

# Microbiomes in Supragingival Biofilms and Saliva of Adolescent Patients with Induced and Naturally Occurring Gingivitis Relative to Gingival Health

**Roland Wirth**

Szegedi Tudományegyetem

**Gergely Maróti**

Szegedi Biológiai Központ

**Lídia Lipták**

Szegedi Tudományegyetem

**Mónika Mester**

Szegedi Tudományegyetem

**Alaa Al Ayoubi**

Szegedi Tudományegyetem

**Bernadett Pap**

Magyar Tudományos Akadémia Szegedi Biológiai Központ

**Melinda MadlÉna**

Szegedi Tudományegyetem

**Janos Minarovits**

Szegedi Tudományegyetem

**Kornél L. Kovács** (✉ [kovacs.kornel@bio.u-szeged.hu](mailto:kovacs.kornel@bio.u-szeged.hu))

Szegedi Tudományegyetem Biológia Intézet <https://orcid.org/0000-0002-3926-0497>

---

## Research

**Keywords:** Gingivitis, Microbiomes, gingival health, saliva

**Posted Date:** August 6th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-45630/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Comparison of the microbiomes in supragingival biofilm and saliva samples collected from juvenile patients developing induced or spontaneous gingivitis with healthy controls.

**Results:** 36 supragingival biofilm samples from 9 adolescent gingivitis patients wearing orthodontic appliances (induced gingivitis), 40 supragingival plaques from 10 patients having spontaneous gingivitis, and 36 control samples from 9 individuals without gingivitis in the same age group were analyzed by 16S rRNA gene amplicon sequencing. Salivary microbiomes of the same persons were characterized by shotgun metagenome sequencing to compare the sessile, i.e. biofilm immobilized communities with planktonic microbiota. The amplicon and whole genome data sets were scrutinized using bioinformatics workflows designed to minimize systemic biases. RDP and RefSeq reference databases were compared in the identification of microbiome members.

The composition and diversity of bacterial communities did not differ extensively between the two groups of gingivitis patients and controls. In spite of the overall similarities, the relative abundance of the genera *Fusobacterium*, *Accermansia*, *Treponema* and *Campylobacter* was prominently higher in samples from gingivitis patients *versus* controls. In contrast, the genera *Lautropia*, *Kingella*, *Neisseria*, *Actinomyces* and *Rothia* were significantly more abundant in controls than in either of the two gingivitis groups.

**Conclusions:** The higher relative abundance of certain gingivitis-associated taxa may either reflect their role in disease pathogenesis or may indicate that gingival inflammation favored the selective overgrowth of distinct bacterial clusters. At any rate, the abundance pattern of certain taxa rather than individual strains shows characteristic features of potential diagnostic value. Stringent bioinformatics treatment of the sequencing data is mandatory to avoid unintentional misinterpretations.

## Background

Gingivitis is a non-destructive, reversible periodontal disease highly prevalent in various human populations [1,2]. In most cases, gingivitis develops as an inflammatory host response elicited by the accumulation of bacterial biofilm, called dental plaque, on the tooth surface [2–5]. Reversibility is an important feature of plaque-induced gingivitis. The alterations of the gingival tissue and the changes in inflammatory mediator levels usually disappeared after removal of the supragingival biofilm in experimental gingivitis models [6,7]. In susceptible individuals, however, persistent gingivitis can lead to periodontitis, a disease associated with irreversible tissue destruction and tooth loss [8–10].

Bacterial biofilm communities are widely considered to play a role in gingival inflammation, because differentially distributed bacterial taxa have been identified in supragingival plaques collected from adults with experimental or naturally occurring gingivitis and healthy individuals, respectively [11–15]. Similarly, changes in the relative abundance of various bacteria were observed in subgingival plaque samples from gingivitis-affected teeth versus non affected teeth [16–18].

Although local plaque accumulation plays a decisive role in the induction of gingivitis, systemic factors may also modulate the gingival inflammatory response [19]. In an experimental gingivitis model, two subpopulations of healthy adults could be distinguished based on the individual variability of the plaque-induced inflammatory response [20]. Thus, important differences in clinical parameters of gingival inflammation between „high responder” and „low responder” individuals were reported, which were characterized by similar amounts of dental plaque deposits [20,21]. It was concluded that metabolic, genetic and environmental factors may affect the bidirectional interactions between the gingival tissue and the bacterial biofilm [21]. Smoking suppressed gingival bleeding on probing, a clinical sign of gingival inflammation in a related study [22]. Other environmental factors may also contribute to the higher gingivitis susceptibility [23]. For example, salivary steroid hormones were also suggested as potential modulators of gingivitis prevalence, especially during adolescence [24].

The propensity to develop gingivitis increases gradually from early childhood to adult age [25,26]. Childhood gingivitis is typically characterized by gingival color change and swelling clinical symptoms, rather than bleeding [27]. Plaque-induced chronic gingivitis is highly prevalent among adolescents and even relative small amounts of plaque may elicit an inflammatory reaction in this age group (12-19 years) [19,28]. It was also observed that the chance of periodontitis development following gingival inflammation is lower in adolescents, compared to adults [29].

Permanent orthodontic treatments are frequently associated with increased presence of oral pathogens, e.g. *Porphyromonas gingivalis*, *Tannerella forsythia*, *Aggregatibacter actinomycetemcomitans*, *Prevotella intermedia*, *Treponema denticola*, which triggers immune response [30]. In a related study *P. intermedia*, *T. denticola*, *Fusobacterium nucleatum* and *Campylobacter rectus* were identified in increased abundance in gingival plaques after 3 months of placement of the braces although the general rearrangement of the core gingival microbiota apparently did not take place [31].

A significant fraction of adolescents and adults may undergo orthodontic treatment to correct crowded, rotated, buried or prominent front teeth [32]. The application of permanent orthodontic appliances (braces) may facilitate plaque accumulation by creating new retention sites, which are difficult to access during teeth cleaning [33,34]. It was also suggested that the surface of orthodontic appliances may favor the build-up of bacterial communities, which differ in composition from the microbiomes of dental plaques causing spontaneous gingivitis [33]. Overall, the published depictions of microbial communities associated with gingivitis are incongruent, at least partly due to methodological differences applied in the various laboratories. Therefore, the utilization of up to date sequencing and bioinformatics workflows are warranted to minimize systemic oversights.

For this reason, and because of the paucity of metagenomics data related to the supragingival biofilm communities of adolescents, we analyzed the microbiomes in supragingival plaque and saliva samples of 15-18 years old patients displaying symptoms of either orthodontic appliance induced or spontaneous gingivitis, and compared them to the microbial communities of age-matched healthy controls. The study included amplicon sequencing of the 16S rRNA gene of dental plaque biofilms and whole genome

sequencing of saliva samples of the same patients. The sequencing data were evaluated employing complex bioinformatics workflows. Both RDP and RefSeq reference databases were tested in the identification of microbiome community members.

## Methods

### Study design and recruitment of participants

The study protocol was approved by the Institutional Review Board of the University of Szeged, Szeged, Hungary. Signed informed consent was obtained from each adult participant enrolled into the study at the Department of Orthodontics and Pediatric Dentistry University of Szeged, Hungary. In case of study participants younger than 18 years old, signed informed consent was obtained from one of the parents. Subjects were divided into 3 study groups: Group A: induced gingivitis patients wearing fixed metal braces (9 patients, 3 males and 6 females, mean age 16.9 year, range: 15-18 years); Group B: spontaneous gingivitis patients diagnosed with plaque-induced gingivitis (10 patients, 7 males and 3 females, mean age 17 year, range: 16-18 years); Group C, healthy individuals without the symptoms of gingivitis (9 patients, 4 males and 5 females, mean age 17 year, range: 15-18 years). None of the participants suffered from known chronic systemic illness and none were treated with antibiotics at least 6 months prior to sampling. All subjects declared having no habitual smoking or drinking dependencies.

The patients' characteristics, including the Modified Gingival Indices (MGI) characteristic for the gingival condition [35,36], are summarized in Supplementary table 1. Fixed orthodontic appliances were placed on the labial tooth surfaces (more than 1.5 mm from the gingival margin) according to the professional rules with an acid-etch composite system (Transbond XT, 3M Unitek, Neuss, Germany), as in an earlier study [37].

### Dental examination and sample collection

Each participant underwent dental examination by a qualified dentist belonging to the staff of the Department of Orthodontics and Pediatric Dentistry. All patients were repeatedly instructed about proper oral hygiene during the orthodontic treatments. The gingival condition of each participant was assessed as described by Lobene et al., using the non-invasive Modified Gingival Index, MGI [35,36,38].

Unstimulated whole saliva samples were collected from the participants by the simple drooling method. The samples were aliquoted and stored at -80 C°. Supragingival plaque (biofilm) samples were taken using sterile paperpoints from the surface of 4 teeth with inflamed gum of each participant diagnosed with gingivitis (Groups A and B) as well as from the surface of the same number of teeth of each healthy controls without the symptoms of gingivitis (Group C). The supragingival plaque samples were stored in sterile plastic tubes (Axygen, MCT-150-C 1.5 mL) at -80 C° until DNA isolation.

# DNA isolation, sequencing and data analysis

## DNA isolation from saliva and supragingival plaque samples

Saliva samples were thawed and 3 mL of each was centrifuged at 13,000 rpm for 5 min. Supragingival plaque samples taken by sterile paperpoints were resuspended in 500  $\mu$ L TE buffer (10 mM Tris, 1 mM EDTA, pH 8) and were also pelleted at 13,000 rpm for 5 min. DNA extractions were carried out from both sample types by using the Macherey-Nagel NucleoSpin Soil DNA kit (Macherey-Nagel: 740780.250). The lysis mixture contained 700  $\mu$ L SL1 and 150  $\mu$ L Enhancer SX lysis solutions. After lysis (bead beating: maximum speed for 5 min), the kit protocol was followed. The quantity of DNA was determined in a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA) and a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, USA). DNA purity was tested by agarose gel electrophoresis and on an Agilent 2200 TapeStation instrument (Agilent Technologies, Santa Clara, USA).

## Next generation-sequencing of supragingival biofilm and saliva samples

Prokaryotic 16S rRNA gene amplification, purification and sequencing were performed as described in "Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System" standard protocol provided by the supplier (Illumina). Briefly: the hypervariable V3-V4 region of the 16S rRNA gene was PCR-amplified by using the forward primer

5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG; and the reverse primer 5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATC

The PCR products were cleaned up by using the Macherey-Nagel NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel: 740609.50) and checked with Agilent TapeStation 2200. Library preparation was done following the instructions of NEBNext Ultra II DNA Library Prep Kit for Illumina (Cat. Num.: E7645L). DNA sequencing was carried out on an Illumina MiSeq machine using V2 sequencing chemistry (MiSeq Reagent Kit v2) (500 cycles). *In vitro* fragment libraries were prepared from the saliva total DNA samples for the whole genome sequencing (shotgun metagenome sequencing) using the NEBNext Ultra II DNA Library Prep Kit by Illumina. Paired-end reads were generated on an Illumina NextSeq sequencer using TG NextSeq 500/550 High Output Kit v2 (300 cycles).

## Bioinformatics pipeline for amplicon sequence analysis of supragingival biofilm samples

A novel bioinformatics pipeline consisting of five modules was constructed to handle the amplicon sequencing data (Figure 1A). In the **Sequence preparation module**, fastq interleaver was used to join paired-end fastq reads from two separate files, one with the left mates and one with the right mates, into

a single file. The **Trimming module** employed Trimmomatic (v.0.36.5) to trimming of the raw sequences (sliding window: 4:20; minlen: 200; leading: 3; trailing: 3) [39]. Raw and processed sequence quality was checked by FastQC (v.0.11.8). Taxonomic classification of amplicon DNA reads was done in the **Taxonomic annotation module** by the sensitive and highly accurate Kraken 2 (v.2.0.8) program using the NCBI RefSeq (genome) and RDP (16S rDNA amplicon) databases [40]. In the **Filtration and normalization module**, both Kraken feature table outputs were filtered by Kraken 2 filter command (confidence threshold: 0.95 indicated ~95% precision to the lowest common ancestor). Normalization to the 16S rRNA gene copy numbers was done through the rrnDB (v.5.6) database [41]. Metagenomeseq (v.1.16.0) was used for cumulative sum scaling and to create normalized and scaled output of microbial abundance matrices [42]. The filtered and normalized microbial abundance matrix was converted to standard biom (JSON formatted, taxon table) format. In the **Statistics and visualization module** Megan6 (v.6.18.1) was used to scrutinize the microbial community and export data for statistic calculations [43]. Rarefaction estimation was performed by Megan6. The average composition of microbial taxa was visualized by the Krona (v.2.6.1) program [44]. The distribution of top 10 most frequent microbes between the three types of samples (A: induced gingivitis; B: spontaneous gingivitis; and C: control) was presented in Circos (v.0.63.9) [45]. For microbial core and diversity calculation MetaCoMET (Metagenomics Core Microbiome Exploration Tool), an interactive web tool, was used [46]. Shannon statistical method helped the estimation of alpha diversity. Emperor program (integrated to MetaCoMET) was employed for principal component analysis. Statistical Analysis of Metagenomic Profiles (STAMP v.2.0.9) computed the abundance differences in the case of whole microbiomes [47]. Dissimilar taxa were identified with two-sided t-test at 0.95 confidence intervals and the results with q-value (corrected p-value) of <0.05 were retained. Minimum difference between proportions was set to 0.3 in STAMP and Storey FDR (False Discovery Rate) was used in order to filter out false positive significant differences. UPGMA (Unweighted Pairwise Grouping Method with Arithmetic mean) with Bray-Curtis method was employed to cluster the samples (sample clusters: serious gingivitis (purple), moderate gingivitis (orange) and healthy (yellow) clusters. Significant differences and the UPGMA clustering of samples were visualized by interactive Tree of Life (iTOL v.5.3) online platform [48].

## **Bioinformatics pipeline of whole metagenome sequence analysis of saliva samples**

Two strategies were applied for the shotgun metagenome data analysis of saliva samples. Read-based metagenomics was combined with a genome centric approach. The workflow of the data analysis is summarized in Figure 1B.

### **Raw sequence filtering**

Galaxy Europe server was employed to pre-process the raw sequences (www.usegalaxy.eu). Low-quality reads were filtered by Prinseq (v.0.20.4), using the following parameters: min. length: 150; min. score: 15;

quality score threshold to trim positions: 20; sliding window used to calculate quality score: 1 [49]. The quality of raw and filtered sequences was checked with FastQC program.

## **Read-based metagenomics**

For taxonomic analysis of read-based metagenomics data the Kraken 2 (v.2.0.8) program was employed using the NCBI RefSeq genome database. The Kraken 2 feature table output were filtered by Kraken 2 filter command (confidence threshold: 0.8 indicated 80% precision to the lowest common ancestor).

### Statistical analysis of read-based metagenomics data

Microbial communities were investigated with MEGAN6 and data were exported for statistic calculations. The efficiency of sequencing was monitored by computing rarefaction curves based on the comparison of the reads with sequence data in RefSeq (NCBI Reference Sequence Database) as well as with rRNA sequences deposited in RDP (Ribosomal Database Project) database. Rarefaction estimation was performed by MEGAN6 [43] (data not shown). The average composition of microbial taxa was visualized via Krona [44]. The distribution of top 10 frequent microbes between the three types of samples was plotted in Circos.

The interactive web tool MetaCoMET (Metagenomics Core Microbiome Exploration Tool) calculated microbial core and diversity. Shannon statistical method was performed to estimate alpha diversity. The Emperor program (integrated in MetaCoMET) did the principal component analysis.

## **Genome-based evaluation of the sequencing data (Binning)**

Filtered sequences produced by Prinseq were co-assembled with Megahit (minimum contig length: 1000 bp, minimum k-mer size: 21, maximum k-mer size 141) [50]. After simplifying the header of contig FASTA file, using the Anvi'o script, Bowtie2 (v.2.3.4) was employed to map back the original sequences to the contigs [51]. Following this step, the Anvi'o (v.5.3 "Margaret") was used in the "metagenomics" workflow [52]. Briefly, during the first stage a contig database was generated, where open reading frames were identified by Prodigal (v.2.6.1) and each contig k-mer frequencies were computed. Then Hidden Markov Model (HMM) of single-copy genes were aligned by HMMER (v.3.0) [53–56]. InterProScan (v5.31-70) and the metagenome classifier Kaiju (v.16.0) was used for the functional and taxonomic annotation of contigs [57–60]. The outputs were imported into the contig database. BAM files made by Bowtie2 helped to profile the contig database, in this way we generated sample-specific information about the contigs, i.e. mean coverage. The sample-specific information was merged together. Three automated binning programs, i.e. CONCOCT (v.1.1.0), METABAT2 (v.2.12.1) and MAXBIN2 (v.2.2.7), were employed to reconstruct microbial genomes from the contigs (Minimum contig length: 2,000) [61–63]. The resulting picture was complemented by the Anvi'o human-guided binning and "anvi-refine" option [64]. All binning results were integrated into the contig database. The Anvi'o interactive interface allowed the visualization and summarization of the data.

## Results

# The adolescent supragingival biofilm microbiota (amplicon sequencing)

The supragingival bacterial communities in 112 supragingival biofilm samples were determined using next generation amplicon sequencing, in which a short, variable region of the 16S rRNA coding marker gene was amplified by PCR before sequencing. 36 samples derived from 9 adolescent patients diagnosed with induced gingivitis, 40 samples from 10 adolescent patients diagnosed with spontaneous gingivitis, and 36 samples from 9 age-matched healthy individuals without the symptoms of gingivitis were included in this study.

## Sequencing depth

In all cases, rarefaction curves reached their asymptotes (data not shown). This corroborated a sufficient sequencing depth to cover all genera in the bacterial communities. In other words, more than enough sequences were obtained from each sample to extract the total taxonomic information of the microbial community in them. Alignment of the quality-filtered reads with sequences in the NCBI RefSeq database allowed the identification of 172 operational taxonomic units (OTUs) in the samples analyzed (Supplementary table 2).

## The core microbial community

The richness or alpha diversity of OTUs was lower in the supragingival bacterial communities of induced and spontaneous gingivitis patients relative to healthy controls, but the difference was not significant (data not shown). Comparison of the microbial communities using PCA (principal component analysis) also showed a general overlap between supragingival plaque samples from gingivitis patients and healthy controls (data not shown).

Slight differences were noted depending on the reference databases employed. Distinct differences were not evident between the study groups at higher taxonomy levels, i.e. genus and beyond. Characteristic alterations took place among the low-abundance members of the community, however, only very few sequence reads could be associated with those taxa, which made statistical comparisons uncertain and likely led to erroneous conclusions. Therefore, the low-abundant species and genera (i.e.: relative abundance <0.01%) were excluded from the subsequent analyses. Annotation of the genera by RDP (Ribosome Database Project) revealed 23 abundant genera (relative abundance  $\geq$ 1%) and 17 rare genera (relative abundance <1%), while using the RefSeq database, 20 abundant species (relative abundance  $\geq$ 1%) and 21 rare species (relative abundance <1%) were identified (Supplementary table 3).

The general similarity of the OTUs in the three study groups allowed the amalgamation of all data and determination of the global microbiota of adolescent supragingival plaque (Figure 2A). *Veillonella parvula* was the most abundant bacterial species in all samples. At genus level, *Prevotella*, *Veillonella*, *Actinomyces*, *Campytophaga* and *Streptococcus* predominated in the supragingival plaques (Figure 2A).

## The most abundant bacterial species in the supragingival biofilms of the distinct study groups

The identified 10 most abundant bacterial species were confirmed by both RDF and RefSeq databases and represented 75 % of the whole community. The following bacterial species were the most prevalent ones in our study (Figure 2, Supplementary table 3): *Veillonella parvula* (Phylum: *Firmicutes*; Genus: *Veillonella*), *Fusobacterium nucleatum* (Phylum: *Fusobacteria*; Genus: *Fusobacterium*), *Rothia dentocariosa* (Phylum: *Actinobacteria*; Genus: *Rothia*), *Haemophilus parainfluenzae* (Phylum: *Proteobacteria*; Genus: *Haemophilus*), *Campylobacter gracilis* (Phylum: *Proteobacteria*; Genus: *Campylobacter*), *Streptococcus sanguinis* (Phylum: *Firmicutes*; Genus: *Streptococcus*), *Campylobacter concisus* (Phylum: *Proteobacteria*; Genus: *Campylobacter*), *Veillonella dispar* (Phylum: *Firmicutes*; Genus: *Veillonella*), *Prevotella oris* (Phylum: *Bacteroidetes*; Genus: *Prevotella*), *Prevotella intermedia* (Phylum: *Bacteroidetes*; Genus: *Prevotella*).

Not surprisingly, the most ubiquitous species belonged in the predominant phyla of the human oral microbiota, i.e. *Firmicutes*, *Fusobacteria*, *Actinobacteria*, *Proteobacteria* and *Bacteroidetes* [65].

## Pair-wise comparison of study groups

Although the composition of the microbial communities in the supragingival biofilm of patients diagnosed with the two types of gingivitis and healthy controls were similar, the ranking order of predominant bacterial species was different in each group (Figure 2B). *V. parvula* dominated the microbiota in induced gingivitis patients, with a relative abundance higher than 57%, followed by *C. gracilis* and species in lower abundance, i.e. *S. sanguinis*, *H. parainfluenzae*, *C. concisus* and *F. nucleatum*. In contrast, in the supragingival biofilm of patients with spontaneous gingivitis, the predominant *V. parvula* was followed by *F. nucleatum* (relative abundance higher than 21%), *P. intermedia* and *C. gracilis*. In healthy controls, the dominant *V. parvula* was accompanied by *R. dentocariosa* and *H. parainfluenzae* and the moderately abundant species *F. nucleatum* and *S. sanguinis* (Figure 2B).

In order to uncover the characteristic and significant differences between the three subject groups, the data sets were compared pairwise (Figure 3). Perhaps the most pronounced of these was the alterations between the diseased and healthy patients. We observed an increased relative abundance of the genera *Fusobacterium*, *Accermansia*, *Treponema* and *Campylobacter* in supragingival plaques of gingivitis patients *versus* controls. In contrast, the genera *Lautropia*, *Kingella*, *Neisseria*, *Actinomyces* and *Rothia*

were substantially more abundant in controls than in either of the two groups of gingivitis patients (Figure 3A). The genus *Megasphaera* showed notable relative distribution changes between the control subjects and induced and spontaneous gingivitis patients. In addition, relative abundances between the two groups of gingivitis patients were also noticeable, which might deserve further studies on large cohorts of subjects.

At the species level, a significantly higher abundance of *C. concisus* was apparent in both gingivitis groups *versus* the controls (Figure 3B). We also observed that the relative abundance of *Candidatus Saccharibacteria oral taxon TM7x*, *R. dentocariosa*, *R. mucilaginoso*, *Lautropia mirabilis* and *H. parainfluenzae* was lower in supragingival biofilms of either gingivitis patients *versus* healthy controls. Interestingly, *Fusobacterium periodonticum* was detected in healthy samples only. Comparison of bacterial species in the two groups of gingivitis patients showed that the relative abundances of *Candidatus Saccharibacteria oral taxon TM7x*, *R. dentocariosa* and *H. parainfluenzae* were significantly higher in patients with induced *versus* spontaneous gingivitis. Other species, including *P. intermedia*, *F. nucleatum*, *Parvimonas micra*, *Dialister pneumosintes*, *C. concisus*, *C. curvus* and *Aggregatibacter segnis* were less abundant in induced gingivitis *versus* spontaneous gingivitis group (Figure 3B). Out of the 13 significantly different species seven were positively confirmed by both RDP (on the taxonomic level of genus) and RefSeq (on the taxonomic level of specie) databases, marked with stars in Figure 3B. These patterns may be useful for diagnostic and future targeted therapy point of view.

## Pathogens in the gingivitis and healthy study groups

Similar to principal component analysis and alpha diversity, clustering of supragingival microbiota using UPGMA (unweighted pair group method with arithmetic mean) did not result in a clear separation the three study groups. Samples from induced gingivitis patients (A), spontaneous gingivitis patients (B) and controls (C) appeared intermingled with each other on the UPGMA tree (see the letters preceding the tooth position numbers in the innermost circle in Figure 4). The samples did not separate into distinct clusters according to the Modified Gingival Index (MGI) reflecting the severity of gingivitis either.

Nevertheless, a different clustering was apparent according to the pathogenic nature of the microbial taxa. Oral microbes have been classified on the basis of their roles in pathogenesis and have been arranged in color coded segments of the “Socransky pyramid” [66,67] (Supplementary figure 1). Species belonging in the purple Socransky complex (mainly *V. parvula*) are indicated by purple ID numbers, i.e. study group letter and tooth position numbers (Figure 4, innermost circle). This sample cluster comprised 52 of the 112 supragingival microbiota included in this study. *F. nucleatum* (25 out of 36 microbiota) predominated in the cluster highlighted in orange, together with other members of the Socransky orange complex [66]. The third cluster, highlighted in yellow, showed a balanced distribution of prevailing species, such as *R. dentocariosa*, *H. parainfluenzae* and *V. dispar*. 12 out of 25 samples were assigned to this cluster (Figure 4).

It is important to note, that no correlation between the pathogenicity of the bacterial members of the supragingival plaques and the health status of the patients (gingivitis or health) could be recognized. We have also investigated the potential relationship between oral health status and the patients' gender, age and orthodontic appliance wearing duration. Although, there was no significant correlation between any of these parameters, a tendency indicating better status of the female patients relative to the male ones and shorter duration of wearing braces versus long exposure was noted. These suggest that there are no substantial differences among the study groups in their general oral health status, hence gingivitis (induced or spontaneous) can be reversed with proper oral hygiene.

## Salivary microbiota of adolescent subjects (whole metagenome sequencing)

In addition to the dental plaques, saliva samples were also collected from the same study participants. In these experiments we wanted to examine the importance of sampling site, i.e. tooth biofilm associated and planktonic microbiota, of gingivitis patients relative to the microbiota of healthy controls belonging in the same age group. Saliva samples contained enough DNA to allow whole metagenome sequencing, thereby we could eliminate the potential random systematic error implied in amplicon sequencing [68]. The analysis and evaluation of sequence data followed a workflow (Figure 1B) similar to the one applied to the amplicon sequencing, although in this case both read-based and genome-based analyses became possible, which extended the information content considerably.

## Read-based analysis of the whole metagenome data

### The most abundant bacteria in the saliva of study groups

The richness and evenness of the salivary microbial communities did not differ significantly between either gingivitis study groups and healthy controls according to the Shannon indices (data not shown). Principal component analysis (PCA) of the salivary microbiota did not reveal distinct clusters either (data not shown). In this respect the whole salivary microbiota exhibited the same pattern as those found in the dental plaque biofilms with amplicon sequencing. Therefore, we could combine all saliva microbiota to generate a global picture of the adolescent saliva microbial landscape (Figure 5A).

The genera *Prevotella*, *Rothia*, *Streptococcus* and *Veillonella* predominated the salivary microbiota. It is noteworthy that *Veillonella* was not a principal genus in saliva, as in the cases of the supragingival biofilm samples (see Figure 2A and 5A).

The relative distribution of the 10 most abundant bacterial genera were determined next. In induced gingivitis, the relative abundance of *Prevotella* was higher than in the two other groups. *Streptococcus* was the most abundant genus in spontaneous gingivitis (Figure 5B). In control saliva samples, the

relative abundance of *Rothia* was comparable with that of *Prevotella*. In each study group, 14 genera accounted for >95% of the taxa identified at the genus level.

## Genome-based evaluation of the saliva sequencing data (binning)

In addition to the read-based data, bioinformatics analysis of saliva samples was accomplished by genome-centric binning (Figure 1B). In this approach, the filtered sequences were first assembled in contigs, which were then distributed into virtual bins, based on their inherent sequence features. Inspection of the genetic content of the individual bins supplied detailed information about taxonomy from a viewpoint distinct of the read-based approach and in many cases about the genes coding for possible specific metabolic pathways.

Genomic fragments belonging in the most abundant taxa of the healthy salivary microbiome [69] could be detected in the saliva of each study group. In line with the results of read-based metagenomics, most of the putative genomes identified by binning belonged to *Prevotella* species, which comprised 8 separate bins, whereas the putative genomes of species belonging in the genera *Veillonella* and *Streptococcus* comprised 4 separate bins, respectively. The genomes of *Actinomyces* and *Rothia* species were detected in 2 distinct bins, each. The genomes of 9 other species as well as the family *Porphyromonadaceae* and the taxon *Candidatus Saccharibacteria TM7x* occupied a single distinct bin (Figure 6). All these genera, were present in both supragingival biofilms and planktonic saliva samples.

The overall similarity of the read-based and genome-based microbiota validated each other, starting from the same saliva sequencing databases the two distinct bioinformatics approaches gave comparable pictures of the microbial communities.

## Discussion

Gingivitis is a non-destructive disease affecting both young and elderly worldwide [19,70]. Typically, the causative agent is the dental plaque and removal of the supragingival biofilm from the surface can revert the inflammatory response elicited by the biofilm-forming bacterial community [3,6,71,72]. When proper oral hygiene is not restored, gingival inflammation may persist and in susceptible individuals it can develop to periodontitis, a destructive inflammatory disease characterized with irreversible changes including loss of the tissues that attach the tooth to the alveolar structures, and alveolar bone loss [8–10]. Metagenomic analysis of the bacterial taxa in dental plaque samples collected from gingivitis-affected teeth and non-affected teeth biofilms revealed differences in the composition and community structure [11–18]. Most of the earlier sequencing studies analyzed pooled plaque samples of adult gingivitis patients or healthy adults. These observations suggested that the relative abundance of various bacteria in supragingival as well as subgingival biofilms may play a role in the progress of gingival inflammation [73].

We applied next-generation sequencing to characterize the bacterial communities in individual supragingival biofilms of adolescent patients diagnosed with either induced or spontaneous gingivitis.

Similar samples were taken from the teeth of age-matched healthy controls. Four distinct supragingival biofilm samples were collected from the surface of four different teeth of each study participant and were analyzed one by one. Furthermore, in order to reveal whether the characteristic features of the supragingival microbes were reflected in the salivary microbiota of the participants, we also identified the predominant bacterial taxa in the non-stimulated saliva samples of the same subjects.

The overall composition and diversity of supragingival biofilm communities was similar in the three study groups (Figures 2A and 5A), although the ranking order of predominant bacterial species was unique for each group (Figures 2B and 5B). *Veillonella parvula*, an early colonizer of the tooth surface predominated the supragingival biofilm microbiota in all three study groups (Figure 2A), whereas *Prevotella* species were the most prevalent in the saliva samples (Figure 5A), indicating their preference for planktonic lifestyle. *V. parvula* was abundant in supragingival plaques in an experimental gingivitis study, and it was suggested that *via* co-aggregation with other salivary bacteria, they played an important role in the formation and growth of multispecies microbial communities [7,74]. In other investigations *Veillonella* was found to be one of the most abundant genera both in supragingival and subgingival plaques of healthy adults [69].

One may envisage gingivitis and periodontitis as successive stages of an inflammatory cascade [75]. In this respect it is noteworthy that *Veillonella* species were implicated in the pathogenesis of periodontitis as members of the so called Socransky "purple complex" [66,67] (Supplementary figure 1). Although the Socransky complexes were denoted originally in subgingival plaque samples of periodontitis patients, the presence or absence of these bacterial complexes is also indicative of gingival health [67]. Early colonizing bacteria on the supragingival tooth surfaces are members of the "yellow", "green" and "purple" complexes, whereas members of the "orange" complex integrate into to the supragingival biofilm later [66,67,76]. Due to its elongated shape and diverse metabolic interactions *Fusobacterium nucleatum*, a member of the "orange" complex, contributes to the organization of supragingival biofilms [77]. Our observations are compatible with the previously suggested role of *Fusobacterium*, *Campylobacter* and *Treponema* in the pathogenesis of gingivitis, whereas the less abundant *Catonella*, *Lachoaerobaculum*, *Schwartzia* and *Akkermansia* genera remain to be characterized in this respect [78,79]. Some genera, including *Lautropia*, *Kingella*, *Neisseria*, *Actinomyces*, *Rothia* and *Megasphaera* were extensively more abundant in the control samples. These genera have been associated with *gingival health* [15,34,67,80–83]. Comparison of relative abundances of bacterial species revealed a substantially higher abundance of *C. concisus* in the two gingivitis groups *versus* controls (Figure 3). *C. concisus*, a member of the "green" complex, is ubiquitous in the oral cavity, it has been found both in periodontal inflammation and periodontal health associated microbiomes [67]. Species belonging in the "orange" complex, such as *P. intermedia*, *F. nucleatum*, *P. micra*, and "purple" complex (*A. segnis*) as well as the new periodontopathogenic species *D. pneumosintes* [84,85] (shown in the grey box in Supplementary figure 1) were more abundant in supragingival biofilms of the non-induced gingivitis group, but not in the induced gingivitis group (Figure 3) [66,67,76,85].

Comparisons at species level did not expose conceivable differences in the relative abundances of the Gram-negative anaerobic periodontal pathogens belonging in the “red” complex (*Porphyromonas gingivalis*, *Treponema denticola* and *Tannerella forsythia*), although the relative abundance of the genus *Treponema* was higher in gingivitis samples (Figure 3).

In contrast, the bacterial species *Candidatus Saccharibacteria oral taxon TM7x*, *R. dentocariosa*, *R. mucilaginosa*, *L. mirabilis* and *H parainfluenzae* were detected in higher relative occurrence in supragingival plaques of the control group, relative to the gingivitis patients (Figure 3). *Candidatus Saccharibacteria oral taxon TM7x* was implicated in periodontal disease, while the increased abundance of the other species in control samples is compatible with their classification as “Associated with Periodontal Health” [15,67]. It is noteworthy, that *Fusobacterium periodonticum* was identified in the control samples only although *F. nucleatum* occurred in all three study groups.

The development of gingivitis in patients undergoing orthodontic therapy has been linked to plaque accumulation at the new retention sites, which are difficult to access for oral hygiene [33,34]. The composition of the bacterial communities in induced gingivitis was also reported distinct from the microbiomes in naturally occurring gingivitis [33]. Our results, based on more advanced bioinformatics analysis, did not corroborate these findings (Figure 4). In spite of the overall similarity, differences were detected at the species level resolution: *Candidatus Saccharibacteria oral taxon TM7x*, and health-associated species *R. dentocariosa* and *H. parainfluenzae* were markedly higher in patients with induced versus non-induced gingivitis (Figure 3B), while *F. nucleatum*, *P. intermedia*, *P. micra*, *C. concisus*, *C. curvus* and *A. segnis* showed opposite tendencies (Figure 3B).

Mapping the salivary bacterial community may be an easy and helpful diagnostic tool of distinct pathological processes [86]. Nevertheless, under the employed stringent evaluation conditions we did not observe major differences between the salivary microbiome of patients with gingivitis compared to controls.

This finding may be related to the salivary microbiome comprising the easily detachable mixture of diverse bacterial communities inhabiting the surface of the tongue, tonsils, and throat as well as the microbes of the supragingival plaque [69].

## Conclusions

Our study indicated that the relative abundance of distinct bacterial taxa in supragingival biofilms may differ noticeably, although not extensively, between induced and non-induced gingivitis patients in spite of the global similarities.

The higher relative abundance of certain bacterial taxa in supragingival biofilms of gingivitis patients *versus* controls signal their involvement in the pathogenesis of the disease and may thus be of diagnostic value to prevent further escalation towards periodontitis.

Mapping salivary microbiome may obscure the variations between healthy and gingivitis status.

At any rate, a stringent and coherent bioinformatics workflow should be employed to detect the relatively small alterations in microbiome composition and taxonomic abundances.

## List Of Abbreviations

MGI – Modified Gingival Index

PCA- Principal Component Analysis

RDP – Ribosomal Database Project

RefSeq – NCBI Reference Sequence database

TE – Tris-EDTA buffer

UPGMA- Unweighted Pairwise Grouping Method

## Declarations

## Ethics approval and consent to participate

The study protocol was approved by the Institutional Review Board of the University of Szeged, Szeged, Hungary. Signed informed consent was obtained from each adult participant enrolled into the study at the Department of Orthodontics and Pediatric Dentistry University of Szeged, Hungary. In case of study participants younger than 18 years old, signed informed consent was obtained from one of the parents.

## Consent for publication

Not applicable.

## Availability of data and material

Raw sequences are available on NCBI Sequence Read Archive (SRA) under the submission number: PRJNA650272.

## Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# Funding

The support of the grant GINOP-2.3.2-15-2016-00011 by the European Regional Development Fund to a project led by JM is kindly acknowledged. Additional projects GINOP-2.2.1-15-2017-00081, GINOP-2.2.1-15-2017-00033, and EFOP-3.6.2-16-2017-00010 also contributed to this work. The grants were managed by the European Union Economic Development and Innovation Operational Programme, National Research, Development and Innovation Office (NKFIH), Hungary. RW and GM received support from the Hungarian schemes (PD132145). This work was also supported by the János Bolyai Research Scholarship (for GM) of the Hungarian Academy of Sciences.

# Authors' contributions

KLK, MM, and JM conceived and designed the study. RW, GM, BP performed the metagenomic studies, LL, MM, AAA participated in the sample collection and results evaluation. RW, KLK and MJ compiled the manuscript drafts. All authors read and approved the final manuscript.

# Acknowledgements

The support of the grant GINOP-2.3.2-15-2016-00011 by the European Regional Development Fund to a project led by JM is kindly acknowledged. Additional projects GINOP-2.2.1-15-2017-00081, GINOP-2.2.1-15-2017-00033, and EFOP-3.6.2-16-2017-00010 also contributed to this work. The grants were managed by the European Union Economic Development and Innovation Operational Programme, National Research, Development and Innovation Office (NKFIH), Hungary. RW and GM received support from the Hungarian schemes (PD132145). This work was also supported by the János Bolyai Research Scholarship (for GM) of the Hungarian Academy of Sciences.

# References

1. Califano J V. Position Paper: Periodontal diseases of children and adolescents. *J Periodontol.* 2003;74:1696–704.
2. Burt B. Position paper: Epidemiology of periodontal diseases. *J Periodontol.* 2005;76:1406–19.
3. Theilade E, Wright WH, Jensen SB, Løe H. Experimental gingivitis in man: II. A longitudinal clinical and bacteriological investigation. *J Periodontal Res.* 1966;1:1–13.
4. Armitage GC. Development of a classification system for periodontal diseases and conditions. *Ann Periodontol.* 1999;4:1–6.
5. Trombelli L, Farina R, Silva CO, Tatakis DN. Plaque-induced gingivitis: Case definition and diagnostic considerations. *J Clin Periodontol.* 2018;45:S44–67.
6. Offenbacher S, Barros S, Mendoza L, Mauriello S, Preisser J, Moss K, et al. Changes in gingival crevicular fluid inflammatory mediator levels during the induction and resolution of experimental

- gingivitis in humans. *J Clin Periodontol.* 2010;37:324–33.
7. Eberhard J, Grote K, Luchtefeld M, Heuer W, Schuett H, Divchev D, et al. Experimental gingivitis induces systemic inflammatory markers in young healthy individuals: A single-subject interventional study. *PLoS One.* 2013;8.
  8. Schätzle M, Loe H, Bürgin W, Anerund A, Boysen H, Lang NP. Clinical course of chronic periodontitis: I. role of gingivitis. *J Clin Periodontol.* 2003;30:909–18.
  9. Lang NP, Schätzle MA, Loe H. Gingivitis as a risk factor in periodontal disease. *J Clin Periodontol.* 2009;36:3–8.
  10. Könönen E, Gursoy M, Gursoy UK. Periodontitis: A multifaceted disease of tooth-supporting tissues. *J Clin Med.* 2019;8:1135.
  11. Kistler JO, Booth V, Bradshaw DJ, Wade WG. Bacterial community development in experimental gingivitis. *PLoS One.* 2013;8.
  12. Huang S, Yang F, Zeng X, Chen J, Li R, Wen T, et al. Preliminary characterization of the oral microbiota of Chinese adults with and without gingivitis. *BMC Oral Health.* 2011;11.
  13. Huang S, Li R, Zeng X, He T, Zhao H, Chang A, et al. Predictive modeling of gingivitis severity and susceptibility via oral microbiota. *ISME J. Nature Publishing Group;* 2014;8:1768–80.
  14. Shaw L, Harjunmaa U, Doyle R, Mulewa S, Charlie D, Maleta K, et al. Distinguishing the signals of gingivitis and periodontitis in supragingival plaque: A cross-sectional cohort study in Malawi. *Appl Environ Microbiol.* 2016;82:6057–67.
  15. Huang S, Li Z, He T, Bo C, Chang J, Li L, et al. Microbiota-based signature of gingivitis treatments: A randomized study. *Sci Rep. Nature Publishing Group;* 2016;6:1–9.
  16. Park OJ, Yi H, Jeon JH, Kang SS, Koo KT, Kum KY, et al. Pyrosequencing analysis of subgingival microbiota in distinct periodontal conditions. *J Dent Res.* 2015;94:921–7.
  17. Schincaglia GP, Hong BY, Rosania A, Barasz J, Thompson A, Sobue T, et al. Clinical, immune, and microbiome traits of gingivitis and peri-implant mucositis. *J Dent Res.* 2017;96:47–55.
  18. Deng K, Ouyang XY, Chu Y, Zhang Q. Subgingival microbiome of gingivitis in Chinese undergraduates. *Chin J Dent Res.* 2017;20:145–52.
  19. Murakami S, Mealey BL, Mariotti A, Chapple ILC. Dental plaque-induced gingival conditions. *J Clin Periodontol.* 2018;45:S17–27.
  20. Trombelli L, Tatakis DN, Scapoli C, Bottega S, Orlandini E, Tosi M. Modulation of clinical expression of plaque-induced gingivitis: II. Identification of “high-responder” and “low-responder” subjects. *J Clin Periodontol.* 2004;31:239–52.
  21. Tatakis DN, Trombelli L. Modulation of clinical expression of plaque-induced gingivitis: I. Background review and rationale. *J Clin Periodontol.* 2004;31:229–38.
  22. Dietrich T, Bernimoulin J-P, Glynn RJ. The effect of cigarette smoking on gingival bleeding. *J Periodontol.* 2004;75:16–22.

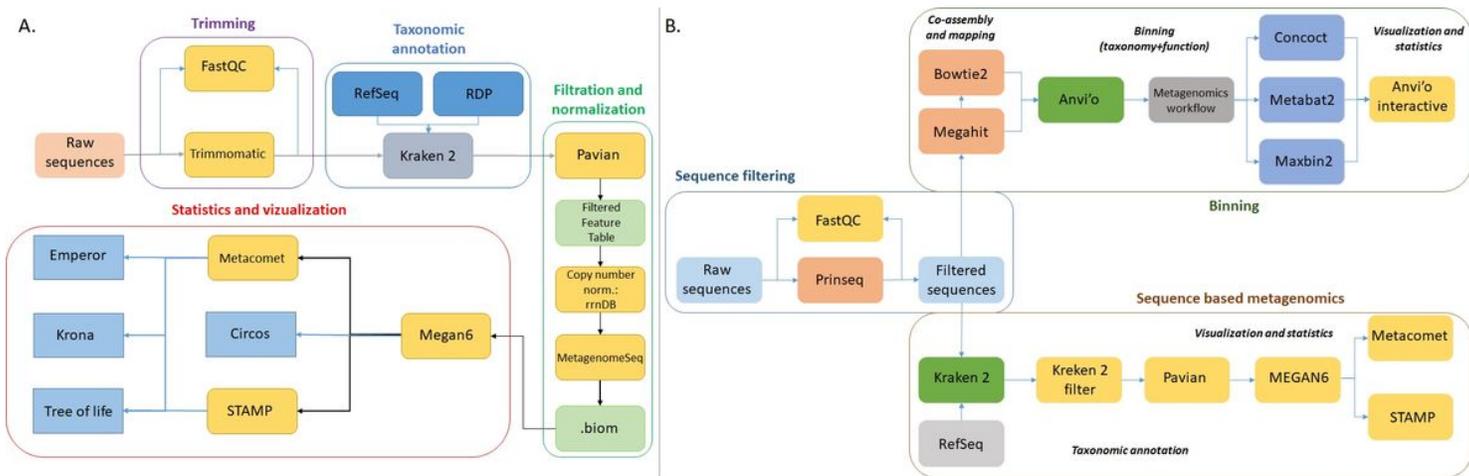
23. Dietrich T, Krall Kaye E, Nunn ME, Van Dyke T, Garcia RI. Gingivitis susceptibility and its relation to periodontitis in men. *J Dent Res.* 2006;85:1134–7.
24. Morishita M, Aoyama H, Tokumoto K, Iwamoto Y. The concentration of salivary steroid hormones and the prevalence of gingivitis at puberty. *Adv Dent Res.* 1988;2:397–400.
25. Matsson L. Development of gingivitis in pre-school children and young adults. *J Clin Periodontol.* 1978;5:24–34.
26. Marsh PD. Dental plaque: Biological significance of a biofilm and community life-style. *J Clin Periodontol.* 2005;32:7–15.
27. Matron L, Goldberg P. Gingival inflammatory reaction in children at different ages. *J. Clin. Periodontol.* 1985. p. 98–103.
28. Oh TJ, Eber R, Wang HL. Periodontal diseases in the child and adolescent. *J Clin Periodontol.* 2002;29:400–10.
29. Al-Ghutaimel H, Riba H, Al-Kahtani S, Al-Duhaimi S. Common periodontal diseases of children and adolescents. *Int J Dent.* 2014;2014:1–7.
30. Gong Y, Lu J, Ding X. Clinical, microbiologic, and immunologic factors of orthodontic treatment-induced gingival enlargement. *Am J Orthod Dentofac Orthop. American Association of Orthodontists;* 2011;140:58–64.
31. Guo R, Lin Y, Zheng Y, Li W. The microbial changes in subgingival plaques of orthodontic patients: A systematic review and meta-analysis of clinical trials. *BMC Oral Health. BMC Oral Health;* 2017;17:1–10.
32. Alhammadi MS, Halboub E, Fayed MS, Labib A, El-Saaidi C. Global distribution of malocclusion traits: A systematic review. *Dental Press J Orthod.* 2018;23:e1–10.
33. Ren Y, Jongsma MA, Mei L, van der Mei HC, Busscher HJ. Orthodontic treatment with fixed appliances and biofilm formation—a potential public health threat? *Clin Oral Investig.* 2014;18:1711–8.
34. Koopman JE, Van Der Kaaij NCW, Buijs MJ, Elyassi Y, Van Der Veen MH, Crielaard W, et al. The effect of fixed orthodontic appliances and fluoride mouthwash on the oral microbiome of adolescents - A randomized controlled clinical trial. *PLoS One.* 2015;10:1–17.
35. Lobene RR, Weatherford T, Ross NM, Lamm RA, Menaker L. A modified gingival index for use in clinical trials. *Clin Prev Dent. United States;* 1986;8:3–6.
36. Lobene RR, Mankodi SM, Ciancio SG, Lamm RA, Charles CH, Ross NM. Correlations among gingival indices: A methodology study. *J Periodontol.* 1989;60:159–62.
37. Lipták L, Szabó K, Nagy G, Márton S, Madléna M. Microbiological changes and Caries-Preventive effect of an innovative varnish containing chlorhexidine in orthodontic patients. *Caries Res.* 2018;52:272–8.
38. He T, Qu L, Chang J, Wang J. Gingivitis models - Relevant approaches to assess oral hygiene products. *J Clin Dent.* 2018;29:45–51.

39. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
40. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol. Genome Biology*; 2019;20:1–13.
41. Roller BRK, Stoddard SF, Schmidt TM. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nat Microbiol. Nature Publishing Group*; 2016;1:1–7.
42. Paulson JN, Stine OC, Bravo HC, Pop M. Robust methods for differential abundance analysis in marker gene surveys. *Nature methods*. 2013;10:1200–2.
43. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, et al. MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput Biol*. 2016;12:1–12.
44. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*. 2011;385:1–9.
45. Connors J, Krzywinski M, Schein J, Gascoyne R, Horsman D, Jones SJ, et al. Circos: An information aesthetic for comparative genomics. *Genome Res*. 2009;19:1639–45.
46. Wang Y, Xu L, Gu YQ, Coleman-Derr D. MetaCoMET: A web platform for discovery and visualization of the core microbiome. *Bioinformatics*. 2016;32:3469–70.
47. Parks DH, Beiko RG. Identifying biologically relevant differences between metagenomic communities. *Bioinformatics*. 2010;26:715–21.
48. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res. Oxford University Press*; 2019;47:256–9.
49. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27:863–4.
50. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31:1674–6.
51. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
52. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*. 2015;3:e1319.
53. Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, et al. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc Natl Acad Sci*. 2013;110:5540–5.
54. Finn DR, Clements J, Eddy RS. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39:W29–W37.
55. Rinke J, Schäfer V, Schmidt M, Ziermann J, Kohlmann A, Hochhaus A, et al. Genotyping of 25 leukemia-associated genes in a single work flow by next-generation sequencing technology with low amounts of input template DNA. *Clin Chem*. 2013;59:1238–50.

56. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V, Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–2.
57. Agarwala V, Khozin S, Singal G, O’Connell C, Kuk D, Li G, et al. Real-world evidence in support of precision medicine: Clinico-genomic cancer data as a case study. *Health Aff*. 2018;37:765–72.
58. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: The protein families database. *Nucleic Acids Res*. 2014;42:222–30.
59. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics*. 2014;30:1236–40.
60. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun*. Nature Publishing Group; 2016;7:1–9.
61. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. CONCOCT: Clustering cONTigs on COverage and ComposiTiON. *arXiv*. 2013;1–28.
62. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*. 2015;3:e1165.
63. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*. 2015;32:605–7.
64. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat Microbiol*. 2018;3:804–13.
65. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu WH, et al. The human oral microbiome. *J Bacteriol*. 2010;192:5002–17.
66. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL. Microbial complexes in subgingival plaque. *J Clin Periodontol*. 1998;25:134–44.
67. Colombo APV, Tanner ACR. The role of bacterial biofilms in dental caries and periodontal and peri-implant diseases: A historical perspective. *J Dent Res*. 2019;98:373–85.
68. Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun*. Elsevier Ltd; 2016;469:967–77.
69. Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol*. 2012;13.
70. Clerehugh V. Periodontal diseases in children and adolescents. *Br Dent J*. 2008;204:469–71.
71. Page RC, Schroeder HE. Pathogenesis of inflammatory periodontal disease. A summary of current work. *Lab Invest*. United States; 1976;34:235–49.
72. Michelet H, Fontaniere B, Magloire H, Durand BM, Tabone E. Extracellular matrix and intermediate filaments in the first stages and repair of experimental gingivitis in man. *J Biol Buccale*. 1991;19:329

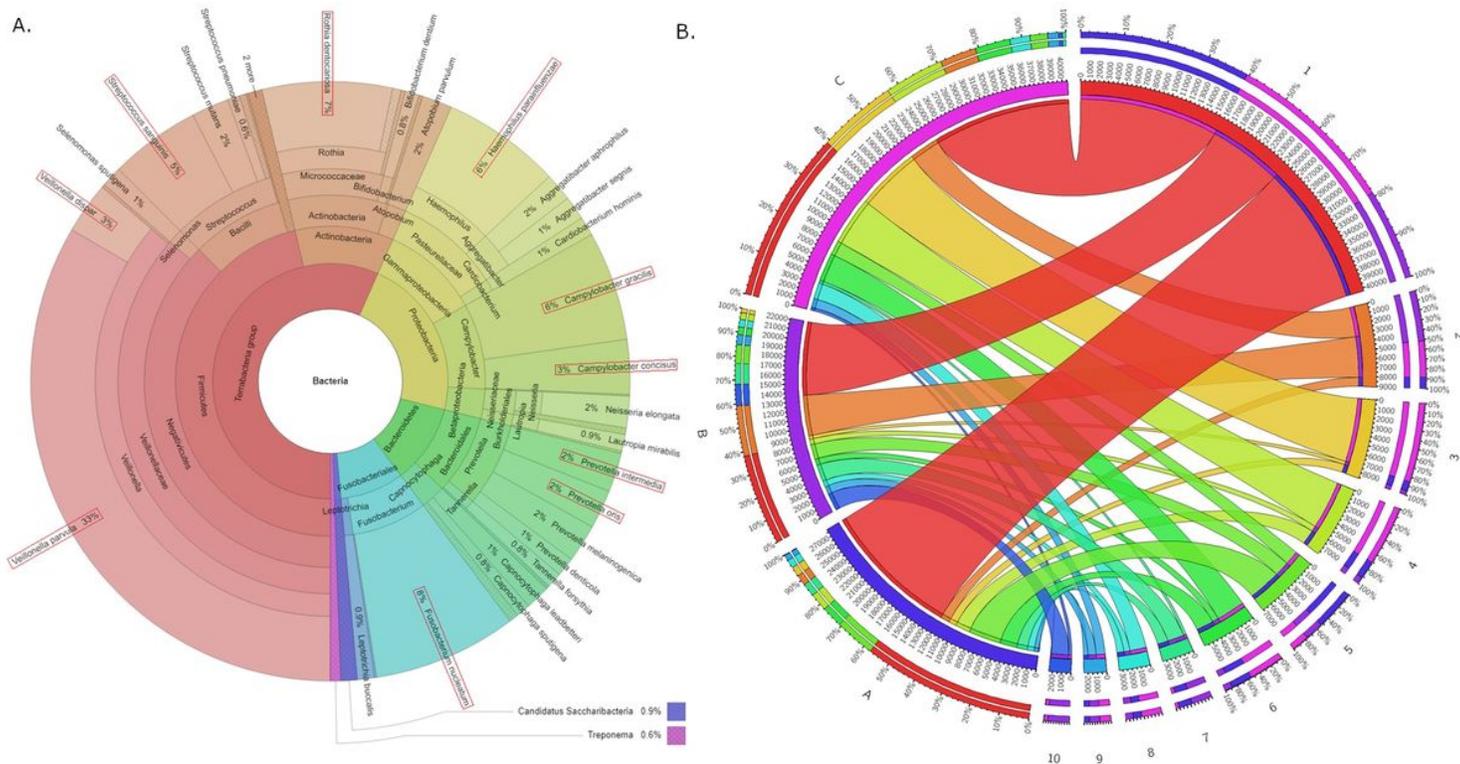
73. Bartold PM, Van Dyke TE. An appraisal of the role of specific bacteria in the initial pathogenesis of periodontitis. *J Clin Periodontol.* 2019;46:6–11.
74. Kolenbrander PE. Multispecies communities: Interspecies interactions influence growth on saliva as sole nutritional source. *Int J Oral Sci.* 2011;3:49–54.
75. Kinane DF, Attstrom R. Advances in the pathogenesis of periodontitis. Group B consensus report of the fifth European workshop in periodontology. *J Clin Periodontol.* 2005;32:130–1.
76. Carrouel F, Viennot S, Santamaria J, Veber P, Bourgeois D. Quantitative molecular detection of 19 major pathogens in the interdental biofilm of periodontally healthy young adults. *Front Microbiol.* 2016;7:1–16.
77. Brennan CA, Garrett WS. *Fusobacterium nucleatum* – symbiont, opportunist and oncobacterium. *Nat Rev Microbiol.* Springer US; 2019;17:156–66.
78. Sharma N, Bhatia S, Sodhi AS, Barta N. Oral microbiome and health. *Microbiology.* 2018;4:42–66.
79. Macuch PJ, Tanner ACR. *Campylobacter* species in health, gingivitis, and periodontitis. *J Dent Res.* 2000;79:785–92.
80. Mervish NA, Hu J, Hagan LA, Arora M, Frau C, Attaie A, et al. Associations of the oral microbiota with obesity and menarche in inner city girls. *J Child Obes.* 2019;4:1–20.
81. Grevich S, Lee P, Leroux B, Ringold S, Darveau R, Henstorf G, et al. Oral health and plaque microbial profile in juvenile idiopathic arthritis. *Pediatr Rheumatol.* 2019;17:1–10.
82. Al-Kamel A, Baraniya D, Al-Hajj WA, Halboub E, Abdulrab S, Chen T, et al. Subgingival microbiome of experimental gingivitis: shifts associated with the use of chlorhexidine and N-acetyl cysteine mouthwashes. *J Oral Microbiol.* Taylor & Francis; 2019;11.
83. Sanz M, Beighton D, Curtis MA, Cury JA, Dige I, Dommisch H, et al. Role of microbial biofilms in the maintenance of oral health and in the development of dental caries and periodontal diseases. Consensus report of group 1 of the Joint EFP/ORCA workshop on the boundaries between caries and periodontal disease. *J Clin Periodontol.* 2017;44:S5–11.
84. Ayala Herrera JL, Apreza Patrón L, Martínez Martínez RE, Domínguez Pérez RA, Abud Mendoza C, Hernández Castro B. *Filifactor alocis* and *Dialister pneumosintes* in a Mexican population affected by periodontitis and rheumatoid arthritis: An exploratory study. *Microbiol Immunol.* 2019;63:392–5.
85. Ferraro CTL, Gornic C, Barbosa AS, Peixoto RJM, Colombo APV. Detection of *Dialister pneumosintes* in the subgingival biofilm of subjects with periodontal disease. *Anaerobe.* 2007;13:244–8.
86. Zhang CZ, Cheng XQ, Li JY, Zhang P, Yi P, Xu X, et al. Saliva in the diagnosis of diseases. *Int J Oral Sci.* Nature Publishing Group; 2016;8:133–7.

## Figures



**Figure 1**

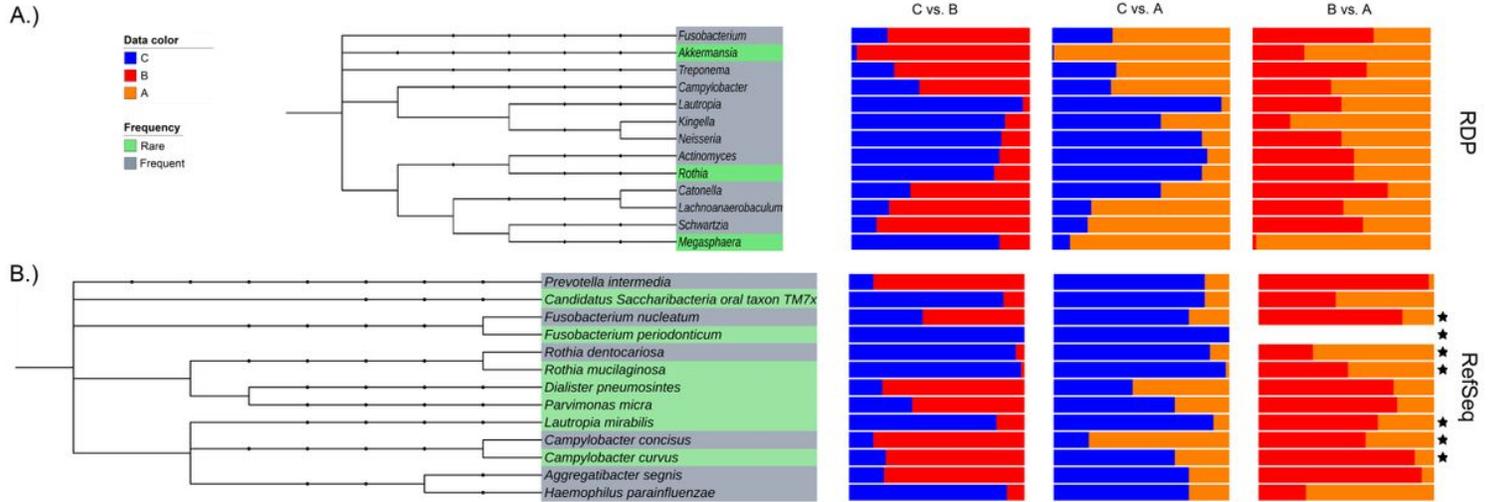
A: Bioinformatics pipeline for amplicon sequence analysis of supragingival biofilm samples. B: Bioinformatics pipeline for metagenome sequencing of saliva samples. For explanation see text.

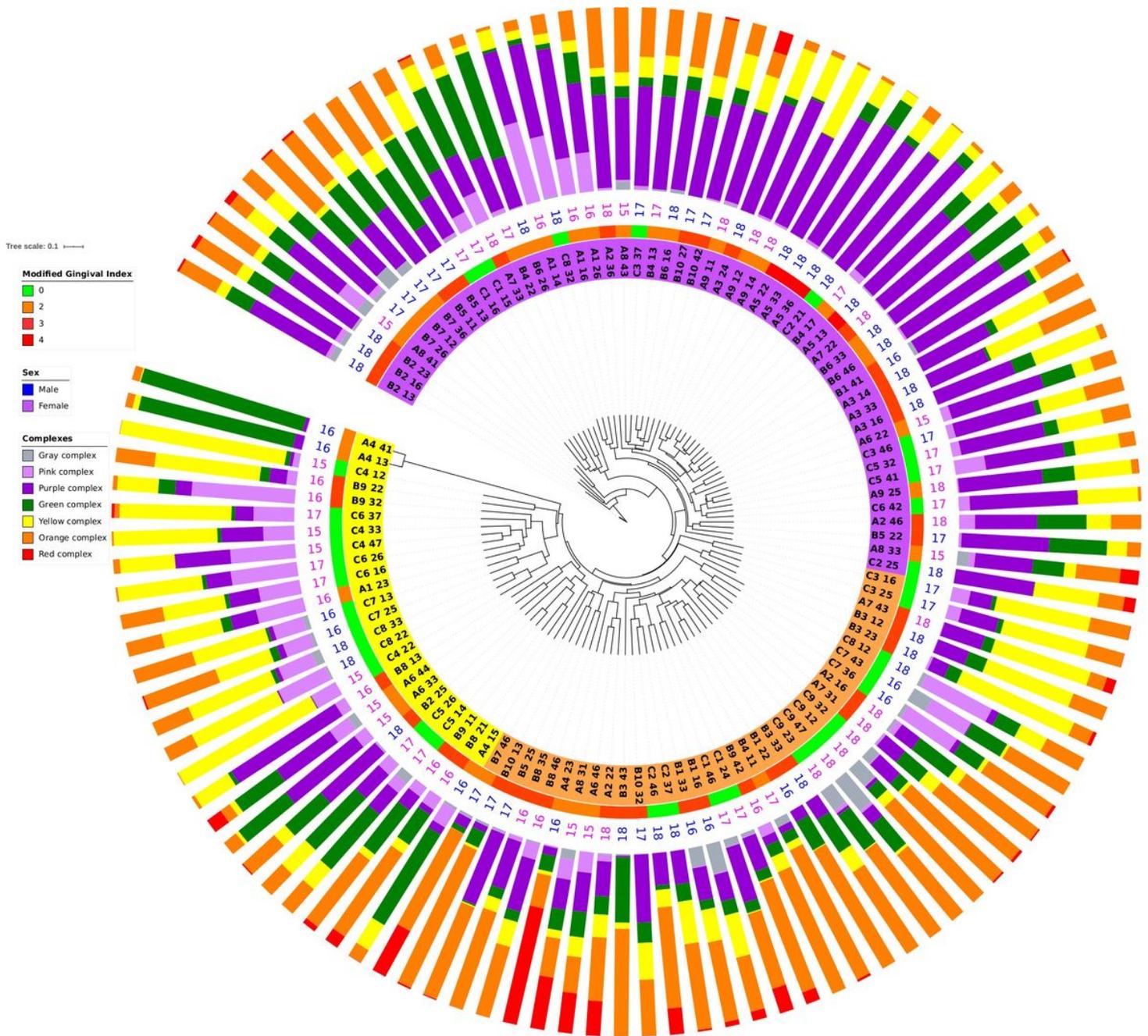


**Figure 2**

(A): Overall composition of the oral microbiota in supragingival biofilm samples derived from gingivitis patients and healthy controls involved in the study. The most abundant species are outlined in red boxes. OTUs were annotated based on the RefSeq NCBI Reference Sequence Database. (B): Relative distribution of the 10 most abundant bacterial species identified in supragingival biofilm samples of the study participants. Designations: 1. Veillonella parvula, 2. Fusobacterium nucleatum, 3. Rothia dentocariosa, 4.

Haemophilus parainfluenzae, 5. Campylobacter gracilis, 6. Streptococcus sanguinis, 7. Campylobacter concisus, 8. Veillonella dispar, 9. Prevotella oris, 10. Prevotella intermedia.

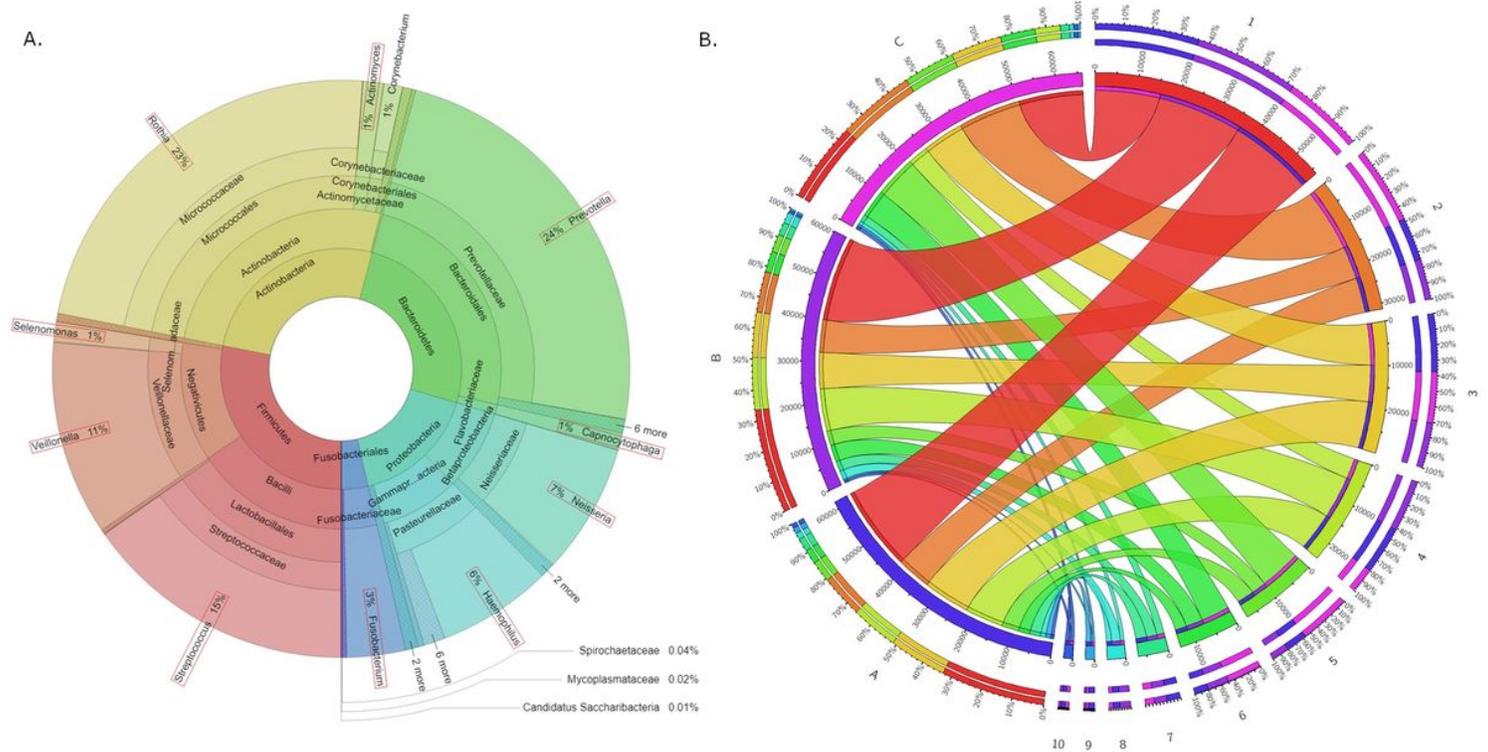




**Figure 4**

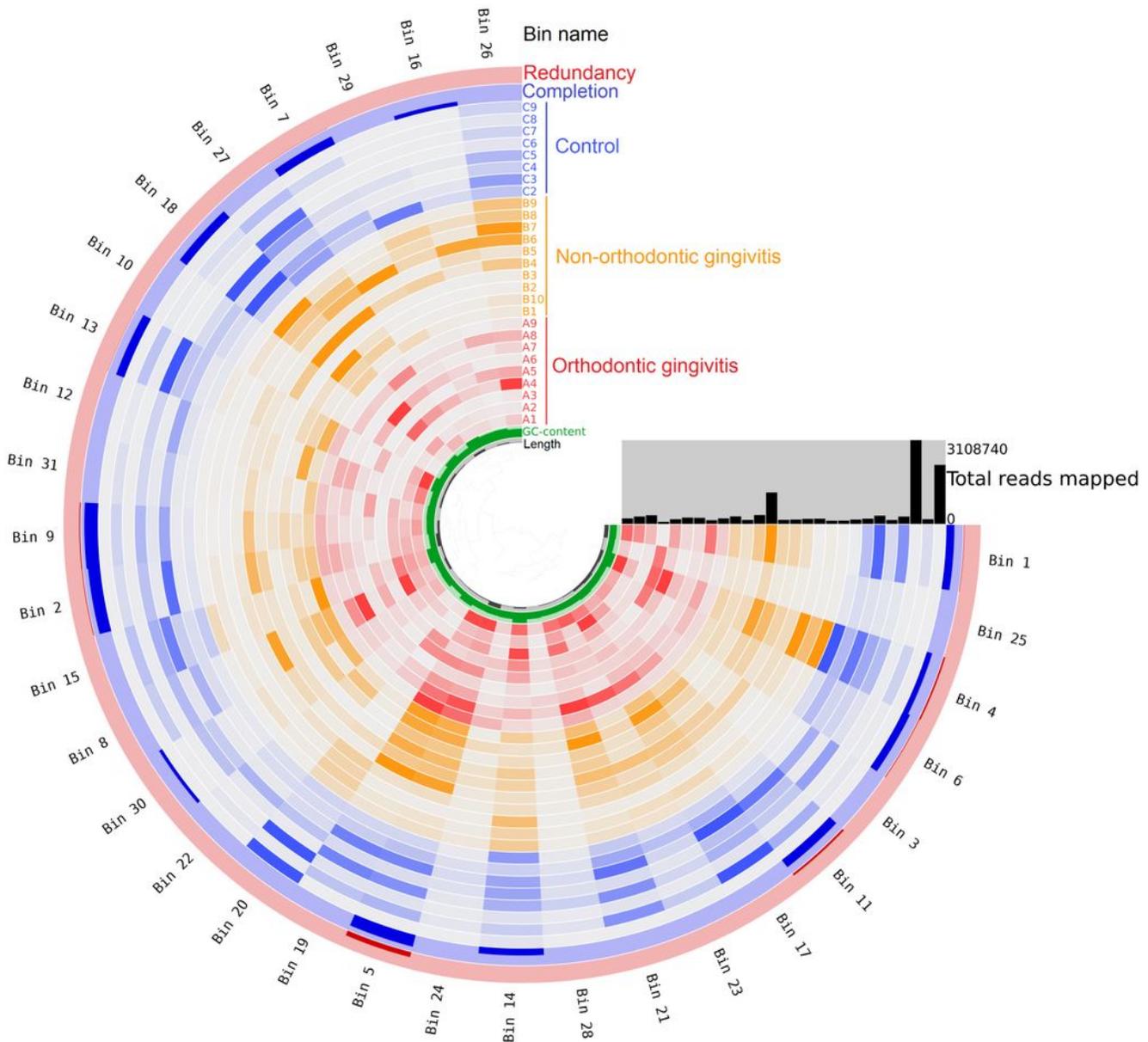
UPGMA (unweighted pair group method with arithmetic mean) cluster analysis of supragingival biofilm metagenomes. The study group designation (capital letters), the numbers marking the anonymous individual patients and sampled tooth position numbers are highlighted in purple, orange and yellow, according to their position on the major branches of the UPGMA tree (innermost circle). Study groups were as follows. (A): samples from induced gingivitis patients; (B): samples from spontaneous gingivitis patients; (C): healthy controls. Modified gingival indices (MGIs) [35,36] are indicated as red, orange and green sectors in the next circle. Predominant bacterial taxa identified in each supragingival biofilm sample are shown as colored columns at the outer edge of the circular diagram above numbers indicating the age of the patients. Color-coding of the bacterial groups was based on their association

with microbial complexes involved in oral pathogenesis, indicating their potential contribution to oral health [66,67,76] (Supplementary figure 1).



**Figure 5**

(A): Overall composition of saliva samples derived from gingivitis patients and healthy controls involved in the study. The most abundant species are outlined in red boxes. OTUs were annotated based on the RefSeq NCBI Reference Sequence Database. (B): Relative distribution of the 10 most abundant bacterial genera identified in the saliva samples of the study participants. Designations: 1. Prevotella, 2. Rothia, 3. Streptococcus, 4. Veillonella, 5. Neisseria, 6. Haemophilus, 7. Fusobacterium, 8. Selenomonas, 9, Capnocytophaga, 10, Actinomyces.



Bin 1: *Prevotella histicola*, Bin 2: *Prevotella pallens*, Bin 3: *Prevotella salivae*, Bin 4: *Unknown Rothia*, Bin 5: *Unknown Veillonella*, Bin 6: *Unknown Streptococcus*, Bin 7: *Lautropia mirabilis*, Bin 8: *Prevotella* sp. C561, Bin 9: *Megasphaera micronuciformis*, Bin 10: *Prevotella nanciensis*, Bin 11: *Unknown Prevotella*, Bin 12: *Prevotella* sp. Oral taxon 306, Bin 13: *Actinomyces graevenitzii*, Bin 14: *Unknown Actinobacteria*, Bin 15: *Actinomyces* sp. ICM47, Bin 16: *Corinebacterium matruchatii*, Bin 17: *Prevotella melaninogenica*, Bin 18: *Unknown Porphyromonadaceae*, Bin 19: *Veillonella atypica*, Bin 20: *Escherichia coli*, Bin 21: *Veillonella dispar*, Bin 22: *Fusobacterium periodonticum*, Bin 23: *Veillonella parvula*, Bin 24: *Streptococcus parasanguinis*, Bin 25: *Streptococcus salivarius*, Bin 26: *Streptococcus pneumoniae*, Bin 27: *Haemophilus parainfluenzae*, Bin 28: *Campylobacter concisus*, Bin 29: *Rothia dentocariosa*, Bin 30: *Candidatus Saccharobacteria oral taxon TM7x*, Bin 31: *Atopobium parvulum*

**Figure 6**

Bacterial taxa identified in saliva samples of induced and non-induced gingivitis patients and controls using genome-based evaluation of sequencing data. The distribution of contigs are plotted in the circular bands. The grouping of contigs was based on sequence-assignments of automated binning programs METABAT2, MAXBIN2 and CONCOCT as well as manually defined bins as presented on the Anvi'o platform. The list of identified bins is given at the bottom part of the figure.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1..docx](#)
- [SupplementaryTable1..docx](#)
- [SupplementaryTable1..docx](#)
- [Supplementarytable2.docx](#)
- [Supplementarytable2.docx](#)
- [Supplementarytable2.docx](#)
- [Supplementarytable3.docx](#)
- [Supplementarytable3.docx](#)
- [Supplementarytable3.docx](#)
- [Supplementfig1Complex.es.tif](#)
- [Supplementfig1Complex.es.tif](#)
- [Supplementfig1Complex.es.tif](#)