Supplementary Methods

Enrollment criteria for human subjects

Included subjects had ≥ 14 teeth, excluding third molars, and ≥ 10 posterior teeth. Patients had not received periodontal treatment before the time of examination, were not suffering from any systemic illness, were not pregnant, nor have they received antibiotics, anticoagulants or non-steroid anti-inflammatory therapy 6-months prior to the study. All subjects were non-smokers. Subjects from the periodontitis group were at least 35 years old, had ≥ 5 teeth with probing depths (PD) ≥ 5 mm and clinical attachment level (CAL) ≥ 4 mm, had BoP in at least 20% of sites and showed radiographic evidence of bone loss. Subjects in the healthy group had no evidence of periodontal disease as determined by at least 90% of sites with PD and CAL ≤ 3 mm, no site with PD > 4 mm and less than 10% of sites presenting BoP. Periodontal examinations were performed by a calibrated skilled clinician (JG).

Sample collection

Subjects were sampled a week after their initial examination. Prior to sampling, supragingival plaque was carefully removed. Subgingival plaque was then collected by introducing a sterile Gracey curette at the bottom of the site and removing the plaque content with a single stroke of the curette. Samples were then placed in a 2 ml tube containing 50 µl of TE buffer (20 mM Tris pH 7.4, 2 mM EDTA) and stored at -80°C.

DNA isolation, 16S rRNA gene library preparation and sequencing

DNA was isolated according to previously described procedures using lysozyme and proteinase K treatment and a DNeasy Blood and Tissue kit (Qiagen) (Diaz et al 2012). DNA recovery from samples varied from 75 to 2790 ng per sample. Amplicon libraries of 16S rRNA gene V1-V2 hypervariable regions were generated in triplicate using universal primers 8F 5'agagtttgatcmtggctcag3' and 361R 5'cyiactgctgcctcccgtag3' (Sundquist et al 2007), which included Roche's 454 Lib-A adapters A and B and a unique multiplex identifier. PCR and library preparation procedures have also been previously described (Diaz et al 2012). Briefly, PCR reactions contained 10 ng of purified DNA, 1 U platinum *iTaq* polymerase (Invitrogen, Carsbad, CA), 1.5 mM $MgCl_2$, 200 μM dNTPs, *iTaq* buffer (1x), 0.5 μM of each forward and reverse primer and molecular grade water to a final volume of 25 μL. Thermal cycler conditions were: initial step at 95°C for 3 minutes; 25 cycles of denaturation at 95°C for 30 seconds, annealing at 50°C for 30 seconds and extension at 72°C for 1 minute; and a final extension step at 72°C for 9 minutes. Negative controls included a DNA isolation control and a PCR control with no added template. Combined triplicate amplicon libraries were sequenced using 454 Titanium chemistry and the 454-GS-FLX sequencing platform (454 Life Sciences, Branford, CT).

Specific details on real-time PCR assays

Standard curves were constructed using 10-fold dilutions of genomic DNA of representative species ranging from 10^2 to 10^8 16S rRNA gene molecules. Genus specific primers included: *Streptococcus* (5'cgacgatacatagccgacctgag3' and 5'tccattgccgaagattccctactg3') (Chalmers et al 2008), *Veillonella* (5'ccgtgatgggatggaaactgc3' and 5'ccttcgccactggtgttcttc3') (Chalmers et al 2008) and *Actinomyces* (5'ggctgcgataccgtgagg3' and 5'tctgcgattactagcgactcc3') (Periasamy et al 2009). Universal primers used to quantify total bacterial load were 5'-tcctacgggaggcagcagt -3' and 5'-ggactaccagggtatctaatcctgtt-3'(Nadkarni et al 2002). Primers were evaluated for their coverage and specificity using ARB (Ludwig et al 2004). We further checked non-specific amplification of genus-specific primers by assessing their ability to amplify genomic DNA from oral species present in our pyrosequencing datasets and showing 2 or less mismatches to the primer pair used. These controls included testing *Streptococcus* primers against *Staphylococcus aureus* ATCC 9144, *Gemella haemolysans* ATCC 10379, *Gemella morbillorum* ATCC 27824, *Abiotrophia defectiva* ATCC 49176 and *Granulicatella adjacens* ATCC 49175. *Veillonella* primers were tested against *Selenomonas sputigena* ATCC 35185, while no taxa had 2 or less mismatches to the *Actinomyces* primers. None of these taxa yielded a PCR product in these assays. Each PCR reaction was performed in a total volume of 20 µl, containing 10 µl of Fast Power SYBR green PCR Master Mix (Applied Biosystems, Foster City, CA), 500 nM of each primer pair, 1 µl DNA template and PCR water. Thermocycler conditions included a denaturation step of 95°C (20 s) and 40 cycles of 95°C (3 s) and 58°C (30 s). Real time PCR amplicon products were analyzed via capillary electrophoresis (Experion) to evaluate presence of non-specific bands. All

samples and standard curves were amplified in triplicate and the mean value of the targeted molecule was used for analysis. Limit of detection for all assays was 10^2 copies.

Sequencing data processing

Sequencing data were processed as previously described (Diaz et al 2012) using MOTHUR (Schloss et al 2009), including steps for denoising, trimming and chimera removal. Individual sequences were assigned to a taxonomy via the RDP classifier (Wang et al 2007), as implemented in MOTHUR, using a bootstrapping threshold of 80%. Template taxonomies included the large RDP reference dataset and the Human Oral Microbiome Database (HOMD), a curated dataset for oral taxa (Dewhirst et al 2010). For Operational Taxonomic Unit (OTU)-based analysis, sequences were clustered using the average neighbor algorithm (Schloss and Westcott 2011) and a 3% dissimilarity cutoff. OTUs were assigned a taxonomy based on the consensus taxonomic assignment for the majority of sequences within each OTU. If a consensus taxonomy was not possible at the species level (based on HOMD), the nearest taxonomical level at which a consensus was reached was reported. In such cases, the representative sequence from the OTU was also compared to the HOMD and if results showed more than 97% similarity to an Oral Taxon (OT), the OT name of the top hit was reported in parentheses as part of the OTU taxonomy. Individually classified sequences were also grouped into phylotypes (from genus to phylum level) based on taxonomic identity.

Sequence libraries were sub-sampled to contain the same number of sequences for α-diversity comparisons. Richness was evaluated by the number of observed OTUs and number of estimated OTUs as calculated with CatchAll (Bunge 2011). Evenness was measured as $E_{Shannon}=D_{Shannon}/\ln(S)$ and diversity was measured by the non-parametric Shannon Index (Chao and Shen 2003).

β-diversity was measured with the Jaccard Index for comparison of communities based on membership and the θ_{YC} distance (Yue and Clayton 2005) for comparison of communities based on structure. A phylogenetic tree was also constructed with CLEARCUT (Evans et al 2006) and communities were compared using UNIFRAC weighted and unweighted metrics (Lozupone and Knight 2005). Principal Coordinate Analysis (PCoA) was performed in MOTHUR and graphs visualized using the rgl application within R (http://www.r-project.org/).

References

Bunge J (2011). Estimating the number of species with catchall. *Pac Symp Biocomput***:** 121-130.

Chalmers NI, Palmer RJ, Jr., Cisar JO, Kolenbrander PE (2008). Characterization of a *Streptococcus* sp.-*Veillonella* sp. community micromanipulated from dental plaque. *J Bacteriol* **190:** 8145-8154.

Chao A, Shen T-J (2003). Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environmental and Ecological Statistics* **10:** 429-443.

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.

Diaz PI, K. DA, Abusleme L, Reese B, Obergfell C, Choquette L, Dongari-Bagtzoglou A, Peterson DE, Terzi E, Strausbaugh LD (2012). Using high throughput sequencing to explore the biodiversity in oral bacterial communities. *Mol Oral Microbiol* **27:** 182-201.

Evans J, Sheneman L, Foster J (2006). Relaxed neighbor joining: a fast distance-based phylogenetic tree construction method. *J Mol Evol* **62:** 785-792.

Lozupone C, Knight R (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71:** 8228-8235.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Forster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, Konig A, Liss T, Lussmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer KH (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32:** 1363-1371.

Nadkarni MA, Martin FE, Jacques NA, Hunter N (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* **148:** 257-266.

Periasamy S, Chalmers NI, Du-Thumm L, Kolenbrander PE (2009). *Fusobacterium nucleatum* ATCC 10953 requires *Actinomyces naeslundii* ATCC 43146 for growth on saliva in a three-species community that includes *Streptococcus oralis* 34. *Appl Environ Microbiol* **75:** 3250-3257.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537-7541.

Schloss PD, Westcott SL (2011). Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl Environ Microbiol* **77:** 3219-3226.

Sundquist A, Bigdeli S, Jalili R, Druzin ML, Waller S, Pullen KM, El-Sayed YY, Taslimi MM, Batzoglou S, Ronaghi M (2007). Bacterial flora-typing with targeted, chip-based Pyrosequencing. *BMC Microbiol* **7:** 108.

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.

Yue JC, Clayton MK (2005). A similarity measure based on species proportions. *Communications in Statistics Theory and Methods* **34:** 2123-2131.