate a taxonomy profile. Shotgun sequencing can be used for additional analyses such as metabolic function profiling, antibiotic resistance gene profiling and etc.

The main goal of this article is defining bacterial species in dead man’s teeth caused periodontitis. Periodontitis is one of the most common diseases of the tissue surrounding the tooth. The human oral cavity is inhabited by approximately 700 species of microorganisms. The primary causative agent resulting in periodontal disease is the bacterial biofilm which grows on the tooth surface leading to the destruction of the supporting and surrounding tooth structures. The red complex has been proposed as a pathogenic consortium causing periodontitis and consisting of *Porphyromonas gingivalis, Tannerella forsythia* and *Treponema denticola* [3].

**Materials and methods**

**Data**

For amplicon analysis we used sequencing data from dental calculus and bone of the dead man Dunheim, Germany obtained by an instrument Roche GS Junior (454). Reads and description can be found by the link:

<https://figshare.com/articles/_Dead_man_s_teeth_dataset/12152040>  
and in the NCBI Short Read Archive:

<https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP029257>  
The GreenGenes database for taxonomic analysis can be found here:

<https://data.qiime2.org/2020.2/common/gg-13-8-99-nb-classifier.qza>  
Metagenome assembly of the calculus sample was download from here:

<https://www.dropbox.com/s/f5j52tliumt6etm/G12_assembly.fna.gz?dl=0>  
and also can be download from MG-RAST server:

<http://metagenomics.anl.gov/mgmain.html?mgpage=download&metagenome=mgm4530391.3>  
Human microbiome data for checkpoint step in our analysis was downloaded from here:  
<https://www.dropbox.com/s/aeq9fsoax68h9qk/SRS014459-Stool.fasta?dl=0>  
<https://www.dropbox.com/s/r6du0tazu7ocyop/SRS014464-Anterior_nares.fasta?dl=0>  
<https://www.dropbox.com/s/xyoy6oo12r9mjry/SRS014470-Tongue_dorsum.fasta?dl=0>  
<https://www.dropbox.com/s/mwxyzjumb6gc9sz/SRS014472-Buccal_mucosa.fasta?dl=0>  
<https://www.dropbox.com/s/ajt786ac2ijnra8/SRS014476-Supragingival_plaque.fasta?dl=0>  
<https://www.dropbox.com/s/6oiowqeg447kq1j/SRS014494-Posterior_fornix.fasta?dl=0>  
and in full variant also can be found on HMP website:

<http://hmpdacc.org/HMASM/>  
*Tannerella forsythia* genome and annotation (NC\_016610.1) was downloaded from the NCBI:

<https://www.ncbi.nlm.nih.gov/nuccore/NC_016610.1>  
  
**Taxonomic analysis**

Quality control for the 16s amplicon sequencing results and taxonomic analysis were performed with the QIIME2 package.

Import of the data and QC were performed with default parameters.

For filtering, according to the results of the QC, parameter n = 148 for --p-trunc-len and

parameter m = 32 for --p-trim-left as total length of the barcode and primer were chosen in

DADA2 pipeline (from QIIME2).

Each step of the analysis was visualized at <https://view.qiime2.org/>.

All the steps described in detail in ***Supplementary materials***.

**Calculation of organism abundances**

To check where the bacteria, found in the calculus sample, likely reside and abound we used MetaPhlAn 3.0.7 (with default parameters). Also, for comparison with the data from the Human Metagenome Project, we drew a heatmap where different bacteria were clustered by the places of their abundance in human organisms using hclust2 tool (full code for preparing the file and drawing can be found in ***Supplementary materials***).

**Comparison of the Tannerella forsythia genomes**

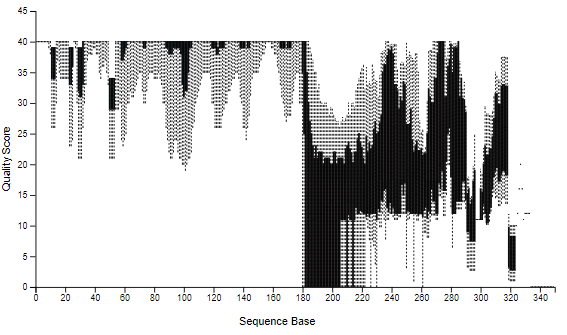
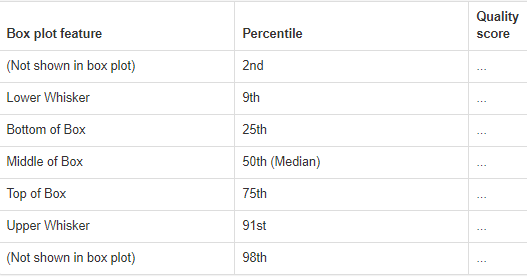
Assembly of the metagenome from calculus was aligned to the modern *Tannerella forsythia* genome using the bwa mem tool.

Preparing for the visualization was performed by samtools utility and visualization itself was done in IGVBrowser.

Intersection of the ancient and modern bacteria was performed with the use of the bedtools utility.  
Functional annotation of the obtained genes was made based on information from UniProt and Pfam databases.

**Results**

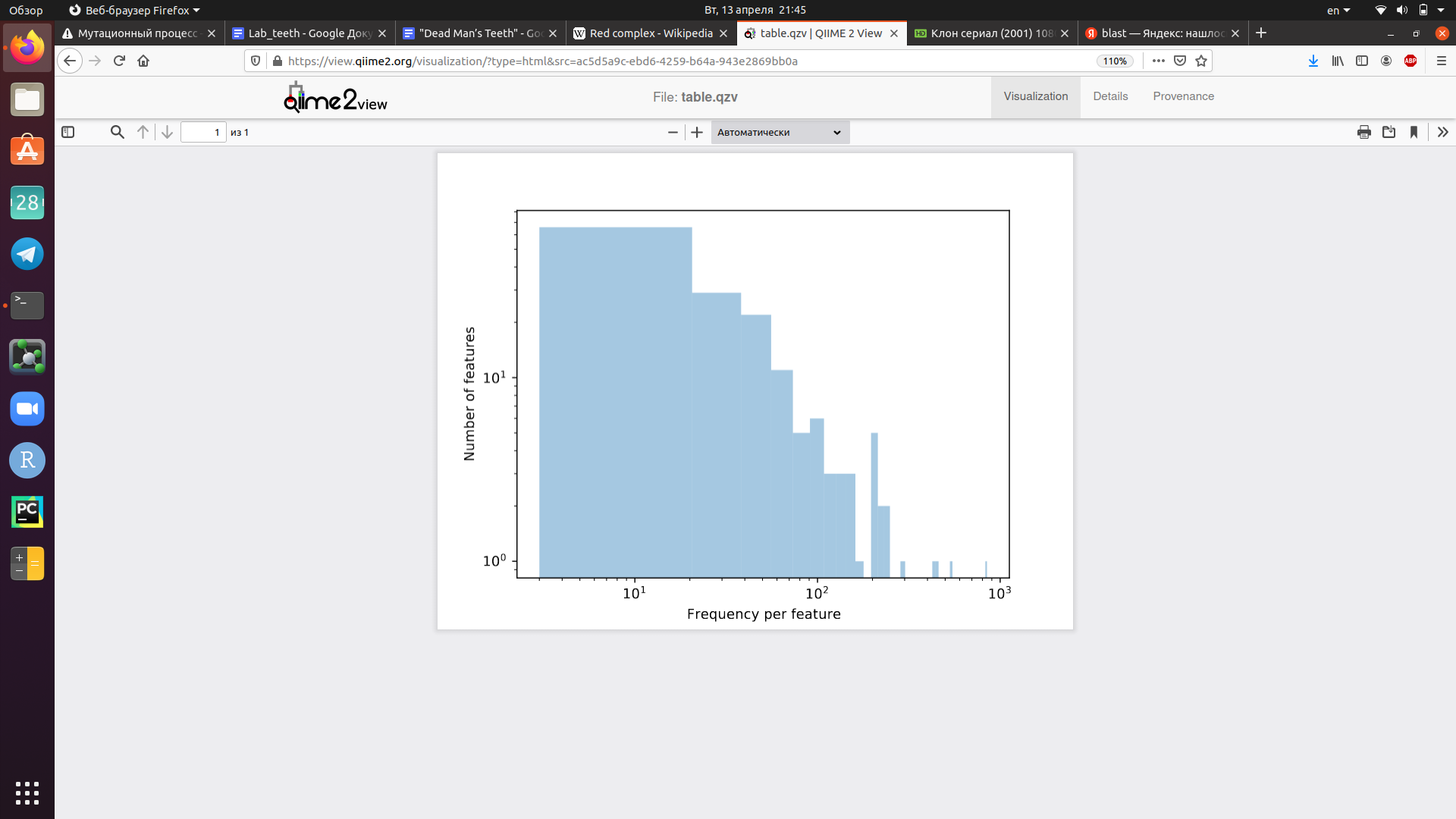
Quality control showed presence of the bad-quality read ends, as can be seen on ***Figure 1***. So, for taxonomic analysis it was decided to filter our reads in order to obtain more accurate and higher-resolved amplicon sequence variants (ASV) rather than operational taxonomic units. Results of the filtration are shown in ***Table 1***.

  
  
***Figure 1.***Distribution of quality scores for the 16s reads.

***Table 1***. - Filtration summary for 16s reads.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sample ID | Input | Filtered | % of input passed | Denoised | Non-chimeric | % of input non-chimeric |
| bone | 5788 | 5589 | 96.56 | 5377 | 5377 | 92.9 |
| calculus | 5362 | 5183 | 96.66 | 5059 | 4817 | 89.84 |

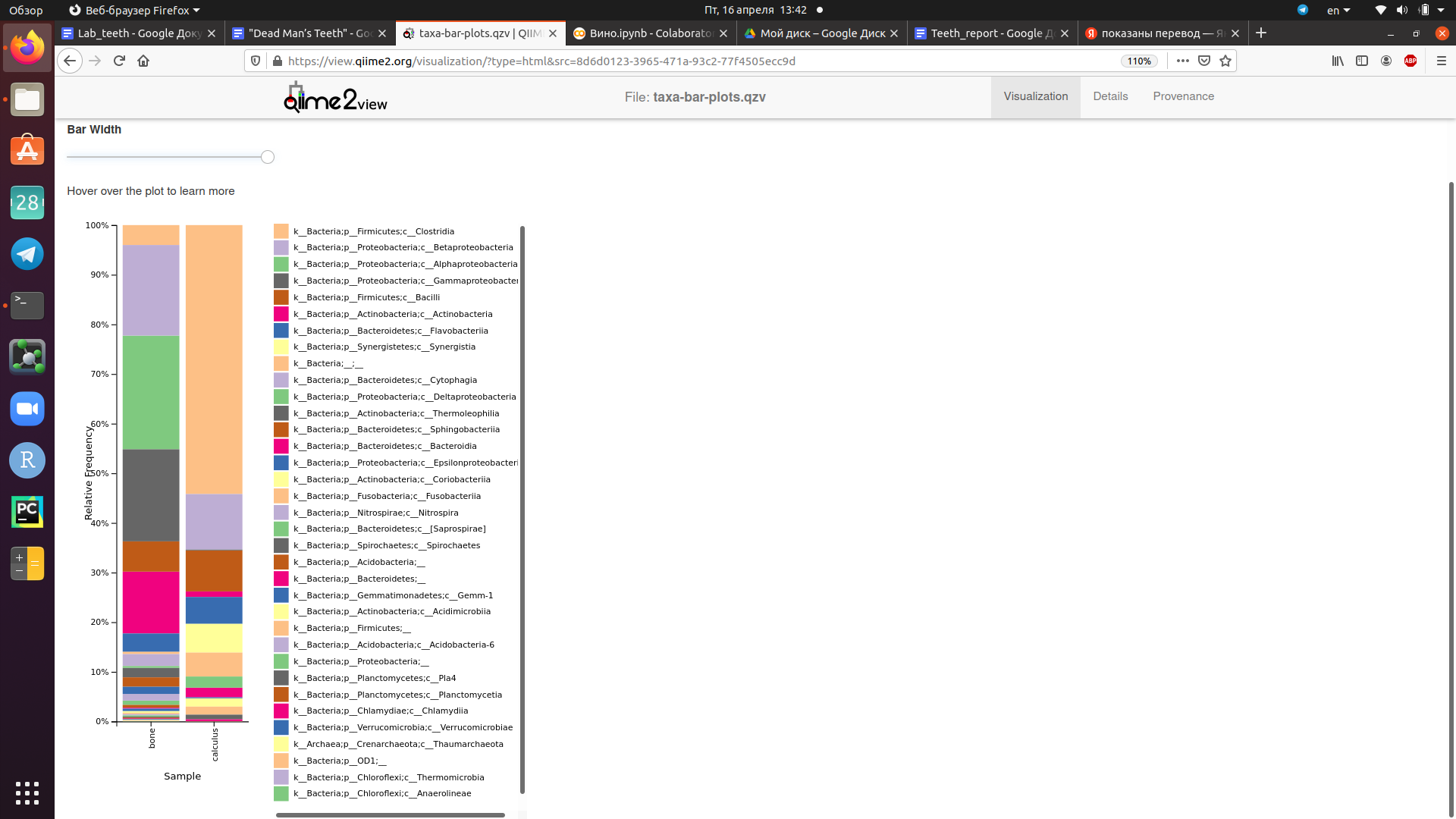
Distribution of the obtained ASV (or features) is displayed on ***Figure 2****.* Some feature statistics can be found in ***Table 2****.* Also, some statistics of the samples can be found in ***Supplementary materials*** (**tables 2-3**).



***Figure 2.***Distribution of the features in both calculus and bone samples.

***Table 2*** - Frequency per feature.

|  |  |
| --- | --- |
| Minimum frequency | 3.0 |
| 1st quartile | 13.0 |
| Median frequency | 29.0 |
| 3rd quartile | 61.5 |
| Maximum frequency | 847.0 |
| Mean frequency | 62.54 |

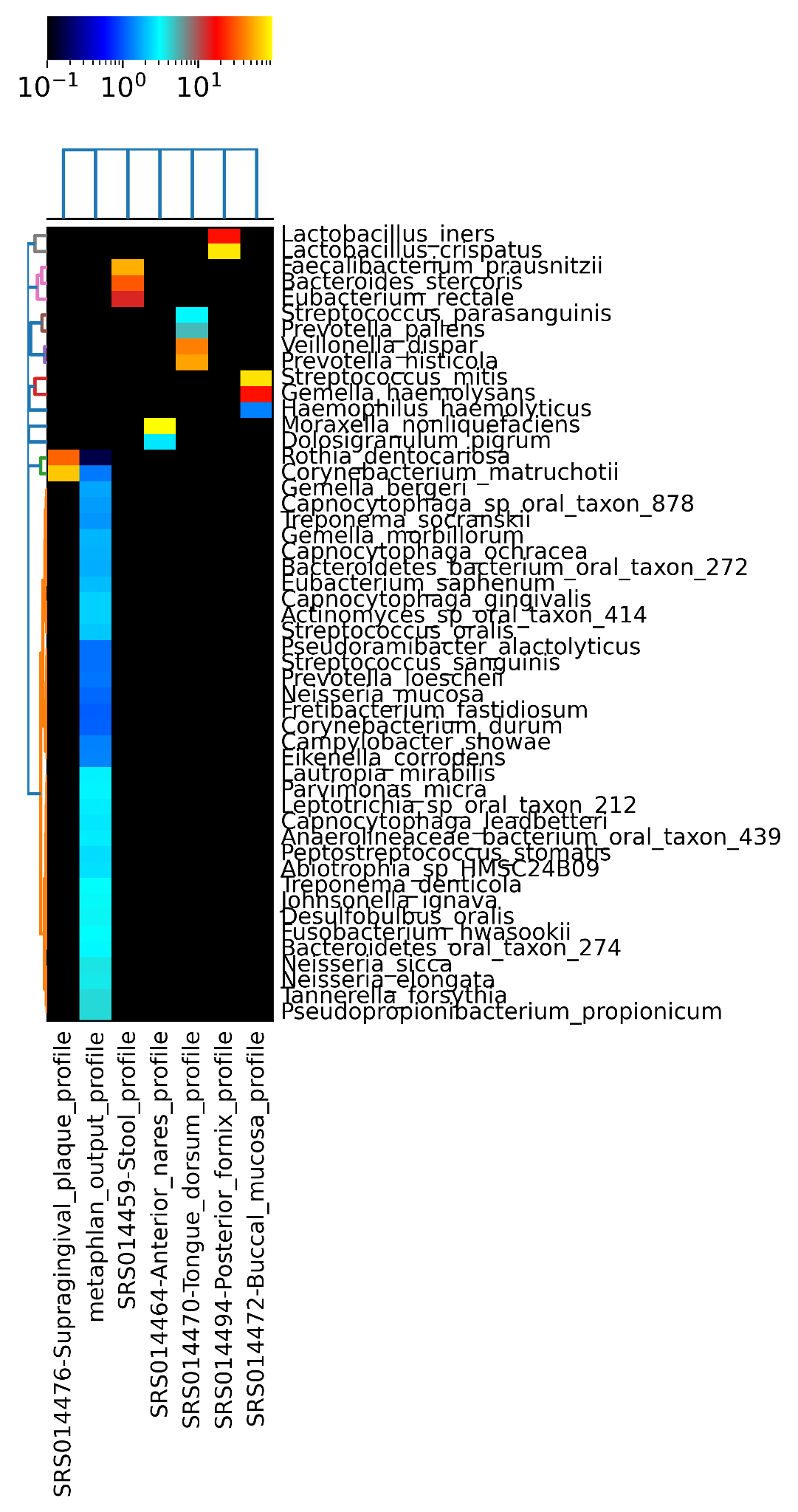
Results of the taxonomic analysis are shown in barplots, ***Figure 3***. Here can be seen proportions of taxa on the class level.  
***Figure 3.***Frequencies of the different bacterial taxa in calculus and bone samples.

Among all found bacteria there were *Tannerella sp*., *Porphyromonas sp*. and *Treponema sp*. in calculus sample, which cause periodontitis nowadays. Proportions of these bacteria are shown in ***Table 3***.

***Table 3*** - Proportions of potentially pathogenic bacteria in the calculus sample.

|  |  |
| --- | --- |
| Bacterium | Proportion, % |
| Tannerella sp. | 0.519 |
| Porphyromonas sp | 0.291 |
| Treponema sp. | 0.353 |
| Treponema socranskii | 0.561 |

To confirm that our sample is not contaminated and represents genuine dental microbial communities we compared our calculus sample metagenome with representative samples of different organism locations from Human Genome Project. The results of comparison are shown as heatmap on***Figure 4***. We can notice that only the Supragingival plaque profile and experimental sample contain the same microorganisms (*Rothia dentocariosa* and *Corynebacterium matruchotii*). Also, none of the representative organisms from other sites were found in our sample, so it was not contaminated.



***Figure 4.***Clustering of the samples from different parts of the human organism and dental calculus sample.

Comparison of the ancient and modern *T. forsythia* genomes revealed new genes, obtained during evolution. These genes are listed in ***Table 5 in the Supplementary materials***.

**Discussion**

The human body is colonized by trillions of microorganisms called microbiota. However composition of microbial communities in various parts of the human organism can differ significantly. In this study we compared two samples - dental calculus (experimental sample) and tooth root (control). Dental calculus contains only representatives of Bacteria while in bone about 0.13 % are representatives of Archaea. The most common bacteria in the calculus sample is k\_\_Bacteria; p\_\_Firmicutes; c\_\_Clostridia; o\_\_Clostridiales; f\_\_[Mogibacteriaceae] (Mogibacteriaceae from Clostridiales order). It was shown in different scientific articles that people who suffer from apical periodontitis have very high-level microbial species belonging to *Mogibacteriaceae* family [4]. The most common bacteria in tooth root sample is k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Betaproteobacteria;o\_\_Burkholderiales;f\_\_Alcaligenaceae. Their role is still unknown.

There are a lot of factors which can influence human microbiome composition - the main ones are oxygen and nutrients availability. Accordingly, microbial content of bone and calculus samples differs as does microbial content of other organism sites, as can be seen on the heatmap, ***Figure 4***. This fact allows scientists to use representative bacteria for the contamination control and study the diseases which have been associated with the human microbiota through abundances of the pathogens.

Previous studies showed that three bacteria of the “red complex” are responsible for severe forms of periodontitis [5]. In the sample from dental calculus we have found genus of all those bacteria through the taxonomic analysis by 16s rRNA genes. Additionally, we determined the presence of *T. denticola* and *T.forsythia* genomesin the metagenome assembly. Also, we found bacterium *Treponema socranskii* which was isolated from patients with periodontitis [6]. These findings tell us that ancient people had suffered from the disease caused by the same bacteria as today.

Comparative analysis of modern *T. forsythia* genome with ancient one revealed about 80 new protein-coding genes. Among them different conjugative transposon proteins, IS110 and other families’ transposases, response regulator etc. The obtained list of genes indicates that a lot of new genes were acquired by recruiting transposable element proteins and using them as cellular proteins. The most interesting new protein is tetracycline resistance ribosomal protection protein. This protein protects the bacterial ribosome from binding with tetracycline antibiotics. This molecule is a translational GTPase and has structure and sequence similarity with elongation factor EF-G [7]. In *Salyers et al.* review was mentioned that conjugative transposons of *Bacteroides* have unique features - the ability to insert in coresident plasmids and mobilize them, to excise and mobilize unlinked integrated elements and to control element transfer functions [8]. So transposable elements can be considered as one of the main drivers of the *T. forsythia* genome evolution.

In addition, it was found new gene coding beta-ketoacyl-ACP synthase III which is involved in the dissociated fatty-acid biosynthesis system in bacteria. This adaptation can be explained by the change of type of food. Modern people eat food which contains more fats and carbohydrates. In support of this theory we also found TIGR04157 family glycosyltransferase as a novel gene. Glycosyltransferases are enzymes that catalyze transfer of the saccharide moieties from an activated sugar to the acceptor and this particular transferase can be obtained by T. forsythia due to the increase of carbohydrate consumption [9].

**Conclusion**

In this research it was shown that ancient people suffered from periodontitis. The development of disease was caused by activity of *T. forsythia and T. denticola.* Also, it was demonstrated that the genome of modern *T. forsythia* contains a lot of new genes. Most of them serve as an adaptation to changing the type of food by the host organism. A lot of these genes could be acquired by T. forsythia via conjugative transposons.

**References**

1. Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. Annu Rev Genet. 2004;38:525-52. doi:10.1146/annurev.genet.38.072902.091216. PMID: 15568985.

2. Durazzi, F., Sala, C., Castellani, G. *et al.* Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci Rep* 11, 3030 (2021). https://doi.org/10.1038/s41598-021-82726-y

3. Mohanty, Rinkee et al. “Red complex: Polymicrobial conglomerate in oral flora: A review.” *Journal of family medicine and primary care* vol. 8,11 3480-3486. 15 Nov. 2019, doi:10.4103/jfmpc.jfmpc\_759\_19

4. Qian, W., Ma, T., Ye, M. *et al.* Microbiota in the apical root canal system of tooth with apical periodontitis. *BMC Genomics* 20, 189 (2019). https://doi.org/10.1186/s12864-019-5474-y

5. Rôças IN, Siqueira JF, Santos KR, Coelho AM (April 2001). ""Red complex" (Bacteroides forsythus, Porphyromonas gingivalis, and Treponema denticola) in endodontic infections: a molecular approach". *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics*. **91** (4): 468–71. doi:10.1067/moe.2001.114379. PMID 11312465

6. Smibert, R. M., Johnson, J. L., & Ranney, R. R. (October 1984). "Treponema socranskii sp. nov. Treponema socranskii subsp. socranskii subsp. nov. Treponema socranskii subsp. buccale subsp. nov., and Treponema socranskii subsp. paredis subsp. nov. Isolated from the Human Periodontia". *International Journal of Systematic Bacteriology*. **34** (4): 457–462. doi:10.1099/00207713-34-4-457

7. Li, W., Atkinson, G., Thakor, N. *et al.* Mechanism of tetracycline resistance by ribosomal protection protein Tet(O). *Nat Commun* 4, 1477 (2013). https://doi.org/10.1038/ncomms2470

8. Salyers, A. A., Shoemaker, N. B., Stevens, A. M., & Li, L. Y. (1995). Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiological reviews*, *59*(4), 579–590.

9. Etzler ME, Varki A, Cummings RL, Esko JD, Freeze HH, Hart GW, eds. (2008). *Essentials of Glycobiology* (2nd ed.). Plainview, N.Y: Cold Spring Harbor Laboratory Press. ISBN 978-0-87969-770-9.