

1       **Microbial Signatures of Oral Dysbiosis, Periodontitis and Edentulism**  
2                               **Revealed by Gene Meter Methodology**

3  
4                               M. Colby Hunter<sup>1</sup>, Alex E. Pozhitkov<sup>2</sup>, and Peter A. Noble<sup>#3</sup>

5  
6 \*Corresponding author, Peter A Noble, Email: [panoble@u.washington.edu](mailto:panoble@u.washington.edu)

7  
8 Authors' affiliations:

- 9  
10       1. Program in Microbiology, Alabama State University, Montgomery, AL 36101  
11       2. Department of Oral Health, University of Washington, Box 3574444, Seattle, Washington  
12           98195-7444 Ph: 206-409-6664  
13       3. Department of Periodontics, University of Washington, Box 3574444, Seattle, Washington  
14           98195-7444 Ph: 206-409-6664

15  
16 Authors' emails:

17  
18 Hunter: [cmooghww@gmail.com](mailto:cmooghww@gmail.com)

19 Pozhitkov: [pozhit@uw.edu](mailto:pozhit@uw.edu)

20 Noble: [panoble@u.washington.edu](mailto:panoble@u.washington.edu)

21  
22

23 **ABSTRACT (N=305 WORDS)**

24

25 Conceptual models suggest certain microorganisms (e.g., the red complex) are indicative of a specific  
26 disease state (e.g., periodontitis); however, recent studies have questioned the validity of these models.  
27 Here, the abundances of 500+ microbial species were determined in 16 patients with clinical signs of  
28 one of the following oral conditions: periodontitis, established caries, edentulism, and oral health. Our  
29 goal was to determine if the abundances of certain microorganisms reflect dysbiosis or a specific clinical  
30 condition that could be used as a signature for dental research. Microbial abundances were determined  
31 by the analysis of 138,718 calibrated probes using Gene Meter methodology. Each 16S rRNA gene was  
32 targeted by an average of 194 unique probes ( $n=25$  nt). The calibration involved diluting pooled gene  
33 target samples, hybridizing each dilution to a DNA microarray, and fitting the probe intensities to  
34 adsorption models. The fit of the model to the experimental data was used to assess individual and  
35 aggregate probe behavior; good fits ( $R^2>0.90$ ) were retained for back-calculating microbial abundances  
36 from patient samples. The abundance of a gene was determined from the median of all calibrated  
37 individual probes or from the calibrated abundance of all aggregated probes. With the exception of  
38 genes with low abundances ( $< 2$  arbitrary units), the abundances determined by the different calibrations  
39 were highly correlated ( $r \sim 1.0$ ). Seventeen genera were classified as signatures of dysbiosis because  
40 they had significantly higher abundances in patients with periodontitis and edentulism when contrasted  
41 with health. Similarly, 13 genera were classified as signatures of periodontitis, and 14 genera were  
42 classified as signatures of edentulism. The signatures could be used, individually or in combination, to  
43 assess the clinical status of a patient (e.g., evaluating treatments such as antibiotic therapies).  
44 Comparisons of the same patient samples revealed high false negatives (45%) for next-generation-  
45 sequencing results and low false positives (7%) for Gene Meter results.

46

47 **KEYWORDS**

48 Gene Meter; calibrated DNA microarrays; periodontitis, edentulism, caries, 454 sequencing, next-  
49 generation sequencing, microbial abundance signatures.

50

51

## 52 INTRODUCTION

53 Periodontitis is an inflammatory disease associated with the tissues that support the teeth. The disease  
54 causes the progressive loss of bone that could result in tooth loss (1). A small assemblage of bacterial  
55 species was once thought to be the cause of periodontal disease. This assemblage consisted of  
56 *Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola*, and was ominously dubbed  
57 the “red complex” (2,3,4). The prevailing idea at the time was that the red complex changed the  
58 signaling pathways of the host. This change made the host vulnerable and led to a change in the relative  
59 abundance of microorganisms thus causing the disease (5,6). However, members of the red complex  
60 have been discovered in people who do not have periodontal disease (8,7,9). Therefore, members of the  
61 red complex could be classified as pathogens or harmless commensals. This fact, along with the finding  
62 that the oral microbiome of humans is much more diverse than previously thought (10,11), has prompted  
63 a paradigm shift in the understanding of the origins of periodontal disease. This new paradigm states  
64 that periodontal disease is not caused by the presence of specific bacteria, but by changes in the  
65 population levels of species in the oral microbiome (12). What begins as a symbiosis between host and  
66 microbes becomes a dysbiosis as the microorganisms transcend beyond the host-imposed boundaries.  
67 The premise of our study is that a change in microbial populations can be determined by the relative  
68 abundances of individual species in patients with different clinical conditions.

69 Oligonucleotide microarray technology has been used to profile microbial communities for quite some  
70 time (13,14,15,16,17). A new version of the methodology, the Gene Meter, was recently introduced  
71 (although not explicitly named “Gene Meter” in refs. 18,19). Unlike conventional DNA microarrays, a  
72 calibration procedure is conducted prior to applying samples onto the microarrays (20,21). The  
73 calibration procedure establishes the microarray probe responses to increasing target concentrations (i.e.,  
74 a dilution series). Individual probes are calibrated or, alternatively, groups of probes collectively  
75 specific to a target (i.e., aggregates) are calibrated. Both approaches of calibration were compared in  
76 this study.

77 Next generation sequencing (NGS) has also been used to profile oral microbial communities (7,9,22,23,  
78 24,25,26,27). This approach involves the massive sequencing of thousands to millions of DNA strands  
79 in a single run (28). While microbial abundances may be quantified by NGS, it remains to be  
80 determined if the quantifications truly represent the quantities of the genes in a biological sample. Only  
81 a limited number of studies have rigorously challenged their quantitative capabilities. In one study, a  
82 complex microbial mixture was spiked with 5 fungal targets including *L. proxima* and *M. noctilunum*  
83 (29). It was shown that depending on the target, the coefficient of determination for the relationship  
84 between actual quantities of the target and its sequencing counts was very low. In another recent study,  
85 NGS is shown to be a semi-quantitative approach for examining fungal populations (30). The results of  
86 the study showed that 454 pyrosequencing counts are, at best, moderately related with spore  
87 concentration measurements by qPCR ( $R^2 \sim 0.5$ ). An RNA-seq spike-in study has also been conducted  
88 (31). Ninety-six RNA sequences of concentrations varying 6 orders of magnitude were mixed and  
89 sequenced and the concentrations were correlated to the number of reads. Although the study design did  
90 not investigate the response of an individual target to concentration, the results suggest that the RNA-seq  
91 sensitivity can vary up to 10-fold depending on the gene target (31). The uncertainty associated with  
92 quantification by sequencing raises questions about the validity of such quantifications and warrants a  
93 well-controlled calibration study, which was done here.

94 The objectives of the study are three-fold: (i) to demonstrate the utility of the Gene Meter approach to  
95 precisely determine 16S rRNA gene abundances using two calibration approaches, (ii) to determine if

96 certain microorganisms have an abundance signature that can be employed to identify specific or general  
97 oral health conditions, and (iii) to compare and contrast two different technologies (Gene Meter and  
98 DNA sequencing) using the same 16 patient samples.

## 99 **MATERIALS AND METHODS**

100 The DNA sequencing, the study design, and sample collection methods have been previously published  
101 (32). The same amplified DNA used for DNA sequencing was also analyzed by the Gene Meter  
102 approach. Splitting the samples in this way enabled direct comparison of results from two distinctly  
103 different molecular methods. To familiarize readers with the previous study, we have provided a brief  
104 overview of patient recruitment, enrollment, and exclusion criteria, sample collection, DNA extraction  
105 and amplification, and DNA sequencing analyses, below.

106 **Patient recruitment, enrollment, and exclusion criteria.** The adult patients were recruited at the  
107 University of Washington, Seattle, USA, and the University of Düsseldorf, Germany and enrolled if they  
108 had one of the following clinical conditions: severe periodontitis, caries, edentulism, and oral health. A  
109 periodontitis case was defined as having at least two interproximal sites at different teeth with clinical  
110 attachment loss (CAL) of 6 mm or greater and at least one interproximal site with probing depth (PD) of  
111 5 mm or greater (33) and a minimum of 20 permanent teeth, not including third molars. Patients were  
112 excluded from the periodontitis group if they had any established caries lesion or wore a removable  
113 partial denture. A caries case was defined as having the following number of teeth with established  
114 caries lesions: 6 or more teeth in subjects 20 to 34 years of age; 4 or more teeth in subjects 35 to 49  
115 years of age; and 3 or more teeth in subjects 50 years of age and older. Established caries was defined as  
116 a class 4 lesion according to the International Caries Detection and Assessment System. The number of  
117 teeth with caries lesion in caries cases was greater than one standard deviation above the mean of caries  
118 extent in respective age group the U.S.A (34). Exclusion criteria for a caries case were interproximal  
119 sites with CAL of 4 mm or greater or PD of 5 mm or greater (33). An edentulous case had to be  
120 completely edentulous in both jaws and their teeth had to be extracted more than one year before the  
121 enrollment in the study. A healthy case was defined as having 28 teeth, not counting third molars, or 24  
122 or more teeth, not counting third molars if premolars had been extracted for orthodontic reasons or were  
123 congenitally missing with no signs of oral disease. Exclusion criteria for a healthy case included:  
124 smoking, loss of permanent teeth due to caries or periodontal disease, any interproximal sites with CAL  
125 of four or greater or PD of 5 mm or greater, or any established caries lesions. Exclusion criteria for all  
126 groups included: oral mucosal lesions, systemic diseases, and use of antibiotics or local antiseptics  
127 within 3 months prior to the study.

128 **Sample collection, DNA extraction and PCR amplification.** For all but the edentulous patients, supra-  
129 and subgingival plaque was collected from sites with the deepest probing depth in each sextant. One  
130 sterile paper point per site was inserted into the deepest aspect of the periodontal pocket or gingival  
131 sulcus. Biofilm from oral mucosae was collected by swiping a sterile cotton swab over the epithelial  
132 surfaces of the lip, left and right buccal mucosae, palate, and dorsum of the tongue (35). Samples were  
133 stored at -80°C.

134 Microbial DNA was isolated from cells by physical and chemical disruption using zirconia/silica beads  
135 and phenol-chloroform extraction in a FastPrep-24 bead beater (36). Prokaryotic 16S rRNA genes were  
136 amplified using universal primers (27F and 1392R) using the GemTaq kit from MGQuest (Cat# EP012).  
137 The PCR program involved a pre-amplification step of 10 cycles with an annealing temperature of 56°C  
138 followed by 20 amplification cycles with annealing temperature 58°C. In each cycle, elongation time  
139 was 1 min 10s, at 72°C. PCR was finalized by extended elongation for 5 min. PCR products were

140 purified with DNA Clean & Concentrator columns (Zymo Research, USA) and quantified using the  
141 NanoDrop (Agilent, USA). Equal quantities of PCR product derived from swab and paper point samples  
142 were pooled together for each patient. For edentulous patients, there were no paper point samples. Each  
143 purified PCR product was sequenced on a Roche 454 Jr. instrument as previously described (32).

144 **DNA sequencing analysis.** The obtained sequences were uploaded to the MG-RAST web server (37).  
145 The MG-RAST pipeline assessed the quality of sequences, removed short sequences (multiplication of  
146 standard deviation of length cutoff of 2.0) and removed sequences with ambiguous bp (non-ACGT;  
147 maximum allowed number of ambiguous base pair was set to 5). The pipeline annotated the sequences  
148 and allowed the integration of the data with previous metagenomic and genomic samples. The RDP  
149 database was used as an annotation source, with minimum sequence identity of 97%, maximum e-value  
150 cutoff of  $10^{-5}$ , and minimum sequence length of 100 nt.

151 **Microarray probe design.** Microarray probes were obtained by tiling along 16S rRNA sequences of  
152 597 oral bacteria (Table S1) using a program written for this purpose. Redundant probes were removed.  
153 Each probe was 25 nt in length. The 16S rRNA gene sequences used to design the probes ranged from  
154 424 to 13,214 nt in length with an average of 1,492 nt. Three of the 16S rRNA sequences also included  
155 5S and 23S rRNA genes. The microarrays were created by NimbleGen (now Roche Inc.).

156 **Sample labeling.** PCR products (above) were purified using a “DNA Clean & Concentrator” kit (Zymo  
157 Research, USA), dried under a flow of dry nitrogen and labeled using ULYSIS direct chemical labeling  
158 kit. The attached dye was Alexa Fluor 546 (i.e., spectral analog of Cy3).

159 **DNA microarray calibration.** Probes on the microarrays were calibrated using a dilution series of  
160 labeled pooled PCR products of all samples. The dilution series for the NimbleGen array was created  
161 using the following quantities of labeled PCR amplicons: 11.99, 7.72, 3.09, 1.54, 0.77, and 0.39  $\mu$ g in  
162 12.5  $\mu$ l. Signal intensities were collected and stored in a database. Two independent calibration  
163 procedures were conducted i.e., calibration of individual probes and calibration of probe aggregates. An  
164 automatic fitting procedure determined the best fit curve – the calibration curve – (e.g., Langmuir,  
165 Linear or Freundlich) as well as the curve parameters from the dilutions and signal intensities. For the  
166 individual probes calibration, signal intensity was that of the individual probe. For the probe aggregate  
167 calibration, signal intensity was a sum of signal intensities of the probes specific to the corresponding  
168 targets. To calculate the abundances of the targets, the calibration curve equations were inverted such  
169 that from the signal intensity a dilution factor could be obtained. The dilution factor is proportional to  
170 the concentration of the target; hence, it was defined as the abundance of the target.

171 Details of the calibration protocols to calculate gene abundance are provided in our recent papers  
172 (18,19). The quantities of samples loaded onto the microarray of processing of the individual patients  
173 were  $3.2 \pm 0.25$   $\mu$ g.

174 **Bioinformatic analyses.** Matching the probes to the targets (*in silico*) and determining the gene  
175 abundances using the unique calibrated probes was conducted using custom-designed programs written  
176 in C++. T-tests were analyzed and histograms were made using SAS JMP 7 and MS Excel.

## 177 **RESULTS**

178 **Microarray probe coverage and selection.** In total, 276,234 probes (25 nt) were synthesized on  
179 NimbleGen microarrays. *In silico* matches of the probes to all 597 oral bacterial sequences (used to  
180 design the probes) revealed 175,206 probes were unique to a single 16S rRNA gene target. To  
181 determine if the microarray design had sufficient coverage, we downloaded the curated core oral

182 microbiome 16S rRNA gene sequence dataset from <http://microbiome.osu.edu> and matched the unique  
183 probes to the 1,045 microbial sequences (38). We found that 66,716 probes were unique to the 1,021  
184 microorganisms in the core oral database. In other words, the unique probes on our microarray had  
185 97.7% coverage of the sequences in the curated core oral microbiome.

186 We further refined the “concept of unique probes” by removing 3 nucleotides on either end (5’- and 3’-)  
187 of the probe (*in silico*) and matching the shorter probes to the 597 16S rRNA bacterial sequences. The  
188 reason for removing these nucleotides was that nucleotides at either end could potentially cross-  
189 hybridize to non-specific targets in the microarray experiments (39). In this experiment, probes were  
190 considered unique if the 19 nt core matched one 16S rRNA gene target in the dataset (i.e., 597 oral  
191 bacterial sequences). Using the 19 nt core, we found 36,488 of the original 175,206 probes could be  
192 susceptible to cross-hybridization, which left 138,718 unique probes that were used in all subsequent  
193 analyses to determine gene abundance. Therefore, the entire sequence (25 nt) of the 138,718 unique  
194 probes were used in all subsequent experiments.

### 195 **Calibration of microarray probes.**

196 All probes were calibrated using a dilution series as outlined in our previous studies (18,19,20,21). The  
197 Langmuir and Freundlich isotherm equations were fitted to the data and the one with the best fit ( $R^2$ )  
198 was retained. Examples of two calibrated probes are shown in [Figure 1](#). Briefly, the fitting algorithm  
199 transforms the data to obtain a straight line; two parameters,  $a$  and  $b$ , are calculated. Depending on the  
200 type of the curve corresponding to the best fit, parameters  $a$  and  $b$  are utilized according to the formula  
201 of the curve. The parameters of the retained equations and fits are shown in [Figure 1](#). Apparently, the  
202 signal intensities of the dilution data for Probe 62 were best explained by Langmuir and the dilution data  
203 for Probe 66 were best explained by Freundlich. For the Freundlich equation,

$$204 \quad y = e^a x^b,$$

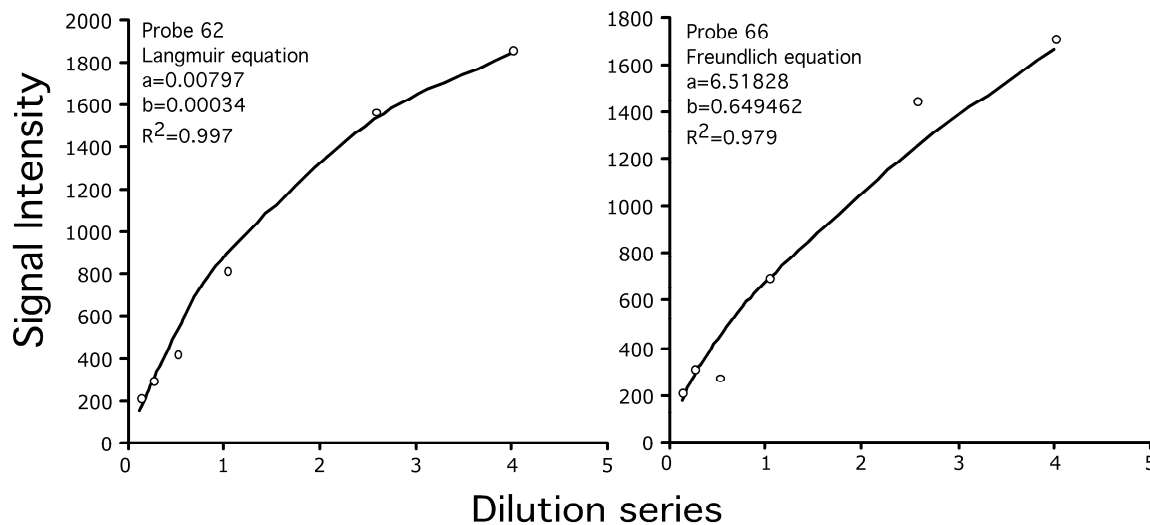
205 for the Langmuir equation,

$$206 \quad y = \frac{b^{-1}(b/a)x}{1+(b/a)x},$$

207 where  $y$  is the signal intensity, and  $x$  is the dilution.

208





209

Figure 1. Calibration curves of two probes. Probe 62 was best calibrated using the Langmuir equation while probe 66 was best calibrated by the Freundlich equation. The equations can be used to back-calculate target concentration (dilution series) from signal intensity.

210

211 These equations were used to back-calculate the gene target concentration from signal intensity values.  
 212 In the case of Probe 62 at PMT=600,  $a=0.00797$ ,  $b=0.00034$ ,  $x$ =dilution factor in the calibration series,  
 213 the signal intensity ( $SI$ ) responds as follows:  $SI=2906.757*0.431679*x/(1+0.431679*x)$ . Inverting  
 214 this equation allows one to calculate the abundance of a gene targeted by this probe. For example, for  
 215 one of the healthy patients, signal intensity was 1793.89 RFU. This yields a gene abundance of  $1793.89$   
 216  $/((2906.757-1793.89)*0.431679)=3.7$  a.u.

217 In the case of Probe 66 at PMT=600 where  $a=6.51828$ ,  $b=0.649462$ ,  $x$ =dilution factor in the calibration  
 218 series, the  $SI$  responds according to Freundlich equation as  $SI=\exp(6.51828)*x^{0.649462}$ . Inverting this  
 219 equation allows one to calculate the abundance of a gene targeted by this probe. For example, for one of  
 220 the healthy patients,  $SI=1268.67$  RFU. This yields a gene abundance of  $(1268.67 /$   
 221  $\exp(6.51828))^{1/0.649462}=2.6$  a.u.

222 Probe aggregates were calibrated same way as individual probes. For the probe aggregates, only linear  
 223 (i.e.,  $y=ax+b$ ) and Freundlich curves were evaluated.

### 224 Individually calibrated probes

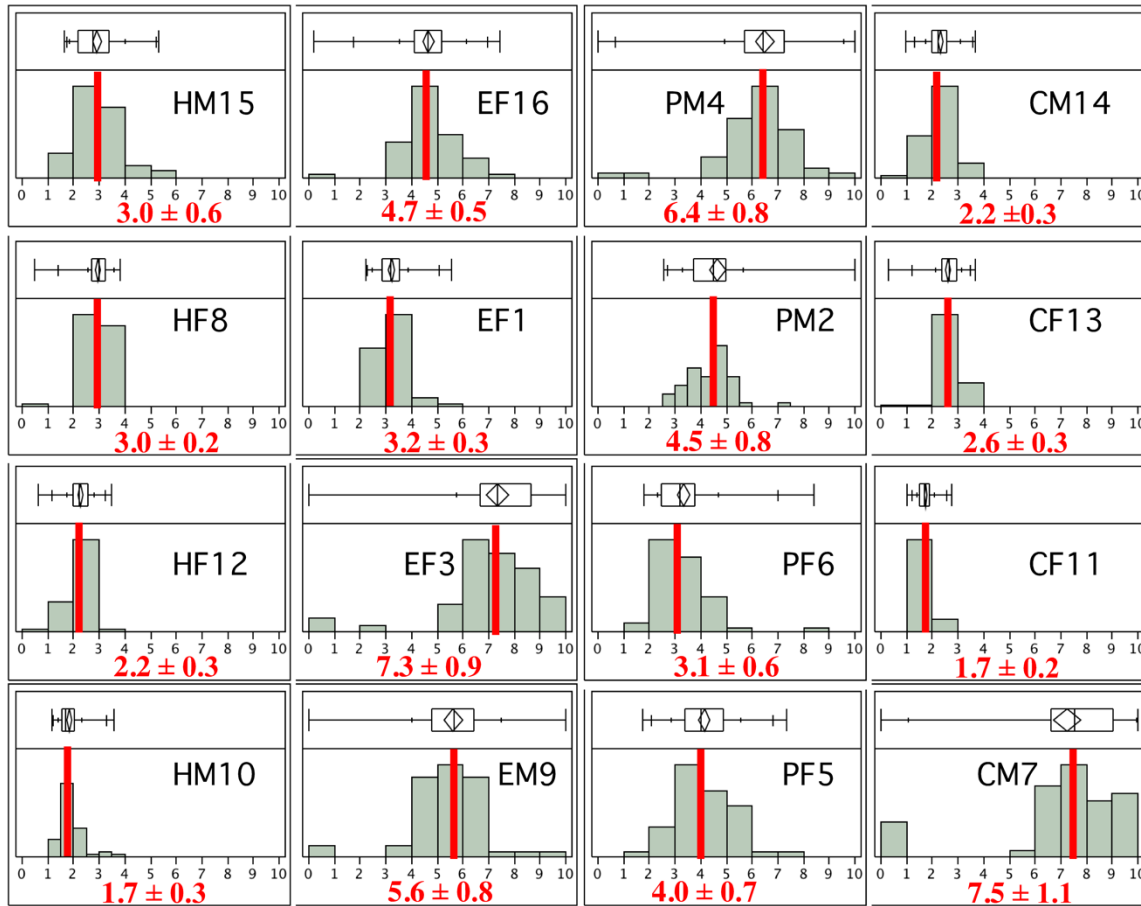
225 Three PMT settings (i.e., 500, 600, 700) were used to assess the signal intensities of the hybridized  
 226 duplexes because we did not know *a priori* which settings were optimal (i.e., the setting that yielded the  
 227 minimum signal saturation with maximum dynamic range). The probes at each setting were  
 228 independently calibrated. In total, 246,232, 246,233, and 207,006 probes were calibrated for the  
 229 microarrays with the 500, 600, and 700 PMT settings, respectively. In terms of percent of the total  
 230 probes, about 89% of the probes were calibrated at the 500 and 600 PMT settings and about 75% of the  
 231 probes at the 700 setting. The lower percent of calibrated probes at the 700 PMT setting was due to  
 232 saturation of the probes.

233 Median abundances  $\pm$  median absolute deviation (MAD) for the 576 rRNA genes and 16 patient  
 234 samples was assembled into a dataset (Table S2). Note that 21 of the original 597 gene targets were not  
 235 targetted by the unique probes. The average number of calibrated probes per 16S rRNA gene was  $194 \pm$

236 214 (avg.  $\pm$  s.d.). The unique probes not yielding abundances for any of the 16 patients were excluded.  
237 Any gene abundance that was greater than 10 a.u. was set to 10 a.u. because we did not expect the  
238 calibrations to forecast accurately beyond the dilution series used to create it. Moreover, 10.0 a.u.  
239 represents 10 times the amount of DNA hybridized to a microarray based on manufacturer's  
240 recommendation. The average gene abundance for the 500, 600, and 700 datasets was  $3.5 \pm 2.0$ ,  $4.1 \pm$   
241  $2.1$ , and  $3.5 \pm 2.0$  arbitrary units (a.u.), respectively. The lowest and highest gene abundances for all  
242 datasets were 0.3 and 10.0 a.u., respectively. The significance of these findings is that gene abundances  
243 could vary by as much as 30-fold and the averaged gene abundances were not significantly affected by  
244 PMT settings. Based on these results, all subsequent analyses of the individualized probes were  
245 performed using the 600 PMT data. The range of abundances for these data was 0.3 to 10 a.u.

246 **Examples: Selected microbial abundances.** Statistical parameters (averages, standard deviations (s.d.),  
247 quartiles, and median) for all 576 rRNA genes in 16 patients were calculated and two examples are  
248 provided below. The histograms and whisker plots of the *Tannerella forsythia* 16S rRNA gene show the  
249 binned distribution of gene abundances in 16 patients based on 100 unique probes (Figure 2). The grey  
250 bars represent the frequencies of the binned abundances and the red bars represent the median  
251 abundances. The whisker plots show the mean, standard deviations, median, and quartiles of the  
252 abundances. The median abundance ( $\pm$  MAD) in the 4 patients with health was  $3.0 \pm 0.6$  a.u.,  $3.0 \pm 0.2$   
253 a.u.,  $2.2 \pm 0.3$  a.u., and  $1.7 \pm 0.3$  a.u. (samples HM15, HF8, HF12 and HM10, respectively). The  
254 average gene abundance (average  $\pm$  s.d.) for patients with health was  $2.4 \pm 0.6$  a.u. The abundance of  
255 this gene in the 4 patients with edentulism, periodontitis, or caries was:  $5.2 \pm 1.7$  a.u.,  $4.5 \pm 1.4$  a.u., and  
256  $3.5 \pm 2.7$  a.u., respectively. Two-tailed T-tests revealed significant differences in the average  
257 abundances for the 4 patients with edentulous and 4 patients with health ( $P < 0.05$ ), and the 4 patients  
258 with periodontitis and 4 patients with health ( $P < 0.05$ ). None of the other paired conditions for this gene  
259 were significantly different from one another (Table 1).





260

Figure 2. Distribution of 16S rRNA gene abundances for *Tannerella forsythia* (GI 289382) in 16 patients based on 50 calibrated probes. Each patient has an identifier. The first letter indicates the condition: H, health; E, edentulism; P, periodontitis; C, caries. The second letter indicates patient sex: M, male; F, female. The third number indicates patient number. The red bar indicates median value. Corresponding median absolute deviation (MAD) values are shown. Two-tailed T-tests showed that *T. forsythia* was at significantly higher abundance in patients with periodontitis (Average  $\pm$  s.d.;  $4.5 \pm 1.4$  a.u) than those in health ( $2.4 \pm 0.6$  a.u) ( $P < 0.05$ ).

261

Table 1. Bacterial species having significantly different abundances for patients with health versus those with edentulism based on individual probes. *n\_probes*, number of probes used to determine the species abundance; Two-sided T-tests were based on four patients with health and four patients with edentulism and  $\alpha = 0.05$ .

Phylum/Class	Genus/Species/GI number	n_probes	Abundance ( $\bar{X} \pm$ s.d.) by:	
			Health	Edentulism
Actinobacteria/Actinobacteria	<i>Actinomyces meyeri</i> (1838945)	74	2.4 $\pm$ 0.5	5.1 $\pm$ 1.7
	<i>Actinomyces odontolyticus</i> (853707)	36	2.2 $\pm$ 0.5	4.4 $\pm$ 1.4
	<i>Actinomyces odontolyticus</i> strain 20536T (6997974)	31	2.4 $\pm$ 0.4	5.5 $\pm$ 1.7

	<i>Actinomyces sp.</i> oral strain Hal-1065 (14537929)	168	2.3 ± 0.4	4.7 ± 1.1
Actinobacteria/Coriobacteria	<i>Atopobium rimae</i> (10719609)	22	2.7 ± 0.6	6.1 ± 2.1
Bacteroidetes/Bacteroidetes	<i>Bacteroides ureolyticus</i> (173919)	387	2.7 ± 0.7	5.6 ± 2.0
	<i>Porphyromonas endodontalis</i> ATCC 35406 (294287)	89	2.7 ± 0.8	5.6 ± 1.9
	<i>Prevotella nigrescens</i> ATCC 33563 (294425)	57	2.4 ± 0.6	5.1 ± 1.8
	<i>Prevotella sp.</i> oral clone AH125 (9988924)	78	2.4 ± 0.6	5.0 ± 1.7
	<i>Prevotella sp.</i> oral clone AO096 (9988919)	98	2.6 ± 0.6	5.4 ± 1.7
	<i>Prevotella sp.</i> oral clone AU069 (9988926)	148	2.6 ± 0.6	5.3 ± 1.8
	<i>Prevotella veroralis</i> ATCC 33779 (294427)	41	2.8 ± 0.8	5.6 ± 1.9
	<i>Tannerella forsythia</i> (289382)	50	2.4 ± 0.6	5.2 ± 1.7
Bacteroidetes/Flavobacteria	<i>Capnocytophaga sp.</i> oral clone BU084 (14537978)	65	2.6 ± 0.6	5.5 ± 1.9
Bacteroidetes/Flavobacteria	<i>Capnocytophaga sp.</i> oral clone X066 (9988942)	44	2.6 ± 0.7	5.7 ± 1.9
Firmicutes/ Bacilli	<i>Streptococcus anginosus</i> strain ATCC33397 (4406240)	53	2.5 ± 0.6	5.3 ± 1.7
	<i>Streptococcus salivarius</i> (176047)	28	2.1 ± 0.4	4.5 ± 1.1
	<i>Streptococcus sp.</i> oral clone FO042 (16612189)	6	2.4 ± 0.6	5.5 ± 1.9
	<i>Streptococcus sp.</i> oral clone FP015 (16612190)	61	2.2 ± 0.5	4.9 ± 1.7
	<i>Streptococcus thoraltensis</i> (2578813)	275	2.6 ± 0.7	5.5 ± 1.9
	<i>Lactobacillus paracasei subsp. paracaseie</i> (1808583)	79	2.5 ± 0.7	5.2 ± 1.7
	<i>Mogibacterium vescum</i> strain:ATCC 700697 (4140242)	83	2.6 ± 0.7	5.6 ± 1.9
Firmicutes/Clostridia	<i>Catonella morbi</i> (1067117)	118	2.5 ± 0.6	5.1 ± 1.7
	<i>Eubacterium sp.</i> oral clone BS091 (14537971)	80	2.7 ± 0.8	5.4 ± 1.8
	<i>Eubacterium sp.</i> oral clone EH006 (14537979)	106	2.7 ± 0.7	5.9 ± 2.1
	<i>Eubacterium sp.</i> oral strain A35MT (9837454)	41	2.5 ± 0.7	5.5 ± 1.8
	<i>Peptostreptococcus sp.</i> oral clone BS044 (14537968)	79	2.6 ± 0.6	5.5 ± 1.7
	<i>Selenomonas sp.</i> oral clone EW076 (14537908)	42	2.4 ± 0.6	4.9 ± 1.6
	<i>Selenomonas sp.</i> oral clone EW084 (14537910)	91	2.7 ± 0.7	5.7 ± 2.0

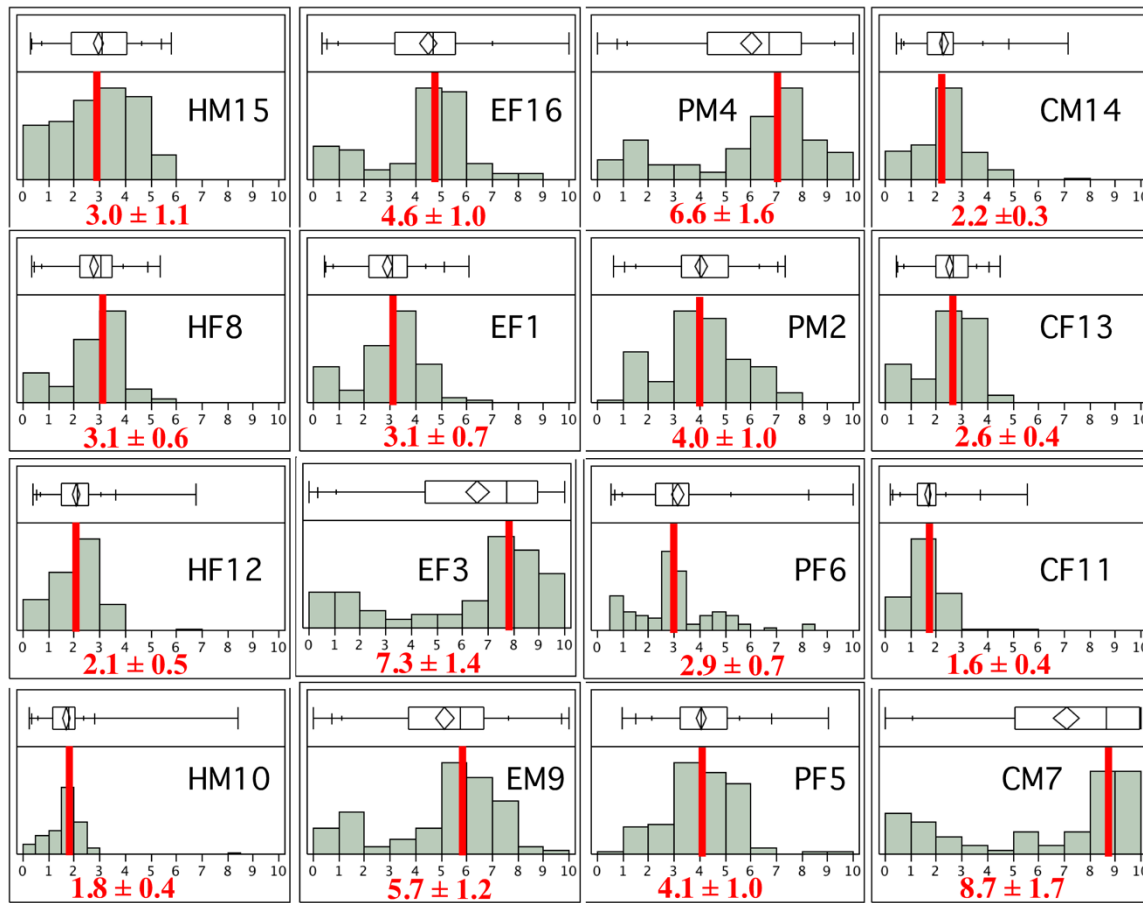
	<i>Selenomonas</i> -like sp. oral strain GAA14 (9837482)	129	2.5 ± 0.6	5.4 ± 1.9
Firmicutes/Cocci	<i>Staphylococcus saccharolyticus</i> (576608)	46	2.4 ± 0.6	5.1 ± 1.8
Firmicutes/Erysipelotrichia	<i>Solobacterium</i> sp. oral clone K010 (9988916)	113	2.7 ± 0.7	5.5 ± 1.9
Firmicutes/Mollicutes	<i>Mycoplasma faucium</i> strain DC333(T) (4883891)	176	2.6 ± 0.7	5.5 ± 1.8
Firmicutes/Negativicutes	<i>Dialister</i> sp. oral clone BS095 (9837480)	99	2.6 ± 0.7	5.3 ± 1.7
	<i>Dialister</i> sp. oral strain GBA27 (9837481)	63	2.7 ± 0.7	5.4 ± 1.8
	<i>Megasphaera</i> sp. oral clone BU057 (14537975)	305	2.5 ± 0.7	5.3 ± 1.8
	<i>Veillonella dispar</i> (793883)	40	2.4 ± 0.6	5.1 ± 1.7
Fusobacteria/Fusobacteria	<i>Leptotrichia</i> sp. oral clone DR011 (14537926)	215	2.6 ± 0.6	5.2 ± 1.7
	<i>Leptotrichia</i> sp. oral clone FB074 (14537951)	114	2.5 ± 0.6	5.5 ± 1.9
Proteobacteria/Betaproteobacteria	<i>Comamonas</i> sp. isolate 158 (3046563)	33	2.6 ± 0.7	5.3 ± 1.7
	<i>Kingella denitrificans</i> (174983)	63	2.6 ± 0.7	5.3 ± 1.7
	<i>Kingella oralis</i> (174982)	184	2.5 ± 0.6	5.3 ± 1.8
	<i>Neisseria</i> sp. oral strain B33KA (10048304)	74	2.7 ± 0.7	5.8 ± 2.0
	<i>Neisseria subflava</i> strain U37 (5327189)	12	2.8 ± 0.9	6.5 ± 1.8
Proteobacteria/Gammaproteobacteria	<i>Klebsiella oxytoca</i> (14549203)	65	2.5 ± 0.5	5.5 ± 1.9
	<i>Klebsiella pneumoniae</i> strain ATCC13884T (3282032)	8	2.3 ± 0.6	4.8 ± 1.5
	<i>Pseudomonas aeruginosa</i> strain:ATCC 27853 (6863010)	30	2.7 ± 0.7	5.6 ± 1.9
Spirochaetes/Spirochaetes	<i>Treponema pectinovorum</i> 8:A:33768 (3757506)	72	2.6 ± 0.6	5.6 ± 2.0
	<i>Treponema pectinovorum</i> OMZ831 (10764812)	108	2.7 ± 0.8	5.5 ± 1.9
	<i>Treponema socranskii</i> (2653628)	77	2.3 ± 0.5	4.9 ± 1.6
	<i>Treponema</i> sp. I:B:C7 (2586374)	16	2.8 ± 0.8	5.8 ± 1.9
	<i>Treponema</i> sp. I:E:U17A (2586380)	37	2.6 ± 0.7	5.4 ± 1.8
	<i>Treponema</i> sp. I:F:D13 (2586376)	70	2.5 ± 0.6	5.1 ± 1.7
	<i>Treponema</i> sp. I:K:T3 (2586382)	35	2.6 ± 0.6	5.5 ± 1.8
	<i>Treponema</i> sp. I:N:D47 (2586384)	113	2.7 ± 0.7	5.7 ± 1.9
	<i>Treponema</i> sp. I:O:AF16 (3046569)	43	2.5 ± 0.6	5.4 ± 1.9
	<i>Treponema</i> sp. I:S:AT39 (3132613)	65	2.6 ± 0.6	5.4 ± 1.8

	<i>Treponema sp.</i> I:T:AT24 (3132614)	105	2.6 ± 0.7	5.3 ± 1.8
	<i>Treponema sp.</i> II:10:D12 (2586353)	69	2.5 ± 0.7	5.3 ± 1.8
	<i>Treponema sp.</i> II:D:G93 (2586355)	89	2.5 ± 0.5	5.1 ± 1.7
	<i>Treponema sp.</i> IV:17B:C21 (2586361)	68	2.4 ± 0.7	5.1 ± 1.6
	<i>Treponema sp.</i> IV:17B:C21 (2586361)	57	2.7 ± 0.7	5.7 ± 1.9
	<i>Treponema vincentii</i> (2764819)	291	2.5 ± 0.5	5.1 ± 1.7
Unidentified	Uncultured human oral bacterium A35 (6671223)	104	2.7 ± 0.7	5.7 ± 1.9

262  
263  
264

265 The histograms and whisker plots of the *Treponema denticola* 16S rRNA gene show the distribution of  
266 abundances in the 16 patients using 94 unique probes (Figure 3). The average abundances of patients  
267 with health, edentulism, periodontitis, and caries were:  $2.5 \pm 0.6$  a.u.,  $5.3 \pm 2.0$  a.u.,  $4.4 \pm 1.6$  a.u., and  
268  $3.8 \pm 3.3$  a.u., respectively. A two-tailed t-test revealed a difference in average abundances for the 4  
269 patients with health and 4 patients with edentulous but only approached significance ( $P \leq 0.06$ ) (Table  
270 2). None of the other paired conditions (i.e., caries versus edentulism, caries versus periodontitis, caries  
271 versus health, edentulism versus periodontitis, or health versus periodontitis) for this gene yielded  
272 differences from one another.

273



274

Figure 3. Distribution of 16S rRNA gene abundances for *Treponema denticola* ATCC 35405 (GI 4580728) in 16 patients based on 94 calibrated probes. Red bar indicates median value. Corresponding MAD values are shown. There is a tendency for *T. denticola* to be more abundant in patients with periodontitis (Average  $\pm$  s.d.;  $4.4 \pm 1.6$  a.u.) than those with health ( $2.5 \pm 0.6$  a.u.) ( $P < 0.09$ ).

Table 2. Bacterial species having significantly different abundances for patients with health versus those with periodontitis based on calibrated individual probes.  $n_{probes}$ , number of probes used to determine the species abundance; Two-sided T-tests were based on four patients with health and four patients with periodontitis and  $\alpha = 0.05$ .

Phylum/Class	Genus/Species/GI number	$n_{probes}$	Abundance ( $\bar{X} \pm$ s.d.) (a.u.) by:	
			Health	Periodontitis
Actinobacteria/Actinobacteria	<i>Actinomyces odontolyticus</i> (853707)	36	2.2 $\pm$ 0.5	4.1 $\pm$ 1.1
	<i>Propionibacterium avidum</i> DSM 4901 (2644976)	53	2.7 $\pm$ 0.7	4.7 $\pm$ 1.3
Bacteroidetes/Bacteroidetes	<i>Prevotella nigrescens</i> ATCC 33563 (294425)	57	2.4 $\pm$ 0.6	4.4 $\pm$ 1.3
	<i>Tannerella forsythia</i> (289382)	50	2.4 $\pm$ 0.6	4.5 $\pm$ 1.4

Bacteroidetes/Flavobacteria	<i>Capnocytophaga sp.</i> oral clone X089 (9988944)	37	2.6 ± 0.6	4.9 ± 1.4
Firmicutes/Bacilli	<i>Streptococcus gordonii</i> (2183315)	14	2.7 ± 0.9	4.7 ± 0.9
	<i>Streptococcus sp.</i> oral clone BW009 (9988906)	11	2.7 ± 0.6	5.1 ± 1.3
	<i>Streptococcus sp.</i> oral strain T1-E5 (14537933)	19	2.6 ± 0.6	4.3 ± 1.1
	<i>Streptococcus sp.</i> oral strain T4-E3 (14537934)	46	2.3 ± 0.5	4.3 ± 1.3
Firmicutes/Clostridia	<i>Selenomonas sp.</i> oral clone AA024 (9837490)	96	2.6 ± 0.7	4.7 ± 1.3
Firmicutes/Cocci	<i>Staphylococcus capitis</i> (576605)	18	2.5 ± 0.5	4.3 ± 1.1
Fusobacteria	<i>Fusobacterium periodonticum</i>	39	2.6 ± 0.7	4.7 ± 1.3
Proteobacteria/Betaproteobacteria	<i>Neisseria flava</i> strain U40 (5327199)	1	2.7 ± 0.8	4.8 ± 0.8
	<i>Neisseria sp.</i> oral clone AP015 (10048301)	15	2.7 ± 0.7	5.2 ± 1.4
	<i>Neisseria sp.</i> oral strain B33KA (10048304)	74	2.7 ± 0.7	5.1 ± 1.6
	<i>Neisseria subflava</i> (5327189)	12	2.8 ± 0.9	5.0 ± 1.3
Proteobacteria/Deltaproteobacteria	<i>Desulfobulbus sp.</i> oral clone CH031 (10048312)	113	2.6 ± 0.6	4.7 ± 1.4
Proteobacteria/Gammaproteobacteria	<i>Klebsiella pneumoniae</i> strain ATCC13884T (3282032)	8	2.3 ± 0.6	4.3 ± 0.9
	<i>Pseudomonas aeruginosa</i> strain LMG 1242T (1907091)	27	2.9 ± 0.8	5.6 ± 1.8
Spirochaetes/Spirochaetes	<i>Treponema pectinovorum</i> 8:A:33768 (3757506)	72	2.6 ± 0.6	4.9 ± 1.5
	<i>Treponema pectinovorum</i> OMZ831 (10764812)	108	2.7 ± 0.8	4.9 ± 1.5
	<i>Treponema socranskii</i> (2653630)	35	2.4 ± 0.5	4.3 ± 1.1
	<i>Treponema sp.</i> Il:C:T1 (2586354)	57	2.7 ± 0.7	5.1 ± 1.4
Unidentified	Human oral bacterium C20 (6671248)	2	2.3 ± 0.5	5.5 ± 1.5

275  
276  
277  
278  
279  
280

**Microbial abundance signatures.** To identify microbial abundance signatures (i.e., 16S rRNA genes) that are unique to a particular condition, we compared the 576 rRNA sequences by condition (Caries, Edentulism, Periodontitis and Health) using two-tailed T-tests with unequal variance at alpha=0.05. No significant differences were found for the Caries versus Edentulism, Caries versus Health, Caries versus



281 Periodontitis, or Edentulism versus Periodontitis conditions. However, specific microbial abundance  
282 signatures were found in Health versus Edentulism and Health versus Periodontitis conditions (Tables 1  
283 and 2), which are described below.

284 **Health versus Edentulism.** Patients with edentulism had higher abundances of 26 genera and 35  
285 bacterial species than patients associated with dental health. The bacterial species included those in the  
286 phyla Actinobacteria, Bacteroidetes, Firmicutes, Fusobacterium, Proteobacteria, and Spirochaetes  
287 (Table 1). The number of unique probes used to determine gene abundances of these microorganisms  
288 ranged from 6 to 387. Similar to the above, highly related strains (e.g., *Actinomyces odontolyticus* GI  
289 853707 and *A. odontolyticus* GI 69977974) yielded similar abundances even though they were based on  
290 different probes (i.e., 36 versus 31 unique probes, respectively). This phenomenon was observed in  
291 different strains and species of the genera *Actinomyces*, *Prevotella*, *Capnocytophaga*, *Streptococcus*,  
292 *Eubacterium*, *Selenomonas*, *Dialister*, *Leptotrichia*, *Kingella*, *Neisseria*, *Klebsiella*, and *Treponema*.

293 **Health versus periodontitis.** Patients with periodontitis had higher abundances of 14 genera and 21  
294 bacterial species than patients with health. The bacterial genera included those found in the phyla  
295 Actinobacteria, Bacteroidetes, Firmicutes, Fusobacterium, Proteobacteria, and Spirochaetes (Table 2).  
296 Note that the number of unique probes used to determine gene abundances varied from one to 113. We  
297 emphasize this point because additional probes targeting a particular 16S rRNA gene presumably  
298 improved the precision. Interestingly, strains of highly related species (e.g., *Streptococcus sp.*; GI  
299 14537933 and GI 14537934) yielded similar abundances even though they were based on different  
300 probes (i.e., 19 versus 46 unique probes, respectively). This phenomenon was also observed in different  
301 strains and species of the genera *Streptococcus*, *Neisseria* and *Treponema*. This finding provides support  
302 for the precision of Gene Meters in terms of gene abundances.

### 303 **Calibrated probe aggregates**

304 The PMT settings affected the number of total genes ( $n=597$ ) that could be calibrated. Specifically, 564,  
305 567, and 572 genes were calibrated for the PMT settings of 500, 600 and 700, respectively. The  
306 abundances for these genes in the 16 patient samples were assembled into a dataset by PMT (Table S3).  
307 As before, gene abundances that were greater than 10 a.u. were set to 10 a.u. because we did not expect  
308 calibration to forecast accurately beyond the dilution series used to create it. The average gene  
309 abundance for the 500, 600, and 700 datasets was  $2.9 \pm 1.8$ ,  $3.4 \pm 1.9$ , and  $4.0 \pm 2.1$  arbitrary units (a.u.),  
310 respectively. The lowest and highest gene abundances for all datasets was 0.03 and 10.0 a.u.,  
311 respectively. Therefore, gene abundances could vary by as much as 300-fold and the averaged gene  
312 abundances were affected by PMT settings, with higher abundances at higher PMT settings. Based on  
313 these results, all subsequent analyses of the aggregated calibrated probes were performed using the 600  
314 PMT setting data.

315 **Microbial abundance signatures.** Microbial abundance signatures (i.e., 16S rRNA genes) were  
316 determined by comparing the 567 rRNA sequences by condition (Caries, Edentulism, Periodontitis and  
317 Health) using two-tailed T-tests with unequal variance at  $\alpha=0.05$ . Although significant differences  
318 were not found for the Caries versus Health comparison, specific microbial abundance signatures were  
319 found for the Caries versus Edentulism, Caries versus Periodontitis, Edentulism versus Health,  
320 Edentulism versus Periodontitis, and Health versus Periodontitis comparisons, which are described  
321 below.

322 **Caries versus Edentulism.** Patients with edentulism had significantly higher abundances of  
323 *Lactobacillus sp.* ( $2.8 \pm 1.21$  a.u.) than patients with caries ( $0.6 \pm 0.42$  a.u.) ( $P<0.03$ ). Hence,  
324 *Lactobacillus sp.* was a putative abundance signature for patients with edentulism.

325 **Caries versus Periodontitis.** Patients with periodontitis had significantly higher abundances of  
 326 *Tannerella forsythia* ( $1.3 \pm 0.58$  a.u.) than patients with caries ( $0.4 \pm 0.28$  a.u.) ( $P < 0.04$ ). Similarly,  
 327 patients with periodontitis had significantly higher abundances of *Fusobacterium nucleatum* ( $1.0 \pm 0.28$   
 328 a.u.) than patients with caries ( $0.4 \pm 0.22$  a.u.) ( $P < 0.01$ ). The high abundances of these microorganisms  
 329 indicates a putative signature for periodontitis as these microorganisms were significantly more  
 330 abundant in patients with periodontitis versus patients with health (see below).

331 **Edentulism versus Health.** Patients with edentulism had significantly higher abundances of 32 genera  
 332 (and two unidentified species) representing 72 different species/strains than patients associated with  
 333 health. The microorganisms included the phyla Actinobacteria, Ascomycota, Bacteroidetes,  
 334 Deinococcus-Thermus, Firmicutes, Fusobacterium, Proteobacteria, Spirochaetes, and Tenericutes  
 335 (Table 3). The phylum Firmicutes contained the most microorganisms ( $n=29$  species) with the  
 336 following classes: Bacilli, Clostridia, Erysipelotrichia, and Negativicutes. The class Bacilli consisted of  
 337 *Bacillus clausii*, and several *Lactobacillus* and *Streptococcus* species. The phylum Actinobacteria  
 338 contained the second most microorganisms ( $n=15$  species) and included the classes Actinobacteria and  
 339 Coriobacterila. Within the class Actinobacteria were the following genera: *Actinomyces*,  
 340 *Bifidobacterium*, *Propionibacterium*, and *Stomatococcus*.

Table 3. Microbial species having significantly different abundances for patients with health versus those with edentulism using calibrated aggregated probes.  $n\_probes$ , number of probes used to determine the species abundance; Two-sided T-tests were based on four patients with health and four patients with edentulism and  $\alpha=0.05$ .

Phylum/Class	Genus/Species/GI number	$n\_probes$	Abundance ( $X \pm s.d.$ ) (a.u.) by:	
			Health	Edentulism
Actinobacteria/Actinobacteria	<i>Actinomyces gerencseriae</i> (1838939)	193	$1.9 \pm 0.36$	$3.1 \pm 0.80$
	<i>Actinomyces israelii</i> (1838944)	314	$1.6 \pm 0.26$	$2.7 \pm 0.67$
	<i>Actinomyces meyeri</i> (1838945)	74	$2.1 \pm 0.41$	$4.4 \pm 1.42$
	<i>Actinomyces odontolyticus</i> (6997974)	31	$2.5 \pm 0.48$	$5.4 \pm 1.73$
	<i>Actinomyces odontolyticus</i> (853707)	36	$1.6 \pm 0.32$	$2.7 \pm 0.54$
	<i>Actinomyces sp.</i> (14537925)	310	$1.7 \pm 0.31$	$3.0 \pm 0.86$
	<i>Actinomyces sp.</i> (14537930)	56	$2.4 \pm 0.52$	$5.0 \pm 1.64$
	<i>Actinomyces sp.</i> (14537912)	82	$2.0 \pm 0.42$	$4.0 \pm 1.26$
	<i>Actinomyces sp.</i> (14537929)	168	$1.8 \pm 0.30$	$3.5 \pm 0.32$
	<i>Actinomyces sp.</i> (2073386)	131	$1.6 \pm 0.21$	$2.6 \pm 0.62$
	<i>Actinomyces viscosus</i> (1838951)	98	$2.2 \pm 0.39$	$3.7 \pm 0.93$

	<i>Bifidobacterium sp. (9837452)</i>	205	2.0 ± 0.33	3.4 ± 0.89
	<i>Propionibacterium avidum (2644976)</i>	53	2.5 ± 0.54	5.4 ± 1.82
	<i>Stomatococcus mucilaginosus (2764819)</i>	291	1.8 ± 0.17	2.8 ± 0.34
Actinobacteria/Coriobacteriia	<i>Atopobium rimae (10719609)</i>	22	2.7 ± 0.62	7.6 ± 2.13
Ascomycota/Saccharomycetes	<i>Candida albicans (2507)</i>	1442	2.2 ± 0.46	4.7 ± 1.56
Bacteroidetes/Bacteroidetes	<i>Porphyromonas-like sp. (9988935)</i>	631	1.9 ± 0.31	3.1 ± 0.82
	<i>Prevotella sp. (9988921)</i>	87	2.6 ± 0.56	5.6 ± 1.94
	<i>Prevotella sp. (14537919)</i>	88	2.6 ± 0.60	5.7 ± 2.02
	<i>Prevotella sp. (9988918)</i>	127	1.9 ± 0.33	3.2 ± 0.85
	<i>Prevotella sp. (9988919)</i>	98	2.2 ± 0.37	3.7 ± 0.97
Bacteroidetes/Flavobacteria	<i>Capnocytophaga sp. (9988943)</i>	101	2.7 ± 0.62	5.9 ± 2.08
Bacteroidetes/Sphingobacteria	<i>Pedobacter sp. (14537940)</i>	418	2.2 ± 0.44	4.4 ± 1.39
Deinococcus-Thermus/Deinococci	<i>Deinococcus sp. (14537950)</i>	790	2.1 ± 0.43	4.2 ± 1.39
Firmicutes/Bacilli	<i>Bacillus clausii (9715774)</i>	27	2.2 ± 0.43	4.7 ± 1.57
	<i>Lactobacillus gasseri (175028)</i>	309	2.1 ± 0.45	4.5 ± 1.54
	<i>Lactobacillus paracasei (1808583)</i>	79	1.7 ± 0.32	2.7 ± 0.71
	<i>Lactobacillus sp. (9988912)</i>	73	0.5 ± 0.13	2.8 ± 1.21
	<i>Paenibacillus sp. (15004607)</i>	149	2.0 ± 0.35	3.6 ± 1.04
	<i>Streptococcus gordonii (2183315)</i>	14	1.8 ± 0.33	2.8 ± 0.65
	<i>Streptococcus mitis (9988909)</i>	6	2.7 ± 0.80	5.7 ± 1.95
	<i>Streptococcus pneumoniae (2183314)</i>	29	1.9 ± 0.40	4.1 ± 1.33
	<i>Streptococcus salivarius (176047)</i>	28	1.9 ± 0.40	4.1 ± 1.10
	<i>Streptococcus sp. (9988907)</i>	91	2.5 ± 0.60	5.4 ± 1.87
	<i>Streptococcus sp. (16612185)</i>	83	2.1 ± 0.45	4.3 ± 1.38
	<i>Streptococcus thoraltensis (2578813)</i>	275	2.3 ± 0.49	4.9 ± 1.60
Firmicutes/Clostridia	<i>Butyrivibrio sp. (9837461)</i>	569	1.8 ± 0.35	3.5 ± 1.10
	<i>Eubacterium limosum (174517)</i>	437	2.1 ± 0.58	4.5 ± 1.58

	<i>Eubacterium minutum</i> (4001750)	60	1.2 ± 0.21	2.0 ± 0.51
	<i>Eubacterium saphenum</i> (1513214)	447	1.7 ± 0.30	3.0 ± 0.77
	<i>Eubacterium sp.</i> (14537976)	196	2.1 ± 0.37	3.6 ± 0.96
	<i>Eubacterium sp.</i> (14537971)	80	1.5 ± 0.43	2.9 ± 0.87
	<i>Eubacterium sp.</i> (14537961)	219	1.9 ± 0.32	3.4 ± 0.88
	<i>Peptostreptococcus anaerobius</i> (175621)	132	1.7 ± 0.28	2.9 ± 0.78
	<i>Peptostreptococcus asaccharolyticus</i> (454180)	439	1.9 ± 0.42	3.5 ± 0.98
	<i>Eubacterium sp.</i> (3861469)	515	2.0 ± 0.35	3.3 ± 0.87
Firmicutes/Erysipelotrichia	<i>Erysipelothrix tonsillarum</i> (7678893)	478	1.8 ± 0.34	3.4 ± 1.03
	<i>Solobacterium sp.</i> (9988916)	113	1.9 ± 0.39	3.5 ± 1.08
Firmicutes/Negativicutes	<i>Selenomonas flueggei</i> (9837496)	95	1.7 ± 0.35	3.3 ± 1.05
	<i>Selenomonas flueggei</i> (9837497)	50	2.2 ± 0.53	4.8 ± 1.50
	Uncultured <i>Selenomonas</i> (14161358)	84	2.7 ± 0.70	5.8 ± 2.00
	Uncultured <i>Veillonella</i> (14161350)	47	1.7 ± 0.44	2.9 ± 0.80
	<i>Veillonella dispar</i> (793883)	40	1.8 ± 0.34	3.4 ± 0.81
Fusobacteria/Fusobacteria	<i>Leptotrichia sp.</i> (16612191)	400	2.4 ± 0.57	5.0 ± 1.69
	<i>Leptotrichia sp.</i> (14537926)	215	2.4 ± 0.50	4.7 ± 1.48
Proteobacteria/Betaproteobacteria	<i>Burkholderia sp.</i> (9988903)	264	2.0 ± 0.35	3.3 ± 0.85
	<i>Kingella denitrificans</i> (174983)	63	2.2 ± 0.36	3.6 ± 0.91
	<i>Neisseria sp.</i> (10048304)	74	2.7 ± 0.57	5.9 ± 2.05
Proteobacteria/Epsilonproteobacteria	<i>Campylobacter sp.</i> (10048314)	239	1.7 ± 0.33	3.0 ± 0.84
Proteobacteria/Gammaproteobacteria	<i>Ewingella americana</i> (903619)	272	2.0 ± 0.39	4.0 ± 1.30
	<i>Klebsiella pneumoniae</i> (3282032)	8	2.2 ± 0.53	4.8 ± 1.61
Spirochaetes/Spirochaetes	<i>Treponema denticola</i> (4580728)	94	1.0 ± 0.16	1.7 ± 0.41

	<i>Treponema pectinovorum</i> (10764812)	108	2.6 ± 0.64	5.6 ± 1.95
	<i>Treponema socranskii</i> (2653628)	77	2.3 ± 0.56	4.7 ± 1.52
	<i>Treponema sp.</i> (3132614)	105	2.7 ± 0.51	4.5 ± 1.50
	<i>Treponema sp.</i> (2586354)	57	2.6 ± 0.65	5.7 ± 1.98
	<i>Treponema sp.</i> (14537906)	128	09 ± 0.16	1.5 ± 0.43
	<i>Treponema sp.</i> (6049684)	85	2.7 ± 0.71	6.0 ± 2.04
	<i>Treponema sp.</i> (2586355)	89	1.4 ± 0.21	2.3 ± 0.56
Tenericutes/Mollicutes	<i>Mycoplasma buccale</i> (4883887)	191	2.6 ± 0.61	5.8 ± 1.95
Unknown	Human oral (6671248)	2	2.7 ± 0.60	4.9 ± 1.29
	Uncultured human (6671223)	104	2.6 ± 0.64	5.9 ± 1.91

341  
342  
343

344 **Edentulism and Periodontitis.** Patients with periodontitis had significantly higher abundances of  
 345 seven microbial species/strains than patients with edentulism (Table 4). The phyla for these bacteria  
 346 included Bacteroidetes, Firmicutes, and Fusobacteria. Notable genera included: *Prevotella*, *Tannerella*,  
 347 and *Fusobacterium*, which are thought to play roles in periodontitis.

Table 4. Microbial species having significantly different abundances for patients with edentulism versus those with periodontitis based on aggregated probes. *n\_probes*, number of probes used to determine the species abundance; Two-sided T-tests were based on four patients with edentulism and four patients with periodontitis and  $\alpha=0.05$ .

Phylum/Class	Genus/Species/GI number	<i>n_probes</i>	Abundance ( $X \pm s.d.$ ) (a.u.) by:	
			Edentulism	Periodontitis
Bacteroidetes/Bacteroidetes	<i>Prevotella denticola</i> (294420)	49	1.5 $\pm$ 0.36	2.1 $\pm$ 0.31
	<i>Prevotella nigrescens</i> (294425)	57	1.0 $\pm$ 0.30	1.7 $\pm$ 0.39
	<i>Prevotella sp.</i> (9988923)	53	0.8 $\pm$ 0.21	1.7 $\pm$ 0.45
	<i>Prevotella veroralis</i> (294427)	41	1.2 $\pm$ 0.29	2.0 $\pm$ 0.42
Firmicutes/Clostridia	<i>Tannerella forsythia</i> (6539450)	94	0.4 $\pm$ 0.13	1.3 $\pm$ 0.58
	<i>Mogibacterium vescum</i> (4140242)	83	0.7 $\pm$ 0.55	2.5 $\pm$ 1.23
Fusobacteria	<i>Fusobacterium nucleatum</i> (4490387)	40	0.5 $\pm$ 0.14	1.0 $\pm$ 0.28

348  
349

350 **Health and Periodontitis.** Patients with periodontitis had higher abundances of 30 genera (five not  
 351 taxonomically identified) and 62 species than patients with health (Table 5). These bacteria were found  
 352 in the phyla Actinobacteria, Bacteroidetes, Chloroflexi, Deferribacteres, Firmicutes, Fusobacteria,  
 353 Proteobacteria, and Spirochaetes. Note that there was a higher abundance of members of the “red  
 354 complex” (specifically, *Porphyromonas gingivalis*, *Tannerella forsythia*, *Treponema denticola*) in  
 355 patients with periodontitis than patients with health.

Table 5. Microbial species having significantly different abundances for patients with health versus those with periodontitis based on aggregated probes. *n\_probes*, number of probes used to determine the species abundance; Two-sided T-tests were based on four patients with health and four patients with periodontitis and  $\alpha=0.05$ .

Phylum/Class	Genus/Species/GI number	<i>n_probes</i>	Abundance ( $X \pm s.d.$ ) (a.u.) by:	
			Health	Periodontitis (n=4)
Actinobacteria/Actinobacteria	<i>Actinomyces sp.</i> (9837444)	31	2.4 $\pm$ 0.52	4.6 $\pm$ 1.45
	<i>Actinomyces sp.</i> (10946538)	116	1.7 $\pm$ 0.28	2.5 $\pm$ 0.52



	<i>Actinomyces odontolyticus</i> (853707)	36	1.6 ± 0.32	2.5 ± 0.48
	<i>Bifidobacterium</i> sp. (9837452)	205	2.0 ± 0.33	3.1 ± 0.71
Bacteroidetes/Bacteriodetes	<i>Bacteroides</i> -like sp. (9988930)	688	2.0 ± 0.40	3.6 ± 1.01
	<i>Bacteroides</i> -like sp. (9988936)	620	1.8 ± 0.52	3.0 ± 0.58
	<i>Porphyromonas endodontalis</i> (294287)	89	2.2 ± 0.50	4.0 ± 1.12
	<i>Porphyromonas gingivalis</i> (509140)	89	1.1 ± 0.29	1.9 ± 0.40
	<i>Porphyromonas</i> -like sp. (9988934)	647	1.8 ± 0.36	3.1 ± 0.71
	<i>Porphyromonas</i> -like sp. (9988935)	631	1.9 ± 0.31	3.0 ± 0.53
	<i>Prevotella bivia</i> (294429)	282	2.0 ± 0.41	3.4 ± 0.93
	<i>Prevotella buccae</i> (294432)	352	1.7 ± 0.35	3.0 ± 0.74
	<i>Prevotella buccalis</i> (294430)	307	1.6 ± 0.43	3.1 ± 0.93
	<i>Prevotella dentalis</i> (1565276)	371	1.9 ± 0.35	3.2 ± 0.80
	<i>Prevotella denticola</i> (294420)	49	1.0 ± 0.19	2.1 ± 0.31
	<i>Prevotella enoeca</i> (3114912)	399	1.9 ± 0.36	3.3 ± 0.90
	<i>Prevotella intermedia</i> (294422)	227	1.3 ± 0.21	2.1 ± 0.39
	<i>Prevotella nigrescens</i> (294425)	57	0.8 ± 0.29	1.7 ± 0.39
	<i>Prevotella nigrescens</i> (509071)	154	1.1 ± 0.20	2.3 ± 0.42
	<i>Prevotella oralis</i> (294434)	245	1.6 ± 0.32	2.8 ± 0.62
	<i>Prevotella oris</i> (294428)	65	1.1 ± 0.17	1.9 ± 0.29
	<i>Prevotella pallens</i> (2108322)	89	1.0 ± 0.21	2.0 ± 0.33
	<i>Prevotella</i> sp. (9988918)	127	1.9 ± 0.33	3.0 ± 0.65
	<i>Prevotella</i> sp. (14537917)	400	2.0 ± 0.44	3.6 ± 0.95
	<i>Prevotella</i> sp. (29165644)	317	1.9 ± 0.36	3.2 ± 0.77
	<i>Prevotella</i> sp. (9988917)	89	1.4 ± 0.24	2.2 ± 0.46
	<i>Prevotella</i> sp. (9988923)	53	0.8 ± 0.44	1.7 ± 0.45
	<i>Prevotella</i> sp.	456	2.0 ± 0.40	3.5 ± 0.83

	<b>(14537973)</b>			
	<b><i>Prevotella sp.</i> (14537923)</b>	229	1.5 ± 0.25	2.3 ± 0.39
	<b><i>Prevotella sp.</i> (14537920)</b>	268	1.2 ± 0.32	1.9 ± 0.13
	<b><i>Prevotella sp.</i> (9988929)</b>	433	1.5 ± 0.35	2.4 ± 0.37
	<b><i>Prevotella sp.</i> (14537960)</b>	112	1.2 ± 0.28	2.1 ± 0.33
	<b><i>Prevotella sp.</i> (14537927)</b>	68	0.8 ± 0.16	1.3 ± 0.10
	<b><i>Prevotella sp.</i> (9988920)</b>	85	1.2 ± 0.19	2.1 ± 0.15
	<b><i>Prevotella tannerae</i> (10039600)</b>	167	1.6 ± 0.28	2.3 ± 0.45
	<b><i>Prevotella veroralis</i> (294427)</b>	41	0.8 ± 0.23	2.0 ± 0.42
	<b><i>Prevotella zooglyphiformans</i> (289379)</b>	276	1.8 ± 0.32	3.0 ± 0.76
	<b><i>Tannerella forsythia</i> (10946530)</b>	393	1.3 ± 0.43	2.2 ± 0.32
	<b><i>Tannerella forsythia</i> (289382)</b>	50	2.3 ± 0.16	3.7 ± 0.47
Bacteroidetes/Flavobacteria	<b><i>Capnocytophaga haemolytica</i> (1199611)</b>	383	2.0 ± 0.39	3.4 ± 0.90
	<b><i>Capnocytophaga sp.</i> (9988944)</b>	37	2.5 ± 0.56	4.5 ± 1.26
	<b><i>Capnocytophaga sputigena</i> (289579)</b>	172	2.0 ± 0.35	3.0 ± 0.62
Chloroflexi	<b><i>TM7 phylum</i> (14537928)</b>	174	1.8 ± 0.47	3.0 ± 0.46
Deferribacteres/Deferribacteres	<b><i>Flexistipes-like sp.</i> (9858891)</b>	137	2.0 ± 0.40	3.6 ± 0.99
Firmicutes/Bacilli	<b><i>Abiotrophia elegans</i> (2460062)</b>	125	2.0 ± 0.40	3.6 ± 0.76
	<b><i>Caryophanon sp.</i> (14537944)</b>	40	2.3 ± 0.51	4.1 ± 1.14
	<b><i>Granulicatella elegans</i> (3123332)</b>	26	2.5 ± 0.57	4.5 ± 1.23
	<b><i>Lactobacillus sp.</i> (9988912)</b>	73	0.5 ± 0.13	1.0 ± 0.25
	<b><i>Staphylococcus epidermidis</i> (1199945)</b>	26	2.6 ± 0.67	4.8 ± 1.45
	<b><i>Staphylococcus xylosus</i> (1199956)</b>	54	1.5 ± 0.25	2.4 ± 0.48
	<b><i>Streptococcus gordonii</i> (2183315)</b>	14	1.8 ± 0.33	2.6 ± 0.51
	<b><i>Streptococcus sobrinus</i> (5578902)</b>	258	2.0 ± 0.32	3.0 ± 0.65

	<i>Streptococcus sp.</i> (9988906)	11	2.7 ± 0.65	4.8 ± 1.40
	<i>Streptococcus sp.</i> (16612184)	45	1.9 ± 0.33	3.5 ± 0.32
Firmicutes/Clostridia	<i>Butyrivibrio sp.</i> (9837461)	569	1.8 ± 0.35	3.1 ± 0.74
	<i>Clostridium tetanomorphum</i> (535115)	525	1.8 ± 0.50	3.1 ± 0.81
	<i>Eubacterium infirmum</i> (3299817)	85	1.7 ± 0.36	2.7 ± 0.66
	<i>Eubacterium saphenum</i> (1513214)	447	1.7 ± 0.30	2.8 ± 0.62
	<i>Eubacterium sp.</i> (9837459)	637	2.0 ± 0.42	3.4 ± 0.94
	<i>Eubacterium sp.</i> (14537976)	196	2.1 ± 0.37	3.3 ± 0.70
	<i>Eubacterium sp.</i> (14537903)	39	1.4 ± 0.25	2.0 ± 0.34
	<i>Eubacterium sp.</i> (14537961)	219	1.9 ± 0.32	3.3 ± 0.73
	<i>Eubacterium sp.</i> (14537982)	367	1.3 ± 0.42	2.5 ± 0.50
	<i>Eubacterium sp.</i> (14537971)	80	1.5 ± 0.43	3.1 ± 0.48
	<i>Mogibacterium diversum</i> (6942242)	114	1.6 ± 0.30	2.8 ± 0.72
	<i>Peptostreptococcus anaerobius</i> (175621)	132	1.7 ± 0.28	2.7 ± 0.59
	<i>Peptostreptococcus asaccharolyticus</i> (454180)	439	1.9 ± 0.42	3.1 ± 0.76
	<i>Peptostreptococcus magnus</i> (454441)	401	1.4 ± 0.38	2.1 ± 0.37
	<i>Peptostreptococcus sp.</i> (9837456)	294	1.8 ± 0.36	3.1 ± 0.83
	<i>Eubacterium sp.</i> (3861469)	515	2.0 ± 0.35	3.0 ± 0.68
Firmicutes/Negativicutes	<i>Dialister sp.</i> (9837481)	63	2.3 ± 0.56	4.0 ± 1.05
	<i>Dialister sp.</i> (9837479)	330	1.3 ± 0.35	2.0 ± 0.32
	<i>Firmicutes sp.</i> (9837472)	204	1.6 ± 0.34	2.4 ± 0.52
	<i>Megasphaera sp.</i> (14537975)	305	2.2 ± 0.49	4.0 ± 1.14
	<i>Megasphaera sp.</i> (9837476)	276	1.8 ± 0.39	3.0 ± 0.75
	<i>Selenomonas sp.</i> (9837490)	96	2.5 ± 0.62	4.7 ± 1.42

	<i>Selenomonas sp.</i> (9837488)	540	2.2 ± 0.47	3.8 ± 1.07
	<i>Veillonella dispar</i> (793883)	40	1.8 ± 0.34	2.7 ± 0.53
Fusobacteria	<i>Filifactor alocis</i> (4127826)	538	1.4 ± 0.28	2.8 ± 0.62
	<i>Fusobacterium nucleatum</i> (43402)	15	2.4 ± 0.61	4.2 ± 1.16
Fusobacteria/Fusobacteria	<i>Leptotrichia sp.</i> (9837509)	240	2.2 ± 0.52	4.0 ± 1.15
	<i>Leptotrichia sp.</i> (9837506)	172	2.2 ± 0.41	3.8 ± 1.05
	<i>Leptotrichia sp.</i> (9837507)	286	2.2 ± 0.45	3.9 ± 1.02
	<i>Leptotrichia sp.</i> (14537981)	326	1.9 ± 0.42	3.1 ± 0.73
Proteobacteria/Betaproteobacteria	<i>Neisseria flava</i> (5327199)	1	2.7 ± 0.76	4.8 ± 0.80
	<i>Neisseria pharyngis</i> (5327180)	12	2.7 ± 0.74	5.0 ± 1.37
	<i>Neisseria sp.</i> (10048301)	15	2.6 ± 0.60	5.1 ± 1.37
Proteobacteria/Epsilonproteobacteria	<i>Campylobacter sp.</i> (10048314)	239	1.7 ± 0.33	2.7 ± 0.61
Proteobacteria/Gammaproteobacteria	<i>Haemophilus segnis</i> (174775)	124	2.5 ± 0.53	4.5 ± 1.29
	<i>Klebsiella pneumoniae</i> (3282032)	8	2.2 ± 0.53	4.2 ± 1.22
Spirochaetes/Spirochaetes	<i>ç</i> (4580728)	94	1.0 ± 0.16	2.2 ± 0.61
	<i>Treponema socranskii</i> (2653630)	35	2.2 ± 0.45	4.0 ± 1.17
	<i>Treponema socranskii</i> (2653627)	47	2.6 ± 0.67	4.8 ± 1.39
	<i>Treponema sp.</i> (2586367)	154	2.5 ± 0.61	4.6 ± 1.35
	<i>Treponema sp.</i> (3132614)	105	2.2 ± 0.51	4.0 ± 1.15
	<i>Treponema sp.</i> (2586364)	77	2.5 ± 0.58	4.6 ± 1.28
	<i>Treponema sp.</i> (2586354)	57	2.6 ± 0.65	4.9 ± 1.45
	<i>Treponema sp.</i> (2586361)	68	1.9 ± 0.40	3.3 ± 0.84
	<i>Treponema sp.</i> (14537906)	128	0.9 ± 0.16	2.3 ± 0.68
	<i>Treponema sp.</i> (2586355)	89	1.4 ± 0.21	2.5 ± 0.44

Unknown	Human oral (6671248)	2	2.7 ± 0.60	4.5 ± 0.96
	Uncultured bacterium (9858892)	252	2.1 ± 0.38	3.1 ± 0.61
	Uncultured bacterium (4530523)	98	1.8 ± 0.30	2.8 ± 0.55
	Uncultured bacterium (4680614)	719	1.4 ± 0.42	2.6 ± 0.45
	Uncultured human (6671255)	14	2.8 ± 0.78	5.0 ± 1.45

356

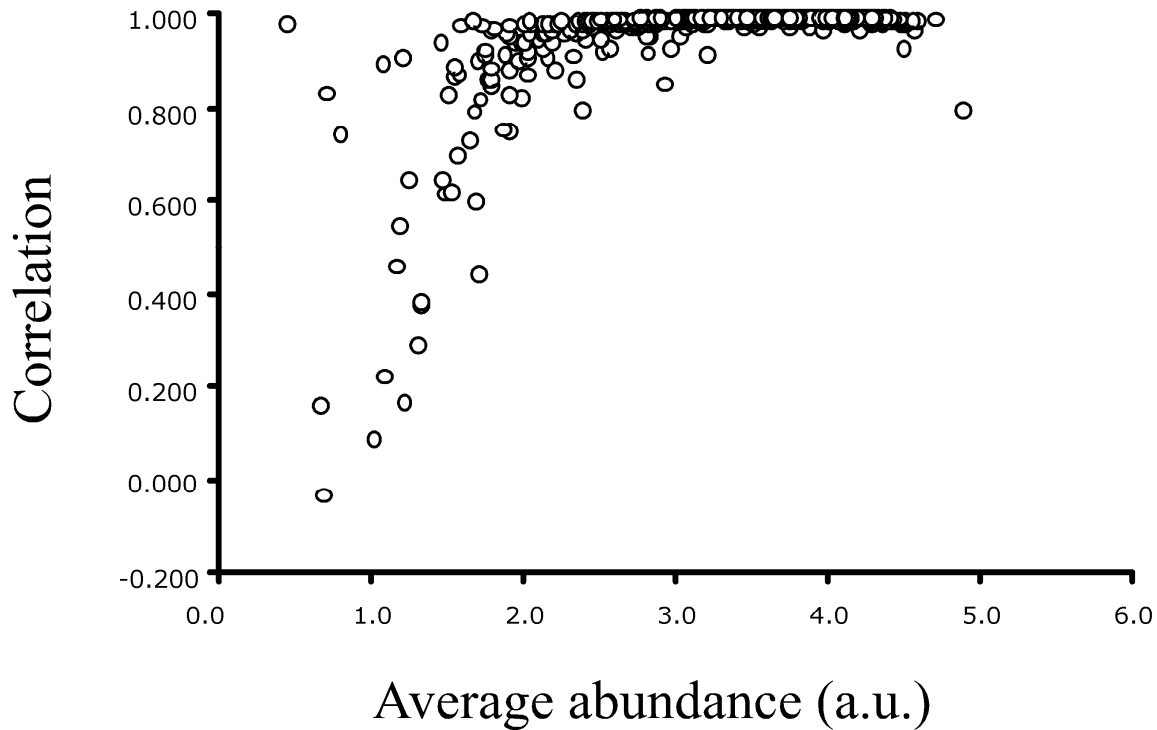
357

358

### 359 **Comparison of calibration methods**

360 Correlation analysis was used to compare calibration methods. While in theory the abundances should  
361 be similar by calibration method and the correlations close to one, the abundances determined by  
362 aggregated probes were anticipated to be more robust than individual probes when the gene target  
363 abundances approached zero. The reason for this is that the sum of all signal intensities obtained from  
364 multiple probes targeting the same gene will always be greater than the signal intensities of individual  
365 probes. At low gene target abundances, the signal intensities of individually calibrated probes approach  
366 the resolution of the scanner, which means the potential for noise in the signal would be minimized in  
367 the aggregated probe results.

368 A plot of the relationship between the correlations of the calibration methods against the average target  
369 abundances is shown in Figure 4. While only 65 out of the possible 567 (~11.5%) microorganisms  
370 (GIs) had correlations of less than 0.95, most microorganisms ( $n=502$ ) (i.e., GIs) had similar abundances  
371 (~88.5% had correlation  $\geq 0.95$ ). Figure 4 shows that when the average abundance was less than 2 a.u.,  
372 the correlations were less than 0.95 (Figure 4). The significance of this finding is that it provides  
373 support that the aggregate probe calibration is more precise than the individual probe calibration.



374

*Figure 4. Correlation of abundances of GIs for two calibration approaches (individual probe calibration and aggregate probes targeting same gene) versus average abundance of the GIs determined using the aggregate calibration approach. Results show the average abundances for GIs are low (< 2.0 a.u.). The correlations are also low, presumably because the signal intensities of the individual probes approach the resolution of the scanner while the sum of the signal intensities of aggregated probes do not.*

375

### 376 **Gene Meter versus DNA sequencing results**

377 In theory, one would expect the same microorganisms to be identified by both approaches (Gene Meters  
378 and DNA sequencing) since the source of DNA was the same PCR amplification products from the  
379 same 16 patient samples.

380 For the Gene Meter approach, we detected 576 microbial targets (564 non-redundant targets) (Figure 5).  
381 Thirty-five of the 564 non-redundant targets could not be taxonomically resolved: 13 were classified as  
382 “unculturable” and 22 could not be identified to the genus level. Subtracting these 35 ambiguous targets  
383 from the 564 non-redundant gene targets left 107 microbial genera.



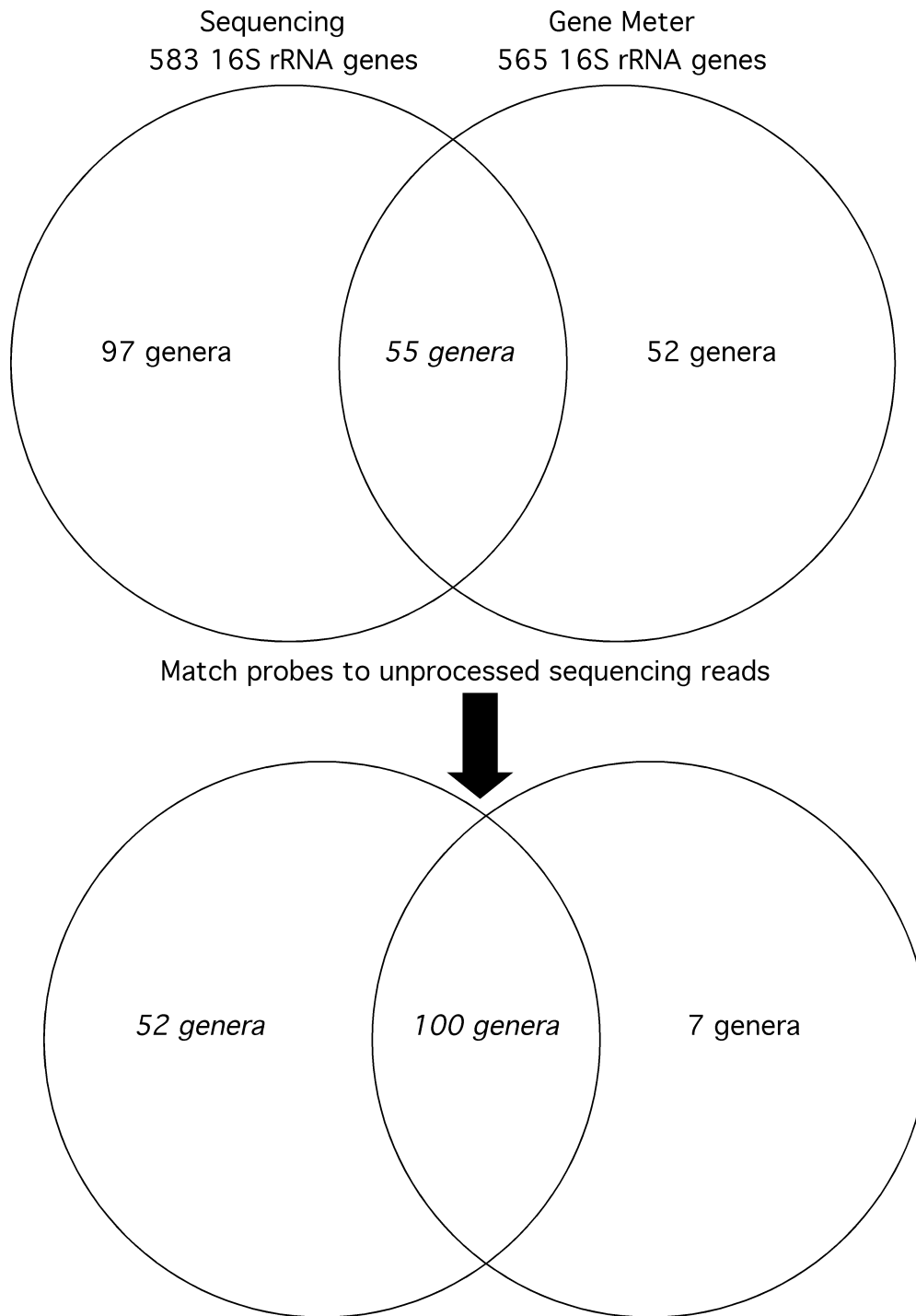


Figure 5. Shared genera by DNA sequencing and Gene Meter for 16 patients. Top shows taxonomic classification based on processed DNA sequencing reads. Bottom shows taxonomic classification based on probe matches of unprocessed DNA sequencing reads and processed DNA sequencing reads.

384

385 For the DNA sequencing approach, we detected 654 targets (583 non-redundant targets) (Figure 5).  
386 Nine of the 583 non-redundant targets yielded ambiguous taxonomy (e.g., “anaerobic bacterium”,

387 “rumen bacterium”) and 17 were plants (e.g., *Nicotiana sylvestris*, *N. tabacum*, *N. tomentosiformism*,  
388 “Onion yellows”, *Oryza nivara*, *O. sativa*, and *Phalaenopsis aphrodite*, *Citrus sinensis*, *Cucumis*  
389 *sativus*, *Dioscorea elephantipes*, *Draba nemorosa*, *Drimys granadensis*, *Heliathus annuus*, *Calycanthus*  
390 *floridus*, *Agrostis stolonifera*, *Amborella trichopoda*, *Calycanthus floridus*, *Gossypium barbadense*, and  
391 *Gossypium barbadense*). Subtracting these 26 ambiguous targets from the 583 non-redundant gene  
392 targets left 152 microbial genera.

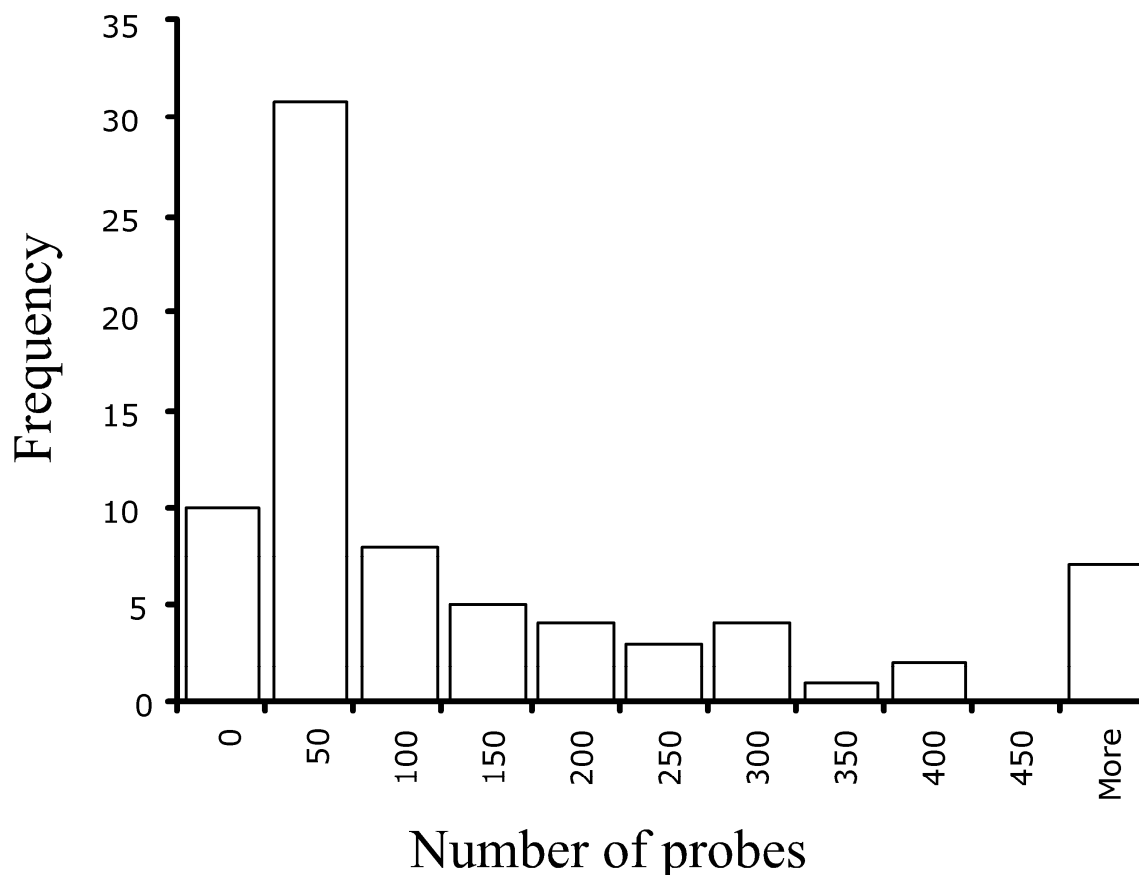
393 The union of the two approaches yielded 55 common microbial genera: *Abiotrophia*, *Actinobacillus*,  
394 *Actinomyces*, *Aggregatibacter*, *Atopobium*, *Bacillus*, *Bacteroides*, *Bergeyella*, *Bifidobacterium*,  
395 *Bradyrhizobium*, *Burkholderia*, *Butyrivibrio*, *Campylobacter*, *Capnocytophaga*, *Clostridium*,  
396 *Corynebacterium*, *Cryptobacterium*, *Desulfovibrio*, *Dialister*, *Eikenella*, *Enterococcus*, *Erysipelothrix*,  
397 *Eubacterium*, *Fusobacterium*, *Gemella*, *Granulicatella*, *Haemophilus*, *Kingella*, *Lactobacillus*,  
398 *Leptotrichia*, *Megasphaera*, *Microbacterium*, *Micrococcus*, *Mitsuokella*, *Moraxella*, *Mycobacterium*,  
399 *Mycoplasma*, *Neisseria*, *Paenibacillus*, *Peptococcus*, *Peptostreptococcus*, *Porphyromonas*, *Prevotella*,  
400 *Propionibacterium*, *Pseudomonas*, *Rothia*, *Selenomonas*, *Slackia*, *Staphylococcus*, *Streptococcus*,  
401 *Tannerella*, *Treponema*, *Variovorax*, *Veillonella*, and *Xanthomonas*.

402 Fifty-two of the 107 genera (~49%) identified by the Gene Meter approach were not identified by the  
403 DNA sequencing approach. The 52 microbial genera included: *Achromobacter*, *Actinobaculum*, *Afipia*,  
404 *Agrobacterium*, *Bartonella*, *Bdellovibrio*, *Brevundimonas*, *Bulleidia*, *Candida*, *Cardiobacterium*,  
405 *Caryophanon*, *Catonella*, *Centipeda*, *Chlamydia*, *Comamonas*, *Deinococcus*, *Delftia*, *Dermabacter*,  
406 *Desulfobulbus*, *Desulfomicrobium*, *Erythromicrobium*, *Eggerthella*, *Enterobacter*, *Escherichia*,  
407 *Ewingella*, *Flavobacterium*, *Flexistipes*, *Helicobacter*, *Holophaga*, *Janthinobacterium*, *Johnsonella*,  
408 *Klebsiella*, *Lautropia*, *Leptothrix*, *Leuconostoc*, *Methanobrevibacter*, *Mogibacterium*, *Ochrobactrum*,  
409 *Olsenella*, *Oribaculum*, *Pedobacter*, *Porphyromonas-like*, *Proteus*, *Rhizobium*, *Simonsiella*,  
410 *Solobacterium*, *Sphingomonas*, *Stenotrophomonas*, *Stomatococcus*, *Suttonella*, and *Tropheryma*.

411 It is important to note that the same genus (and species) could have different GI numbers. In the work  
412 below, the GI number was used (rather than the genus name) because there were instances when a genus  
413 with a certain GI was detected by both approaches but the same genus with a different GI number was  
414 not detected (details below). Altogether, the 52 genera were represented by 76 GI numbers. Since the  
415 average gene abundance maxima for these GIs was 8.6 a.u. (close to the maximum of 10 a.u.) in at least  
416 one patient, the microorganisms should have been identified in the DNA sequencing results. Putative  
417 reasons why they were not include: (i) the processing of the DNA sequencing reads, (ii) inadequate read  
418 depth, and (iii) false positive results, which are discussed below.

419 **Hidden jewels in the unprocessed sequencing reads.** To identify microbial species, DNA sequencing  
420 reads are subjected to the various filtering procedures (e.g., singleton removal, minimum read lengths,  
421 similarities to existing rRNA databases, e.g., 97% similarity, 100 bp min alignment). The filtering can  
422 potentially remove 16S rRNA genes of microbial species that are actually present in the sample but not  
423 reported in the final taxonomic assignment. To demonstrate this phenomenon, one could match the  
424 unique probe sequences of each microbial species (not observed in the final taxonomic assignment)  
425 against all unfiltered DNA sequence reads (i.e., before taxonomic assignments are made). If any of the  
426 unique probes happen to match the unprocessed DNA sequences for a particular microbial species, one  
427 could infer that filtering procedures were responsible for not identifying the microorganism in the final  
428 DNA sequencing results. In other words, the rRNA genes of the microbial species were actually present  
429 in the unfiltered DNA sequence reads and indeed detected by the Gene Meter approach. However, they  
430 are not present in the final filtered DNA sequencing results.

431 Using custom-designed C++ programs and *Actinomyces* sp. (GI 2073386) as a positive control (since the  
432 16S rRNA gene of this microorganism was found in all patient samples), we matched the unique probes  
433 against the DNA sequencing reads of each patient sample. We found that 67 of the 76 GIs (~88%) were  
434 present in the DNA sequencing reads. The 45 genera included: *Achromobacter*, *Actinobaculum*,  
435 *Agrobacterium*, *Bartonella*, *Bdellovibrio*, *Bulleidia*, *Candida*, *Cardiobacterium*, *Caryophanon*,  
436 *Catonella*, *Centipeda*, *Chlamydia*, *Deinococcus*, *Delftia*, *Dermabacter*, *Desulfobulbus*,  
437 *Desulfomicrobium*, *Eggerthella*, *Enterobacter*, *Erythromicrobium*, *Ewingella*, *Flavobacterium*,  
438 *Flexistipes*, *Helicobacter*, *Holophaga*, *Janthinobacterium*, *Johnsonella*, *Klebsiella*, *Lautropia*,  
439 *Leptothrix*, *Leuconostoc*, *Methanobrevibacter*, *Mogibacterium*, *Ochrobactrum*, *Olsenella*, *Oribaculum*,  
440 *Pedobacter*, *Porphyromonas-like*, *Proteus*, *Simonsiella*, *Solobacterium*, *Sphingomonas*, *Stomatococcus*,  
441 *Suttonella*, and *Tropheryma*. A histogram of the frequencies of the unique probes for each of the 76 GIs  
442 revealed that most organisms had 50 or more unique probes matching to the DNA sequencing reads,  
443 with some having up to 400 matching probes (Figure 6).



444 *Figure 6. Number of probes of 76 microorganisms (GIs) belonging to the 52 genera (Figure 4)*  
445 *matching unprocessed DNA sequencing reads. Most microorganisms had 50 or more probe matches to*  
446 *sequencing reads, indicating they were in fact present in the patient samples (even though they were not*  
447 *detected by MG-RAST presumably due to filtering).*

448 The significance of this result is that many microbial species -- not taxonomically identified in DNA  
449 sequencing results -- were actually present in the patient samples. Moreover, the inclusion of these  
450 results into the union of the two approaches increased the number of common microbial genera from 55  
451 to 100 (Figure 4). In other words, 100 of the 107 genera (~93%) identified by the Gene Meter approach

452 were found in the processed and unprocessed DNA sequencing reads. These results indicate false  
453 negatives in the taxonomic assignment of the DNA sequencing results due to filtering and false positives  
454 in the Gene Meter results (since 7 were not identified in the sequencing reads).

455 **Microorganisms identified by the Gene Meter approach but not found in the unfiltered**  
456 **sequencing reads.** Nine of the 76 GIs (~12%) not detected in any of the unfiltered DNA sequencing  
457 reads included: *Afipia sp.* ( $n=191$  probes), *Brevundimonas diminuta* ( $n=224$  probes), *Caryophanon sp.*  
458 ( $n=53$  probes), *Comamonas sp.* ( $n=19$  probes), *Delftia sp.* ( $n=4$  probes), *Escherichia coli* ( $n=17$  probes),  
459 *Rhizobium loti* ( $n=335$  probes), *Simonsiella muelleri* ( $n=31$  probes), and *Stenotrophomonas maltophilia*  
460 ( $n=68$  probes). The number of unique probes is shown because a microbial species with low number  
461 might not be detected in the unfiltered reads simply because of the low probability of finding a match. It  
462 should be noted that one species of *Leptothrix* (GI 14537943) with  $n=38$  probes did not match any  
463 unfiltered sequence reads – however, another species of *Leptothrix* (GI 14537937) with  $n=361$  probes  
464 did match 96 sequencing reads. These findings indicate that the number of unique probes did influence  
465 whether or not a particular species was found in the unfiltered sequencing reads.

466 In our study, microorganisms with a low number of unique probes ( $n=38$  or less) included: *Comamonas*  
467 *sp.*, *Delftia sp.*, *Simonsiella muelleri*, and *Escherichia coli*. We concluded that these microorganisms  
468 were probably not detected in the unfiltered sequencing reads because they had too few probes.

469 The remaining 5 species, detected by the Gene Meter approach but not detected in the unfiltered  
470 sequencing reads, had a high number of unique probes (i.e., *Afipia sp.*, *Brevundimonas diminuta*,  
471 *Caryophanon sp.*, *Rhizobium loti*, and *Stenotrophomonas maltophilia*). In theory, these microorganisms  
472 should have been detected in the unfiltered sequencing reads – but they were not. There are two  
473 possible reasons for this phenomenon. First, there might not have been sufficient read depth of the  
474 unfiltered sequencing data. According to our previous paper (Table S1 in Ref. 32), the read depth varied  
475 by patient sample with the lowest being 2,064 reads for a healthy patient and the highest being 23,904  
476 reads for a caries patient. *Afipia sp.* for example, had high abundance in one of the caries patients (8.4  
477 a.u.) and one of the patients with edentulism (7.6 a.u.) (determined by the Gene Meter approach) -- yet,  
478 the read counts for these samples were 11,532 and 18,515, respectively, indicating this microorganism  
479 should have been identified in the unfiltered DNA sequencing reads. Hence, there is weak support for  
480 the argument since the highest abundances (by the Gene Meter approach) occurred in samples that also  
481 had moderate to high read counts.

482 Second, these microorganisms might not have been found in the unfiltered sequencing reads because  
483 they are false positives of the Gene Meter approach due to non-specific hybridization to other gene  
484 targets. We tested for the potential of non-specific hybridizations by removing the first three nucleotides  
485 on the 5'- and 3'-ends from the unique probes (because they might hybridize to non-specific DNA  
486 targets) and matching the short probes (19 nt) against the unprocessed sequencing reads. We found that  
487 several probes of *Caryophanon sp.*, and *Rhizobium loti* and *Stenotrophomonas maltophilia* matched the  
488 unfiltered sequencing reads. Specifically, two unique probes (1015327 and 1015321) for *Caryophanon*  
489 *sp.* matched unfiltered sequence reads in a patient with caries and another patient with health. These  
490 probes have the potential to non-specifically bind to 16S rRNA sequences of other species in the patient  
491 samples and therefore they are false positives.

492 Ninety-six of the 151 genera (~64%) identified by DNA sequencing were not detected by the Gene  
493 Meter approach included: *Acetitomaculum*, *Acetivibrio*, *Acetobacterium*, *Acidithiobacillus*, *Acidovorax*,  
494 *Actinoplanes*, *Aerococcus*, *Agrococcus*, *Alistipes*, *Aminobacterium*, *Amycolatopsis*, *Anaerostipes*,  
495 *Aquitalea*, *Arcanobacterium*, *Arthrobacter*, *Bavariicoccus*, *Bergeriella*, *Blautia*, *Brachymonas*,

496 *Brenneria*, *Brevibacterium*, *Butyricimonas*, *Carnobacterium*, *Cellulophaga*, *Cellulosimicrobium*,  
497 *Chelonobacter*, *Chitinophaga*, *Chryseobacterium*, *Collinsella*, *Cryobacterium*, *Cytophaga*,  
498 *Dechloromonas*, *Desulfonisa*, *Desulfosporosinus*, *Desulfotomaculum*, *Enterorhabdus*,  
499 *Exiguobacterium*, *Finegoldia*, *Gallibacterium*, *Gardnerella*, *Geobacillus*, *Globicatella*, *Gordonibacter*,  
500 *Hespellia*, *Jonesia*, *Kocuria*, *Kytococcus*, *Lactococcus*, *Mannheimia*, *Megamonas*, *Melissococcus*,  
501 *Mobiluncus*, *Myroides*, *Odoribacter*, *Pantoea*, *Parabacteroides*, *Paraprevotella*, *Pasteurella*,  
502 *Pectinatus*, *Phascolarctobacterium*, *Promicromonospora*, *Pseudobutyrvibrio*, *Pseudonocardia*,  
503 *Pyramidobacter*, *Ralstonia*, *Renibacterium*, *Rhodococcus*, *Riemerella*, *Rikenella*, *Robinsoniella*,  
504 *Roseburia*, *Ruminococcus*, *Saccharopolyspora*, *Sanguibacter*, *Sebaldella*, *Serratia*, *Spirochaeta*,  
505 *Sporomusa*, *Streptobacillus*, *Streptomyces*, *Syntrophococcus*, *Tenacibaculum*, *Terrabacter*,  
506 *Tetragenococcus*, *Thermoanaerobacter*, *Thermomonospora*, *Thermus*, *Thiobacillus*, *Tissierella*,  
507 *Trichococcus*, *Vagococcus*, *Weeksella*, *Weissella*, and *Xylanimicrobium*. These microorganisms were  
508 not detected because the microarray probes did not target their 16S rRNA genes.

509 **An attempt to calibrate DNA sequencing data.** We investigated the feasibility of calibrating the 454  
510 sequencing reads by diluting pooled patient samples with DNA from salmon sperm. We chose salmon  
511 sperm because the sequences are dissimilar to microbial 16S rRNA genes and can easily be  
512 distinguished. In theory, DNA sequencing reads should yield similar linear equations for the dilution  
513 series but, in reality, amplification and sequencing biases vary for different targets and one would expect  
514 the slopes of the dilution series to be dissimilar (Figure 7). Our results show that the slope of *Neisseria*  
515 sp. was 69.0, while that of *Porphyromonas gingivalis* was 2818.2. Similarly, the slope of *Eubacterium*  
516 *brachy* was 102.7, while that of *Veillonella* sp. was 397.3. The importance of these findings are two-  
517 fold. First, it provides proof of the extreme biases of sequencing reads generated using NGS. Second,  
518 the results suggest that it might be possible to obtain precise abundances of 16S rRNA genes using NGS  
519 by calibrating 16S rRNA gene reads, before determining gene abundances, similar to what was done  
520 with the DNA microarray output using the Gene Meter approach. To our knowledge, this is the first  
521 study to provide preliminary evidence that gene abundances might be precisely determined by next-  
522 generation sequencing data.

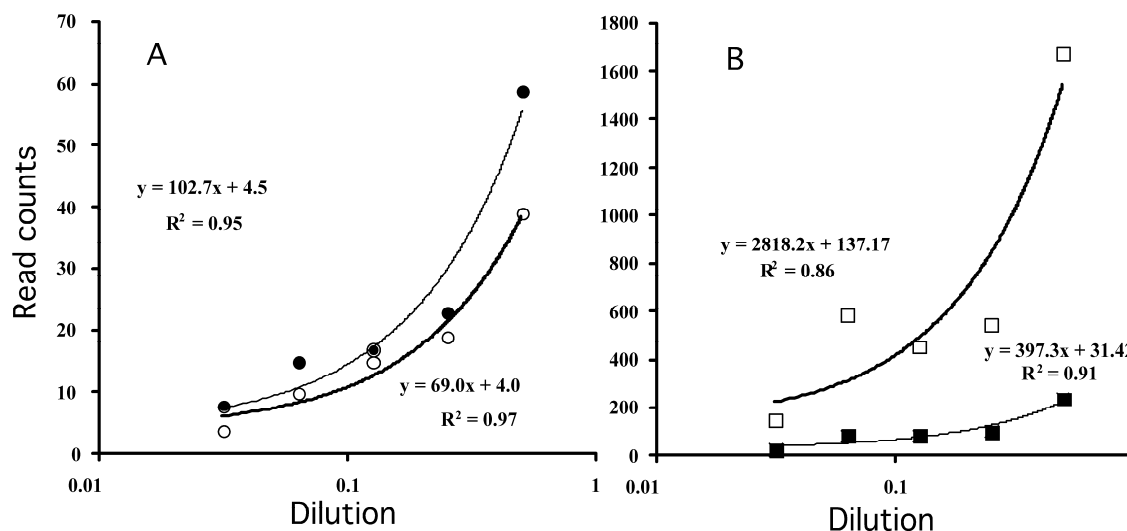


Figure 7. Calibration of 454 reads using the Gene Meter approach. Panel A: *Neisseria* sp., open circles; *Eubacterium brachy*, closed circles. Panel B: *Porphyromonas gingivalis*, open squares; *Veillonella* sp., closed squares.



523 s

## 524 **DISCUSSION**

525

### 526 **DNA microarrays - too early to abandon**

527 A common perception among researchers is that high-throughput DNA sequencing is the best approach  
528 to characterize microbial communities because other approaches, such as DNA microarrays are noisy.  
529 Indeed, conventional DNA microarrays are noisy (41) because the signal is not calibrated and the output  
530 has to be normalized for sample comparison. The Gene Meter approach solves the noise problem by  
531 calibrating all probes on a microarray before testing biological samples. The calibration identifies  
532 probes with noisy behavior because they will not fit the adsorption models. Noisy probes are not used in  
533 subsequent analyses. The Gene Meter approach also does not require any normalization of the output  
534 because the same amount of DNA is loaded onto every microarray. In other words, the abundances of  
535 different samples can be directly compared without introducing biases associated with other methods  
536 such as DNA sequencing and conventional DNA microarrays. It is for these reasons that we believe that  
537 the Gene Meter approach is superior to existing approaches – particularly when one desires to determine  
538 the abundance of a gene in a pool of genes.

539

540 It should be noted that DNA sequencing approaches (such as the one used in our previous study; 32)  
541 have problems too, that limit their ability to determine gene abundance. For example, an extra PCR  
542 amplification step is required in DNA sequencing (e.g., emulsion PCR) and that adds biases to the  
543 interpretation of the output. The Gene Meter approach does not need any extra amplification step – the  
544 labeled DNA is added directly to the microarray. Another problem in DNA sequencing (as shown in  
545 this study) is the post-processing of the reads sometimes filter out 16S rRNA genes that are actually  
546 present in biological samples (specifically, 40 out of 107 genera in this study). The fact that many of the  
547 unique probes of the 40 unidentified genera matched the unprocessed sequencing reads suggest that the  
548 taxonomic assignments in other sequencing studies might have grossly underestimated the actual  
549 diversity of microorganisms in biological samples.

550

551 We recognize that quantitative PCR (qPCR) could also have been used to determine the abundance of  
552 16S rRNA genes in our study. However, like DNA sequencing, this approach is subject to PCR  
553 amplification biases and normalization steps. In addition, qPCR can only investigate a limited number  
554 of gene targets at one time (i.e., it is not high throughput), while the Gene Meter approach is high  
555 throughput but limited to the number of probes on microarray surface.

556

557 Similar to all approaches, the Gene Meter approach is not without its problems. For instance, seven  
558 genera identified by the Gene Meter approach were not detected in either the processed or unprocessed  
559 DNA sequencing reads. Three of the 7 genera were classified as false positives as they had the potential  
560 to cross-hybridize to the 16S rRNA genes of other species. Although there are several additional  
561 explanations for these phenomena (e.g., insufficient DNA sequencing depth; need for better probe  
562 design), the Gene Meter approach still provides precise target abundances for many genera and is  
563 superior to other approaches because normalization is not required.

564

565 A simple solution to prevent normalization biases and improve upon abundance determinations for DNA  
566 sequencing is to calibrate the instrument before analyzing samples. This study was the first to  
567 demonstrate that a DNA sequencing instrument can be calibrated using a dilution series (Figure 7).

568 Future studies are now warranted to demonstrate the utility of calibration for DNA sequencing  
569 instruments using biological samples.

570

### 571 **Calibration by aggregate probes works optimally**

572 In early Gene Meter studies (18,19), the dilution series were calibrated using the averaged signal  
573 intensities of all replicated probes specific to a gene target. The averaging of the signal reduced noise by  
574 minimizing the effects of outliers. Since we did not use identically replicated probes in the present  
575 study, we calibrated the DNA microarray two ways using individual probes or aggregated probes. We  
576 sought to determine which of the two calibrations was optimal.

577 Calibration of the individual probes involved modeling the signal intensities of an individual probe in a  
578 dilution series. The abundance of the gene in a biological sample was determined by back-calculating  
579 the abundances of all probes which target that specific gene using the calibration and then averaging (or  
580 determining the median) the gene abundances. For example, consider three calibrated probes targeting  
581 the same gene. The abundance of a specific gene by calibrated probe A was 1 a.u., probe B was 2 a.u.  
582 and probe C was 3 a.u. Therefore, the final abundance of the gene was the average or median, 2 a.u.

583 Calibration of the aggregated probes involved summing the signal intensities of all probes in a dilution  
584 series that target a specific gene and determining the model that best fits the experimental data. For  
585 example, if the signal intensities of probes A, B, and C for a specific gene were 300, 500, and 700  
586 relative fluorescence units (RFU) at one dilution, then the signal intensity used to calibrate the gene  
587 would be 1500 RFU. The abundance of a gene in a biological sample was determined by summing the  
588 signal intensities of all probes that target the gene and back-calculating the relative abundance from the  
589 calibration.

590 What distinguishes the two calibration approaches is that the aggregate probe approach uses the sum of  
591 the signal intensities of all probes as input to the calibration model while the individual probe approach  
592 uses only the signal intensity of an individual probe as input to the calibration model. Also, in the  
593 aggregate probe approach, the abundance is directly determined from the model, while in the individual  
594 probe approach, the final gene abundance is determined by the averaged (or median) abundance of all  
595 probes targeting that gene.

596 We showed that calibrations of aggregated probes were better than individual probes when the gene  
597 target abundances were low. The reason for this phenomenon is that the low abundance genes are more  
598 affected by signal noise because some readings were close to the level of resolution of the microarray  
599 scanner. The sum of the signal intensities of all probes targeting a specific gene minimizes this problem  
600 because the noise is not averaged. Our results showed that if the gene target abundance was greater than  
601 2 a.u. (Figure 4), there were no significant differences in the gene abundances by the calibration  
602 approach.

603 Interestingly, we found more differences in gene abundances by patient condition using aggregated  
604 probes (Tables 3, 4 and 5) than individual probes (Table 1 and 2). This finding suggests that several  
605 microorganisms do have differences in abundances by condition -- but they occur at low abundances (<2  
606 a.u.). In other words, the differences by condition could only be detected using the aggregated probe  
607 approach. Therefore, the aggregated probes was used for examining microbial species differences by  
608 condition (below).

### 609 **Abundance signatures for dysbiosis**

610 The following 17 genera had significantly higher abundances in patients with periodontitis and patients  
611 with edentulism when contrasted with patients with health: *Actinomyces*, *Bifidobacterium*, *Butyrivibrio*,  
612 *Campylobacter*, *Capnocytophaga*, *Eubacterium*, *Klebsiella*, *Lactobacillus*, *Leptotrichia*, *Neisseria*,  
613 *Peptostreptococcus*, *Porphyromonas*, *Prevotella*, *Selenomonas*, *Streptococcus*, *Treponema*, and  
614 *Veillonella*. At the species level this included: *Actinomyces odontolyticus*, *Klebsiella pneumoniae*,  
615 *Peptostreptococcus anaerobius*, *Streptococcus gordonii*, *Treponema socranskii*, and *Veillonella dispar*.  
616 We classified these microorganisms as ‘signatures of dysbiosis’ in the human oral microbiome because  
617 they occurred in more than one condition. They were excluded from consideration to a single condition  
618 below (e.g., a signature for periodontitis).

619 These microorganisms have previously been reported to be involved in dysbiosis in human oral studies.  
620 For example, *Actinomyces* sp. are found in edentulous patients (42). *Bifidobacterium* are found in the  
621 mouths of edentulous patient wearing dentures (43) and associated with root decay and periodontal  
622 disease (44). *Butyrivibrio*, *Campylobacter*, *Eubacterium*, *Prevotella*, *Selenomonas*, *Streptococcus*,  
623 *Leptotrichia*, and *Treponema* are associated with peri-implant communities (45). *Eubacterium*,  
624 *Prevotella* and *Lactobacillus* are major genera found in edentulous patients (46). *Prevotella*,  
625 *Fusobacterium*, *Leptotrichia*, *Streptococcus*, *Neisseria*, and *Veillonella* were found in edentulous infants  
626 (24,47,48). *Peptostreptococcus* has been associated with periodontal pockets of partially edentate  
627 patients (49) and is associated with soft tissue inflammation in edentate patients (42). *Prevotella* and  
628 *Campylobacter* have been observed in edentate patients (42). *Treponema* *Campylobacter*,  
629 *Peptostreptococcus*, and *Porphyromonas* are associated with periodontitis (50, 51). *Veillonella*,  
630 *Streptococcus* and *Candida* are often found in the saliva of edentulous patients (52).

631 At the species level, *Actinomyces odontolyticus* has been found in failed dental implants (53). *Klebsiella*  
632 *pneumoniae* has been frequently observed in patients with diseased implants (54) and patients with  
633 periodontitis (55,56). *Peptostreptococcus anaerobius* has been found inside root canals (50).  
634 *Streptococcus gordonii* has been shown to be an important player in the development of periodontitis as  
635 it is involved in co-adhesion and metabolic interactions with other pathogens (53,57). *Treponema*  
636 *socranskii* has been associated with peri-implant diseases (58,59), lesions of teeth with apical  
637 periodontitis (60), infected root canals and abscesses (61). *Veillonella dispar* is an earlier biofilm  
638 colonizer (62) and involved in periodontal disease (63).

639

## 640 **Abundance signatures for periodontitis**

641 The following 13 genera had significantly higher abundances in patients with periodontitis than those  
642 with health: *Abiotrophia*, *Bacteroides-like*, *Caryophanon*, *Clostridium*, *Dialister*, *Filifactor*,  
643 *Fusobacterium*, *Granulicatella*, *Haemophilus*, *Megasphaera*, *Mogibacterium*, *Staphylococcus*, and  
644 *Tannerella*. At the species level this included: *A. elegans*, *C. tetanomorphum*, *F. alocis*, *F. nucleatum*,  
645 *G. elegans*, *H. segnis*, *M. diversum*, *S. epidermidis*, *S. xylosum*, and *T. forsythia*. Note that some genera  
646 could not be resolved to the species level. These microorganisms were classified as microbial signatures  
647 of periodontitis in the oral microbiome.

648 The following microorganisms have been previously shown to be associated with periodontal disease:  
649 *Abiotrophia elegans* (64,65), *Bacteroides-like* (66), *Dialister* (67,68,69), *Filifactor alocis*  
650 (70,71,72,73,74), *Fusobacterium nucleatum* (15,42,75,76,77,78), *Granulicatella elegans* (79),  
651 *Haemophilus* (70,80,81), *Haemophilus segnis* (82,83) *Megasphaera* (4,84), *Mogibacterium* (85,86),  
652 *Staphylococcus epidermidis* (87), *Tannerella forsythia* (71,77,89,90). These findings support our  
653 findings that these microorganisms are signatures of periodontitis in the human oral microbiome.



654 Although the genera *Clostridium Mogibacterium* and *Staphylococcus* have been associated with  
655 periodontal disease (10,91), there is no evidence in the literature suggesting that *Clostridium*  
656 *tetanomorphum*, *Mogibacterium diversum*, *Staphylococcus xylosus* or the genus *Caryophanon* has been  
657 associated with periodontitis. Of note, *Mogibacterium diversum* has been associated with caries (88)  
658 and the genera *Caryophanon* has been found in the gingival flora of dogs (92). Recall that our earlier  
659 results suggest that *Caryophanon* might be a false positive because it was not found in the sequencing  
660 reads and some probes were shown to have the potential to hybridize with other species.

661 A probable reason these microorganisms were not recognized in the literature is because they occurred  
662 at low abundances and might be overlooked using alternative technologies (e.g., NGS). For example, of  
663 the 525 *Clostridium tetanomorphum* probes, the abundance in patients with health was  $1.8 \pm 0.50$  a.u.  
664 versus  $3.1 \pm 0.81$  a.u. in patients with periodontitis. Similarly, of the 114 *Mogibacterium diversum*  
665 probes, the abundance in patients with health was  $1.6 \pm 0.30$  a.u. versus  $2.8 \pm 0.72$  a.u. in patients with  
666 periodontitis, and of the 54 *Staphylococcus xylosus* probes, the abundance in patients with health was  
667  $1.5 \pm 0.25$  a.u. versus  $2.4 \pm 0.48$  a.u. in patients with periodontitis (Table 5). All comparisons are  
668 statistically significant and all species have been previously found in human oral cavities.

669

## 670 **Abundance signatures for edentulism**

671 The following 14 genera had significantly higher abundances in patients with edentulism than patients  
672 with health: *Atopobium*, *Bacillus*, *Burkholderia*, *Candida*, *Deinococcus*, *Erysipelothrix*, *Ewingella*,  
673 *Kingella*, *Mycoplasma*, *Paenibacillus*, *Pedobacter*, *Propionibacterium*, *Solobacterium*, and  
674 *Stomatococcus*. At the species level, this included: *Atopobium rimae*, *Bacillus clausii*, *Candida albicans*,  
675 *Erysipelothrix tonsillarum*, *Ewingella americana*, *Kingella denitrificans*, *Mycoplasma buccale*,  
676 *Propionibacterium avidum*, and *Stomatococcus mucilaginosus*. While high abundances of *Atopobium*  
677 species, *Candida albicans* and *Bacillus sp.* have been previously reported in edentulous patients (93), we  
678 could not find any evidence for the high abundance of the other genera and species in patients with  
679 edentulism in the literature. Hence, this is the first study to show higher abundance of these  
680 microorganisms in patients with edentulism.

681

## 682 **Technology rules our understanding of the oral microbiome.**

683 Much of what is known about the oral microbiome is based on the technology available at the time of  
684 publication. For example, the idea behind the ‘red complex’ originated from the checkerboard DNA–  
685 DNA hybridization technique, which was used to examine the microbial composition of plaque in  
686 patients in health and periodontitis (94,95), the salivary microbiota levels in relation to periodontal status  
687 (96), the relationship of cigarette smoking to the composition of the subgingival microbiota (97,98), the  
688 differences between the subgingival microbiota in patients from different geographic locations (99), the  
689 relationship of ethnic/racial group, occupational and periodontal disease status (100), and effects of  
690 different periodontal therapies (101,102). While the technique is rapid, sensitive, and relatively  
691 inexpensive, its major shortcoming was cross-hybridization (103). The implementation of new  
692 technologies such as DNA microarrays and next-generation-sequencing improved upon our  
693 understanding of the human oral microbiome (28) – but these technologies are not really quantitative.  
694 Gene Meter methodology allows researcher to precisely determine microbial abundances, which allows  
695 for statistical comparisons of different oral clinical conditions. We used these comparisons to find  
696 abundance signatures for dysbiosis, periodontitis and edentulism. These signatures could be used,

697 individually or in combination, to assess the clinical status of a patient (e.g., evaluating treatments such  
698 as antibiotic therapies, and oral microbial transplantations, e.g., 105).

699

## 700 ACKNOWLEDGEMENTS

701

## 702 CONTRIBUTIONS

703

704 Experimental design: PAN and AP.

705 Laboratory Work: MCH, PAN, and AP.

706 Bioinformatic Work: PAN and AP.

707 Manuscript writing: MCH, PAN, and AP.

708

## 709 FUNDING

710

711 This work was supported by: the International Team for Implantology (grant number  
712 954\_2013), the University of Washington Hach Memorial Fund, and a training grant  
713 from the National Institutes of Health (grant number 5 T90 D021984-03).

## 714 LITERATURE CITED

715

- 716 1. Armitage GC. Development of a classification system for periodontal diseases and conditions. *Ann*  
717 *Periodontol.* 1999; 4:1-6. PMID: 10863370
- 718 2. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A, Wade WG. The  
719 human oral microbiome. *J Bacteriol.* 2010; 192:5002-17. doi: 10.1128/JB.00542-10. PMID:  
720 20656903
- 721 3. Holt SC, Ebersole JL. *Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia*:  
722 the "red complex", a prototype polybacterial pathogenic consortium in periodontitis. *Periodontol*  
723 2000, 2005; 38:72-122. PMID: 15853938
- 724 4. Kumar PS, Griffen AL, Barton JA, Paster BJ, Moeschberger ML, Leys EJ. New bacterial species  
725 associated with chronic periodontitis. *J Dent Res.* 2003; 82:338-44. PMID: 12709498
- 726 5. Darveau RP. Periodontitis: a polymicrobial disruption of host homeostasis. *Nat Rev Microbiol.*  
727 2010; 8:481-90. doi: 10.1038/nrmicro2337. PMID: 20514045
- 728 6. Hajishengallis G, Liang S, Payne MA, Hashim A, Jotwani R, Eskan MA, McIntosh ML, Alsam A,  
729 Kirkwood KL, Lambris JD, Darveau RP, Curtis MA. Low-abundance biofilm species orchestrates  
730 inflammatory periodontal disease through the commensal microbiota and complement. *Cell Host*  
731 *Microbe.* 2011; 10:497-506. doi: 10.1016/j.chom.2011.10.006. PMID: 22036469
- 732 7. Diaz PI, Dupuy AK, Abusleme L, Reese B, Oberfell C, Choquette L, Dongari-Bagtzoglou A,  
733 Peterson DE, Terzi E, Strausbaugh LD. Using high throughput sequencing to explore the  
734 biodiversity in oral bacterial communities. *Mol Oral Microbiol.* 2012; 27:182-201. doi:  
735 10.1111/j.2041-1014.2012.00642.x. PMID: 22520388
- 736 8. Bik EM, Long CD, Armitage GC, Loomer P, Emerson J, Mongodin EF, Nelson KE, Gill SR, Fraser-  
737 Liggett CM, Relman DA. Bacterial diversity in the oral cavity of 10 healthy individuals. *ISME J.*  
738 2010; 4:962-74. doi: 10.1038/ismej.2010.30. PMID: 20336157
- 739 9. Liu B, Faller LL, Klitgord N, Mazumdar V, Ghodsi M, Sommer DD, Gibbons TR, Treangen TJ,  
740 Chang YC, Li S, Stine OC, Hasturk H, Kasif S, Segrè D, Pop M, Amar S. Deep sequencing of the  
741 oral microbiome reveals signatures of periodontal disease. *PLoS One.* 2012; 7:e37919. doi:  
742 10.1371/journal.pone.0037919. PMID: 22675498
- 743 10. Griffen AL, Beall CJ, Campbell JH, Firestone ND, Kumar PS, Yang ZK, Podar M, Leys EJ. Distinct  
744 and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing.  
745 *ISME J.* 2012; 6:1176-1185. PMID: 22170420
- 746 11. Zaura E, Keijsers BJ, Huse SM, Crielaard W. Defining the healthy "core microbiome" of oral  
747 microbial communities. *BMC Microbiol.* 2009; 15:259. PMID: 22313693
- 748 12. Hajishengallis G, Lamont RJ. Beyond the red complex and into more complexity: the polymicrobial  
749 synergy and dysbiosis (PSD) model of periodontal disease etiology. *Molecul Oral Microbiol,* 2012;  
750 27:409–419. PMID: 23134607
- 751 13. Huyghe A, Francois P, Charbonnier Y, Tangomo-Bento M, Bonetti EJ, Paster BJ, Bolivar I, Baratti-  
752 Mayer D, Pittet D, Schrenzel J; Geneva Study Group on Noma (GESNOMA). Novel microarray  
753 design strategy to study complex bacterial communities. *Appl Environ Microbiol.* 2008; 74:1876-85.  
754 doi: 10.1128/AEM.01722-07. PMID: 18203854

- 755 14. Iwai S, Fei M, Huang D, Fong S, Subramanian A, Grieco K, Lynch SV, Huang L. Oral and airway  
756 microbiota in HIV-infected pneumonia patients. *J Clin Microbiol.* 2012; 50:2995-3002. doi:  
757 10.1128/JCM.00278-12. PMID: 22760045
- 758 15. Topcuoglu N, Kulekci G. 16S rRNA based microarray analysis of ten periodontal bacteria in patients  
759 with different forms of periodontitis. *Anaerobe.* 2015; 35(Pt A):35-40. doi:  
760 10.1016/j.anaerobe.2015.01.011. PMID: 25638399
- 761 16. Roth SB, Jalava J, Ruuskanen O, Ruohola A, Nikkari S. Use of an oligonucleotide array for  
762 laboratory diagnosis of bacteria responsible for acute upper respiratory infections. *J Clin Microbiol.*  
763 2004; 42:4268-74. PMID: 15365022
- 764 17. Pozhitkov AE, Bailey KD, Noble PA. Development of a statistically robust quantification method  
765 for microorganisms in mixtures using oligonucleotide microarrays. *J Microbiol Methods.* 2007;  
766 70:292-300. PMID: 17553581
- 767 18. Pozhitkov A.E., P.A. Noble, J. Bryk, D. Tautz. A revised design for microarray experiments to  
768 account for experimental noise and the uncertainty of probe response. *PlosOne* 2014; 9: e91295  
769 PMID: 24618910
- 770 19. Harrison A., Binder, H. Buhot, A., Burden, C.J., Carlon, E., Gibas, C., Gamble, L.J. Halperin, A.,  
771 Hooyberghs, J., Kreil, D.P., Levicky, R., Noble, P.A., Ott, A., Pettitt, B.M., Tautz, D., Pozhitkov,  
772 AE. Physico-chemical foundations underpinning microarray and next-generation sequencing  
773 experiments. *Nucl Acids Res.* 2013; 41: 2779-96. PMID: 23307556
- 774 20. Hunter M.C., Pozhitkov A.E. and Noble P.A. Accurate predictions of postmortem interval using  
775 linear regression analyses of gene meter expression data. 2016;  
776 <http://www.biorxiv.org/content/early/2016/06/12/058370>
- 777 21. Pozhitkov A.E., Neme R., Domazet-Loo T., Leroux B.G., Soni S., Tautz D, and Noble P.A.  
778 Thanatotranscriptome: genes actively expressed after organismal death. 2016;  
779 <http://www.biorxiv.org/content/early/2016/06/11/058305>
- 780 22. Nasidze I, Li J, Quinque D, Tang K, Stoneking M. Global diversity in the human salivary  
781 microbiome. *Genome Res.* 2009; 19:636-43. doi: 10.1101/gr.084616.108. Epub 2009 Feb 27. PMID:  
782 19251737
- 783 23. Ahn J, Yang L, Paster BJ, Ganly I, Morris L, Pei Z, Hayes RB. Oral microbiome profiles: 16S rRNA  
784 pyrosequencing and microarray assay comparison. *PLoS One.* 2011; 6:e22788. doi:  
785 10.1371/journal.pone.0022788. PMID: 21829515
- 786 24. Cephas KD, Kim J, Mathai RA, Barry KA, Dowd SE, Meline BS, Swanson KS. Comparative  
787 analysis of salivary bacterial microbiome diversity in edentulous infants and their mothers or  
788 primary care givers using pyrosequencing. *PLoS One.* 2011; 6:e23503. doi:  
789 10.1371/journal.pone.0023503. PMID: 21853142
- 790 25. Simón-Soro A, Tomás I, Cabrera Rubio R, Catalan MD, Nyvad B, Mira A. Microbial geography of  
791 the oral cavity. *J Dent Res.* 2013; 92:616-21. doi: 10.1177/0022034513488119. PMID: 23674263
- 792 26. Xu H, Hao W, Zhou Q, Wang W, Xia Z, Liu C, Chen X, Qin M, Chen F. Plaque bacterial  
793 microbiome diversity in children younger than 30 months with or without caries prior to eruption of  
794 second primary molars. *PLoS One.* 2014; 9:e89269. doi: 10.1371/journal.pone.0089269. PMID:  
795 24586647
- 796 27. Sato Y, Yamagishi J, Yamashita R, Shinozaki N, Ye B, Yamada T, Yamamoto M, Nagasaki M,  
797 Tsuboi A. Inter-individual differences in the oral bacteriome are greater than intra-day fluctuations  
798 in individuals. *PLoS One.* 2015; 10:e0131607. doi: 10.1371/journal.pone.0131607. PMID: 26121551

- 799 28. Pozhitkov AE, Beikler T, Flemmig T, Noble PA. High-throughput methods for analysis of the  
800 human oral microbiome. *Periodontol* 2000. 2011;55:70-86. doi: 10.1111/j.1600-0757.2010.00380.x.  
801 PMID: 21134229
- 802 29. Amend Anthony S, Seifert Keith A., Bruns Thomas D. Quantifying microbial communities with 454  
803 pyrosequencing: does read abundance count? *Molecular Ecol.* 2010; 19:124. PMID: 21050295
- 804 30. Dannemiller, N. Lang-Yona, N. Yamamoto, Y. Rudich, J. Peccia. Combining real-time PCR and  
805 next-generation DNA sequencing to provide quantitative comparisons of fungal aerosol populations.  
806 *Atmospheric Environment*, 2014: 84:113–121.
- 807 31. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, Gingeras TR, Oliver B. Synthetic spike-  
808 in standards for RNA-seq experiments. *Genome Res.* 2011; 21:1543-51. PMID: 21816910
- 809 32. Pozhitkov A.E., Leroux B., Randolph T.W., Beikler T., Flemmig T.F., and Noble P.A. Towards  
810 microbiome transplant as a therapy for periodontitis: an exploratory study of periodontitis microbial  
811 signature contrasted by oral health, caries and edentulism. *BMC Oral Health* 2015; 15:125 PMID:  
812 26468081
- 813 33. Page RC, Eke PI. Case definitions for use in population-based surveillance of periodontitis. *J*  
814 *Periodontol.* 2007; 78:1387–99. PMID: 17608611
- 815 34. Dye BA, Tan S, Smith V, Lewis BG, Barker LK, Thornton-Evans G, et al. Trends in oral health  
816 status: United States, 1988–1994 and 1999–2004. National Center for Health Statistics. *Vital Health*  
817 *Stat.* 2007; 11:248.
- 818 35. Beikler T, Schnitzer S, Abdeen G, Ehmke B, Eisenacher M, Flemmig TF. Sampling strategy for  
819 intraoral detection of periodontal pathogens before and following periodontal therapy. *J Periodontol.*  
820 2006; 77:1323–3132. PMID: 16881801
- 821 36. Flemmig TF, Arushanov D, Daubert D, Rothen M, Mueller G, Leroux BG. Randomized controlled  
822 trial assessing efficacy and safety of glycine powder air polishing in moderate to deep periodontal  
823 pockets. *J Periodontol.* 2012; 83:444–52. PMID: 21861637
- 824 37. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens  
825 R, Wilke A, Wilkening J, Edwards RA. The metagenomics RAST server - a public resource for the  
826 automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinf.* 2008; 9:386. PMID:  
827 18803844
- 828 38. Griffen AL, Beall CJ, Firestone ND, Gross EL, Difranco JM, Hardman JH, Vriesendorp B, Faust  
829 RA, Janies DA, Leys EJ. CORE: a phylogenetically-curated 16S rDNA database of the core oral  
830 microbiome. *PLoS One.* 2011 Apr 22;6(4):e19051. doi: 10.1371/journal.pone.0019051. PMID:  
831 21544197
- 832 39. Pozhitkov A, Noble PA, Domazet-Lozo T, Nolte AW, Sonnenberg R, Staehler P, Beier M, Tautz D.  
833 Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of  
834 oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res.* 2006;  
835 17;34:e66. PMID: 16707658
- 836 40. Diaz PI, Chalmers NI, Rickard AH, Kong C, Milburn CL, Palmer RJ Jr, Kolenbrander PE.  
837 Molecular characterization of subject-specific oral microflora during initial colonization of enamel.  
838 *Appl Environ Microbiol.* 2006; 72:2837-2848. PMID: 16597990
- 839 41. Tu Y, Stolovitzky G, Klein U. Quantitative noise analysis for gene expression microarray  
840 experiments. *Proc Natl Acad Sci USA.* 2002; 99:14031. PMID: 12388780



- 841 42. Danser MM, vanWinkelhoff AJ, van der Velden U. Periodontal bacteria colonizing oral mucous  
842 membranes in edentulous patients wearing dental implants. *J Periodontol.* 1997; 68:209-16.
- 843 43. Mantzourani M, Gilbert SC, Fenlon M, Beighton D. Non-oral bifidobacteria and the aciduric  
844 microbiota of the denture plaque biofilm. *Mol Oral Microbiol.* 2010; 25:190-9. doi: 10.1111/j.2041-  
845 1014.2009.00565.x. PMID: 20536746
- 846 44. Dina MN, Mărgărit R, Andrei OC. Pontic morphology as local risk factor in root decay and  
847 periodontal disease. *Rom J Morphol Embryol.* 2013; 54:361-4. PMID: 23771082
- 848 45. Kumar PS, Mason MR, Brooker MR, O'Brien K. Pyrosequencing reveals unique microbial  
849 signatures associated with healthy and failing dental implants. *J Clin Periodontol.* 2012; 39:425-33.  
850 doi: 10.1111/j.1600-051X.2012.01856.x. PMID: 22417294
- 851 46. Könönen E, Asikainen S, Alaluusua S, Könönen M, Summanen P, Kanervo A, Jousimies-Somer H.  
852 Are certain oral pathogens part of normal oral flora in denture-wearing edentulous subjects? *Oral*  
853 *Microbiol Immunol.* 1991; 6:119-22. PMID: 1945487
- 854 47. Teles FR, Teles RP, Sachdeo A, Uzel NG, Song XQ, Torresyap G, Singh M, Papas A, Haffajee AD,  
855 Socransky SS. Comparison of microbial changes in early redeveloping biofilms on natural teeth and  
856 dentures. *J Periodontol.* 2012; 83:1139-48. doi: 10.1902/jop.2012.110506. PMID: 22443543
- 857 48. Könönen E, Asikainen S, Jousimies-Somer H. The early colonization of gram-negative anaerobic  
858 bacteria in edentulous infants. *Oral Microbiol Immunol.* 1992; 7:28-31. PMID: 1528621
- 859 49. vanWinkelhoffAJ, Loos BG, van der Reijden WA, van der Velden U. *Porphyromonas gingivalis*,  
860 *Bacteroides forsythus* and other putative periodontal pathogens in subjects with and without  
861 periodontal destruction. *J Clin Periodontol.* 2002; 29:1023-8. PMID: 12472995
- 862 50. Piovano, S. Bacteriology of most frequent oral anaerobic infections. *Anaerobe* 1999; 5:221-227
- 863 51. vanWinkelhoff AJ, Goené RJ, Benschop C, Folmer T. Early colonization of dental implants by  
864 putative periodontal pathogens in partially edentulous patients. *Clin Oral Implants Res.* 2000;  
865 11:511-20. PMID: 11168244
- 866 52. Sato M, Hoshino E, Nomura, S, Ishioka K. Salivary microflora of geriatric edentulous persons  
867 wearing dentures. *Microbial ecology in health and disease.* 1993; 6: 293-299.
- 868 53. Sakanaka A, Kuboniwa M, Takeuchi H, Hashino E, Amano A. Arginine-ornithine antiporter ArcD  
869 controls arginine metabolism and interspecies biofilm development of *Streptococcus gordonii*. *J Biol*  
870 *Chem.* 2015; 290:21185-98. doi: 10.1074/jbc.M115.644401. PMID: 26085091
- 871 54. Sarkonen N, Könönen E, Eerola E, Könönen M, Jousimies-Somer H, Laine P. Characterization of  
872 *Actinomyces* species isolated from failed dental implant fixtures. *Anaerobe.* 2005;11:231-7. PMID:  
873 16701573
- 874 55. Ardila CM, Fernández N, Guzmán IC. Antimicrobial susceptibility of moxifloxacin against gram-  
875 negative enteric rods from Colombian patients with chronic periodontitis. *J Periodontol.* 2010;  
876 81:292-9. doi: 10.1902/jop.2009.090464. PMID: 20151809
- 877 56. Gonçalves MO, Coutinho-Filho WP, Pimenta FP, Pereira GA, Pereira JA, Mattos-Guaraldi AL,  
878 Hirata R Jr. Periodontal disease as reservoir for multi-resistant and hydrolytic enterobacterial  
879 species. *Lett Appl Microbiol.* 2007;44:488-94. PMID: 17451514
- 880 57. Hajishengallis G, Lamont RJ. Dancing with the stars: how choreographed bacterial interactions  
881 dictate nososymbiocity and give rise to keystone pathogens, accessory pathogens, and pathobionts.  
882 *Trends Microbiol.* 2016; 24:477-89. doi: 10.1016/j.tim.2016.02.010.

- 883 58. Máximo MB, de Mendonça AC, Renata Santos V, Figueiredo LC, Feres M, Duarte PM. Short-term  
884 clinical and microbiological evaluations of peri-implant diseases before and after mechanical anti-  
885 infective therapies. *Clin Oral Implants Res.* 2009; 20:99-108. doi: 10.1111/j.1600-  
886 0501.2008.01618.x. PMID: 19126114
- 887 59. Persson GR, Renvert S. Cluster of bacteria associated with peri-implantitis. *Clin Implant Dent Relat*  
888 *Res.* 2014;16:783-93. doi: 10.1111/cid.12052. PMID: 23527870
- 889 60. Rosa TP, Signoretti FG, Montagner F, Gomes BP, Jacinto RC. Prevalence of *Treponema* spp. in  
890 endodontic retreatment-resistant periapical lesions. *Braz Oral Res.* 2015;29. pii: S1806-  
891 83242015000100228. doi: 10.1590/1807-3107BOR-2015.vol29.0031. PMID: 25627883
- 892 61. Montagner F, Jacinto RC, Signoretti FG, Gomes BP. *Treponema* species detected in infected root  
893 canals and acute apical abscess exudates. *J Endod.* 2010;36:1796-9. doi: 10.1016/j.joen.2010.08.008.  
894 PMID: 20951290
- 895 62. Mashima I, Nakazawa F. The influence of oral *Veillonella* species on biofilms formed by  
896 *Streptococcus* species. *Anaerobe.* 2014; 28:54-61. doi: 10.1016/j.anaerobe.2014.05.003. PMID:  
897 24862495
- 898 63. Moon JH, Lee JH, Lee JY. Subgingival microbiome in smokers and non-smokers in Korean chronic  
899 periodontitis patients. *Mol Oral Microbiol.* 2015;30:227-41. doi: 10.1111/omi.12086. PMID:  
900 25283067
- 901 64. Rego RO, Oliveira CA, dos Santos-Pinto A, Jordan SF, Zambon JJ, Cirelli JA, Haraszthy V. Clinical  
902 and microbiological studies of children and adolescents receiving orthodontic treatment. *Am J Dent.*  
903 2010 Dec;23(6):317-23. PMID: 21344829
- 904 65. Mikkelsen L, Theilade E, Poulsen K. *Abiotrophia* species in early dental plaque. *Oral Microbiol*  
905 *Immunol.* 2000;15:263-8. PMID: 11154413
- 906 66. Olsvik B, Flynn MJ, Tenover FC, Slots J, Olsen I. Tetracycline resistance in *Prevotella* isolates from  
907 periodontally diseased patients is due to the tet(Q) gene. *Oral Microbiol Immunol.* 1996; 11:304-8.  
908 PMID: 9028255
- 909 67. Lee MY, Kim YJ, Gu HJ, Lee HJ. A case of bacteremia caused by *Dialister pneumosintes* and  
910 *Slackia exigua* in a patient with periapical abscess. *Anaerobe.* 2016; 38:36-8. doi:  
911 10.1016/j.anaerobe.2015.11.006. PMID: 26612007
- 912 68. Gonçalves C, Soares GM, Faveri M, Pérez-Chaparro PJ, Lobão E, Figueiredo LC, Baccelli GT,  
913 Feres M. Association of three putative periodontal pathogens with chronic periodontitis in Brazilian  
914 subjects. *J Appl Oral Sci.* 2016; 24:181-5. doi: 10.1590/1678-775720150445. PMID: 27119767
- 915 69. Collins JR, Arredondo A, Roa A, Valdez Y, León R, Blanc V. Periodontal pathogens and  
916 tetracycline resistance genes in subgingival biofilm of periodontally healthy and diseased Dominican  
917 adults. *Clin Oral Investig.* 2016; 20:349-56. doi: 10.1007/s00784-015-1516-2. PMID: 26121972
- 918 70. Colombo AP, Boches SK, Cotton SL, Goodson JM, Kent R, Haffajee AD, Socransky SS, Hasturk H,  
919 Van Dyke TE, Dewhirst F, Paster BJ. Comparisons of subgingival microbial profiles of refractory  
920 periodontitis, severe periodontitis, and periodontal health using the human oral microbe  
921 identification microarray. *J Periodontol.* 2009; 80:1421-32. doi: 10.1902/jop.2009.090185. PMID:  
922 19722792
- 923 71. Oliveira RR, Fermiano D, Feres M, Figueiredo LC, Teles FR, Soares GM, Faveri M. Levels of  
924 candidate periodontal pathogens in subgingival biofilm. *J Dent Res.* 2016; 95:711-8. doi:  
925 10.1177/0022034516634619. PMID: 26936213

- 926 72. Basic A, Dahlén G. Hydrogen sulfide production from subgingival plaque samples. *Anaerobe*. 2015;  
927 35(Pt A):21-7. doi: 10.1016/j.anaerobe.2014.09.017. PMID: 25280920
- 928 73. Aruni AW, Mishra A, Dou Y, Chioma O, Hamilton BN, Fletcher HM. *Filifactor alocis*--a new  
929 emerging periodontal pathogen. *Microbes Infect*. 2015; 17:517-30. doi:  
930 10.1016/j.micinf.2015.03.011. PMID: 25841800
- 931 74. Chen H, Liu Y, Zhang M, Wang G, Qi Z, Bridgewater L, Zhao L, Tang Z, Pang X. A *Filifactor*  
932 *alocis*-centered co-occurrence group associates with periodontitis across different oral habitats. *Sci*  
933 *Rep*. 2015; 5:9053. doi: 10.1038/srep09053. PMID: 25761675
- 934 75. Ahn SH, Song JE, Kim S, Cho SH, Lim YK, Kook JK, Kook MS, Lee TH. NOX1/2 activation in  
935 human gingival fibroblasts by *Fusobacterium nucleatum* facilitates attachment of *Porphyromonas*  
936 *gingivalis*. *Arch Microbiol*. 2016; 198:573-83. doi: 10.1007/s00203-016-1223-7. PMID: 27071620
- 937 76. Mendes RT, Nguyen D, Stephens D, Pamuk F, Fernandes D, Van Dyke TE, Kantarci A. Endothelial  
938 cell response to *Fusobacterium nucleatum*. *Infect Immun*. 2016; 84:2141-8. doi: 10.1128/IAI.01305-  
939 15. PMID: 27185790
- 940 77. Nickles K, Scharf S, Röllke L, Mayer I, Mayer M, Eickholz P. Detection of subgingival periodontal  
941 pathogens--comparison of two sampling strategies. *Clin Oral Investig*. 2016; 20:571-9. doi:  
942 10.1007/s00784-015-1530-4. PMID: 26193958
- 943 78. Bui FQ, Johnson L, Roberts J, Hung SC, Lee J, Atanasova KR, Huang PR, Yilmaz Ö3 Ojcius DM.  
944 *Fusobacterium nucleatum* infection of gingival epithelial cells leads to NLRP3 inflammasome-  
945 dependent secretion of IL-1 $\beta$  and the danger signals ASC and HMGB1. *Cell Microbiol*. 2016;  
946 18:970-81. doi: 10.1111/cmi.12560. PMID: 26687842
- 947 79. Lourenço TG, Heller D, Silva-Boghossian CM, Cotton SL, Paster BJ, Colombo AP. Microbial  
948 signature profiles of periodontally healthy and diseased patients. *J Clin Periodontol*. 2014; 41:1027-  
949 36. doi: 10.1111/jcpe.12302. PMID: 25139407
- 950 80. Park OJ, Yi H, Jeon JH, Kang SS, Koo KT, Kum KY, Chun J, Yun CH, Han SH. Pyrosequencing  
951 Analysis of Subgingival Microbiota in Distinct Periodontal Conditions. *J Dent Res*. 2015; 94:921-7.  
952 doi: 10.1177/0022034515583531. PMID: 25904141
- 953 81. Takeshita T, Matsuo K, Furuta M, Shibata Y, Fukami K, Shimazaki Y, Akifusa S, Han DH, Kim  
954 HD, Yokoyama T, Ninomiya T, Kiyohara Y, Yamashita Y. Distinct composition of the oral  
955 indigenous microbiota in South Korean and Japanese adults. *Sci Rep*. 2014; 4:6990. doi:  
956 10.1038/srep06990. PMID: 25384884
- 957 82. Petsios A, Nakou M, Manti F. Microflora in adult periodontitis. *J Periodontal Res*. 1995; 30:325-31.  
958 PMID: 7494174
- 959 83. Kamma JJ, Nakou M, Manti FA. Predominant microflora of severe, moderate and minimal  
960 periodontal lesions in young adults with rapidly progressive periodontitis. *J Periodontal Res*. 1995;  
961 30:66-72. PMID: 7722848
- 962 84. Kumar PS, Griffen AL, Moeschberger ML, Leys EJ. Identification of candidate periodontal  
963 pathogens and beneficial species by quantitative 16S clonal analysis. *J Clin Microbiol*. 2005;  
964 43:3944-55. PMID: 16081935
- 965 85. Marchesan JT, Morelli T, Moss K, Barros SP, Ward M, Jenkins W, Aspiras MB, Offenbacher S.  
966 Association of *Synergistetes* and *Cyclodipeptides* with periodontitis. *J Dent Res*. 2015; 94:1425-31.  
967 doi: 10.1177/0022034515594779. PMID: 26198391



- 968 86. Camelo-Castillo AJ, Mira A, Pico A, Nibali L, Henderson B, Donos N, Tomás I. Subgingival  
969 microbiota in health compared to periodontitis and the influence of smoking. *Front Microbiol.* 2015;  
970 6:119. doi: 10.3389/fmicb.2015.00119. PMID: 25814980
- 971 87. Fujii R, Saito Y, Tokura Y, Nakagawa KI, Okuda K, Ishihara K. Characterization of bacterial flora  
972 in persistent apical periodontitis lesions. *Oral Microbiol Immunol.* 2009; 24:502-5. doi:  
973 10.1111/j.1399-302X.2009.00534.x. PMID: 19832803
- 974 88. Nakazawa F, Poco SE Jr, Sato M, Ikeda T, Kalfas S, Sundqvist G, Hoshino E. Taxonomic  
975 characterization of *Mogibacterium diversum* sp. nov. and *Mogibacterium neglectum* sp. nov.,  
976 isolated from human oral cavities. *Int J Syst Evol Microbiol.* 2002; 52(Pt 1):115-22. PMID:  
977 11837293
- 978 89. Wong BK, McGregor NR, Butt HL, Knight R, Liu LY, Darby IB. Association of clinical parameters  
979 with periodontal bacterial haemolytic activity. *J Clin Periodontol.* 2016; 43:503-11. doi:  
980 10.1111/jcpe.12554. PMID: 27105613
- 981 90. Lanza E, Magan-Fernandez A, Bermejo B, de Rojas J, Marfil-Alvarez R, Mesa F. Complementary  
982 clinical effects of red complex bacteria on generalized periodontitis in a caucasian population. *Oral*  
983 *Dis.* 2016; 22:430-7. doi: 10.1111/odi.12471. PMID: 26948988
- 984 91. VieiraColombo AP, Magalhães CB, Hartenbach FA, Martins do Souto R, Maciel da Silva-  
985 Boghossian C. Periodontal-disease-associated biofilm: A reservoir for pathogens of medical  
986 importance. *Microb Pathog.* 2016; 94:27-34. doi: 10.1016/j.micpath.2015.09.009. PMID: 26416306
- 987 92. Nyby MD, Gregory DA, Kuhn DA, Pangborn J. Incidence of *Simonsiella* in the oral cavity of dogs. *J*  
988 *Clin Microbiol.* 1977; 6:87-8. PMID: 886011
- 989 93. O'Donnell LE, Robertson D, Nile CJ, Cross LJ, Riggio M, Sherriff A, Bradshaw D, Lambert M,  
990 Malcolm J, Buijs MJ, Zaura E, Crielaard W, Brandt BW, Ramage G. The oral microbiome of  
991 denture wearers is influenced by levels of natural dentition. *PLoS One.* 2015; 10:e0137717. doi:  
992 10.1371/journal.pone.0137717. PMID: 26368937
- 993 94. Ximenez-Fyvie LA, Haffajee AD, Socransky SS. Comparison of the microbiota of supra- and  
994 subgingival plaque in subjects in health and periodontitis. *J Clin Periodontol.* 2000; 27:648–657.  
995 PMID: 10983598
- 996 95. Ximenez-Fyvie LA, Haffajee AD, Socransky SS. Microbial composition of supra- and subgingival  
997 plaque in subjects with adult periodontitis. *J Clin Periodontol.* 2000; 27:722–732. PMID: 11034118
- 998 96. Darout IA, Albandar JM, Skaug N, Ali RW. Salivary microbiota levels in relation to periodontal  
999 status, experience of caries and miswak use in Sudanese adults. *J Clin Periodontol;* 2002; 29:411–  
1000 420. PMID: 12060423
- 1001 97. Bostrom L, Bergstrom J, Dahlen G, Linder LE. Smoking and subgingival microflora in periodontal  
1002 disease. *J Clin Periodontol;* 2001; 28:212–219. PMID: 11284533
- 1003 98. Haffajee AD, Socransky SS. Relationship of cigarette smoking to the subgingival microbiota. *J Clin*  
1004 *Periodontol* 2001; 28:377–388. PMID: 11350499
- 1005 99. Haffajee AD, Bogren A, Hasturk H, Feres M, Lopez NJ, Socransky SS. Subgingival microbiota of  
1006 chronic periodontitis subjects from different geographic locations. *J Clin Periodontol;* 2004; 31:996–  
1007 1002. PMID: 15491316
- 1008 100. Craig RG, Boylan R, Yip J, Bangboye P, Koutsoukos J, Mijares D, Ferrer J, Imam M,  
1009 Socransky SS, Haffajee AD. Prevalence and risk indicators for destructive periodontal diseases in 3  
1010 urban American minority populations. *J Clin Periodontol.* 2001; 28: 524–535. PMID: 11350519

- 1011 101. Feres M, Haffajee AD, Allard K, Som S, Socransky SS. Change in subgingival microbial  
1012 profiles in adult periodontitis subjects receiving either systemically-administered amoxicillin or  
1013 metronidazole. *J Clin Periodontol*. 2001; 28:597–609. PMID: 11422580
- 1014 102. Sakellari D, Belibasakis G, Chadjipadelis T, Arapostathis K, Konstantinidis A. Supragingival  
1015 and subgingival microbiota of adult patients with Down’s syndrome. Changes after periodontal  
1016 treatment. *Oral Microbiol Immunol*. 2001; 16:376–382. PMID: 11737662
- 1017 103. Socransky SS, Haffajee AD. Periodontal microbial ecology. *Periodontol 2000*. 2005; 38:135-87.
- 1018 104. Socransky SS, Smith C, Martin L, Paster BJ, Dewhirst FE, Levin AE. Checkerboard DNA–DNA  
1019 hybridization. *Biotechniques*. 1994; 17:788–792. PMID: 15853940
- 1020 105. Pozhitkov AE, Leroux BG, Randolph TW, Beikler T, Flemmig TF, Noble PA. Towards  
1021 microbiome transplant as a therapy for periodontitis: an exploratory study of periodontitis microbial  
1022 signature contrasted by oral health, caries and edentulism. *BMC Oral Health*. 2015; 15:125. doi:  
1023 10.1186/s12903-015-0109-4. PMID: 26468081
- 1024

1025 **Microbial Signatures of Oral Dysbiosis, Periodontitis and Edentulism**  
1026 **Revealed by Gene Meter Methodology**

1027  
1028 M. Colby Hunter, Alex E. Pozhitkov, and Peter A. Noble

1029  
1030 **Supplementary Materials**

1031  
1032 Table S1. GI numbers for 16S rRNA gene sequences of 597 oral bacteria.

1033 Table S2. Median abundances for the 576 rRNA genes and 16 patient samples based on individual probe  
1034 calibration.

1035 Table S3. Abundances for the 567 rRNA genes and 16 patient samples based on aggregate probe  
1036 calibration.

1037

1038

1039