Microbiome differential calculus for 1000-years-old teeth

Michael Greenberbg, Polina Vaganova
Bioinformatics Institute, Saint Petersburg, Russia

Abstract.
Over the past decade, metagenomics has revolutionised microbial ecology by allowing direct genetic analysis of genomes within environmental samples. This study uses metagenomic data from ancient dental root and calculus samples to detect periodontal disease and compare results between shotgun and 16S sequencing. Using amplicon sequencing of 16S rRNA genes and shotgun metagenomics, we identify the microbial composition of these ancient samples. Our results highlight the potential of metagenomics to reveal historical microbial populations and provide insights into the evolution of *Tannerella forsythia*.

Keywords: Metagenomic, microbiome, periodontit, microbial pompeii, shotgun-sequencing .

1 Introduction

One of the most remarkable events in the field of microbial ecology in the last decade has been the discovery and development of metagenomics. Metagenomics is defined as the direct genetic analysis of genomes contained within an environmental sample. Metagenomics typically involves two specific sequencing strategies: amplicon sequencing, most commonly of the 16S rRNA gene as a phylogenetic marker; or shotgun sequencing, which captures the full breadth of DNA within a sample.[1] The use of the 16S ribosomal RNA gene as a phylogenetic marker has proven to be an efficient and cost-effective strategy for microbiome analysis.[2] However, 16S rRNA gene amplicon sequencing is typically limited to taxonomic classification, depending on the database and classifiers used, and thus provides incomplete information.[2] Furthermore, the choice of the 16S rRNA gene region targeted for sequencing appears to be one of the major factors underlying technical differences in the resulting microbiome composition.[3] Shotgun metagenomics, on the other hand, offers the advantage of species- and strain-level classification of bacteria. It also allows researchers to directly determine functional aspects of the bacteria in the samples,[4] and enables the exploration of previously unknown microbial life that would otherwise remain unclassifiable.[5] However, the relatively high cost of shotgun metagenomics and more demanding bioinformatic requirements have prevented its use for large-scale microbiome analysis.[1]

Recently, metagenomic analysis of ancient DNA reads sequenced from ancient human samples has enabled researchers to obtain information about historical human populations with great depth. In this study, we provide a metagenomic analysis of 1000 year old dental roots and calculus to find ancient periodontal disease and provide a comparison between shotgun and 16S sequencing results.

2 Materials and methods

In our study, we used V5 16S ribosomal RNA reads obtained from different samples using the Roche GS Junior (454) platform, together with whole metagenome shotgun sequencing of dental calculus from the G12 sample. Data pre-processing and taxonomic assignment were performed using the dada2[6] (version 1.30.0) and phyloseq[7] (version 1.46.0) R packages. Subsequent analysis was performed using the MicrobiomeAnalyst online tool.

Kraken2[8] (version 2.1.3) was used for taxonomic assignment of the premade G12 metagenome assembly. The Pavian online tool was used to visualise the kraken2 results. To compare the *Tannerella* genomes between the G12 sample and the modern *Tannerella forsythia* strain (genome accession number NC_016610.1), BWA[9] (version 0.7.17) was used for alignment. Sorting and indexing of the alignment file was performed using SAMtools[10] (version 1.19.2). BEDtools[11] (version 2.31.1) was used for the identification of genes unique to the modern strain.

3 Results

The beta diversity between calculus and root samples is shown in figure 1. We observed statistically significant differences between both types of samples (PERMANOVA, p-value = 0.007). However, the alpha diversities (Figure 2) for both sample types did not show a significant difference (T-test, p-value = 0.86103). This suggests that the overall operational taxonomic unit (OTU) compositions between the groups are similar (or the test isn't sensitive enough), but the distribution of OTUs between the samples is different.

16S classification results for different samples are shown at figure 3. We found members of "the red complex" (*Porphyromonas gingivalis*, *Tannerella forsythia*, *Treponema denticola*) using 16S data in 2 read libraries (fig. 4). These libraries correspond to the B61 and G12 samples.

Shotgun sequence data profiling revealed 39 bacterial phylum, identified bacteria are shown at figure 5. 16S classification results for G12 sample revealed only 7 bacterial phylum (fig. 6), 6 of them were presented in shotgun classification results for G12 sample and *Epsilonbacteraeota* phyla was unique for 16S data.

The assembled G12 metagenome was aligned to the *Tannerella forsythia* reference genome. By intersecting resulting BED file with the GFF3 annotation of the reference genome, we were able to identify genes unique to the modern strain. All annotated genes are presented in supplementary section. Among these genes, we discovered the presence of tetracycline resistance ribosomal protection protein, as well as several multidrug resistance (MDR) genes such as MMPL family transporter. Interestingly, the most abundant type of new *Tannerella* genes were found to be associated with transposons (various IS-family mobile elements) and the conjugation process (conjugative transposon proteins such as TraJ and TraK, and conjugative transfer proteins).

4 Discussion

We found the "red complex" bacteria in two subjects. This group of bacteria associated with severe forms of periodontal disease and usually found in periodontal pockets.[12] We can therefore assume that both subjects suffered from periodontal disease.

The alpha and beta diversities indicate that the overall OTU compositions between calculus and root are similar, but the distribution of OTUs between samples differs. However, it should be noted that certain genus are present exclusively in either the root (such as *Tissierella*) or the calculus (such as *Streptococcus*). This may be due to differences in bacterial living conditions, with less aerobic conditions in the root and more aerobic conditions in the calculus. However, further investigation is required to confirm this hypothesis.

Our observations demonstrate the evolution of one of the members of the red complex, *Tannerella forsythia*. Additional genes involved in both specific (tetracycline resistance ribosomal protection protein) and non-specific (MMPL family transporter) antibiotic resistance have emerged in the genome of this bacterium. One of the possible mechanisms for the emergence of these novelties is the expansion of the repertoire of genes associated with conjugation and mobile elements, significantly increasing the potential for horizontal gene transfer and contributing to bacterial evolution.[13,14]

References

1 X. C. Morgan and C. Huttenhower, "Chapter 12: Human microbiome analysis," *PLoS Computational Biology* **8**, e1002808 (2012).

2 M. G. I. Langille, J. Zaneveld, J. G. Caporaso, *et al.*, "Predictive functional profiling of microbial communities using 16s rrna marker gene sequences," *Nature Biotechnology* **31**, 814–821 (2013).

3 A. Hiergeist, J. Gläsner, U. Reischl, *et al.*, "Analyses of intestinal microbiota: Culture versus sequencing: Figure 1," *ILAR Journal* **56**, 228–240 (2015).

4 A. M. Walsh, F. Crispie, O. O'Sullivan, *et al.*, "Species classifier choice is a key consideration when analysing low-complexity food microbiome data," *Microbiome* **6** (2018).

5 C. Rinke, P. Schwientek, A. Sczyrba, *et al.*, "Insights into the phylogeny and coding potential of microbial dark matter," *Nature* **499**, 431–437 (2013).

6 B. J. Callahan, P. J. McMurdie, M. J. Rosen, *et al.*, "Dada2: High-resolution sample inference from illumina amplicon data," *Nature Methods* **13**, 581–583 (2016).

7 P. J. McMurdie and S. Holmes, "phyloseq: An r package for reproducible interactive analysis and graphics of microbiome census data," *PLoS ONE* **8**, e61217 (2013).

8 D. E. Wood, J. Lu, and B. Langmead, "Improved metagenomic analysis with kraken 2," *Genome Biology* **20** (2019).

9 H. Li and R. Durbin, "Fast and accurate short read alignment with burrows–wheeler transform," *Bioinformatics* **25**, 1754–1760 (2009).

10 P. Danecek, J. K. Bonfield, J. Liddle, *et al.*, "Twelve years of SAMtools and BCFtools," *GigaScience* **10** (2021). giab008.

11 A. R. Quinlan and I. M. Hall, "Bedtools: a flexible suite of utilities for comparing genomic features," *Bioinformatics* **26**, 841–842 (2010).

12 K. Saidath, R. Mohanty, S. Asopa, *et al.*, "Red complex: Polymicrobial conglomerate in oral flora: A review," *Journal of Family Medicine and Primary Care* **8**(11), 3480 (2019).

13 A. J. Weisberg and J. H. Chang, "Mobile genetic element flexibility as an underlying principle to bacterial evolution," *Annual Review of Microbiology* **77**, 603–624 (2023).

14 F. de la Cruz, *The Role of Conjugation in the Evolution of Bacteria*, 133–138. American Society of Microbiology (2012).

5 Supplementary

Genes present only in modern *Tannerella forsythia* strain.

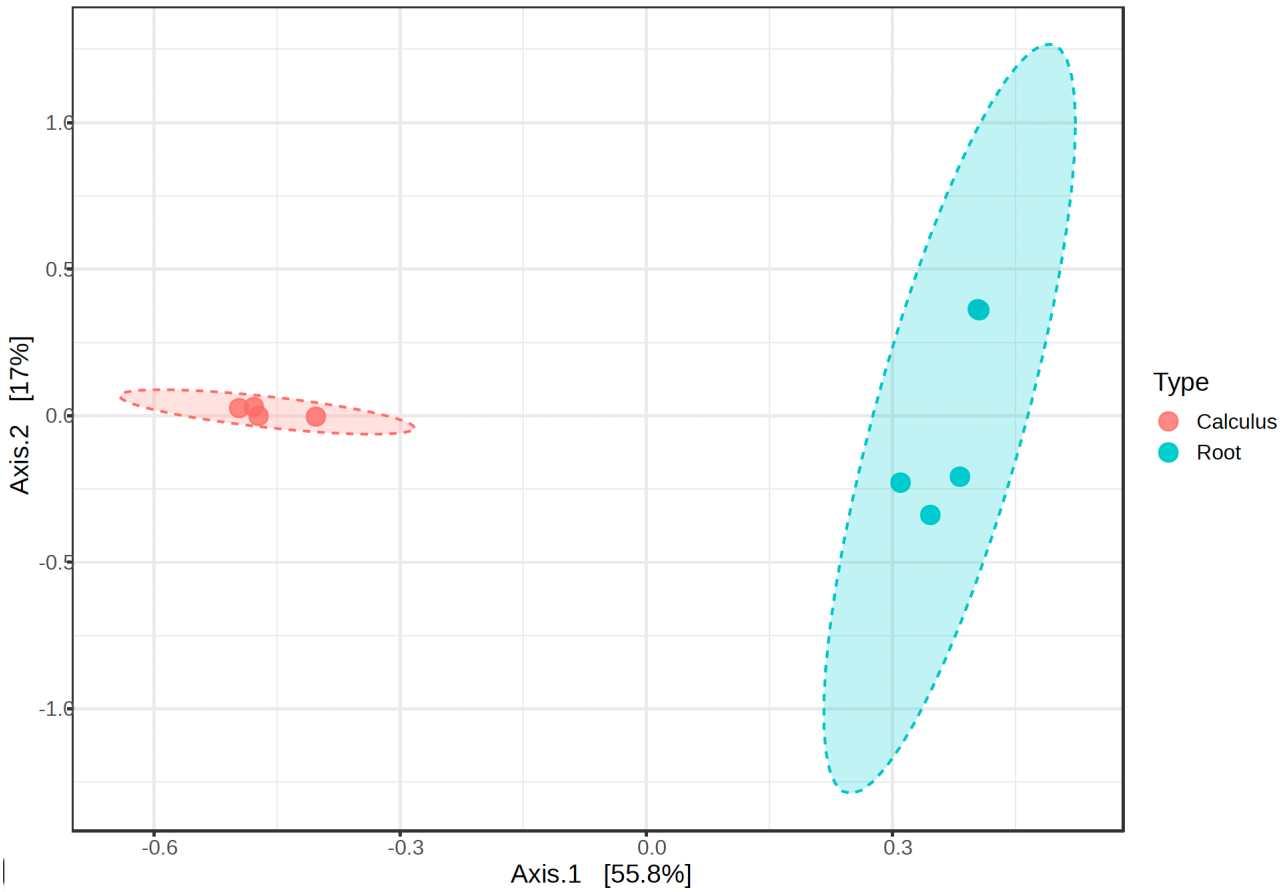

Fig 1: Beta diversity between root and calculus.

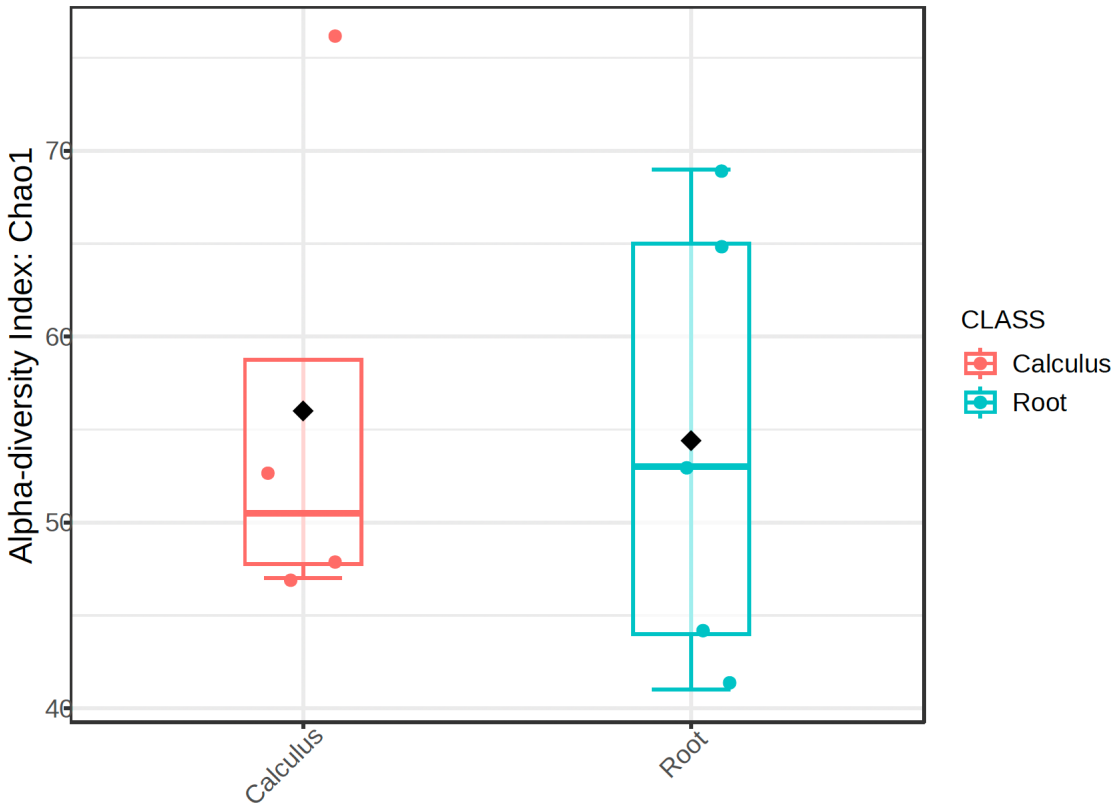

Fig 2: Alpha diversity in root and calculus.

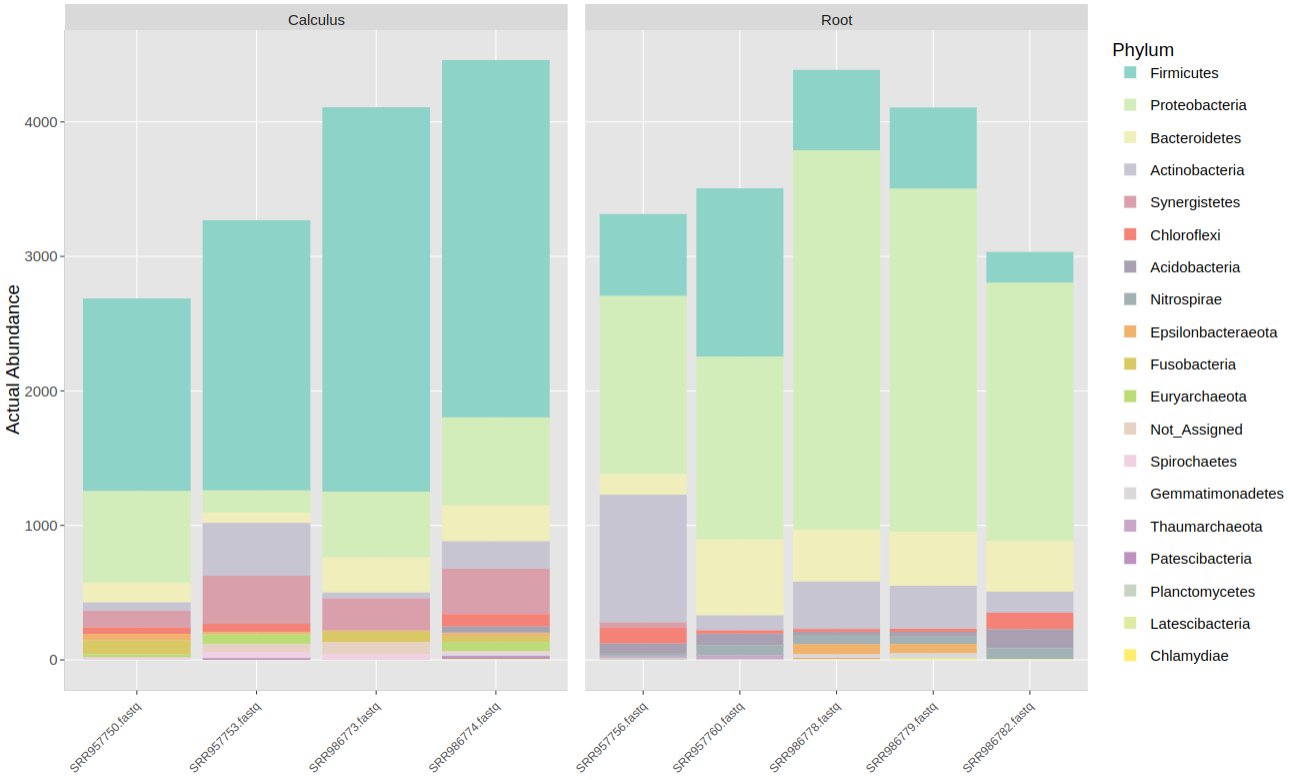

Fig 3: 16S classification results for different samples.

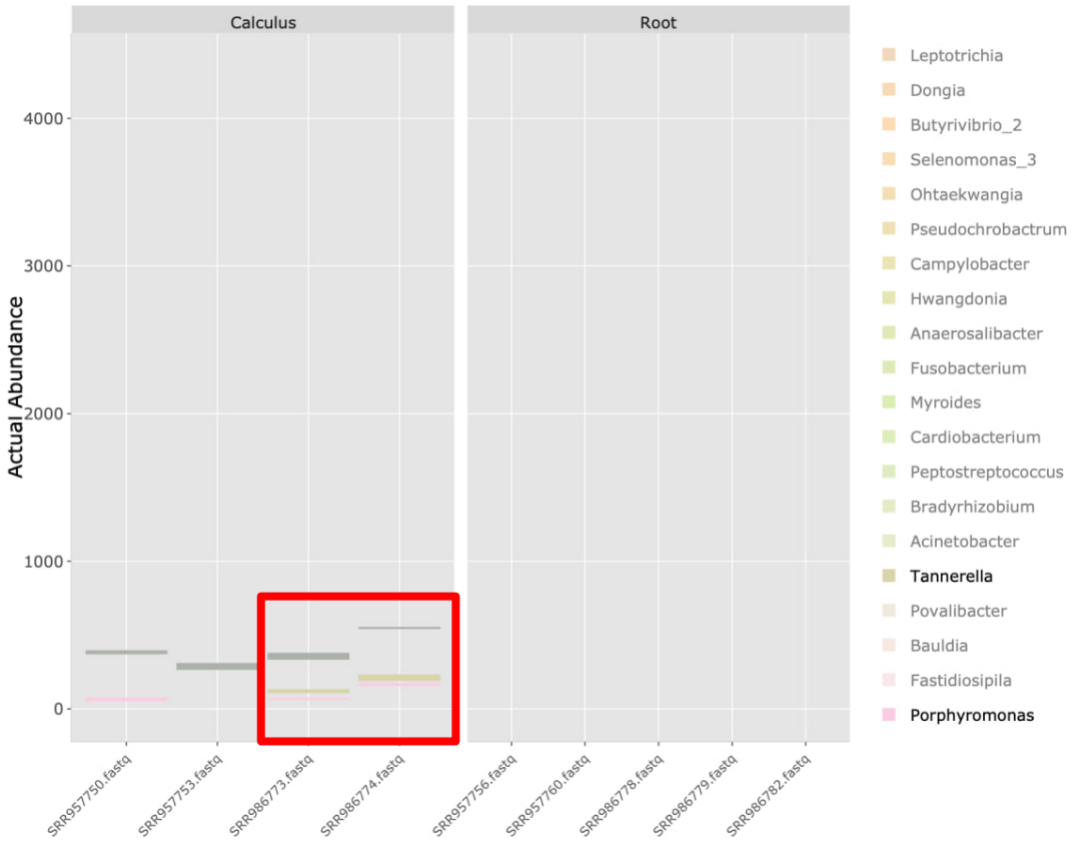

Fig 4: Bacteria from "the red complex" were found in two read libraries (red frame) in 16S data, related to different people.

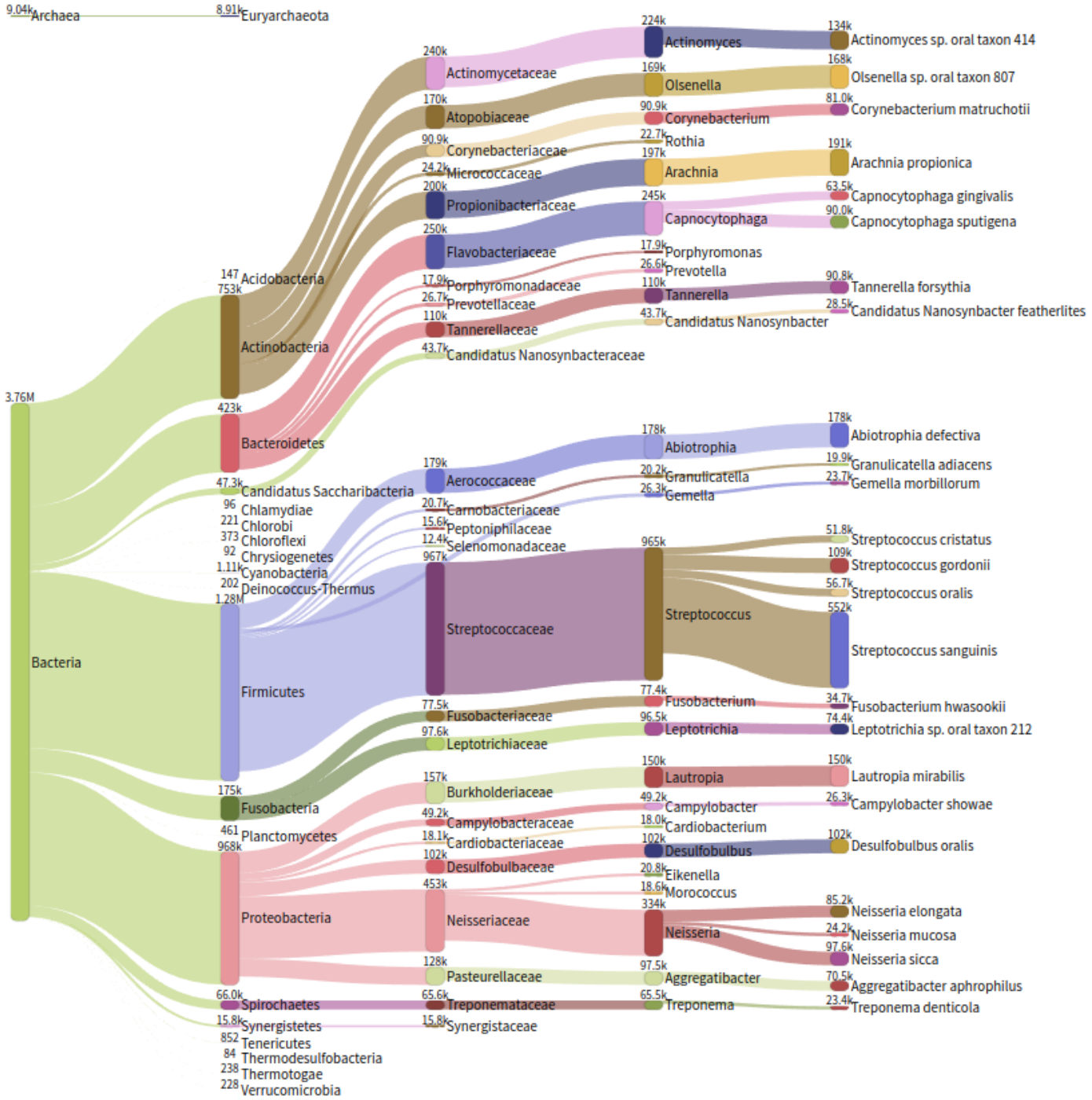

Fig 5: Sankey diagram of shotgun classification results.

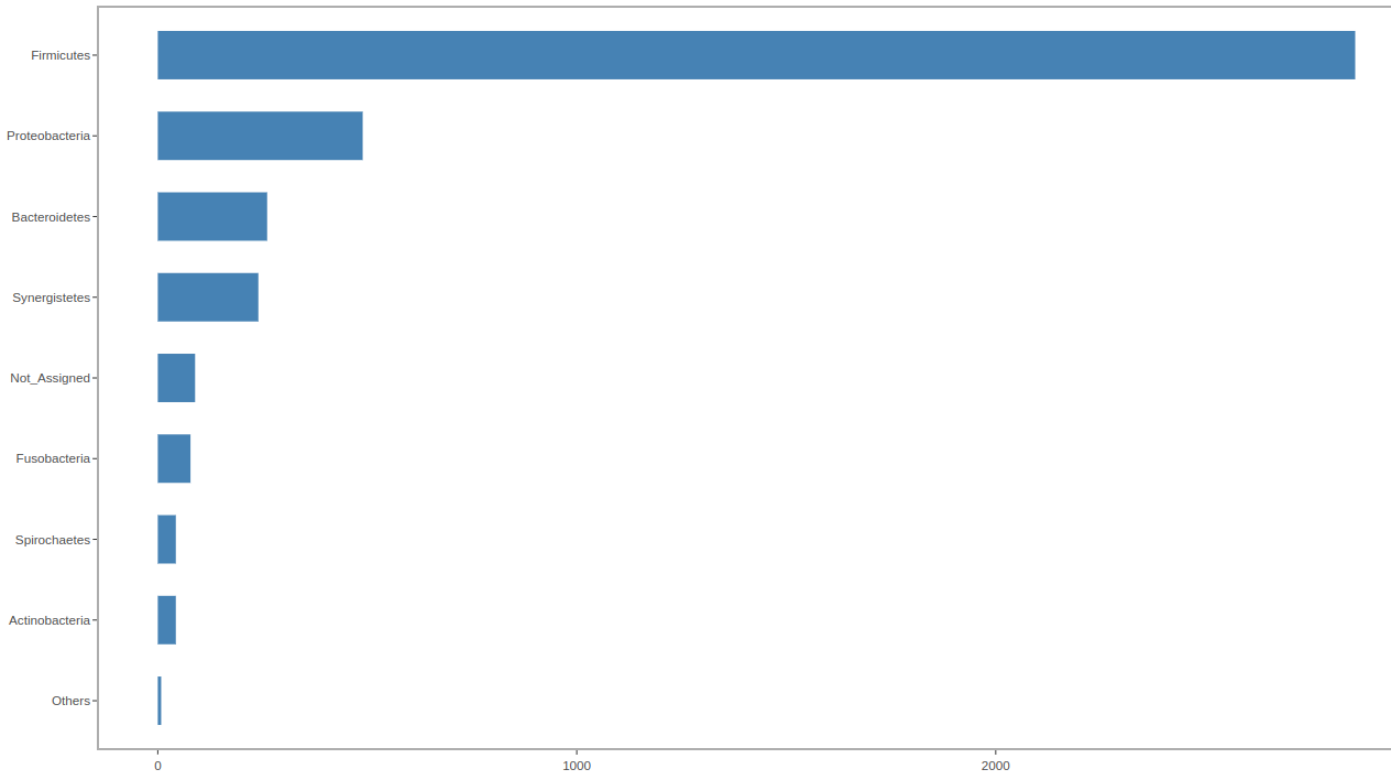

Fig 6: Bacterial phylum in 16S classification results for G12.