

High-resolution taxonomic profiling of the subgingival microbiome for biomarker discovery and periodontitis diagnosis

Szymon P. Szafranski, Melissa L. Wos-Oxley, Ramiro Vilchez-Vargas, Ruy Jáuregui, Iris Plumeier, Frank Klawonn, Jürgen Tomasch, Christa Meisinger, Jan Kühnisch, Helena Sztajer, Dietmar H. Pieper, Irene Wagner-Döbler

Angaben zur Veröffentlichung / Publication details:

Szafranski, Szymon P., Melissa L. Wos-Oxley, Ramiro Vilchez-Vargas, Ruy Jáuregui, Iris Plumeier, Frank Klawonn, Jürgen Tomasch, et al. 2015. "High-resolution taxonomic profiling of the subgingival microbiome for biomarker discovery and periodontitis diagnosis." *Applied and Environmental Microbiology* 81 (3): 1047–58.
<https://doi.org/10.1128/aem.03534-14>.



High-Resolution Taxonomic Profiling of the Subgingival Microbiome for Biomarker Discovery and Periodontitis Diagnosis

Szymon P. Szafranski,^a Melissa L. Wos-Oxley,^b Ramiro Vilchez-Vargas,^{b*} Ruy Jáuregui,^b Iris Plumeier,^b Frank Klawonn,^{c,d} Jürgen Tomasch,^a Christa Meisinger,^e Jan Kühnisch,^f Helena Sztajer,^a Dietmar H. Pieper,^b Irene Wagner-Döbler^a

Microbial Communication Research Group, Department of Molecular Infection Biology, Helmholtz Centre for Infection Research (HZI), Braunschweig, Germany^a; Microbial Interactions and Processes Research Group, Department of Medical Microbiology, Helmholtz Centre for Infection Research (HZI), Braunschweig, Germany^b; Bioinformatics and Statistics Research Group, Department of Cellular Proteomics, Helmholtz Centre for Infection Research (HZI), Braunschweig, Germany^c; Department of Computer Science, Ostfalia University of Applied Sciences, Wolfenbüttel, Germany^d; National Cohort Study Center Augsburg, Helmholtz Centre Munich, Germany^e; Department of Conservative Dentistry, Ludwig-Maximilians-University, Munich, Germany^f

The oral microbiome plays a key role for caries, periodontitis, and systemic diseases. A method for rapid, high-resolution, robust taxonomic profiling of subgingival bacterial communities for early detection of periodontitis biomarkers would therefore be a useful tool for individualized medicine. Here, we used Illumina sequencing of the V1-V2 and V5-V6 hypervariable regions of the 16S rRNA gene. A sample stratification pipeline was developed in a pilot study of 19 individuals, 9 of whom had been diagnosed with chronic periodontitis. Five hundred twenty-three operational taxonomic units (OTUs) were obtained from the V1-V2 region and 432 from the V5-V6 region. Key periodontal pathogens like *Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia* could be identified at the species level with both primer sets. Principal coordinate analysis identified two outliers that were consistently independent of the hypervariable region and method of DNA extraction used. The linear discriminant analysis (LDA) effect size algorithm (LEfSe) identified 80 OTU-level biomarkers of periodontitis and 17 of health. Health- and periodontitis-related clusters of OTUs were identified using a connectivity analysis, and the results confirmed previous studies with several thousands of samples. A machine learning algorithm was developed which was trained on all but one sample and then predicted the diagnosis of the left-out sample (jackknife method). Using a combination of the 10 best biomarkers, 15 of 17 samples were correctly diagnosed. Training the algorithm on time-resolved community profiles might provide a highly sensitive tool to detect the onset of periodontitis.

Chronic periodontitis affects roughly 750 million people, making it one of the most prevalent infections worldwide (1). Its outcome can ultimately be tooth loss, but it is additionally associated with several systemic conditions, such as atherosclerotic cardiovascular disease (2). Periodontitis is the result of dynamic interactions between a highly complex oral microbial community and the host immune system, and its etiology is not causally understood (3). The polymicrobial synergy and dysbiosis model (4) takes into account both the extensive research on individual periodontal pathogens and the recent next-generation sequencing results, which suggest that periodontitis is triggered by keystone pathogens like *Porphyromonas gingivalis* that are able to manipulate the immune response of the host. Hence, the host is no longer able to control the growth of the periodontal biofilm, tipping the balance from homeostasis to dysbiosis. Commensals can then become accessory pathogens that act synergistically with keystone pathogens and form a proinflammatory community that elicits a nonresolving and tissue-destructive host response (4).

Cultivation methods and various cultivation-independent molecular methods have all been used to study the composition of the polymicrobial periodontal community (5). Next-generation sequencing opens a new window which allows the species composition of microbial communities to be determined in unprecedented depth and has shown that the oral microbial community is the second most diverse one after the gut microbial community (6).

In the context of individualized medicine, it will be desirable to detect changes in the periodontal community that are related to host genetics, life style, and infections and to identify biomarkers for periodontitis at the earliest possible stage to initiate treatment

and prevent the disease from entering the chronic stage. The present study was part of the pretest 1 phase of the German National Cohort (GNC), which is conducted as an interdisciplinary, population-based longitudinal cohort study (7). In the GNC, a total of 200,000 individuals (aged 20 to 69) will be recruited, and their health will be monitored for a period of 30 years to study the development of chronic diseases (cardiovascular diseases, cancer, diabetes, and neurodegenerative, pulmonary, and infectious diseases). A subpopulation of 40,000 individuals will be subject to in-depth analysis protocols. The major aim of the pretest 1 phase

Citation Szafranski SP, Wos-Oxley ML, Vilchez-Vargas R, Jáuregui R, Plumeier I, Klawonn F, Tomasch J, Meisinger C, Kühnisch J, Sztajer H, Pieper DH, Wagner-Döbler I. 2015. High-resolution taxonomic profiling of the subgingival microbiome for biomarker discovery and periodontitis diagnosis. *Appl Environ Microbiol* 81:1047–1058. doi:10.1128/AEM.03534-14.

Address correspondence to Irene Wagner-Döbler, Irene.Wagner-Doebler@helmholtz-hzi.de.

* Present address: Ramiro Vilchez-Vargas, Laboratory of Microbial Ecology and Technology (LabMET), Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.03534-14>.

TABLE 1 Demographic and clinical characteristics of the individuals studied

Characteristic	Value for individuals with indicated oral status	
	Healthy (<i>n</i> = 10)	Periodontitis (<i>n</i> = 9)
Mean age ± SD (yr)	52.4 ± 12.3	56.9 ± 11.8
No. who were female (%)	60	66.7
PPPD ^a (mm)	2–3	4–10
Mean no. of teeth sampled ± SD	2.4 ± 0.5	2.4 ± 0.9
Diagnosis ^b	Periodontally healthy	Chronic periodontitis

^a PPPD, periodontal pocket probing depth.

^b Based on criteria defined in reference 15.

was to test the feasibility of methods that might be used later in the main study.

Here, we chose Illumina MiSeq 16S rRNA gene amplicon sequencing because it has lower costs than pyrosequencing and greater sequencing depth and lower error rates than other platforms (8). On the other hand, it produces short reads, making taxonomic assignment more difficult (9–13).

Since the various regions of the rRNA gene have different powers of taxonomic resolution, we compared community profiles obtained by independently amplifying the V1-V2 and V5-V6 hypervariable regions. We also determined the effect of successively more aggressive DNA extraction. Many statistical analyses can be applied to detect biomarkers. Here, we used the linear discriminant analysis (LDA) effect size algorithm (LEfSe) pipeline that has been developed for complex biological data sets. Finally, we developed a pipeline for diagnosis of periodontitis based on the composition of the microbial communities. For this goal, we used the 10 best biomarkers and applied the jackknife (leave-one-out) method. The entire pipeline was automated and can be applied to stratify periodontal microbial communities or microbial communities associated with other disease outcomes in a large set of samples, making it a useful approach for individualized medicine.

MATERIALS AND METHODS

Participants and ethics statement. The present study was designed in cooperation between the Infection and Dental Health working groups in the German National Cohort. The study protocol was approved by the local ethics committee (Bayerische Landesärztekammer, Munich, Germany), and written consent was obtained from all participants. All participants of this substudy were participants of the so-called pretest 1 study of the GNC, Study Center Augsburg, Germany. A random population sample was drawn by the local registration office in Augsburg, Bavaria (Germany) according to distinct selection criteria (age range of 20 to 69 years and main residency in Augsburg). Each participant was invited by regular mail and, if available, by telephone. In case of no response, a reminder letter was sent. Finally, 19 subjects followed the invitation to have a dental examination and microbiological analysis of the subgingival plaque. Demographic and clinical characteristics of the individuals studied are presented in Table 1.

Standardization of the clinical examination. Each participant was examined by a dentist in a medical praxis. Prior to the dental examination, all participants had the opportunity to brush their teeth. The examination equipment consisted of a dental mirror with a plane surface, a blunt community periodontal index (CPI) probe (CP-11.5B6; Hu-Friedy, Chicago, IL, USA), and a light-emitting-diode (LED) lamp (model 17.9; Rupprich, Görlitz, Germany). The dental examination included, first, registration of the dental status according to the decay-missing-filled (DMF) index (14), and second, measurement of the pocket probing depth at six points per

tooth in mm. Periodontitis was diagnosed in subjects with probing depths of >3.5 mm (15).

Standardization of sample collection and processing. After removal of supragingival plaque with cotton pellets, subgingival samples were collected by insertion of a sterile paper point (ISO 40; VDW, Munich, Germany) for 30 s into the deepest pocket. Cotton rolls were used to avoid saliva contamination. One paper point was sampled per tooth, and 1 to 4 teeth were sampled for each subject (see Data Set S1 in the supplemental material). Paper points originating from the same subject were placed together in an Eppendorf tube, immediately frozen, and stored at –80°C until extraction.

DNA was extracted from the paper points in two consecutive steps, increasing the aggressiveness of the lysis protocol. In the first extraction (enzymatic and mechanical [EM] extraction method) the paper points in one Eppendorf tube were incubated in lysis buffer containing 10 mM Tris, 1 mM EDTA, pH 8.0, 2.5 mg/ml lysozyme (Sigma, Steinheim, Germany), and 50 U/ml mutanolysin (Sigma, Steinheim, Germany) at 25°C for 1 h on a shaking incubator at 350 rpm. Afterwards, 700 µl of the second lysis buffer, containing 100 mM Tris-HCl, 100 mM EDTA, 100 mM NaCl, 1% (wt/vol) polyvinylpyrrolidone, and 2% (wt/vol) cetrimeronium bromide, pH 8, were added, together with 50 mg of acid-washed glass beads, and the mixture vortexed at full speed 10 times for 30 s with 30-s intervals on ice. Samples were centrifuged at 14,000 rpm for 1 min at 4°C. Pellets were stored on ice until the second extraction. Supernatants were mixed with one volume phenol-chloroform-isoamyl alcohol (25:24:1) and centrifuged, and the aqueous phase was washed with one volume of chloroform. DNA (designated as EM derived) was precipitated from the aqueous phase with 1 volume of isopropyl alcohol at room temperature and stored in water at –20°C.

For the second extraction (FastPrep [FP] method), the pellets left after the first extraction were processed with the FastDNA spin kit for soil, following the manufacturer's instructions. Cells were lysed in a FastPrep-24 instrument (MP Biomedicals, Santa Ana, CA, USA) for 45 s at 6.0 m s⁻¹. DNA (designated as FP derived) was eluted and stored in water at –20°C. The DNA concentrations in the extracts were quantified using a PicoGreen double-stranded DNA (dsDNA) quantitation kit (Molecular Probes, Leiden, The Netherlands).

Preparation of 16S rRNA gene amplicon libraries, sequencing, and data processing. The hypervariable regions of the 16S rRNA gene were amplified using primers 27F and 338R for the V1-V2 region (9) and primers 807F and 1050R (16) for the V5-V6 region. Primers were barcoded and provided with adapters for Illumina sequencing as described previously (9). For each periodontal pocket sample, two DNA extractions were prepared (EM and FP methods), and two sets of primers were used for each DNA (V1-V2 and V5-V6). Thus, four amplicon libraries were obtained for each of the 19 samples, yielding a total of 76 libraries, which were sequenced using 150-bp paired-end sequencing chemistry on an Illumina MiSeq platform. A total of 1.4 million sequence reads were obtained. Data set quality filtering and definition of operational taxonomic units (OTUs) were performed based on the method previously described (9). Briefly, first, the raw reads were quality trimmed. By sliding a window of 10% of the read length along the sequence, the average quality was assessed. Read fragments from the 3' end of the reads that had a PHRED score of the fastq file (Q value) below 10 were removed. Reads that after trimming were shorter than 140 bp, had an N character in their sequence, any mismatches within the primers and barcodes, or more than 10 homopolymer stretches were removed. Next, reads were demultiplexed and sorted. Reads were trimmed conservatively to 120 nt. The paired ends were subsequently matched to give 240 nt. The Mothur programs unique.seqs and pre.cluster were used to collapse and precluster the reads, allowing 2 mismatches over a 240-nt length between the primary sequence and any other to define an OTU. The data set was then filtered to consider only those OTUs that were present in an abundance of >0.001% for the whole experiment and were present (i) in at least one sample at a relative abundance of >1% of the total sequences of that sample, (ii) in at least 2% of samples at a relative

abundance of >0.1% for a given sample, or (iii) in at least 5% of samples at any abundance level. The OTUs that passed these thresholds were entered in an Excel table and used as the basis for all subsequent analyses.

Taxonomy was assigned to the OTUs based on the best match to the reference sequences from the Human Oral Microbiome Database (17) and according to the naive Bayesian classifier provided by the Ribosomal Database Project (18). OTUs were blasted against Human Oral Microbiome Database (HOMD) 16S rRNA RefSeq version 13.2 using the online HOMD 16S rRNA Sequence Identification Tool (<http://www.homd.org>) with default settings. Species and genus names were assigned based on the highest score (bits) for a score of >368 (allowing only two mismatches) and a score of >332 (allowing 12 mismatches), respectively, and were manually checked (see Fig. S11 in the supplemental material). In case of ambiguous assignment, the taxonomy was assigned to the lowest taxonomic level that was common for all references identified as the best match. Taxonomy was further confirmed using the RDP naive Bayesian rRNA Classifier (<http://rdp.cme.msu.edu>). The taxonomy of OTUs with a score of <332 (more than 12 mismatches) with any HOMD reference sequence (42 V1-V2 OTUs and 41 V5-V6 OTUs) was assigned solely using the RDP naive Bayesian rRNA Classifier (<http://rdp.cme.msu.edu>) with a bootstrap cutoff of 50%. For coherence, taxonomic nomenclature as in HOMD was used, except that OTUs belonging to the family *Veillonellaceae* were placed into the order *Selenomonadales* (class *Negativicutes*) instead of into the order *Clostridiales* (class *Clostridia*), as recently suggested (19).

Statistical analysis. R (20) and PRIMER version 6 with PERMANOVA+ add-on software (PRIMER-E, Plymouth Marine Laboratory, Plymouth, United Kingdom) (21) were used to analyze the data. Rarefaction curves were calculated on unstandardized, OTU-level data using the vegan package (22). A community similarity matrix was generated using the Bray-Curtis coefficient by comparing the standardized, log-transformed abundances of each taxon in regard to every pairwise combination of all samples. Principal coordinate analysis (PCoA) was performed on the Bray-Curtis similarities matrix with the PERMANOVA+ PCO routine.

The patterns generated between 19 communities after profiling the V1-V2 region of the 16S rRNA gene were compared to that generated using the V5-V6 region of the 16S rRNA gene by a Mantel-like test (the PRIMER routine RELATE) using Spearman's rank correlation coefficient with 999 permutations (21). Similar comparisons were performed for profiles from the two DNA extraction steps.

Sample diversity was calculated using the Shannon diversity index, where $H' = -\sum_i p_i \ln(p_i)$ and p_i is the relative abundance of the i^{th} OTU (21). Welch's two-sample t test was used to test the differences between the Shannon diversity values of samples from healthy individuals and those with periodontitis (20).

A linear discriminant analysis (LDA) effect size (LefSe) pipeline (23), available at <http://huttenhower.sph.harvard.edu/galaxy/>, was used for biomarker detection. As a first step, the nonparametric factorial Kruskal-Wallis (KW) rank sum test was used to detect taxa with significant differential abundances. Biological consistency was subsequently investigated with a set of pairwise tests among subclasses using the (unpaired) Wilcoxon rank sum test. Finally, LDA was used to estimate the effect size of each differentially abundant trait. Alpha values of 0.05 were used for the KW rank sum test, and a threshold of 2.0 was chosen for logarithmic LDA scores.

The 50 most abundant OTUs were selected, and their abundances were correlated across the 19 samples using Spearman rank correlation and visualized using group-average agglomerative hierarchical clustering (24). The bias of arbitrarily choosing a particular correlation coefficient level to group species can be eliminated by using the similarity profile (SIMPROF) permutation test, which seeks statistically significant evidence of genuine clusters in the absence of *a priori* groups (21); this test was performed at the 99% significance level using 9,999 simulations to generate the dendrogram and 10,000 permutations to generate a permutation distribution.

Based on the receiver operating characteristics (ROC) curve, the area under the curve (AUC) was computed to judge the potential of an OTU as a biomarker. This calculation was performed for two different abundance cutoffs: only OTUs present at a relative abundance of >1% or at >0.1% in at least one sample were studied. Due to the large number of OTUs considered as potential biomarker candidates, it was necessary to investigate whether higher AUC values originated from random variations or from true biological correlations. For this purpose, for each OTU, the values of the 19 patients were randomly shuffled and the AUC values for these random data were also calculated. AUC values larger than 0.5 for the random data occur only by chance. In order to see how much the AUC values of the true data deviated from the random data, box plots of these AUC value distributions were generated. The ROC curve of the OTU with the highest AUC value was plotted together with the ROC curve corresponding to the best AUC value of the random data. Furthermore, permutation tests were carried out. The above-described procedures for generating random data and the calculation of the best AUC value for the random table were repeated 1,000 times to carry out a permutation test. In this way, an empirical distribution of the corresponding best random AUC values was obtained. A P value for the significance of the best AUC value for the original data was obtained from this permutation test. Thus, the P value defines the fraction of random best AUC values that were larger than the best AUC value for the original data.

To elucidate which biomarkers could distinguish the healthy individuals from those with periodontitis, standard classifiers (random forests) were evaluated based on a jackknife (leave-one-out) method. Per cycle, one sample was taken out of the data set, and the AUC values for the OTUs were calculated with the remaining samples. The 10 OTUs with the best AUC values were then used to train the classifier. The trained classifier was applied to the sample that had been removed from the data set to predict whether the individual was healthy or had periodontitis. This procedure was repeated for each individual. In this way, a prediction model for the diagnosis (healthy or periodontitis) of each individual without involving that individual's particular sample in the construction of the classifier was developed. Confusion matrices showing the numbers of correctly and wrongly classified individuals were calculated.

RESULTS AND DISCUSSION

Quality control and taxonomy assignment of reads. The workflow for processing the reads and for taxonomy assignment is shown in Fig. 1A. Sequence data underwent quality control, removal of rare sequences, and clustering into operational taxonomic units (OTUs) as previously described (9). Briefly, 60.8% of the sequence reads passed the quality control, resulting in a total of 0.84 million reads (see Data Set S1 in the supplemental material). For each individual, an average of 44,412 (standard deviation, 6,470) 240-bp sequence reads was generated, which were derived from 4 community profiles per patient, i.e., V1-2_EM (primers V1-V2, EM DNA extraction), V1-2_FP (primers V1-V2, FP DNA extraction), V5-6_EM (primers V5-V6, EM DNA extraction), and V5-6_FP (primers V5-V6, FP DNA extraction). With V1-V2 primers, 584 OTUs were obtained from the periodontal pocket samples, and 495 OTUs were obtained with V5-V6 primers.

Taxonomy was assigned using the Human Oral Microbiome Database (HOMD, <http://www.homd.org>) (17). This approach takes advantage of the curated phylogeny-based database for the human oral bacteria, including described species, unnamed isolates, and uncultivated phylotypes. Ninety-two percent of all OTUs could be assigned at the species or genus level using the HOMD database. Those sequences that showed more than 12 mismatches to any HOMD reference were assigned according to the Ribosomal Database Project (RDP) Classifier (<http://rdp.cme.msu.edu/classifier>) (18), thus al-

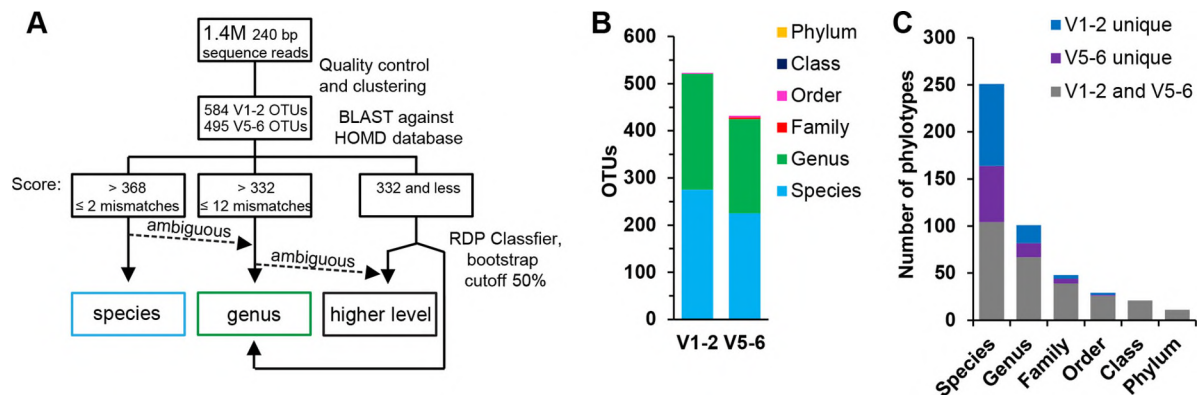


FIG 1 Global comparison of taxonomic diversity obtained by Illumina sequencing of the V1-V2 and V5-V6 hypervariable regions of the 16S ribosomal gene. (A) Workflow for taxonomy assignment. (B) Numbers of OTUs taxonomically assigned to the species, genus, family, order, class, and phylum levels. (C) Numbers of phylotypes identified.

lowing the identification of species or genera that are atypical for the oral cavity.

Detection and removal of paper point-associated OTUs. Recently, sterile paper points were identified as a contamination source in the microbial communities of periodontal pocket clinical samples (25). Members of the genera *Enterococcus* and *Exiguobacterium* were identified in blanks sequenced as controls. Their occurrence in the clinical samples was highly correlated, and their overamplification was most pronounced in samples containing small amounts of DNA, reaching up to 46% of all reads (25). Both genera were also detected in our study. Overall, 3.5% of the total reads were assigned to the genera *Enterococcus* and *Exiguobacterium*. While they were almost absent in samples from individuals with periodontitis, they reached, on average, 5.8% of all reads in samples from healthy individuals, similar to the observations described previously (25). Their presence would therefore have biased the comparison between samples from healthy and diseased individuals significantly. Since paper point-associated *Enterococcus-Exiguobacterium* OTUs were additionally correlated with other OTUs in the previous study (25), we searched for such phylotypes using Spearman's rank correlation coefficient ($\rho > 0.7$ and $\rho > 0.77$ for V1-V2 and V5-V6, respectively). The data in Fig. S1 in the supplemental material show that in all four profiles analyzed, a cluster of highly correlated OTUs that included *Enterococcus* and *Exiguobacterium* OTUs was detected. They had, on average, only 94.5% identity to the best match in HOMD, which indicates that they were not typical inhabitants of the oral cavity. The OTUs retained had, on average, 99% identity to the best match in HOMD. In all, approximately 11% of total reads were manually removed from the data set, resulting in 523 V1-V2 OTUs and 432 V5-V6 OTUs that were then analyzed further. The original data set, the curated data set, and the list of OTUs that were removed are shown in Data Set S1 in the supplemental material. Data Set S2 provides the normalized reads per OTU.

Comparison of taxonomic resolution between hypervariable regions V1-V2 and V5-V6 of the 16S rRNA gene. A total of 0.37 million V1-V2 reads were clustered into 523 OTUs, and 0.39 million V5-V6 reads were grouped into 432 OTUs (Fig. 1B). Rarefaction curves showed that saturation was reached at $>10,000$ reads per sample (see Fig. S2A in the supplemental material). The taxonomic diversity of 523 species-level OTUs was in the range found

by other next-generation-sequencing studies of periodontal pocket samples (26–29). However, the number of reads per sample was higher in our study, suggesting that Illumina amplicon sequencing covers the diversity of the community better.

We compared the two data sets for V1-V2 and V5-V6 with respect to the presence or absence of species and higher taxa. In total, putative representatives of 251 species, 101 genera, 48 families, 29 orders, 21 classes, and 11 phyla were identified by both regions (Fig. 1C). Overall, 99% of all OTUs were assigned to at least at the genus level and 52% were assigned to the species level, including unnamed species that are designated as a human oral taxon (HOT) (17). Species-level taxonomy assignment was possible for 275 V1-V2 OTUs but for only 225 V5-V6 OTUs. While 104 species were assigned by both V1-V2 and V5-V6 amplicons, 87 species were assigned only with V1-V2 primers and 60 were only detected with V5-V6 primers (see Data Set S1 in the supplemental material). Among the most important species known to be associated with periodontal disease, *P. gingivalis*, *Tannerella forsythia*, and *Treponema denticola* were identified by both regions. We analyzed whether species-level taxonomic identification was indeed possible using 240 nt of the V1-V2 or the V5-V6 region, respectively, of the 16S rRNA sequence. To this end, the full 16S rRNA gene sequences of species of the genera *Fretibacterium*, *Porphyromonas*, *Tannerella*, and *Treponema* available in the HOMD database were used to construct phylogenetic trees. Then, two additional trees were constructed in which the V1-V2 OTUs and the V5-V6 OTUs that were assigned to those genera in our study were aligned with the reference sequences (see Data Set S1). This analysis demonstrated that within the genus *Porphyromonas*, species-level identification was possible for *P. gingivalis* and *Porphyromonas asaccharolytica*, while *Porphyromonas endodontalis* and *Porphyromonas catoniae* could not be differentiated from HOT phylotypes closely related at the species level using either hypervariable region (see Fig. S5 in the supplemental material). The analysis for the genus *Treponema*, which harbors almost 50 HOT species-level phylotypes and 10 validly described species, showed that *T. denticola* could be assigned to the species level by one specific OTU using the V1-V2 amplicon and by two V5-V6 OTUs (see Fig. S6, S7, and S8). The HOMD database contains *T. forsythia* and three unnamed species-level HOT phylotypes which can be distinguished using full-length 16S rRNA gene sequences

(see Fig. S9). It was possible to distinguish between all four species using 240 nt of either the V1-V2 or the V5-V6 hypervariable region.

The genus *Fretibacterium* contains only one validly described species, *Fretibacterium fastidiosum*, and additionally, seven species-level HOT phylotypes. They can be distinguished based on their full-length 16S rRNA gene sequences, as well as by only 240 nucleotides in the hypervariable V1-V2 region (see Fig. S10 in the supplemental material). Six of them, including *F. fastidiosum*, can also be differentiated based on 240 nt from the V5-V6 hypervariable region. The data show that the V1-V2 region is slightly more diagnostic than the V5-V6 region for most of the species studied, in accordance with recent analyses of the taxonomic precision of the hypervariable regions of the 16S rRNA gene (30, 31). The authors also confirm our finding that there is no hypervariable region that is equally diagnostic for all genera under investigation; hence, using a combination of different regions is recommended.

As expected, the differences between taxonomic assignments using the V1-V2 or V5-V6 region were less pronounced at the higher taxonomic levels. The majority of genera (68 of 101), families (39 of 48), and orders (26 of 29) were assigned by both V1-V2 and V5-V6 amplicons. All 21 classes were detected by both V1-V2 and V5-V6 amplicons.

We then investigated the effect of the amplified region on the abundance of taxonomic classes in our samples (Fig. 2A). No difference was observed in the average relative abundance of all classes except five: *Synergistetes* were almost undetectable with V5-V6 primers, while they comprised more than 1% using V1-V2 primers. *Deltaproteobacteria*, *Fusobacteria*, and *Erysipelotrichia* were more abundant in the V1-V2 profiles, while *Clostridia* were more than four times more abundant in the V5-6_FP than in the V1-2_EM data set. Significant differences in the community profiles generated by different sets of primers have been observed before. For example, *Bacteroidetes* were found to be almost absent while *Proteobacteria* and *Synergistetes* were at least two times more abundant in V4 profiles than in V1-V2 profiles (28).

To summarize, sequencing of different regions of the 16S rRNA gene as performed here provides taxonomic resolution of a wider range of taxa and eliminates bias introduced by the PCR procedure. Therefore, congruent conclusions that can be drawn from community profiles obtained with different primer sets are more reliable.

Identification of hard-to-lyse phylotypes. We then investigated how the community structure was affected by sequentially more aggressive cell lysis. In the first DNA extraction step (EM), only a mild mechanical lysis using bead beating was applied. In the second extraction step (FP), the pellet from the first extraction was treated using the FastPrep instrument, which has been shown to crack even very recalcitrant soil bacteria and to retrieve up to 20-fold more DNA from gut samples than other methods (32). Unexpectedly, the abundances of all but three classes of bacteria were the same in the EM and FP DNA preparations. This suggests that only a small fraction of the DNA was removed during the first extraction (EM) and a proportional number of genomes was obtained in the second extraction (FP). Differences were found for *Actinobacteria*, *Clostridia*, and *Deltaproteobacteria*, which were more abundant in FP profiles than in EM profiles independent of the region amplified, suggesting that these classes contain a large fraction of hard-to-lyse phylotypes.

To confirm this, we looked for OTUs that were on average at

least 4-fold more abundant in the FP profiles than in their EM counterparts and could be detected in at least one community at an abundance of $\geq 1\%$. Thirty-one and 33 such OTUs could be detected in the V1-V2 and V5-V6 analysis, respectively (see Data Set S1 in the supplemental material). The majority of them belonged to the classes *Actinobacteria* and *Clostridia*, but representatives of 3 other classes were also found. The species with the highest fold change in abundance between FP and EM profiles was *Olsenella* sp. HOT 807 (*Actinobacteria*) for the V1-V2 primer set (73-fold) and *Actinomyces* sp. (*Actinobacteria*) for the V5-V6 primer set (58-fold). A member of the *Clostridia*, *Eubacterium nodatum* HOT 694, that has previously been shown to be a biomarker of periodontal disease (33) was 11-fold more abundant in the V5-6_FP profile, emphasizing the importance of the aggressive cell lysis for the detection of markers of periodontitis.

Periodontal pocket microbial community composition in health and disease. The periodontal microbial communities from 9 individuals diagnosed to have periodontitis and 10 healthy individuals are shown for the 4 profiles created for each sample on the class level (Fig. 2B). To assess the influence on community similarity of the DNA extraction method and the region amplified, a Mantel-like test (the PRIMER routine RELATE) was applied to the standardized OTU abundance matrices (see Data Set S2 in the supplemental material). The Spearman's rank correlation coefficient (ρ) values were in the range of 0.84 to 0.92, suggesting good overall correlation between profiles. The Shannon index of diversity revealed a slightly higher diversity for periodontitis samples (see Fig. S2B), consistent with the results of previous studies (26, 28, 29, 34).

Subsequently, we will concentrate on the results of the V1-2_FP profile, which combines high taxonomic resolution with aggressive cell lysis. The results for the other three profiles can be found in the supplemental material. Ordination analysis was performed to distinguish communities from individuals diagnosed with periodontitis from those that were healthy. Since some of the species with low abundance might have an important role in shifting the community activity toward a dysbiotic stage (4), log-transformed data were used, which down weights the importance of the abundant taxa. Figure 3 shows the PCoA results for the V1-2_FP profile on the OTU level. Samples from healthy individuals were separated from those with periodontitis on the levels of OTU, genus, and class and for all types of profiles (i.e., independent of the primers and extraction methods used) (see Fig. S3 in the supplemental material). However, among the individuals with periodontitis, there were always two outliers whose communities resembled those of healthy subjects. They were individuals 1 and 11. For the analysis of biomarkers, those two samples were excluded.

Identification of biomarkers by the AUC method. Each OTU was considered independently as a candidate for a biomarker. In order to judge the potential of an OTU as a biomarker, the area under the curve (AUC) of the receiver operating characteristics (ROC) curve was computed. Due to the large number of potential biomarker candidates (up to 523), it was necessary to investigate whether higher AUC values originated from random variations or from true biological correlations. Box plots of the results show that the distribution of the AUC values of the original data tends to contain significantly more high values than the random data (see Fig. S4A in the supplemental material). This is also confirmed by the plots of the corresponding ROC curves (see Fig. S4B).

The analysis was consistent across taxa (data not shown), but

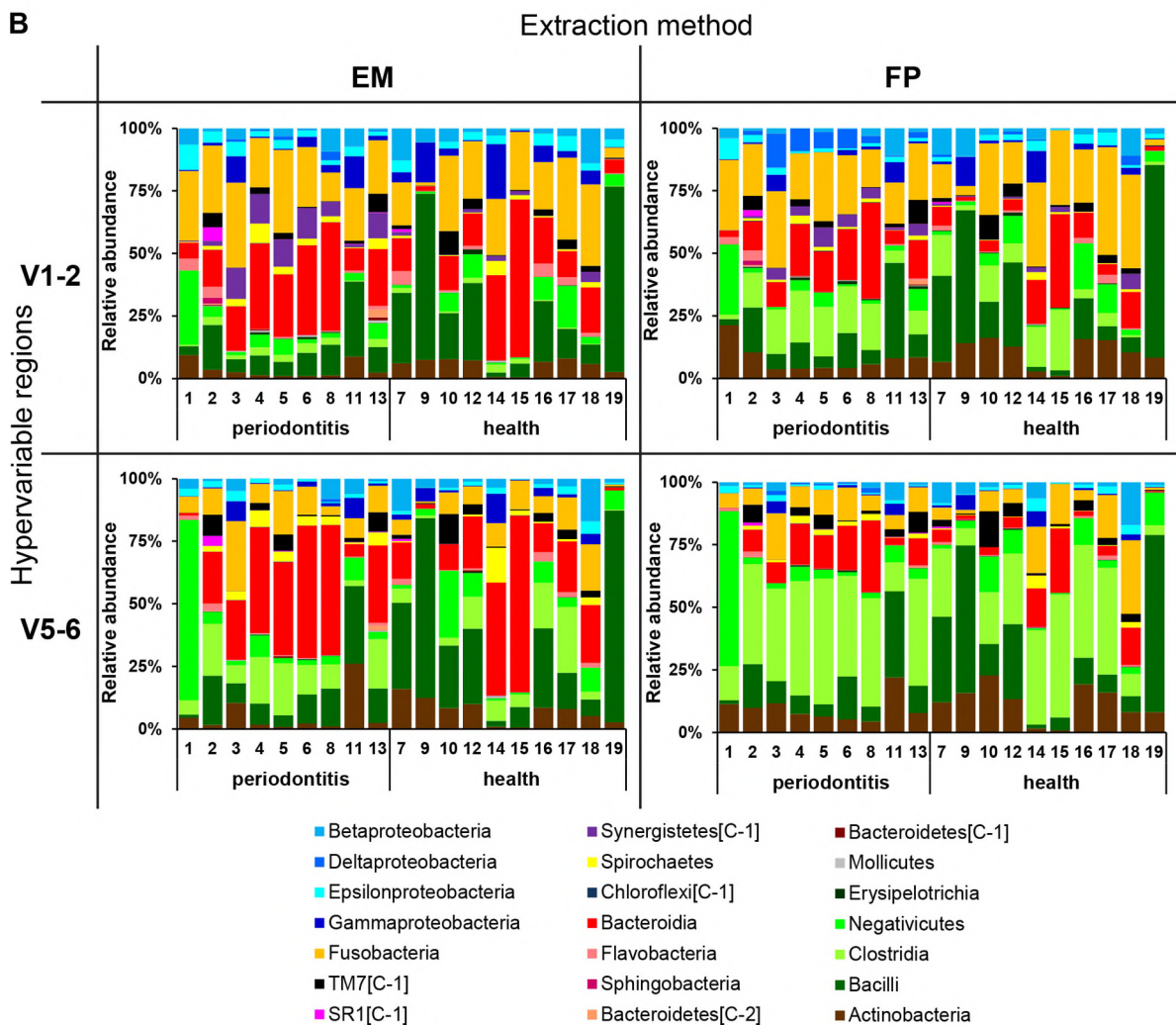
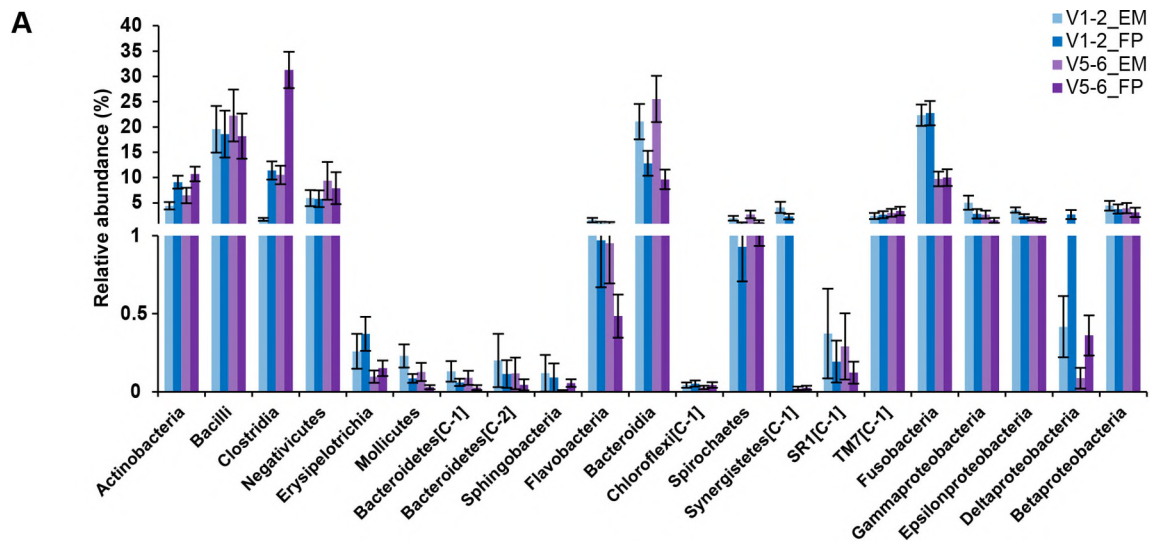


FIG 2 Microbial community composition of periodontal pockets as assessed by 16S rRNA gene amplicon sequencing. (A) Abundance of taxonomic classes averaged over all 19 individuals for the four different community profiles analyzed. The error bars show the standard errors of the means. (B) Composition of the periodontal communities in 19 individuals on the class level. Four different community profiles are shown for each sample, differing with respect to the DNA extraction method (EM and FP) and hypervariable region of the 16S rRNA gene (V1-V2 and V5-V6) amplified. Profiles from healthy subjects and those diagnosed with periodontitis are grouped on the right and left sides of the subpanels, respectively. See Materials and Methods for details on the FP and EM methods.

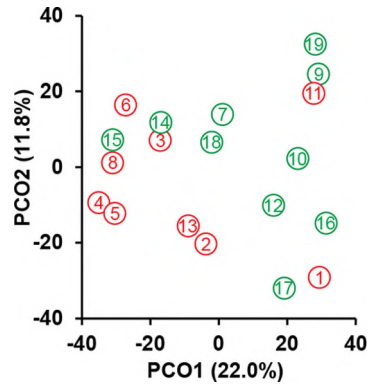


FIG 3 Principal coordinate analysis of the periodontal pocket microbial communities in health and periodontitis. The percentages of the total variation that are explained by the first two axes are indicated. The data shown are for the V1-2_FP profile (hypervariable region V1-V2, DNA extracted by the FP method; see Materials and Methods for details) on the OTU level.

due to the increase in the number of potential biomarkers from class (21) to genus (101) to OTU (523), the statistical power decreased with lower taxonomic levels. On the OTU level, *Chloroflexi* [G-1] (nomenclature as in HOMD [<http://www.homd.org>]) was identified as the only biomarker for periodontitis that had statistical significance (Table 2; see also Data Set S3 in the supplemental material). This uncultivated bacterium had a very low abundance (<0.5%), but it was found with both the V1-V2 and the V5-V6 primers. It has been found in the oral cavity before (17) and has also been linked to periodontitis previously (29). A 96% identical phylotype was recently discovered in the canine oral microbiome (35). There are three draft genome sequences for *Chloroflexi* in the HOMD database, but they are in progress and additional information is not yet available.

Identification of biomarkers based on the LefSe pipeline.

The LefSe pipeline has been especially designed for biomarker discovery in multidimensional complex data sets. LefSe analyzes the abundances of all phylotypes, testing whether values in health and disease are differentially distributed, and then the LDA model ranks the phylotypes based on relative differences between two

TABLE 2 The 10 best biomarkers obtained by the AUC method for the V1-2_FP community profile^a of the subgingival microbiome

Rank of the biomarker	Taxonomic affiliation	AUC	<i>p</i> value
1	<i>Chloroflexi</i> _[G-1]_sp_HOT_439_V1-2_347	0.98	0.036
2	<i>Treponema_socranskii_ss_socranskii</i> _HOT_769_V1-2_130	0.96	0.066
3	<i>Peptostreptococcaceae_XI</i> _[G-4]_sp_HOT_369_V1-2_123	0.94	0.134
4	<i>Fusobacterium_sp</i> _V1-2_386	0.93	0.191
5	<i>Fusobacterium_nucleatum_ss_vincentii</i> _HOT_200_V1-2_68	0.93	0.191
6	<i>Porphyromonas_sp</i> _V1-2_445	0.93	0.191
7	<i>Streptococcus_sp</i> _V1-2_324	0.93	0.191
8	<i>Porphyromonas_sp</i> _V1-2_404	0.92	0.276
9	<i>Actinomyces_sp</i> _V1-2_190	0.92	0.276
10	<i>Mogibacterium_sp</i> _V1-2_145	0.91	0.338

^a V1-2_FP community profile was obtained by amplifying sequences from the V1-V2 hypervariable region of the rRNA gene after the FastPrep (FP) method was used for DNA extraction.

conditions, taking into account both variability and discriminatory power (23). Here, we performed the analysis on the levels of OTU, genus, and class.

Eight classes of bacteria were identified as being linked to periodontitis, namely, *Bacteroidetes* [C-1], *Bacteroidia*, *Chloroflexi*, *Clostridia*, *Deltaproteobacteria*, *Mollicutes*, *Spirochaetes*, and *Synergistetes* (see Data Set S3 in the supplemental material). All of them have been identified previously as biomarkers of periodontitis (26, 28, 29, 34). In total, 34 genera were identified as being associated with periodontitis and 6 were linked to oral health, also confirming previous studies (26, 28, 29, 34, 36).

In total, 94 V1-V2 OTUs and 72 V5-V6 OTUs were significantly more abundant in samples from individuals with periodontitis (see Data Set S3 in the supplemental material). Figure 4 shows the biomarkers identified for the V1-2_FP profile. *Porphyromonas gingivalis* HOT 619 had the highest LDA score among all OTUs. *P. gingivalis* is the best understood periodontal pathogen and has been shown to trigger changes in the diversity and abundance of the oral microbial community that lead to dysbiosis and unresolved inflammation; thus, it is a keystone pathogen for periodontitis (37, 38). It forms the so-called red complex together with two other species, namely, *T. denticola* and *T. forsythia* (4, 39). In our study, 9 OTUs classified as *Treponema* spp. and 5 OTUs classified as *Tannerella* spp. were identified as biomarkers. *Tannerella* spp. had the fourth highest LDA score in the V1-2_FP profile.

In total, 13 OTUs classified as *Fretibacterium* spp. had significant LDA scores (see Data Set S3 in the supplemental material). *Fretibacterium fastidiosum* HOT 363 had the second highest LDA score in the V1-2_EM profile, and four other OTUs assigned to the genus *Fretibacterium* were also found among the top 10 OTUs with the highest LDA scores. Members of this genus were linked to periodontitis previously but not at such a high level of significance (28, 29, 40–42).

Several important biomarkers required the aggressive DNA extraction method for identification, e.g., *Desulfobulbus* sp. HOT 41 (see Data Set S3). Single-cell sequencing of this uncultured member of the *Deltaproteobacteria* had revealed large gene clusters encoding virulence factors of known human pathogens, suggesting the importance of this organism for the etiology of periodontitis (43). Four species within the order *Clostridiales*, i.e., *Peptostreptococcaceae* XIII [G-4] sp., *Eubacterium* XI [G-6] sp., *Peptostreptococcaceae* XI [G-4] sp. HOT 369, and *Peptostreptococcus stomatitis* HOT 113 were also linked to periodontitis and are also hard-to-lyse organisms.

Among the OTUs with significant LDA scores, *Chloroflexi* [G-1] was also found, confirming the biomarker detection by the AUC method. Interestingly, *Aggregatibacter actinomycetemcomitans* was not identified as a biomarker of periodontitis in our study, although it is generally considered a major cause of aggressive periodontitis (44). *A. actinomycetemcomitans* amounted to 5% in sample 3 (periodontitis), and had a very low abundance (<0.05%) in several other samples, both from healthy and from diseased individuals. Carriage is most strictly correlated with periodontitis in localized aggressive childhood periodontitis (44); in our samples, we had only one individual with aggressive periodontitis and she was 70 years old.

To summarize, periodontitis biomarkers identified in previous studies were fully confirmed by our analysis. The red complex pathogen *P. gingivalis* was identified as the best marker for periodontitis, followed by species-level OTUs of *Tannerella* spp. and

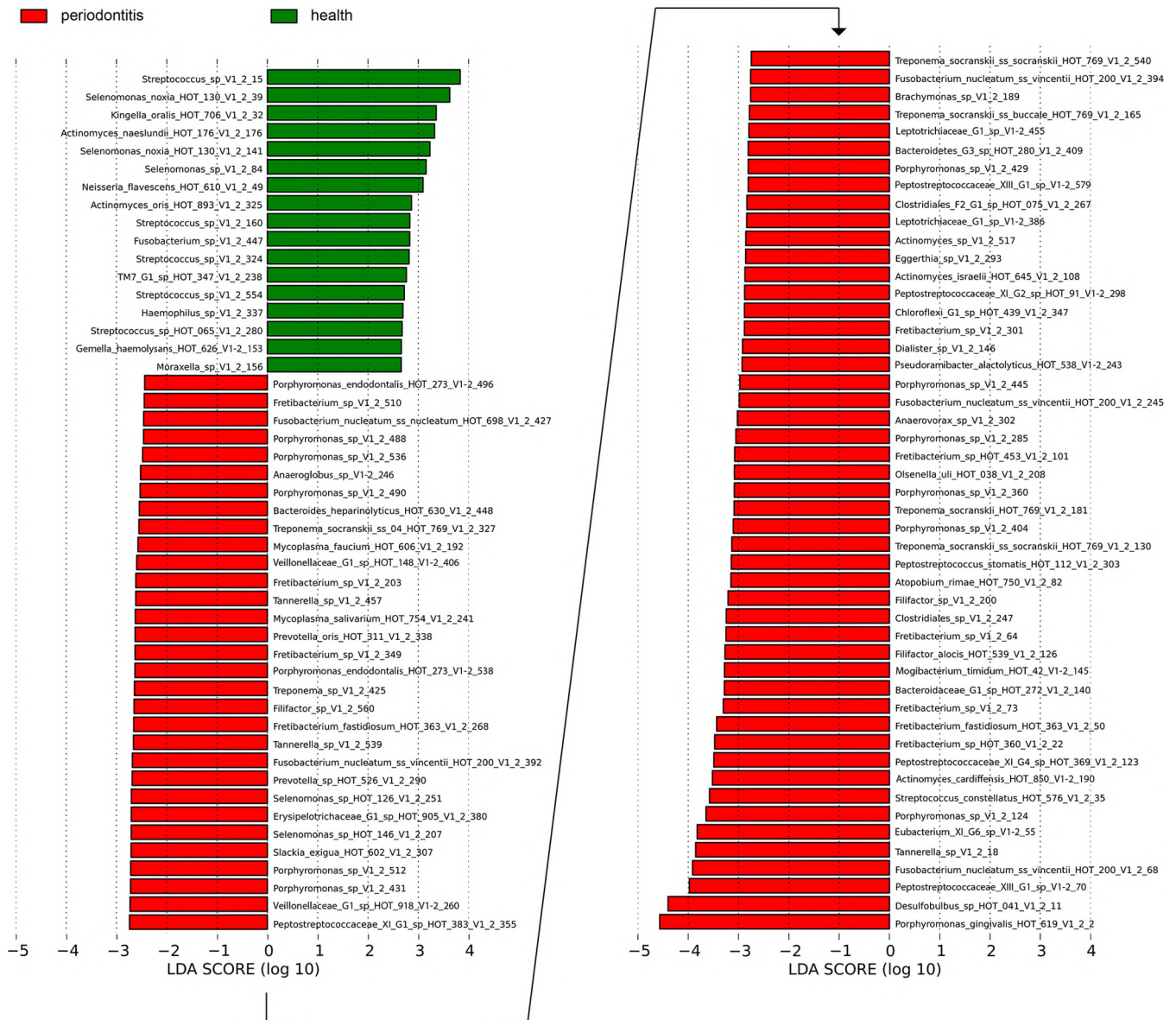


FIG 4 Identification of biomarkers based on the linear discriminant analysis (LDA) and effect size (LefSe) pipeline. The OTUs with an LDA score of >2 from the V1-2_FP profile are shown.

Treponema spp. Aggressive cell lysis is important for hard-to-lyse biomarkers, e.g., *Desulfolobus* sp. HOT 41 and *Eubacterium nodatum* HOT 694. The relevance of *Fretibacterium* spp. for periodontitis was further enhanced by our study. The two methods used for biomarker discovery (AUC and LefSe) obtained similar results, but the LefSe pipeline provided more statistical power.

Community connectivity. In a polymicrobial community, the interactions between the individual species are key to understanding the shift from a healthy community toward dysbiosis (45). An interaction network constructed from checkerboard and microarray analysis of subgingival plaque samples (3,079 periodontitis samples and 89 samples from healthy individuals) identified one highly connected community in individuals with disease and two distinct communities in healthy individuals (46). Within the periodontitis community, four bacterial modules were obtained and

were shown to be similar to those of a previous study of 13,261 subgingival plaque samples from 185 individuals (47). The brown module was reproduced in both studies and consisted of the red complex pathogens and several additional biomarkers of periodontitis (46). Although these analyses were confined to cultivated species, they provide an excellent reference to test whether similar modules would be detectable in our community profiles. We analyzed Spearman's correlation between abundance patterns using the 50 most abundant OTUs (similar to the analysis in reference 46), accounting for 64% and 72% of all reads in the V1-V2 and V5-V6 community profiles, respectively.

The results presented in Fig. 5 show that two clusters were formed by the OTUs. The top cluster contained subclusters previously found in samples from healthy individuals, while the bottom cluster contained subclusters previously found in periodon-

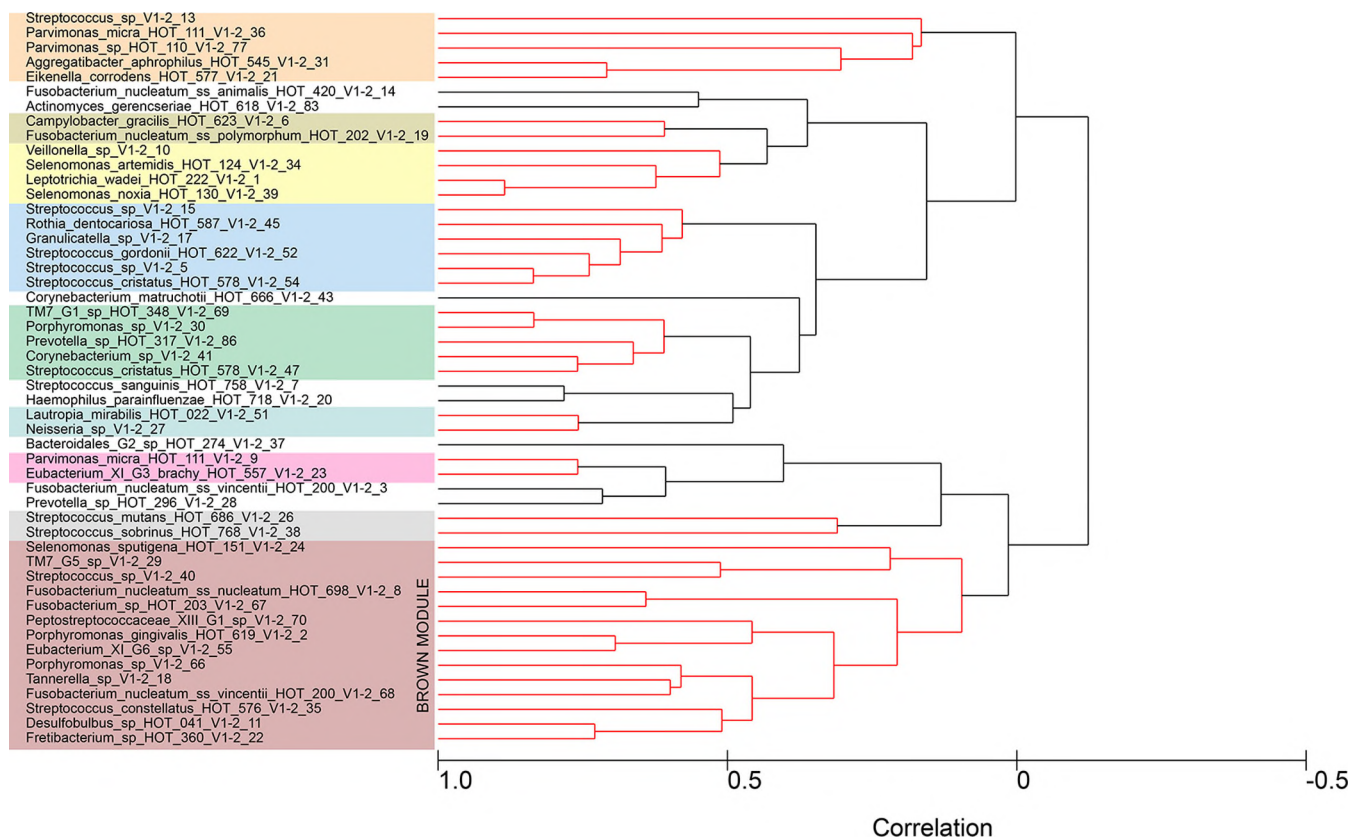


FIG 5 Hierarchical cluster analysis of OTUs. Correlation of the abundances of the 50 most abundant OTUs was calculated across the 19 samples. Hierarchical cluster analysis was performed using Spearman rank correlation. Red lines and colored boxes indicate groups of OTUs where clustering was statistically significant using the SIMPROF routine in PRIMER 6 (at a significance level of 99%). The V1-2_FP data set was used for this analysis.

titis samples (46); thus, the two branches might reflect the shift from healthy to dysbiotic communities. The brown module (see above) was fully reproduced in our analysis (depicted in red in Fig. 5). It consisted of *P. gingivalis*, *Tannerella* sp. (best match to *T. forsythia*, with 98.3% identity), and *Eubacterium* sp. (best match to *E. nodatum*, with 98.8% identity). OTUs classified as *T. denticola* and *Treponema socranskii* were not among the 50 most abundant species, but their abundance was highly correlated with that of *P. gingivalis* (ρ between 0.42 and 0.44). In this module, we additionally found strains that had previously been located in other modules, e.g., *Fusobacterium nucleatum* and *Streptococcus constellatus* HOT 576 (46) (47). Additionally, strains that had not been located in any of the modules previously were also found in the red complex cluster, namely, *Selenomonas sputigena* HOT 151, TM7 [G-5] sp., *Desulfobulbus* sp. HOT 41, and *Fretibacterium* sp. HOT 360. Thus, the brown module, comprising periodontitis-associated co-occurring microbes described previously on the basis of thousands of samples, was fully reproduced using Illumina sequencing of V1-V2 and V5-V6 amplicons from 19 individuals.

Diagnosis of periodontitis by the jackknife method. In order to see how well combinations of biomarkers could distinguish the healthy individuals from those with periodontitis, the 10 best biomarkers of each profile were selected and standard classifiers (random forests) were constructed. The classifiers, including the selection of the biomarkers, were evaluated based on a jackknife (leave-one-out) method (Fig. 6). To determine the influence of

OTUs with very low abundance, which might contain sequencing errors, the analysis was conducted using all OTUs (using the abundance-filtering criteria described in Materials and Methods) or only the 300, 200, 100, and 50 most abundant OTUs of the data set.

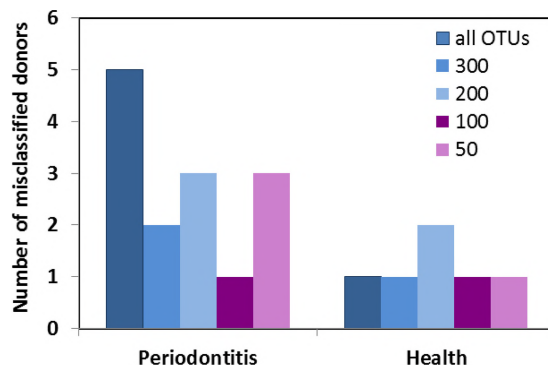


FIG 6 Predictive power of the 10 best biomarkers for diagnosis of health versus periodontitis demonstrated by the number of misclassified subjects in each category. The full V1-2_FP data set (all OTUs) or only the 300, 200, 100, or 50 most abundant OTUs were used to select the best biomarkers based on the AUC value. The jackknife method was applied for selecting the 10 best biomarkers and for training the data set, and significance was calculated with the random forest method.

The best result was obtained using the 100 most abundant OTUs. Using this data set, one sample derived from a healthy individual and one from an individual with periodontitis were misdiagnosed, and thus, 15 of 17 samples were correctly classified. Up to six samples were misclassified using 50, 200, 300, or all OTUs. The data show that, indeed, a compromise has to be found between too low resolution of community composition (50 OTUs) and too high resolution, which might be hampered by sequencing errors. Considering samples 1 and 11, which were excluded from the biomarker analyses, overall, 15 of 19 individuals were thus correctly diagnosed based on the composition of their periodontal communities. Given the complexity of the periodontal microbiome and the frequently observed outlier communities, which previously rendered such a diagnosis impossible, this is a very high fraction. In a larger patient cohort, the outliers would increase in frequency and probably form subgroups that might be correlated with clinical, genetic, or lifestyle parameters. Thus, the jackknife algorithm could be trained on different types of periodontitis communities and its predictive power would be improved accordingly.

Periodontitis samples that were always misclassified as healthy based on the taxonomic composition of the microbial community were samples 1, 3, and 11. The individuals who provided those samples lacked typical periodontal pathogens and, thus, harbored a relatively benign microbial community, but they suffered from chronic periodontitis. For example, various species of *Veillonella* comprised approximately 20% of all OTUs in sample 1. Sample 3 was dominated by *Fusobacterium nucleatum* and *Bacteroidales* [G-2] sp. HOT 274 (approximately 13% each). In these individuals, disease development could have been triggered by additional factors, e.g., smoking or genetic susceptibility. Alternatively, the 16S rRNA sequences might fail to reflect the pathogenicity of the strains, since it is only a taxonomic biomarker. Strains of a species that have the same 16S rRNA sequence can have massively different pathogenicity, and similar virulence factors have been found in phylogenetically distant species (see below).

A clinically healthy individual that was always misclassified as having chronic periodontitis based on the composition of the periodontal microbial community was represented by sample 7. This community was dominated by OTUs that were neither good nor bad according to our analyses, i.e., they were not among the biomarkers for health or periodontitis identified by the LefSe algorithm. However, sample 7 contained two important periodontitis-associated OTUs among the top most abundant taxa: *Fusobacterium_nucleatum_ss_vincentii_HOT_200_V1-2_3*, comprising 3.7% of the community, and *Eubacterium_XI_[G-6]_sp_V1-2_55*, which is the 6th best biomarker of periodontitis (Fig. 4) and comprised 2.7% of the community.

The aim of biomarker discovery in periodontitis must be to predict the risk of developing the disease at an early stage, before clinical symptoms manifest themselves. This could be approached in a prospective cohort study like the GNC where the composition of the microbial community and the development of periodontitis can be monitored in the participants, together with life style, genetic, and physiological parameters, over 30 years. In the long run, a taxonomic profile of the periodontal community might be able to predict the risk for developing periodontitis.

Although deep sequencing of the oral microbiome is already revolutionizing our understanding of caries (48) and periodontitis (4), two of the most prevalent human diseases, a real under-

standing of the shift from the healthy community to the dysbiotic stage is unlikely to be detected by sequencing of the 16S ribosomal marker gene alone. There is no strict correlation between phylogeny and function. For example, well characterized virulence factors from *P. gingivalis*, the best studied periodontal pathogen, have been found in taxonomically distant members of the oral community, e.g., *A. actinomycetemcomitans*, *Prevotella intermedia*, *T. denticola*, and *Neisseria cinerea* (49). Conversely, closely related strains of, e.g., *Tannerella*, can have widely different virulence properties (50). Deep-sequencing studies of the 16S rRNA marker gene have expanded the list of periodontitis-associated species-level phylotypes significantly but have contributed little to an understanding of the etiology of periodontal disease, including the unresolved question of cause and effect (5, 51). Thus, the primary function of ribosomal gene amplicon sequencing is to reliably and sensitively stratify the samples into subgroups. Patient age, race, immune status, smoking, diabetes, etc., have already been shown to have a profound influence on the oral microbiome, but how these differences relate to disease progression is not known. Using the method developed here, sample stratification can be performed at high throughput. As a next step, metatranscriptomics (52, 53) or other specialized analyses can then be applied to a smaller number of representative samples.

Conclusions. We have shown that Illumina 16S rRNA gene amplicon sequencing of clinical samples from periodontal pockets is a fast approach to obtain high-resolution taxonomic abundance profiles of periodontal microbial communities. Aggressive DNA extraction and the use of two hypervariable regions of the 16S rRNA gene independently increase the robustness and reliability of the results. We developed an automated analysis pipeline that can be used for basic research, epidemiological studies, or the development of diagnostic tools.

ACKNOWLEDGMENTS

We are grateful for the cooperation of the Augsburg study center of the German National Cohort and especially to Jakob Linsaisen. We thank all patients who contributed samples and all clinicians who helped in taking the samples and in preserving, cataloguing, and shipping them. We thank Anneke Thien, Braunschweig, for taking subgingival pocket samples from IWD for testing DNA extraction methods. We thank Robert Geffers and Michael Jarek for help with data analysis and continuous support in optimizing the sequencing pipeline. Finally, we express our sincere and very special thanks to Frank Pessler, Twincore, Braunschweig. Informal discussions with Frank essentially triggered the whole study. Throughout its course, Frank provided many interesting ideas and continuously stimulated interactions.

REFERENCES

1. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V, Abraham J, Ackerman I, Aggarwal R, Ahn SY, Ali MK, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Bahalim AN, Barker-Collo S, Barrero LH, Bartels DH, Basáñez MG, Baxter A, Bell ML, Benjamin EJ, Bennett D, Bernabé E, Bhalla K, Bhandari B, Bikbov B, Bin Abdulhak A, Birbeck G, Black JA, Blencowe H, Blore JD, Blyth F, Bolliger I, Bonaventure A, Boufous S, Bourne R, Boussinesq M, Braithwaite T, Brayne C, Bridgett L, Brooker S, Brooks P, et al. 2012. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380:2163–2196. [http://dx.doi.org/10.1016/S0140-6736\(12\)61729-2](http://dx.doi.org/10.1016/S0140-6736(12)61729-2).
2. Lockhart PB, Bolger AF, Papapanou PN, Osinbowale O, Trevisan M, Levison ME, Taubert KA, Newburger JW, Gornik HL, Gewitz MH, Wilson WR, Smith SC, Jr, Baddour LM, American Heart Association Rheumatic Fever, Endocarditis, and Kawasaki Disease Committee of the Council on

- Cardiovascular Disease in the Young, Council on Epidemiology and Prevention, Council on Peripheral Vascular Disease, Council on Clinical Cardiology. 2012. Periodontal disease and atherosclerotic vascular disease: does the evidence support an independent association? A scientific statement from the American Heart Association. *Circulation* 125:2520–2544. <http://dx.doi.org/10.1161/CIR.0b013e31825719f3>.
3. Teles R, Teles F, Frias-Lopez J, Paster B, Haffajee A. 2013. Lessons learned and unlearned in periodontal microbiology. *Periodontol* 2000 62:95–162. <http://dx.doi.org/10.1111/prd.12010>.
 4. Hajishengallis G, Lamont RJ. 2012. Beyond the red complex and into more complexity: the polymicrobial synergy and dysbiosis (PSD) model of periodontal disease etiology. *Mol Oral Microbiol* 27:409–419. <http://dx.doi.org/10.1111/j.2041-1014.2012.00663.x>.
 5. Wade WG. 2013. The oral microbiome in health and disease. *Pharmacol Res* 69:137–143. <http://dx.doi.org/10.1016/j.phrs.2012.11.006>.
 6. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM, FitzGerald MG, Fulton RS, Giglio MG, Hallsworth-Pepin K, Lobos EA, Madupu R, Magrini V, Martin JC, Mitreva M, Muzny DM, Sodergren EJ, Versalovic J, Wollam AM, Worley KC, Wortman JR, Young SK, Zeng Q, Aagaard KM, Abolude OO, Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL, Anderson S, Appelbaum E, Arachchi HM, Armitage G, Arze CA, Ayvaz T, Baker CC, Begg L, Belachew T, Bhonagiri V, Bihan M, Blaser MJ, Bloom T, Bonazzi V, Brooks J, Buck GA, Buhay CJ, Busam DA, Campbell JL, et al. 2012. Structure, function and diversity of the healthy human microbiome. Human Microbiome Project Consortium. *Nature* 486:207–214. <http://dx.doi.org/10.1038/nature11234>.
 7. Wichmann HE, Kaaks R, Hoffmann W, Jockel KH, Greiser KH, Linseisen J. 2012. The German National Cohort. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsforschung* 55:781–787. (In German.) <http://dx.doi.org/10.1007/s00103-012-1499-y>.
 8. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 30:434–439. <http://dx.doi.org/10.1038/nbt.2198>.
 9. Camarina-Silva A, Jauregui R, Chaves-Moreno D, Oxley AP, Schaumburg F, Becker K, Wos-Oxley ML, Pieper DH. 2014. Comparing the anterior nares bacterial community of two discrete human populations using Illumina amplicon sequencing. *Environ Microbiol* 16:2939–2952. <http://dx.doi.org/10.1111/1462-2920.12362>.
 10. Degnan PH, Ochman H. 2012. Illumina-based analysis of microbial community diversity. *ISME J* 6:183–194. <http://dx.doi.org/10.1038/ismej.2011.74>.
 11. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79:5112–5120. <http://dx.doi.org/10.1128/AEM.01043-13>.
 12. Mizrahi-Man O, Davenport ER, Gilad Y. 2013. Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS One* 8:e53608. <http://dx.doi.org/10.1371/journal.pone.0053608>.
 13. Ong SH, Kukkillaya VU, Wilm A, Lay C, Ho EX, Low L, Hibberd ML, Nagarajan N. 2013. Species identification and profiling of complex microbial communities using shotgun Illumina sequencing of 16S rRNA amplicon sequences. *PLoS One* 8:e60811. <http://dx.doi.org/10.1371/journal.pone.0060811>.
 14. World Health Organization. 1997. Oral health surveys: basic methods, 4th ed. World Health Organization, Geneva, Switzerland. <http://apps.who.int/iris/handle/10665/41905>.
 15. Ainamo J, Barmes D, Beagrie G, Cutress T, Martin J, Sardo-Infirri J. 1982. Development of the World Health Organization (WHO) community periodontal index of treatment needs (CPITN). *Int Dent J* 32:281–291.
 16. Bohorquez LC, Delgado-Serrano L, Lopez G, Osorio-Forero C, Klepac-Ceraj V, Kolter R, Junca H, Baena S, Zambrano MM. 2012. In-depth characterization via complementing culture-independent approaches of the microbial community in an acidic hot spring of the Colombian Andes. *Microb Ecol* 63:103–115. <http://dx.doi.org/10.1007/s00248-011-9943-3>.
 17. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A, Wade WG. 2010. The human oral microbiome. *J Bacteriol* 192:5002–5017. <http://dx.doi.org/10.1128/JB.00542-10>.
 18. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267. <http://dx.doi.org/10.1128/AEM.00062-07>.
 19. Marchandin H, Teyssier C, Campos J, Jean-Pierre H, Roger F, Gay B, Carlier JP, Jumas-Bilak E. 2010. *Negativicoccus succinivorans* gen. nov., sp. nov., isolated from human clinical samples, emended description of the family Veillonellaceae and description of *Negativicutes* classis nov., Selenomonadales ord. nov. and Acidaminococcaceae fam. nov. in the bacterial phylum Firmicutes. *Int J Syst Evol Microbiol* 60(Pt 6):1271–1279. <http://dx.doi.org/10.1099/ijs.0.013102-0>.
 20. R Development Core Team. 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
 21. Clarke KR, Warwick RM. 2001. Change in marine communities: an approach to statistical analysis and interpretation, 2nd ed. PRIMER-E Ltd., Plymouth, United Kingdom.
 22. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens HH, Wagner H. 2013. *vegan*: Community Ecology Package. R package version 2.0-10. <http://cran.r-project.org/web/packages/vegan/index.html>.
 23. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. 2011. Metagenomic biomarker discovery and explanation. *Genome Biol* 12:R60. <http://dx.doi.org/10.1186/gb-2011-12-6-r60>.
 24. Kolde R. 2013. *heatmap*: Pretty Heatmaps. R package version 0.7.7. <http://cran.r-project.org/web/packages/heatmap/index.html>.
 25. van der Horst J, Buijs MJ, Laine ML, Wismeijer D, Loos BG, Crielaard W, Zaura E. 2013. Sterile paper points as a bacterial DNA-contamination source in microbiome profiles of clinical samples. *J Dent* 41:1297–1301. <http://dx.doi.org/10.1016/j.jdent.2013.10.008>.
 26. Ge X, Rodriguez R, Trinh M, Gunsolley J, Xu P. 2013. Oral microbiome of deep and shallow dental pockets in chronic periodontitis. *PLoS One* 8:e65520. <http://dx.doi.org/10.1371/journal.pone.0065520>.
 27. Bik EM, Long CD, Armitage GC, Loomer P, Emerson J, Mongodin EF, Nelson KE, Gill SR, Fraser-Liggett CM, Relman DA. 2010. Bacterial diversity in the oral cavity of 10 healthy individuals. *ISME J* 4:962–974. <http://dx.doi.org/10.1038/ismej.2010.30>.
 28. Griffen AL, Beall CJ, Campbell JH, Firestone ND, Kumar PS, Yang ZK, Podar M, Leys EJ. 2012. Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *ISME J* 6:1176–1185. <http://dx.doi.org/10.1038/ismej.2011.191>.
 29. Abusleme L, Dupuy AK, Dutzan N, Silva N, Burleson JA, Strausbaugh LD, Gamonal J, Diaz PI. 2013. The subgingival microbiome in health and periodontitis and its relationship with community biomass and inflammation. *ISME J* 7:1016–1025. <http://dx.doi.org/10.1038/ismej.2012.174>.
 30. Guo F, Ju F, Cai L, Zhang T. 2013. Taxonomic precision of different hypervariable regions of 16S rRNA gene and annotation methods for functional bacterial groups in biological wastewater treatment. *PLoS One* 8:e76185. <http://dx.doi.org/10.1371/journal.pone.0076185>.
 31. Kumar PS, Brooker MR, Dowd SE, Camerlengo T. 2011. Target region selection is a critical determinant of community fingerprints generated by 16S pyrosequencing. *PLoS One* 6:e20956. <http://dx.doi.org/10.1371/journal.pone.0020956>.
 32. Kennedy NA, Walker AW, Berry SH, Duncan SH, Farquarson FM, Louis P, Thomson JM, Satsangi J, Flint HJ, Parkhill J, Lees CW, Hold GL. 2014. The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS One* 9:e88982. <http://dx.doi.org/10.1371/journal.pone.0088982>.
 33. Haffajee AD, Teles RP, Socransky SS. 2006. Association of *Eubacterium nodatum* and *Treponema denticola* with human periodontitis lesions. *Oral Microbiol Immunol* 21:269–282. <http://dx.doi.org/10.1111/j.1399-302X.2006.00287.x>.
 34. Kistler JO, Booth V, Bradshaw DJ, Wade WG. 2013. Bacterial community development in experimental gingivitis. *PLoS One* 8:e71227. <http://dx.doi.org/10.1371/journal.pone.0071227>.
 35. Dewhirst FE, Klein EA, Thompson EC, Blanton JM, Chen T, Milella L, Buckley CM, Davis IJ, Bennett ML, Marshall-Jones ZV. 2012. The canine oral microbiome. *PLoS One* 7:e36067. <http://dx.doi.org/10.1371/journal.pone.0036067>.
 36. Junemann S, Prior K, Szczepanowski R, Harks I, Ehmke B, Goesmann A, Stoye J, Harmsen D. 2012. Bacterial community shift in treated periodontitis patients revealed by ion torrent 16S rRNA gene amplicon se-

- quencing. PLoS One 7:e41606. <http://dx.doi.org/10.1371/journal.pone.0041606>.
37. Hajishengallis G, Liang S, Payne MA, Hashim A, Jotwani R, Eskan MA, McIntosh ML, Alsam A, Kirkwood KL, Lambris JD, Darveau RP, Curtis MA. 2011. Low-abundance biofilm species orchestrates inflammatory periodontal disease through the commensal microbiota and complement. *Cell Host Microbe* 10:497–506. <http://dx.doi.org/10.1016/j.chom.2011.10.006>.
 38. Hajishengallis G, Darveau RP, Curtis MA. 2012. The keystone-pathogen hypothesis. *Nat Rev Microbiol* 10:717–725. <http://dx.doi.org/10.1038/nrmicro2873>.
 39. Holt SC, Ebersole JL. 2005. *Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia*: the “red complex,” a prototype polybacterial pathogenic consortium in periodontitis. *Periodontol* 2000 38:72–122. <http://dx.doi.org/10.1111/j.1600-0757.2005.00113.x>.
 40. You M, Mo S, Watt RM, Leung WK. 2013. Prevalence and diversity of Synergistetes taxa in periodontal health and disease. *J Periodont Res* 48:159–168. <http://dx.doi.org/10.1111/j.1600-0765.2012.01516.x>.
 41. Belibasakis GN, Ozturk VO, Emingil G, Bostanci N. 2013. Synergistetes cluster A in saliva is associated with periodontitis. *J Periodont Res* 48:727–732. <http://dx.doi.org/10.1111/jre.12061>.
 42. Vartoukian SR, Palmer RM, Wade WG. 2009. Diversity and morphology of members of the phylum “synergistetes” in periodontal health and disease. *Appl Environ Microbiol* 75:3777–3786. <http://dx.doi.org/10.1128/AEM.02763-08>.
 43. Campbell AG, Campbell JH, Schwientek P, Woyke T, Sczyrba A, Allman S, Beall CJ, Griffen A, Leys E, Podar M. 2013. Multiple single-cell genomes provide insight into functions of uncultured Deltaproteobacteria in the human oral cavity. *PLoS One* 8:e59361. <http://dx.doi.org/10.1371/journal.pone.0059361>.
 44. Henderson B, Ward JM, Ready D. 2010. *Aggregatibacter (Actinobacillus) actinomycetemcomitans*: a triple A* periodontopathogen? *Periodontol* 2000 54:78–105. <http://dx.doi.org/10.1111/j.1600-0757.2009.00331.x>.
 45. Shade A, Handelsman J. 2012. Beyond the Venn diagram: the hunt for a core microbiome. *Environ Microbiol* 14:4–12. <http://dx.doi.org/10.1111/j.1462-2920.2011.02585.x>.
 46. Duran-Pinedo AE, Paster B, Teles R, Frias-Lopez J. 2011. Correlation network analysis applied to complex biofilm communities. *PLoS One* 6:e28438. <http://dx.doi.org/10.1371/journal.pone.0028438>.
 47. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL, Jr. 1998. Microbial complexes in subgingival plaque. *J Clin Periodontol* 25:134–144. <http://dx.doi.org/10.1111/j.1600-051X.1998.tb02419.x>.
 48. Gross EL, Beall CJ, Kutsch SR, Firestone ND, Leys EJ, Griffen AL. 2012. Beyond *Streptococcus mutans*: dental caries onset linked to multiple species by 16S rRNA community analysis. *PLoS One* 7:e47722. <http://dx.doi.org/10.1371/journal.pone.0047722>.
 49. Wang J, Qi J, Zhao H, He S, Zhang Y, Wei S, Zhao F. 2013. Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. *Sci Rep* 3:1843. <http://dx.doi.org/10.1038/srep01843>.
 50. Beall CJ, Campbell AG, Dayeh DM, Griffen AL, Podar M, Leys EJ. 2014. Single cell genomics of uncultured, health-associated *Tannerella* BU063 (Oral Taxon 286) and comparison to the closely related pathogen *Tannerella forsythia*. *PLoS One* 9:e89398. <http://dx.doi.org/10.1371/journal.pone.0089398>.
 51. Wade WG. 2011. Has the use of molecular methods for the characterization of the human oral microbiome changed our understanding of the role of bacteria in the pathogenesis of periodontal disease? *J Clin Periodontol* 38(Suppl 11):7–16. <http://dx.doi.org/10.1111/j.1600-051X.2010.01679.x>.
 52. Duran-Pinedo AE, Chen T, Teles R, Starr JR, Wang X, Krishnan K, Frias-Lopez J. 2014. Community-wide transcriptome of the oral microbiome in subjects with and without periodontitis. *ISME J* 8:1659–1672. <http://dx.doi.org/10.1038/ismej.2014.23>.
 53. Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M. 2014. Metatranscriptomics of the human oral microbiome during health and disease. *mBio* 5(2):e01012-14. <http://dx.doi.org/10.1128/mBio.01012-14>.