The Secrets of a Long-Gone Smile

Evgeny Kovalenko[1] and Anya Zhurakovskaya[1]

[1]Bioinformatics Institute

Abstract

Metagenomics is a branch of bioinformatics that studies microbial communities, their species composition and the relationship between them. In particular cases scientists can study communities of microorganisms that lived thousands of years ago. In the current work, a community of microorganisms that lived on people's teeth several centuries ago was analyzed. It has been shown that the bacteria that cause periodontitis in our time were also the source of the disease in the past. Genome analysis also showed that the change of *Treponema denticola* is primarily associated with transposone evolution.

Introduction

Metagenomics is the study of the structure and function of entire nucleotide sequences isolated and analyzed from all the organisms (typically microorganisms) in a bulk sample. Metagenomics is often used to study a specific community of microorganisms, such as those residing on human skin, in the soil or in a water sample [1]. Investigating microbial diversity, population structure, genetic and evolutionary and cooperative relationships metagenomics allows to reveal hidden abundance of species, understand relationship inside consortium and suggest possible consequences of this interaction [2].

Recovery of DNA sequences longer than a few thousand base pairs from environmental samples is very difficult. Two approaches are primarily used to identify individual species or Operational Taxonomic Unit: 16S rRNA analysis and shotgun sequencing [3].

According to shotgun sequencing, DNA is randomly divided on many short sequences and then reconstructs into a consensus sequence. Shotgun metagenomics uses for profiling of taxonomic composition and recovering of whole genome sequences [4]. 16S rRNA gene sequencing bases on marker gene analysis - 16S rRNA contains conservative and variable regions acceptable for microorganism identification [5].

In our investigation we explore samples of dental calculus of ancient people and study the history of oral diseases in humans. Dental calculus preserve bacterial DNA so these samples are a starting point for metagenomic research based on analysis of V5 16S ribosomal RNA. In result our research we figure out what bacterium triggered periodontitis of ancient people and traced it evolution till our days.

Methods

Amplicon sequencing

Raw DNA reads were obtained from NCBI databases (SRA: SRP029257). The DNA was extracted from the material underneath the dental calculus of teeth found at the monastic site in Dunheim, Germany.

Raw reads were demultiplexed, clustered into amplicon sequence variants (ASV), and compared to taxonomic databases using QIIME2 package [6]. To visualize the taxonomic composition of metagenomic samples we used the online version of MicrobiomeAnalyst [7].

Shotgun sequencing

To evaluate changes between ancient and modern pathogen that causes periodontal disease we used metagenomic data from affected individual [8] and compared it to *Tannerella forsythia* genome. Assembly of metagenomic data was aligned to the reference genome using Burrows–Wheeler algorithm [9]. To identify uncovered regions we used BEDTools suite (intersect command with -v flag) [10].

Results

Amplicon sequencing

Analysis of read amount demonstrated that samples totally includes 46248 reads (5138.67 on average). Then we stripped barcode and filtered reads for further analysis. We discovered the number of amplicon sequence variant(ASV) - analogue of the traditional OTUs via clusterisation. The number of ASV was 465. After that we analysed 16S rRNA sequence and determined which taxonomic groups amplicons belong to. The "red complex" bacteria (*Porphyromonas gingivalis* and *Tannerella forsythia* and *Treponema denticola*) appears in dental calculus samples of both healthy and affected by peridontitis teeth (Figure 2). The only bacteria that was not detected in healthy teeth is *T. forsythia* as shown on Figure 3.

Compiled on: April 17, 2023.
Draft manuscript prepared by the author.

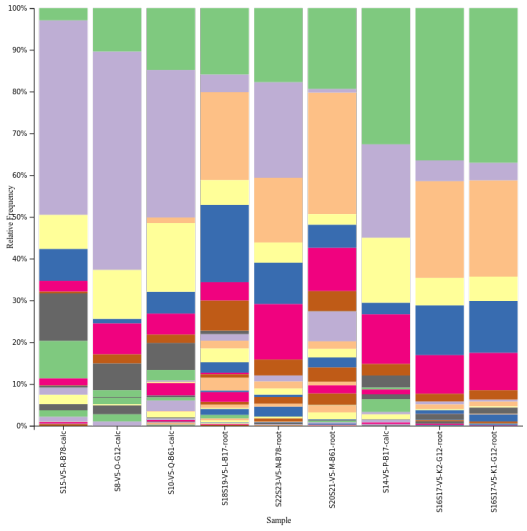

Figure 1. The distribution of microorganism species in dental calculus of different people

Shotgun sequencing

The comparison between the reference genome of *T. forsythia* and the genome obtained from metagenomic assembly reveals the presence of certain genes that have emerged through evolution. In total, there are 189 such genes. According to annotation, most of these genes are assotiated with transposase activity.

Discussion

We investigated the minority of microbial diversity in order to explore representatives of "red complex" in the tooth of sick individuals. The range of search was directed to *P.gingivalis*, *T. forsythia* and *T.denticola*. Every samples includes *P.gingivalis* and *T.denticola* in calculus (root don't contains these bacteria hence it is a heritage of fossils). However, *T.denticola* presents only in samples of ill people that may indicate that the presence of all three bacteria is necessary for the appearance of periodontal disease (or, which is also likely, *T.denticola* acts as a trigger for the onset of the disease).

The comparison of ancient and modern *T. forsythia* genomes shows that in the span of 1000 years this bacteria aquired a lot of mobile elements and transposons.

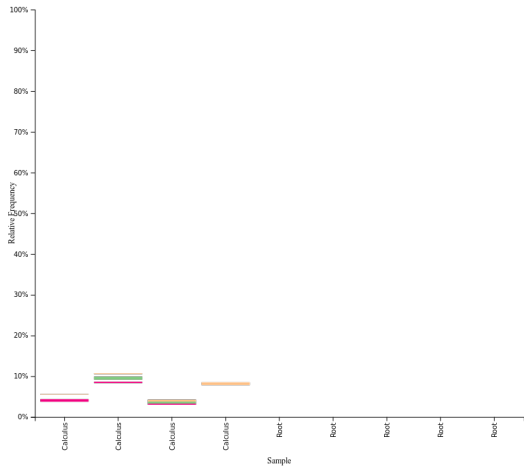

Figure 2. The appearance of harmful bacteria in dental calculus and root (green - *Treponema denticola*, pink - *Porphyromonas gingivalis*, yellow - *Tannerella forsythia*)

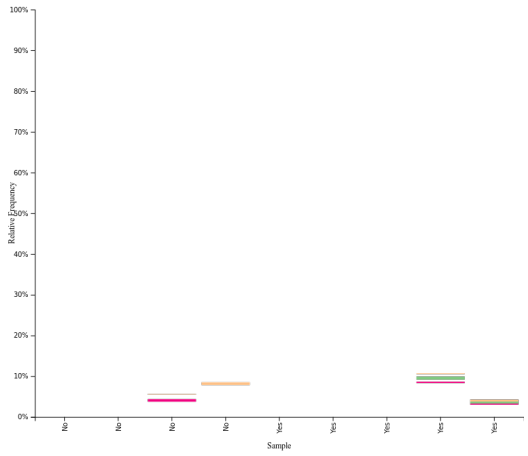

Figure 3. The distribution of investigated species among healthful and sick people (green - *Treponema denticola*, pink - *Porphyromonas gingivalis*, yellow - *Tannerella forsythia*)

References

1. NIH. The New Science of Metagenomics. National Research Council (US) Committee on Metagenomics: Challenges and Functional Applications 2007;.
2. Lu Z, FengXin C, Zhan Z. Advances in Metagenomics and Its Application in Environmental Microorganisms. Frontiers in Microbiology 2021;12.
3. Schloss P, Handelsman J. Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. Applied and Environmental Microbiology 2005;71(3):1501–1506.
4. Quince C, Walker A, Simpson J, et al. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 2017;35:833–844.
5. Johnson J, Spakowicz D, Hong B, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun 2019;10:5029.
6. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, Scalable and Exten-

sible Microbiome Data Science using QIIME 2. Nature Biotechnology 2019;37(8):852–857.
7. Chong J, Liu P, Zhou G, Xia JJ. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. Nature Protocols 2020 01;15:1–23.
8. Rast;. https://www.mg-rast.org/mgmain.html?mgpage=download&metagenome=mgm4530391.3.
9. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009 05;25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
10. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010 01;26(6):841–842. https://doi.org/10.1093/bioinformatics/btq033.