

# SCIENTIFIC REPORTS



OPEN

## Comparative metagenomics reveals taxonomically idiosyncratic yet functionally congruent communities in periodontitis

Received: 12 April 2016  
Accepted: 15 November 2016  
Published: 19 December 2016

Shareef M. Dabdoub\*, Sukirth M. Ganesan\* & Purnima S. Kumar

The phylogenetic characteristics of microbial communities associated with periodontitis have been well studied, however, little is known about the functional endowments of this ecosystem. The present study examined 73 microbial assemblages from 25 individuals with generalized chronic periodontitis and 25 periodontally healthy individuals using whole genome shotgun sequencing. Core metabolic networks were computed from taxa and genes identified in at least 80% of individuals in each group. 50% of genes and species identified in health formed part of the core microbiome, while the disease-associated core microbiome contained 33% of genes and only 1% of taxa. Clinically healthy sites in individuals with periodontitis were more aligned with sites with disease than with health. 68% of the health-associated metagenome was dedicated to energy utilization through oxidative pathways, while in disease; fermentation and methanogenesis were predominant energy transfer mechanisms. Expanded functionality was observed in periodontitis, with unique- or over-representation of genes encoding for fermentation, antibiotic resistance, detoxification stress, adhesion, invasion and intracellular resistance, proteolysis, quorum sensing, Type III/IV secretion systems, phages and toxins in the disease-associated core microbiome. However, different species or consortia contributed to these functions in each individual. Several genes, but not species, demonstrated robust discriminating power between health and disease.

Periodontitis, a microbially induced disease that destroys the structures anchoring the tooth to the jawbone, is the sixth most prevalent disease in the world, affecting over 700 million adults worldwide<sup>1</sup>. The consequences of untreated disease are tooth loss, poor nutritional status, loss of speech and masticatory function. With the annual cost of periodontal treatment exceeding 15 billion dollars in the USA alone, this disease poses a significant health burden<sup>2</sup>. Additionally, emerging evidence implicates periodontitis in the pathogenic pathways of several systemic diseases<sup>3,4</sup>, and therefore, the consequences of untreated periodontitis may extend beyond the oral cavity.

Periodontitis is a site-specific disease, with the disease affecting many, but not all teeth in an individual with disease. Also, disease progression occurs through recurrent bursts of destruction followed by varying periods of quiescence. This temporally and spatially haphazard mode of disease progression has been named the random burst model<sup>5</sup>. Several reasons have been proposed to explain the random burst hypothesis, including the quality of plaque, presence of plaque retentive factors and local inflammatory response<sup>6</sup>, however, the mechanism underlying this arbitrariness is poorly understood; and poses a barrier to effective clinical treatment and disease prevention.

An ecological shift in the indigenous microbiome towards dysbiosis is known to play a primary role in the etiology of this disease. While the taxonomic profiles of eubiotic and dysbiotic periodontal communities are well-studied<sup>7–15</sup>, little is known about their genome signatures, and whether community level functional alterations accompany compositional shifts in the transition to and in the state of disease. Even less is known about the selection pressures exerted by a disease-associated environment on the pre-existing microbiome. Also, although several lines of evidence have suggested a role for viruses, archaea and fungi in disease etiology<sup>16–18</sup>, their contributions to the functionality of the periodontal ecosystem has never been explored.

Division of Periodontology, College of Dentistry, The Ohio State University, Columbus, Ohio, USA. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to P.K. (email: kumar.83@osu.edu)

The primary purpose of the present investigation, therefore, was to better understand the functional potentials encoded within health-compatible and disease-associated periodontal microbiomes, using a comprehensive metagenomic approach and computational bioinformatics to characterize correlations between the mycobiome, virome, archaeome and bacteriome. A secondary aim was to investigate if periodontitis is the result of a site-specific dysbiosis or a global shift in the subgingival microbial ecosystem.

## Methods

**Subject and site selection.** Approval for this study was obtained from the Office of Responsible Research Practices at The Ohio State University and the study was conducted in accordance with the approved guidelines. Twenty-five periodontally healthy never-smokers (attachment loss  $\leq 1$ , probing pocket depths  $\leq 3$ , gingival index  $\leq 1$ ) and twenty-five never-smokers with generalized moderate to severe chronic periodontitis (attachment loss  $\geq 5$ , probing pocket depths  $\geq 5$ , gingival index  $> 1$  in 30% or more sites) were recruited following clinical and radiographic examination and informed consent obtained. Exclusion criteria for both groups included diabetes, HIV infection, use of immunosuppressant medications, bisphosphonates, or steroids, antibiotic therapy, or oral prophylactic procedures within the preceding 3 months, and fewer than 20 teeth in the dentition. Sample size was estimated using the HMP package in R<sup>19</sup>, based on an 80% power to detect an effect size of at least 0.20 using weighted UniFrac distances as the primary outcome variable, assuming a two-sided significance level of 0.05.

**Sample collection.** From healthy subjects, samples were collected and pooled from 15 mesial sites on teeth with CAL  $\leq 1$  mm, PD  $\leq 3$  mm, GI  $\leq 1$  and no BOP (shallow-healthy) using sterile endodontic paper-points (Caulk-Dentsply, Milford, DE, USA). From the disease group, subgingival plaque from four nonadjacent proximal sites with attachment loss (CAL)  $\geq 5$  mm, probe depths (PD)  $\geq 6$  mm, bleeding on probing (BOP), and Loe and Silness gingival index (GI)  $\geq 2$  was collected using 15 paper points and pooled (deep-diseased). Samples were similarly acquired from four sites with CAL  $\leq 1$  mm, PD  $\leq 3$  mm, GI  $\leq 1$  and no BOP and separately pooled (shallow-diseased).

**DNA isolation and sequencing.** Bacterial DNA was isolated from paper points, using Qiagen DNA MiniAmp kit (Qiagen, Valencia, CA, USA) and quantified using Qubit fluorometer. Library generation was completed using an Illumina TruSeq kit according to the manufacturer's instructions. Briefly, genomic DNA was sheared enzymatically yielding an average fragment size of 500 base pairs. The fragment ends were blunted and adenylated, before ligation of barcodes and sequencing adaptors. Quantified and pooled libraries were clustered on the Illumina MiSeq (Illumina, Inc., San Diego, California), and 150 bp paired-end sequencing was performed in a commercial facility (Molecular Research LP, Shallowater, TX).

**Metagenomic Analysis.** Trimmed and filtered sequences were uploaded to the MG-RAST metagenomics analysis pipeline (version 3.3.6)<sup>20,21</sup> (Argonne National Laboratory) for quality processing and basic functional analysis. The MG-RAST API<sup>22</sup>, and the custom Python library we developed to access it and analyze/visualize results, were used throughout the analysis process to download relevant data and pipeline results (available for download at <http://github.com/smdabdoub/PyMGRAS>). Comparisons of functional potential between clinical groups were made in the context of the KEGG (Kyoto Encyclopedia of Genes and Genomes)<sup>23</sup> and the SEED<sup>24</sup> ontological hierarchies and statistical analysis of differential functional potential was performed using R and DESeq<sup>25</sup>. Taxonomic identities for archeal, fungal, and viral sequences were assigned using the Lowest Common Ancestor (LCA) alignment to the M5NR database<sup>26</sup>. Bacterial rRNA genes were filtered from the metagenome by BLAT<sup>27</sup> search against 90% clustered SILVA<sup>28</sup>. Rarefaction curves were generated from the 16S data and used to estimate sequencing coverage (Supplementary Fig. 1A). Taxonomic profiles were generated by comparing these filtered rDNA sequences to the HOMD database<sup>29</sup> using the QIIME (version 1.8) and PhyloToAST (version 1.2) pipelines as previously described<sup>30</sup>. A core microbiome was computed for health and disease to include all s-OTUs (core phylome) and functional genes (core metagenome) present in at least 80% of the subjects in a group.

Contributions of each species to community function were assessed by network analysis of KEGG-annotated bacterial genome sequences. Taxonomic identifiers were assigned to the genes by alignment to the full set of 1528 genomes in the Human Oral Microbiome Database (HOMD) using Bowtie 2 (version 2.2.5)<sup>31</sup>. Sequences with multiple equivalent matches (within 10% alignment score) were assigned to the lowest common ancestor (LCA). The LSU genes of fungal sequences were similarly isolated, and identified by comparison to the SILVA database housed within MG-RAST. Species assignment was made at 97% identity<sup>32</sup>. Function-taxonomy networks were visualized using force directed networks (Fruchterman-Reingold<sup>33</sup>) and Force Atlas 3D<sup>34</sup>. Network-wide degree of specialization ( $H_2'$ ) was determined as a standardized degree of entropy, with 0.0 representing extreme generalization and 1.0 extreme specialization<sup>35,36</sup>. Specialization was calculated using version 2.05 of the R package bipartite<sup>37</sup>. The ability of genes to discriminate between health and disease was examined using a machine-learning algorithm (RandomForest package in R). The robustness of the classifier was evaluated using ROC curves (ROCR package in R). Two-thirds of the dataset was used to train the algorithm, which was tested on the remaining data. This was iterated 10 times and the mean 'importance' computed for each marker gene. The importance classification was used to select marker genes based on the methodology of Diaz-Uriarte<sup>38</sup>. For each iteration of the test, a 'confusion table' was created for each of the marker genes based on the number of correctly classified and misclassified samples; and this data was used to compute sensitivity and specificity.

## Results

An average of 3.32 million sequences per sample (range 1.97–4.53 million) were obtained for the shallow-healthy, shallow-diseased, and deep-diseased groups respectively. On average 77.22% of the shallow-healthy, 70.16% of the deep-diseased and 62.96% of the shallow-diseased sequences were human sequences. This is line with other investigations using similar approaches to study the subgingival microbiome<sup>13,39,40</sup>. Average coverage per sample ranged from 38% to 93% based on Nonpareil<sup>41</sup>, and was not significantly different between groups ( $p > 0.05$ , Tukey HSD, Supplementary Figure 1B). These sequences represented 4837 functionally annotated genes of bacterial, fungal, viral and archeal origin. The distribution of the sequences in each sample is shown in Supplementary Figure 2. Overall, 95% or more of the sequences belonged to bacteria. Viruses, archaea and fungi were identified in the samples, but were variably distributed. An average of 7% (range from 2.5–9.3%) of the sequences encoded for rRNA.

**Energy efficiency and functional equitability are central characteristics of a health-compatible microbial ecosystem.** 3348 functions were identified from periodontally healthy subjects, of which 1811 were present in the core healthy metagenome ( $\geq 80\%$  of the healthy group); suggesting that in health, more than half of the microbial metagenome is conserved among all individuals.

The most abundant functional group represented metabolic potential. Within this group 29% of genes encoded for protein metabolism, with genes responsible for protein, amino acid, RNA and DNA biosynthesis predominating this group (Fig. 1A,B and Supplementary Table 2). The second most abundant functionality was carbohydrate metabolism (22%). Within this framework, the primary pathways were related to aerobic metabolism of monosaccharides, organic acids, polysaccharides and one-carbon sugars, with an abundance of co-factors as catalysts (Fig. 1B). 17% of the core genes contributed to oxidative phosphorylation (aminotransferases, TCA cycle, pentose phosphate shunt, electron transport (cytochrome and ubiquinone families) and membrane transport (Fig. 1B and Supplementary Table 2)).

The second most abundant group of functions encoded in health (28%) was related to virulence lifestyle, and included antimicrobial resistance genes (AMRs), cell-signaling, competence, peptidoglycan biosynthesis and non-siderophore type iron transport (Fig. 1B and Supplementary Table 2). Additionally, a large number of phages, transposons, and gram-negative cell wall components (ranging from 1.3% to 5.7% of the metagenome) were also observed in health.

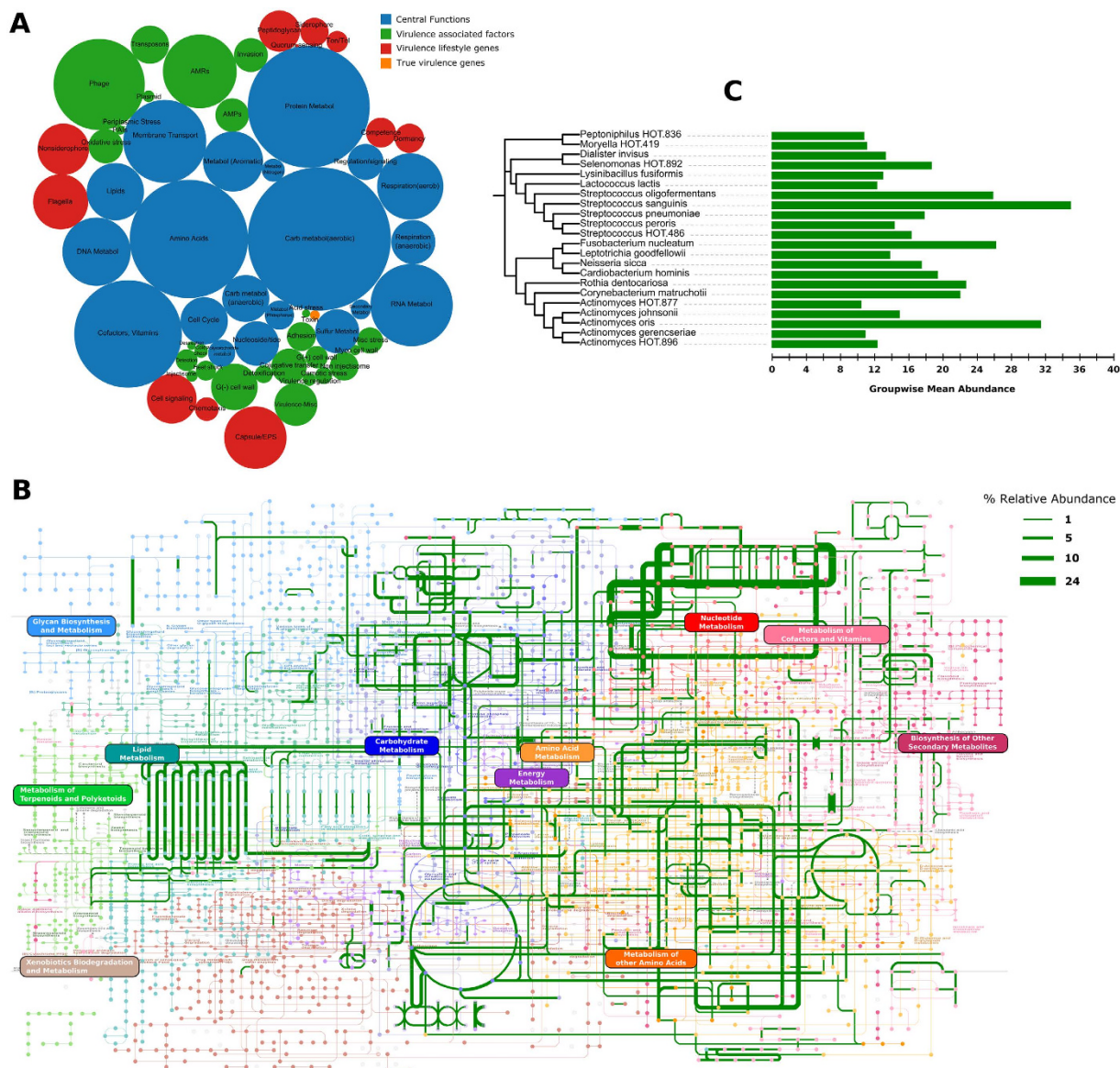
75% of the community membership was made up of 46 species belonging to the genera *Streptococcus*, *Veillonella*, *Actinomyces*, *Corynebacterium*, *Neisseria*, *Fusobacterium* and *Selenomonas*, of which 22 belonged to the core (Fig. 1C). There were no viruses or archaeobacteria in the core; however, the fungal species *Candida albicans* was identified in 87% of healthy individuals. Co-occurrence networks demonstrated strong correlations (Spearman's  $\rho > 0.8$ ,  $p < 0.05$ ) between *Actinomyces gerencseriae*, *A. oris*, *A. johnsonii*, *Selenomonas noxia*, *S. sputigena*, *S. artemides*, *Streptococcus sanguinis*, *S. oralis*, *Fusobacterium nucleatum* and *Veillonella parvula*. Each of the species contributed  $879 \pm 137$  genes. While there was 60% overlap in the genes that each species contributed, there was a 98% overlap in the functions encoded by these genes, indicating that these species contribute to similar functions in the health-compatible microbiome (functional equitability or functional generalization). Some level of functional specialization was observed, with 6 species contributing genes encoding for motility and chemotaxis. Together, these 6 species (*Campylobacter rectus*, *C. showae*, *C. curvus*, *Selenomonas noxia*, *Eubacterium yurii*, and *Centipeda periodontii*) formed 0.019% of the core microbiome, and hence, were rare taxa within the microbiome.

**Global dysbiosis in disease.** In addition to collecting samples from sites with clinical disease (deep-diseased), samples were collected from clinically healthy sites in subjects with disease (shallow-diseased) and from healthy subjects (shallow-healthy). The shallow-diseased sites were clinically similar to the shallow-healthy sites (Supplementary Table 1). If periodontitis were a result of a site-specific dysbiosis, then we would expect to see marked differences between deep-diseased and shallow-diseased samples, as well as significant similarities between shallow-healthy and shallow-diseased samples.

Relative abundances of functional genes and taxa in the three groups were compared using DESeq2<sup>25</sup>. Shallow-healthy and deep-diseased sites demonstrated significant differences in nearly 2000 genes ( $p < 0.05$ , FDR adjusted Wald test), 1730 of which were part of the core metagenomes (Fig. 2A). Shallow-healthy and shallow-diseased sites differed in 1000 gene abundances (Fig. 2B), again, 880 of these belonged to the core. Few differences were detected between shallow-diseased and deep-diseased sites (Fig. 2C).

Phylogenetically, both deep-diseased and shallow-diseased sites demonstrated a significantly higher alpha diversity (Chao) when compared to shallow-healthy sites ( $p < 0.05$ , Tukey HSD, Fig. 3A). Higher levels of several species (notably those belonging to the genera *Porphyromonas*, *Fusobacterium*, *Fretibacterium*, *Filifactor*, *Parvimonas*, *Selenomonas*, *Treponema* and *Kingella*), and lower levels of health-compatible species were observed in disease ( $p < 0.05$ , FDR adjusted Wald test, Fig. 3B and Supplementary Table 4). Shallow-diseased sites also exhibited lower levels of *Dialister invisus*, *Tannerella forsythia*, *Fusobacterium nucleatum* and *Fretibacterium* HOT.452 when compared to deep-diseased sites. There were no significant differences in the abundances of human viral species between groups (Fig. 4A). In shallow-healthy sites, gram-positive phages, especially those belonging to the genera *Streptococcus*, *Enterococcus* and *Lactobacillus* predominated ( $p < 0.05$ , FDR adjusted Wald test), while in disease (both shallow-diseased and deep-diseased sites) gram-negative phages associated with *Prevotella*, *Burkholderia*, *Campylobacter*, *Haemophilus* and *Aggregatibacter* were significantly more abundant (Fig. 4A and B). The abundances of the archaeobacterial species *Methanobrevibacter oralis*, *M. smithii*, *Methanomassiliicoccus luminyensis* and *Methanosphaera stadtmaniae* were significantly higher in disease when compared to health (Fig. 4C). Shallow-diseased and shallow-healthy sites demonstrated higher levels of *Candida albicans* when compared to deep-diseased sites (Fig. 4D).

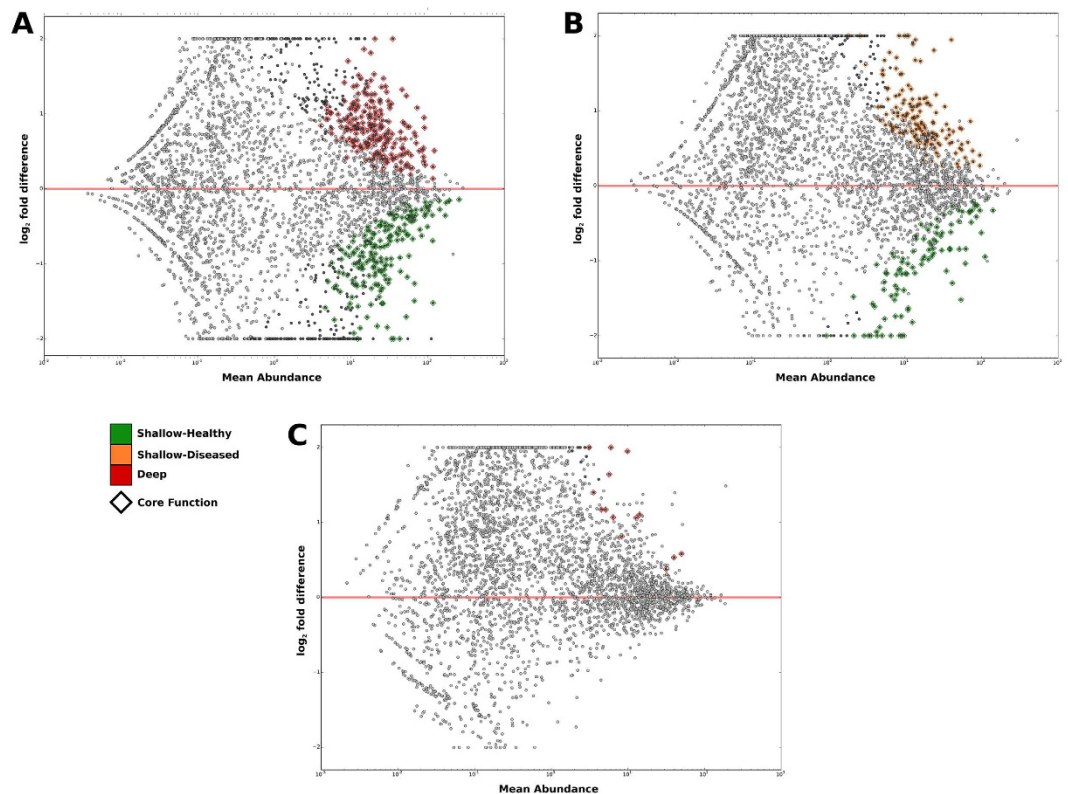




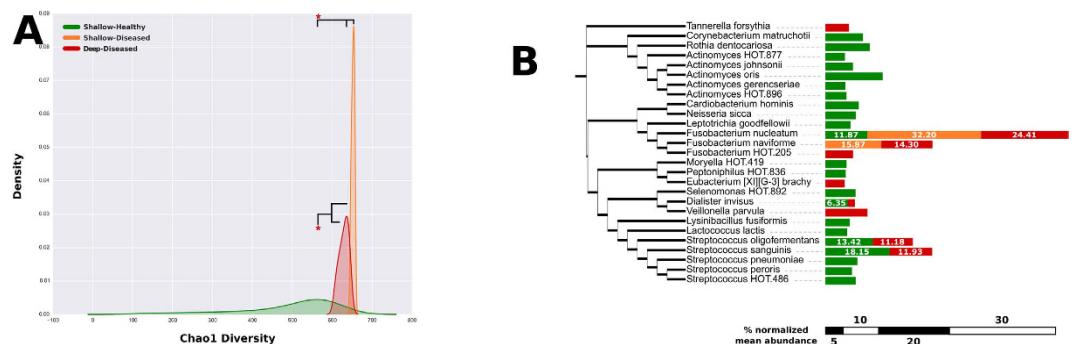
**Figure 1. Predominant functionalities in health.** (A) Shows a circle-packing graph of core genes grouped into higher order functions. Circles are sized by relative abundances of genes contributing to each function. (B) Shows a KEGG map of the core metabolic pathways in the health-associated microbiome. The lines are sized by log fold abundances. The genes and pathways used to create this map are presented in Supplementary Table 2. (C) Shows a selected group of species that contributed to these functions. The species shown here belonged to the core microbiome (80% or more of healthy individuals). The green bars represent the relative abundances of the species in all samples.

**Expanded functional capabilities and functional specialization in disease.** Out of the 4837 functional units identified in the present investigation, 2612 were common to both health and disease, 1489 were uniquely observed in disease (of which 1249 belonged to the disease-associated core microbiome) and 736 uniquely in health (657 in the health-compatible core microbiome). The common genes predominantly encoded for central functions such as carbohydrate and protein metabolism, aerobic respiration, protein and amino acid synthesis and virulence. However, the abundances of these common genes were significantly different between groups ( $p < 0.05$ , FDR adjusted Wald test, Fig. 2A and B, and Supplementary Table 3). The genes unique to disease encoded for fermentation, antibiotic resistance, detoxification stress, adhesion, invasion and intracellular resistance, proteolysis, quorum sensing, Type III and IV secretion systems, phages and toxins and superantigens (Fig. 5 and Supplementary Table 3). Significant over-representation ( $p < 0.05$ , FDR adjusted Wald test) of genes associated with oxidation of primary alcohols, butyrate, isovalerate, propionate, acetate, glycolate and aromatic compounds was observed in disease (Fig. 5, and Supplementary Table 3). Furthermore, while shallow-healthy samples demonstrated high levels of aerobic reductases, anaerobic reductases were almost exclusively found in disease, and constituted 1.2% of this metagenome. Strong and significant correlations ( $r^2 > 0.7$ ,  $p < 0.05$ ) were observed





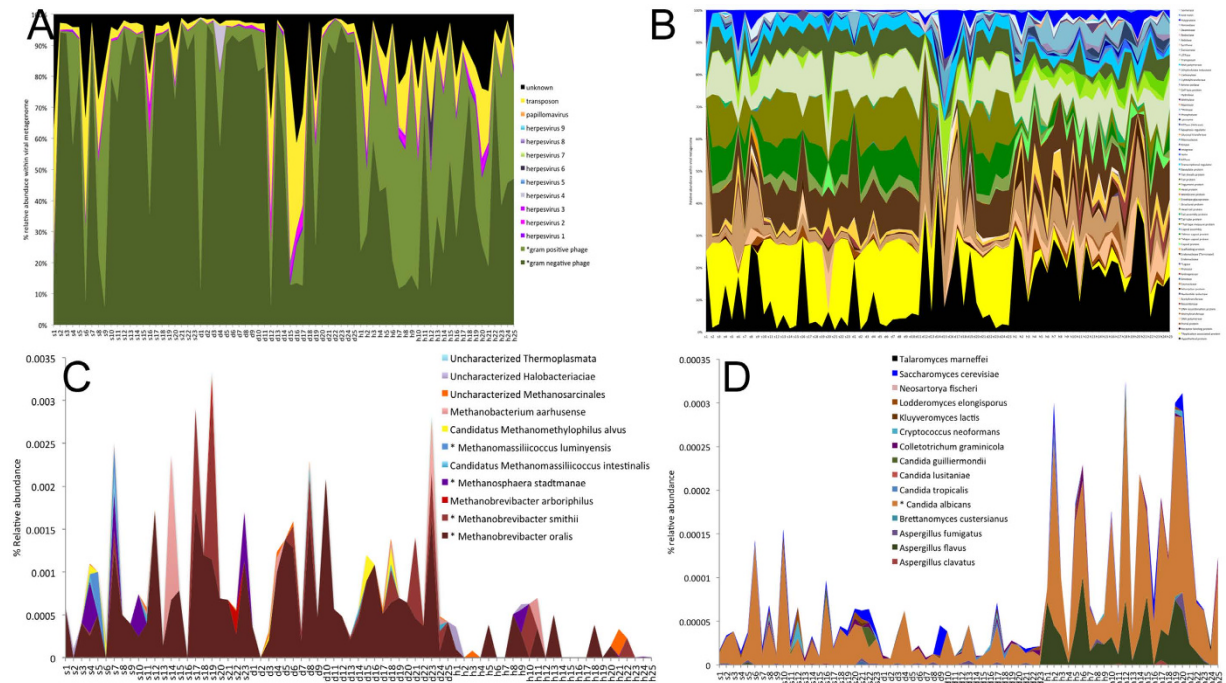
**Figure 2. Bland-Altman plots of metagenomic differences between healthy, shallow and deep sites.** Relative mean abundances of genes were plotted against log differences in abundance between groups. (A) Shows differences between healthy subjects and deep sites in subjects with periodontitis. Genes that were significantly overrepresented in deep sites ( $p < 0.05$ , FDR adjusted Wald test) are in red, those whose levels were significantly greater in health are in green. The central red line represents a log fold difference of zero. (B) Shows differential abundances between shallow sites (in orange) in subjects with periodontitis and healthy subjects (in green), while comparisons between shallow and deep sites in subjects with periodontitis are shown in (C). The genes and functions that were used to create these plots are shown in Supplementary Table 3.



**Figure 3. Alpha and beta phylogenetic diversity in health and disease.** Kernel plots of Chao diversity index of healthy subjects and shallow and deep sites of subjects with periodontitis are shown in (A). Significant differences ( $p < 0.05$ , Tukey HSD) are indicated by an asterisk (\*). Taxonomic differences between core microbiomes of health and disease are shown in (B). The bars represent the mean relative abundances of each species in each group.

between bacterial fermentation and archaeal methanogenesis genes (especially coenzyme F420, coenzyme B, coenzyme M, methanofuran, and methanopterin) in shallow-diseased and deep-diseased but not shallow-healthy sites (Supplementary Figure 3). Very few sulfate-reducing genes were identified in disease.

49 genes contributing to flagellar motility and 14 genes encoding for chemotaxis were identified in the core disease microbiome. Of these, 28 genes, especially those contributing to assembly of the filament, hook, basal body, rods and rings (FlhA/B, FlgG/C, FliG/I/J/K/L/S, FlgC/G) were either over-represented or uniquely represented in deep-diseased and shallow-diseased sites when compared to shallow-healthy samples ( $p < 0.05$ , FDR



**Figure 4. Non-bacterial members of the metagenomes of periodontal health and disease.** Distribution of viral species by sample is shown in (A) and viral genes in (B). Relative abundances of species belonging to the archaeobacterial kingdom in each sample is shown in (C,D) shows the distribution of species belonging to the fungal kingdom by sample. Genes and taxa that were significantly different between deep-diseased and shallow-healthy sites ( $p < 0.05$ , FDR adjusted Wald test) are indicated by an asterisk (\*) in the legend.

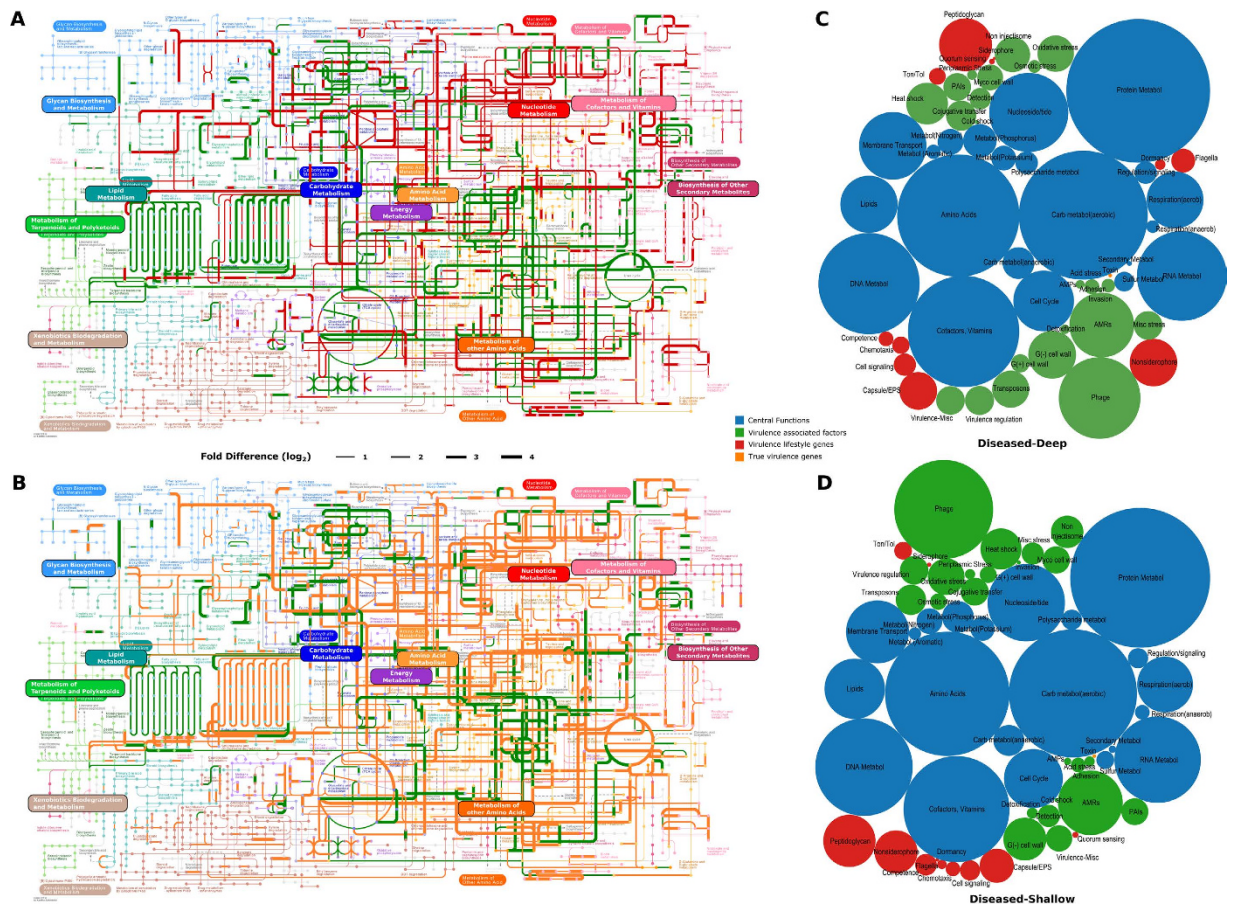
adjusted Wald test, Fig. 5, and Supplementary Table 3). Type III secretion systems (FlhA/B and FliI) were also significantly higher in disease when compared to health. Flagellar proteins play important roles not only in bacterial motility, but also in adhesion, Type III secretion and virulence. Flagellins belong to the family of PAMPS and stimulate innate and adaptive immune responses through TLR5<sup>42</sup>.

Overall, 28% of the genes in the shallow-healthy, 36.8% in shallow-diseased and 38.2% in deep-diseased sites encoded for virulence. However, in health, these genes accounted for 8.9% of the genome abundance, while in disease, 33.1% of the genomic content was attributable to virulence. The greatest differences between health and disease were observed in antibiotic resistance, iron acquisition, and gram negative and gram-positive cell wall components ( $p < 0.05$ , FDR adjusted Wald test, Fig. 5 and Supplementary Table 3). The abundances of efflux pumps ranged from 0.07 to 0.6% of the genomic content, and a statistically significant correlation ( $r^2 = 0.61$ ,  $p = 0.0036$ ) was observed between efflux pumps and gram-negative phages in disease, but not in health (Supplementary Figure 4C). EPs are a three-component system (comprising an outer membrane protein, a periplasmic fusion protein and an inner membrane protein), which displace toxic compounds (including antibiotics) from gram-negative bacteria. Although EPs are an intrinsic part of the genome of several gram-negative bacteria, within an ecosystem these genes are typically acquired either through horizontal gene transfer or through mutations<sup>43,44</sup>. A similar correlation was also observed between LPS, metal resistance and efflux pumps (Supplementary Figure 4A and B). LPS is most known for being a powerful antigen that triggers a florid inflammatory response, however, its role in the bacterium is one of barrier function. Our data corroborate previous studies in the literature suggesting a synergy between cell exclusion and efflux in mediating antibiotic resistance<sup>45</sup>.

Iron is important for bacterial survival since it facilitates electron transport, nucleotide synthesis, peroxide reduction and other essential cellular functions. The host typically sequesters iron by complexing it as hemoglobin, or by storage proteins such as ferritin and lactoferrin (nutritional immunity). Bacteria take up iron either directly from heme and heme-containing compounds using surface receptors and ABC transporters, or indirectly using high affinity small-molecule chelators known as siderophores. Bacteria also use iron availability as a metric to sense their environment; iron-deprivation leads to expression of several outer membrane proteins, siderophores, hemolysins and toxins, while availability of iron promotes pathogen expansion and cellular invasion. Iron transport genes accounted for 1.3% of the health-associated genome and 5% of the disease-associated metagenome. Both siderophoric and non-siderophoric transport mechanisms were significantly different between health and disease.

Lipid-A, the lipid moiety of LPS, is a powerful antigen that elicits a florid pro-inflammatory host response. Lipid-A synthesis is known to be upregulated in the presence of hemin<sup>46</sup>. Genes responsible for Lipid-A synthesis were preferentially enriched in disease when compared to health, with *Capnocytophaga*, *Campylobacter*, *Fusobacterium*, *Porphyromonas*, *Prevotella*, *Tannerella*, and *Treponema*, as major contributors. The combined enrichment of both iron acquisition genes and Lipid A in disease suggests that this microbiome has the





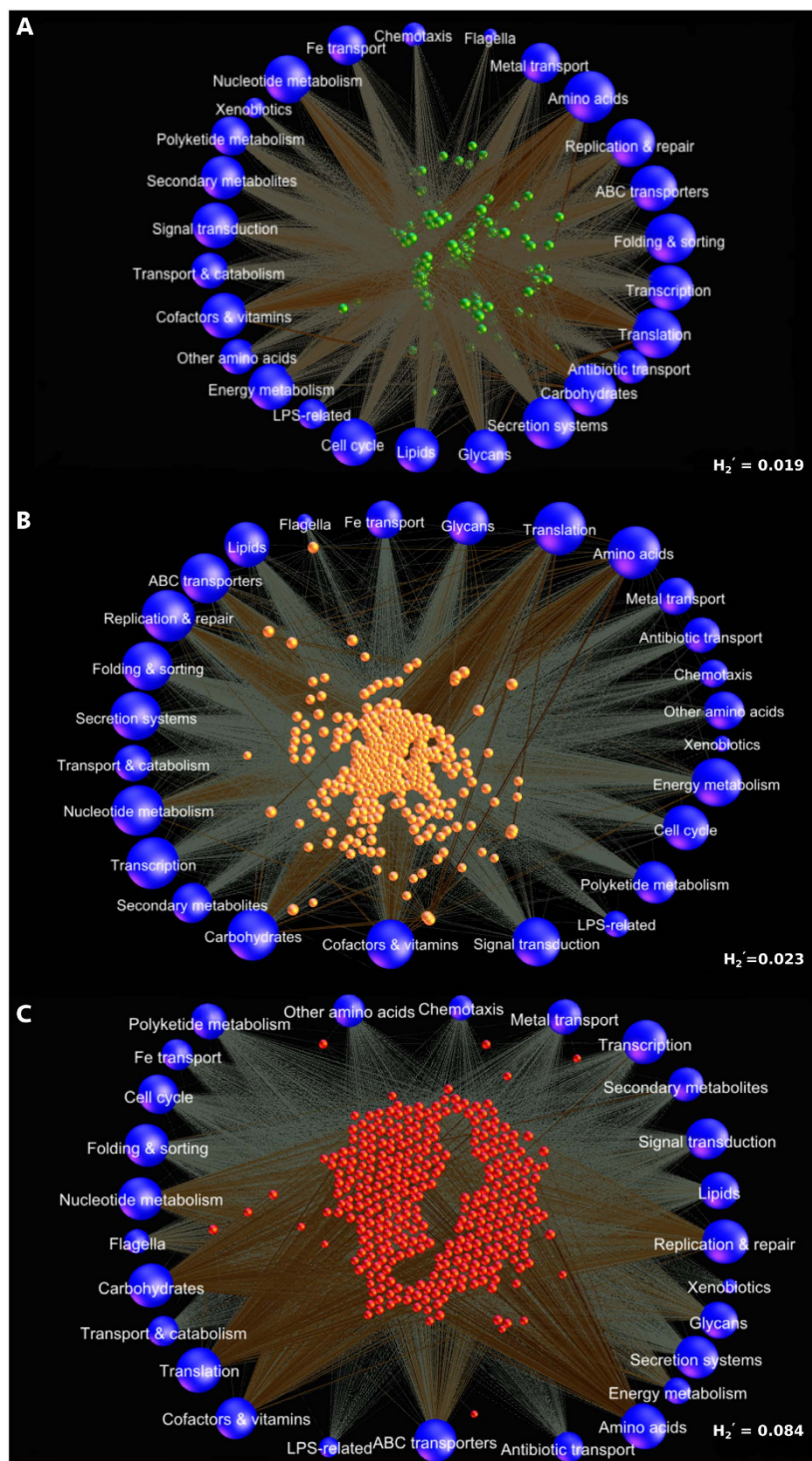
**Figure 5. Metabolic differences between health and disease.** KEGG maps of differences between healthy subjects and deep sites are shown in (A). The pathways are sized by relative abundances (Log scale) of genes contributing to the functionality. (B) shows a circle-packing graph of core genes in deep-diseased sites grouped into higher order functions. Circles are sized by relative abundances of genes contributing to each function. Differences between healthy subjects and shallow sites are shown in (C) and core genes in shallow-diseased sites in (D). The genes and functions that were used to create these maps are shown in Supplementary Table 3.

potential for a virulent transcriptional profile in the presence of blood (i.e. in conditions of inflammation, such as gingivitis).

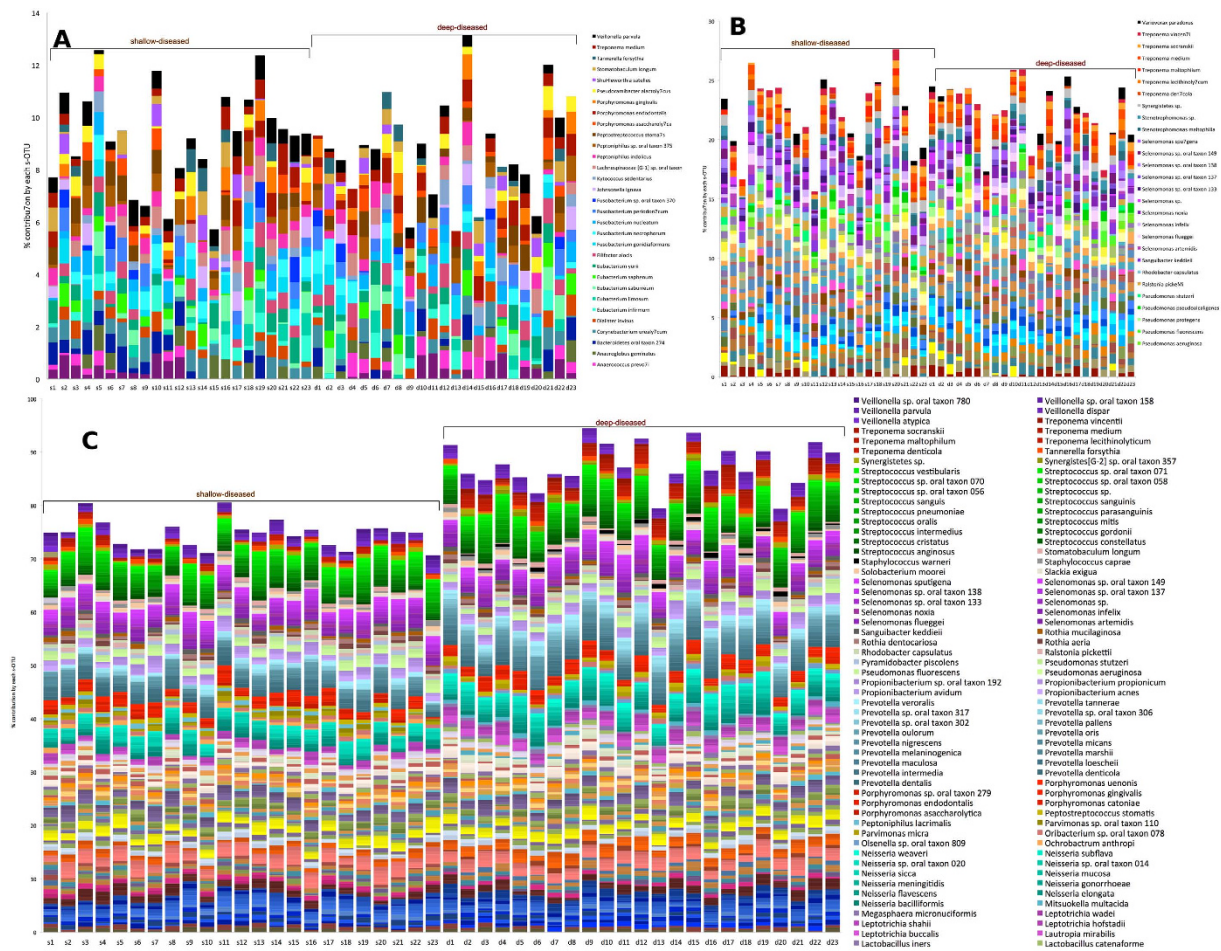
Genes responsible for management of oxidative stress formed 4.12% of shallow-healthy sites and 0.98% of disease ( $p < 0.05$ , FDR adjusted Wald test, Fig. 5, and Supplementary Table 3). The disease-associated microbiome, on the other hand, demonstrated a greater abundance of rubrerythrin and sigma factors. Glutathione is an important redox-buffering compound that protects bacterial cells from osmotic stress, electrophiles and oxidative stress and by acting as an electron donor during reduction of lipid peroxides and hydroperoxides, and for scavenging reactive oxygen. The present investigation suggests that the health-compatible microbiome is well equipped to handle oxidative stresses through the glutathione, and that this functionality is not as marked in disease. Rubrerythrin is a non-haem iron compound that protects anaerobic bacteria such as *P. gingivalis* from reactive oxygen and nitrogen species, both of which are produced during a neutrophil-mediated host response<sup>47</sup>. This mechanism enables growth and tissue invasion by the organism. Sigma factors are dissociable subunits of RNA polymerase. Recent evidence indicates that these factors may play a major role in enabling bacterial transition from a free-living state to host invasion<sup>48</sup> by regulating the expression of several virulence genes. Thus, the data indicate that the disease-associated microbiome possesses the capabilities for host tissue invasion in response to environmental stress.

A significantly higher degree of functional specialization was evident in disease when compared to health ( $p < 0.05$ , Tukey HSD of  $H_2'$ , Fig. 6). 547 bacterial, viral, and archeobacterial s-OTUs were identified in disease ( $148 \pm 23$  in each sample). Each species contributed an average of 867 genes, however the range varied from 97 genes (Archeobacteria, *Treponema*, *Synergistes*, TM5) to 1256 genes (*Streptococcus*, *Neisseria*, *Actinomyces*, *Filifactor*, *Dialister*, *Porphyromonas*, *Fusobacterium*, *Eubacterium*). Also, while more than 400 species contributed genes encoding for respiration, protein and carbohydrate metabolism, 32 species contributed flagellar genes, 69 contributed genes for glycan synthesis and 76 species encoded for LPS.





**Figure 6. Functional contributions of bacterial species in the subgingival metagenome.** Force-directed networks of bacterial species and their contribution to metabolic pathways in health (A), shallow sites (B) and deep sites (C). Each network graph contains nodes (circles) and edges (lines). Nodes in the center of each network represent species-level OTU's in healthy (green), deep sites (red) and shallow sites (orange) and nodes on the outer edge represent the functional contributions of these species. Edges represent the number of genes contributed by each species to each functional family. Only significant correlations between species and their functional contributions ( $p < 0.05$ , t-test) with a coefficient of at least 0.75 are shown. The data used to create these networks are presented in Supplementary Table 5. Few species-level nodes can be seen in health, with equal number of edges connecting each of these species to the functional nodes. Both deep and shallow sites demonstrate larger numbers of species-level nodes than health. Moreover, while many functions species are connected to their cognate species by large numbers of edges, certain functions have contributions only from a few species. This is numerically indicated by the degree of functional specialization ( $H_2'$ ).



**Figure 7. Phylogenetic distribution of functional potential in health and disease.** Distribution of taxa encoding for fermentation (A), flagella (B) and iron acquisition (C) by sample. 23 paired samples of shallow and deep sites in subjects with periodontitis and 25 samples from periodontally healthy subjects are shown.

**Taxonomically idiosyncratic yet functionally congruent communities in disease.** Of the 547 s-OTUs identified in disease, only 9 were found in the core microbiome associated with disease (Fig. 3B). On the other hand, the core metagenome of deep-diseased sites comprised of 1207 out of 3855 functional units, while that of shallow-diseased sites comprised 1211 out of 4137 functional units, indicating that one third of the functionality is conserved among all sites in an individual with disease.

The predominant bacterial species responsible for fermentative pathways were *Anaerococcus lactolyticus*, *A. prevotii*, *Anaeroglobus geminatus*, *Bacteroidetes* oral taxon 274, *Corynebacterium urealyticum*, *Dialister invisus*, *Eubacterium infirmum*, *E. limosum*, *E. saburreum*, *E. saphenum*, *E. yurii*, *Filifactor alocis*, *Fusobacterium gonidiaformans*, *F. necrophorum*, *F. nucleatum*, *F. periodonticum*, *Fusobacterium* sp. oral taxon 370, *Johnsonella ignava*, *Kytococcus sedentarius*, *Lachnospiraceae* [G-1] sp. oral taxon, *Peptoniphilus indolicus*, *Peptoniphilus* sp. oral taxon 375, *Peptostreptococcus stomatis*, *Porphyromonas asaccharolytica*, *P. endodontalis*, *P. gingivalis*, *Pseudoramibacter alactolyticus*, *Shuttleworthia satelles*, *Stomatobaculum longum*, *Tannerella forsythia*, *Treponema medium*, and *Veillonella parvula*. However, different bacterial consortia contributed to fermentation in each subject, even between shallow and deep sites within each subject (Fig. 7A). Similarly, although flagellar genes formed part of the core microbiome of disease, several different species contributed to these functions in each sample. For example, in certain individuals, the Treponemes were the dominant contributors of flagellar function, while in certain others it was the Selenomonads or Campylobacters (Fig. 7B). Although there were no statistically significant differences in the abundances of iron acquisition genes between shallow-diseased and deep-diseased sites, in shallow-diseased sites, this was attributable to species belonging to the genera *Neisseria*, *Bifidobacterium*, *Porphyromonas*, *Selenomonas*, *Actinomyces* and *Streptococcus*, while in deep-diseased sites of the same individuals *Prevotella*, *Lactobacillus*, *Fusobacterium* and *Treponema* contributed to a large fraction of this functionality (Fig. 7C).

All three red complex bacteria (*Porphyromonas gingivalis*, *Treponema denticola* and *Tannerella forsythia*) were detected in 23 out of the 73 samples. At least one of these species was detected in 45 samples. Principal co-ordinate analysis of the functional genes did not reveal significant clustering of the samples based on presence or absence of these species. Moreover, a Random Forest machine-learning algorithm did not identify any species or consortia in the core microbiomes of health or disease that could discriminate between health and disease.



Gene	Functional role	Association
N-methylhydantoinase (ATP-hydrolyzing) (EC 3.5.2.14)	Amino acid derivatives	Disease
Putrescine transport ATP-binding protein PotG (TC 3.A.1.11.2)	Amino acid derivatives	Disease
Heterodisulfide reductase	Anaerobic respiratory reductases	Disease
Deoxyribonuclease YjjV	Carbohydrate Metabolism	Health
Dihydrolipoamide dehydrogenase (EC 1.8.1.4)	Carbohydrate Metabolism	Health
Hydroxypyruvate reductase (EC 1.1.1.81)	Carbohydrate Metabolism	Health
Iron-containing alcohol dehydrogenase	Carbohydrate Metabolism	Health
Pyruvate oxidase (EC 1.2.3.3)	Carbohydrate Metabolism	Health
Chromosome (plasmid) partitioning protein ParB-2	Cell Division	Health
Flavodoxin 2	Cofactors, Vitamins, Prosthetic Groups, Pigments	Disease
CRISPR-associated RAMP Cmr4	CRISPs	Disease
CRISPR-associated RecB family exonuclease Cas4b	CRISPs	Disease
Stage V sporulation protein	Dormancy and Sporulation	Disease
UDP-2,3-diacetylglucosamine hydrolase (EC 3.6.1.–)	Gram-Negative cell wall components	Disease
N-acetylmannosaminyltransferase (EC 2.4.1.187)	Gram-Positive cell wall components	Disease
Haemin uptake system permease protein	Iron Acquisition	Disease
Membrane fusion protein (MFP) component of efflux pump	Membrane Transport	Disease
Na(+) H(+) antiporter subunit A	Membrane Transport	Disease
Nudix hydrolase	Phage regulation of gene expression	Disease
Co/Zn/Cd efflux system membrane fusion protein	Resistance to antibiotics and toxic compounds	Disease
Cobalt-zinc-cadmium resistance protein CzcD	Resistance to antibiotics and toxic compounds	Disease
rRNA adenine N-6-methyltransferase (EC 2.1.1.48)	Resistance to antibiotics and toxic compounds	Disease
BatC (Bacteroides aerotolerance operon)	Respiration	Disease
Cytochrome c oxidase polypeptide I (EC 1.9.3.1)	Respiration	Disease
Cold shock protein CspC	Stress Response	Health
Ferric siderophore transport system, biopolymer transport protein ExbB	Ton and Tol transport systems	Disease
Flagellar biosynthesis protein FliQ	Virulence, Disease and Defense	Disease
Flagellar biosynthesis protein FliS	Virulence, Disease and Defense	Disease
Flagellar motor rotation protein MotB	Virulence, Disease and Defense	Disease
Hemolysin III	Virulence, Disease and Defense	Disease
Inner membrane protein CreD	Virulence, Disease and Defense	Disease

**Table 1. Candidate marker genes.** The ability of genes to discriminate between health and disease. The genes that were identified using random forest and their predicted function are shown.

However, the algorithm did identify 31 functional genes within the core microbiomes of health and disease that were capable of discriminating between health and disease. Together, this panel of genes was capable of classifying subjects into the health and disease categories with 99% sensitivity and 100% specificity (Table 1).

## Discussion

There have been several investigations in recent years that have examined the functional potential of the subgingival microbiome in health and disease<sup>12–14,40,49</sup>. While some of these investigations have used a targeted DNA-array based approach to examine selected functions<sup>14</sup>, others have been limited by small sample sizes in making statistical comparisons<sup>12,13,40,49</sup>. Moreover, all of these studies have focused only on the bacteriome. Since there is considerable evidence in the literature that viruses, archaea and fungi are common inhabitants of the subgingival microbial ecosystem and may play a role in health or disease<sup>17,18,50</sup>, we examined the metagenomes of all these domains using a comprehensive open-ended approach on a sample size large enough to permit robust statistical inferences to be made. There is little evidence in the literature on what fraction of the subgingival microbiome is comprised of viruses and fungi. In the present investigation, only 5% of the genomic abundance was attributable to these taxonomic clades, since we did not specifically enrich for these organisms. While it is possible that these taxa were underrepresented in the present investigation, previous investigations have reported that the proportions of viruses in complex microbial communities can be accurately estimated without enrichment strategies<sup>51</sup>. Moreover, other investigations have reported similar proportions of these organisms<sup>13</sup>.

Since the oral cavity is an open microbial ecosystem with transient members (allochthonous constituents) and stable colonizers (autochthonous community), we examined the core metagenome<sup>52</sup> of health and disease to minimize the effect of allochthonous species and genes on the analysis. We defined the core metagenome as that which is found in 80% or more of individuals. This is a far more conservative definition than that used by previous investigations<sup>10,52</sup>, and ensured that the genes under investigation were truly representative of the subgingival metagenome.

The term ecosystem describes a community of living organisms interacting as a system and linked to each other through energy transfer and nutritional flow<sup>53</sup>. Our data suggest that the health-compatible microbiome



is a highly energy efficient ecosystem, with 68% of the genome dedicated to energy acquisition, transfer and utilization. The genomic framework is set up for energy acquisition mainly through carbohydrate metabolism. Energy transfer is mediated mainly through oxidative phosphorylation; the high levels of aminotransferases and the robust glutamate pathway point to the citric acid cycle as a preponderant energy transfer mechanism. This is possibly facilitated by high oxygen tensions that prevail in the healthy gingival sulcus<sup>54</sup>. Strong co-occurrence patterns among known nutritional and structural symbionts (e.g., Veillonella, Streptococci, Actinomyces and Fusobacteria) attest to functional cooperativity in this system.

On the other hand, in periodontitis (both shallow and deep sites), fermentation and methanogenesis are the predominant pathways for energy acquisition. Effective fermentation requires the presence of a 'hydrogen sink', that is, sulfate-reducing or methanogenic species<sup>55</sup>. The robust archeal presence, sparsity of sulfate-reducing genes and correlations between abundances of bacterial fermentation genes and archeal methanogenesis genes suggest syntrophic interspecies hydrogen transfer between archaeobacteria and eubacteria in periodontal disease, which corroborates previous hypotheses that the presence of archaea may promote colonization by fermenters<sup>17,56</sup>. Fermentation of one mole of glucose yields 2–4 molecules of ATP when compared to aerobic respiration, which yields 32–36 ATP. Also, the end products of fermentation, short chain fatty acids such as butyrate, propionate and isobutyrate for example, have been strongly associated with periodontitis<sup>57,58</sup>. Thus, our data suggest that the disease-associated microbiome lacks the capability for efficient energy processing, forcing this community to do 'hard work' rather than 'smart work' for survival, and that the by-products thus created may contribute to disease etiology.

Recent paradigms of disease pathogenesis have promulgated the 'pathobiont' hypothesis, which states that disease occurs due to expansion of certain members of the indigenous microbiome rather than acquisition of new species<sup>59,60</sup>. To investigate this, the transition from health to disease was modeled by comparing the core microbiomes of healthy subjects to sites with and without clinical disease in subjects with disease. Phylogenetically and functionally, diversity progressively increased from shallow-healthy to shallow-diseased to deep-diseased sites; and this was due to both increase in abundances of certain indigenous members and functions as well as addition of new members and their associated functions. Importantly, a central feature of health was that all microbial members contributed genes that perform the functions required by this ecosystem. Thus, the health-compatible ecosystem appears to be a generalist microbiome. By contrast, disease is dominated by 'specialist organisms', which encode for novel metabolic functions (e.g., proteolysis, fermentation, methanogenesis) or virulence factors (e.g., motility, communication, stress response, iron acquisition, antibiotic resistance) not seen in health. Furthermore, functional cooperativity between bacteria, viruses and archaea is more readily observed in disease than in health and, while a strong core microbiome was detected in health, it was conspicuously absent in disease. Taken together, the microbial heterogeneity, the predominance of specialist species and the presence of novel functions that correlate with the requirements of the environment suggest that many more microbial events underlie the etiology of periodontitis than simple pathobiont expansion.

It has previously been shown that a gradient exists in the levels of pro-inflammatory cytokines, oxygen tension and antioxidant capacity between shallow-healthy, shallow-diseased and deep-diseased sites<sup>61</sup>. Our data suggest that a progressive regime shift occurs in the microbial ecosystem from health to disease, which is reflective of the gradients in the local microenvironment. This, of course, leads to the question whether the disease microbiome is the cause or the product of the disease. To answer this, we compared shallow-diseased sites with deep-diseased sites. The core microbiomes of shallow-diseased sites were phylogenetically distinct from, but functionally more aligned with deep-diseased than with shallow-healthy sites (especially in energy processing, virulence, chemotaxis, stress response and phage-mediated transfers), indicating that the microbiomes of these sites do indeed, have the potential to induce disease. This observation that shallow-disease sites possess similar pro-inflammatory abilities to disease active sites serves in part, to explain the random burst model of disease activity. However, since the potential to cause disease does not equate to disease causation, this needs to be corroborated by longitudinal investigations of the microbial metatranscriptome during shifts from health to disease.

Several decades of research have explored the possibility of identifying species that would serve as markers or predictors of disease<sup>62–67</sup>. However, the present investigation demonstrates that while a tremendous functional complementarity exists in disease, this microbiome is taxonomically heterogeneous. Further, 30 genes found in the core-disease microbiome (corresponding to 14 distinct functions) were capable of discriminating between health and disease. Also, these genes were present in both shallow-diseased and deep-diseased sites, indicating that any site may be sampled to screen subjects for disease. Thus, our data suggest that a gene-centric rather than a species-centric approach to identifying markers and predictors may be more fruitful.

## Conclusions

The central characteristics of the health-compatible subgingival microbial community are energy efficiency and functional equitability. Fewer functions are encoded within this healthy microbiome, and the general functional potential is distributed across most species, while expansion of functional capabilities can be observed in disease, within certain species or consortia contributing a few, unique functions. Disease is also characterized by taxonomic heterogeneity and functional congruence. Importantly, sites without clinical disease in subjects with disease are functionally more aligned with sites with disease than with healthy sites, indicating that they may be more at-risk-for-harm than previously believed.

## References

1. Kassebaum, N. J. *et al.* Global burden of severe periodontitis in 1990–2010: a systematic review and meta-regression. *J Dent Res* **93**, 1045–1053, doi: 10.1177/0022034514552491 (2014).
2. Brown, L. J., Johns, B. A. & Wall, T. P. The economics of periodontal diseases. *Periodontol* **2000** **29**, 223–234 (2002).

3. Mercado, F. B., Marshall, R. I. & Bartold, P. M. Inter-relationships between rheumatoid arthritis and periodontal disease. A review. *J Clin Periodontol* **30**, 761–772 (2003).
4. Seymour, G. J. *et al.* Relationship between periodontal infections and systemic disease. *Clin Microbiol Infect* **13** Suppl 4, 3–10, doi: 10.1111/j.1469-0691.2007.01798.x (2007).
5. Goodson, J. M. *et al.* Patterns of progression and regression of advanced destructive periodontal disease. *J Clin Periodontol* **9**, 472–481 (1982).
6. Gemmell, E., Yamazaki, K. & Seymour, G. J. Destructive periodontitis lesions are determined by the nature of the lymphocytic response. *Critical reviews in oral biology and medicine: an official publication of the American Association of Oral Biologists* **13**, 17–34 (2002).
7. Paster, B. J. *et al.* Bacterial diversity in human subgingival plaque. *J Bacteriol* **183**, 3770–3783, doi: 10.1128/JB.183.12.3770-3783.2001 (2001).
8. Kumar, P. S., Griffen, A. L., Moeschberger, M. L. & Leys, E. J. Identification of candidate periodontal pathogens and beneficial species using quantitative 16S clonal analysis. *J Clin Microbiology* **43** (2005).
9. Kumar, P. S. *et al.* New bacterial species associated with chronic periodontitis. *J Dent Res* **82**, 338–344 (2003).
10. Abusleme, L. *et al.* The subgingival microbiome in health and periodontitis and its relationship with community biomass and inflammation. *ISME J* **7**, 1016–1025, doi: 10.1038/ismej.2012.174 (2013).
11. Griffen, A. L. *et al.* Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *The ISME journal* **6**, 1176–1185, doi: 10.1038/ismej.2011.191 (2012).
12. Shi, B. *et al.* Dynamic changes in the subgingival microbiome and their potential for diagnosis and prognosis of periodontitis. *MBio* **6**, e01926–01914, doi: 10.1128/mBio.01926-14 (2015).
13. Wang, J. *et al.* Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. *Scientific reports* **3**, 1843, doi: 10.1038/srep01843 (2013).
14. Li, Y. *et al.* Phylogenetic and functional gene structure shifts of the oral microbiomes in periodontitis patients. *ISME J* **8**, 1879–1891, doi: 10.1038/ismej.2014.28 (2014).
15. Jünemann, S. *et al.* Bacterial Community Shift in Treated Periodontitis Patients Revealed by Ion Torrent 16S rRNA Gene Amplicon Sequencing. *PLoS ONE* **7**, e41606, doi: 10.1371/journal.pone.0041606 (2012).
16. Michalowicz, B. S. *et al.* Human herpesviruses and Porphyromonas gingivalis are associated with juvenile periodontitis. *J Periodontol* **71**, 981–988 (2000).
17. Lepp, P. W. *et al.* Methanogenic Archaea and human periodontal disease. *Proc Natl Acad Sci USA* **101**, 6176–6181, doi: 10.1073/pnas.0308766101 (2004).
18. Krom, B. P., Kidwai, S. & ten Cate, J. M. Candida and Other Fungal Species: Forgotten Players of Healthy Oral Microbiota. *Journal of Dental Research* **93**, 445–451, doi: 10.1177/0022034514521814 (2014).
19. La Rosa, P. S. *et al.* Hypothesis Testing and Power Calculations for Taxonomic-Based Human Microbiome Data. *PLoS ONE* **7**, e52078, doi: 10.1371/journal.pone.0052078 (2012).
20. Wilke, A. *et al.* The MG-RAST metagenomics database and portal in 2015. *Nucleic Acids Res* **44**, D590–594, doi: 10.1093/nar/gkv1322 (2016).
21. Tang, W. *et al.* Building Scalable Data Management and Analysis Infrastructure for Metagenomics (2013).
22. Wilke, A. *et al.* A RESTful API for accessing microbial community data for MG-RAST. *PLoS Comput Biol* **11**, e1004008, doi: 10.1371/journal.pcbi.1004008 (2015).
23. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30 (2000).
24. Overbeek, R. *et al.* The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* **42**, D206–214, doi: 10.1093/nar/gkt1226 (2014).
25. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi: 10.1186/s13059-014-0550-8 (2014).
26. Wilke, A. *et al.* The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatics* **13**, 141, doi: 10.1186/1471-2105-13-141 (2012).
27. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res* **12**, 656–664, doi: 10.1101/gr.229202. Article published online before March 2002 (2002).
28. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590–596, doi: 10.1093/nar/gks1219 (2013).
29. Chen, T. *et al.* The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database: the journal of biological databases and curation* **2010**, baq013, doi: 10.1093/database/baq013 (2010).
30. Mason, M. R. *et al.* The subgingival microbiome of clinically healthy current and never smokers. *ISME J*, doi: 10.1038/ismej.2014.114 (2014).
31. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359, doi: 10.1038/nmeth.1923 (2012).
32. Forum on Microbial Health, Board on Global Health & Medicine, I. o. In *Microbial Ecology in States of Health and Disease: Workshop Summary The National Academies Collection: Reports funded by National Institutes of Health* (2014).
33. Fruchterman, T. M. & Reingold, E. M. Graph drawing by force-directed placement (1991).
34. Bastian, M., Heymann, S. & Jacomy, M. Gephi: an open source software for exploring and manipulating networks. *ICWSM* **8**, 361–362 (2009).
35. Bluthgen, N., Menzel, F. & Bluthgen, N. Measuring specialization in species interaction networks. *BMC Ecol* **6**, 9, doi: 10.1186/1472-6785-6-9 (2006).
36. Dormann, C. F., Fründ, J., Blüthgen, N. & Gruber, B. Indices, graphs and null models: analyzing bipartite ecological networks (2009).
37. Dormann, C. F., Gruber, B. & Fründ, J. Introducing the bipartite package: analysing ecological networks. *interaction* **1**, 0.2413793 (2008).
38. Díaz-Uriarte, R. & Alvarez de Andrés, S. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* **7**, 1–13, doi: 10.1186/1471-2105-7-3 (2006).
39. Liu, B. *et al.* Deep Sequencing of the Oral Microbiome Reveals Signatures of Periodontal Disease. *PLoS One* **7**, e37919, doi: 10.1371/journal.pone.0037919 (2012).
40. Duran-Pinedo, A. E. *et al.* Community-wide transcriptome of the oral microbiome in subjects with and without periodontitis. *ISME J* **8**, 1659–1672, doi: 10.1038/ismej.2014.23 (2014).
41. Rodriguez, R. L. & Konstantinidis, K. T. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics* **30**, 629–635, doi: 10.1093/bioinformatics/btt584 (2014).
42. Haiko, J. & Westerlund-Wikstrom, B. The role of the bacterial flagellum in adhesion and virulence. *Biology (Basel)* **2**, 1242–1267, doi: 10.3390/biology2041242 (2013).
43. Martinez, J. L. & Baquero, F. Mutation frequencies and antibiotic resistance. *Antimicrob Agents Chemother* **44**, 1771–1777 (2000).
44. Levy, S. B. The 2000 Garrod lecture. Factors impacting on the problem of antibiotic resistance. *The Journal of antimicrobial chemotherapy* **49**, 25–30 (2002).
45. Nikaido, H. & Pages, J. M. Broad-specificity efflux pumps and their role in multidrug resistance of Gram-negative bacteria. *FEMS Microbiol Rev* **36**, 340–363, doi: 10.1111/j.1574-6976.2011.00290.x (2012).

46. Al-Qutub, M. N. *et al.* Hemin-dependent modulation of the lipid A structure of Porphyromonas gingivalis lipopolysaccharide. *Infect Immun* **74**, 4474–4485, doi: 10.1128/IAI.01924-05 (2006).
47. Mydel, P. *et al.* Roles of the host oxidative immune response and bacterial antioxidant rubrerythrin during Porphyromonas gingivalis infection. *PLoS Pathog* **2**, e76, doi: 10.1371/journal.ppat.0020076 (2006).
48. Boor, K. J. Bacterial stress responses: what doesn't kill them can make them stronger. *PLoS biology* **4**, e23 (2006).
49. Jorth, P. *et al.* Metatranscriptomics of the human oral microbiome during health and disease. *MBio* **5**, e01012–01014, doi: 10.1128/mBio.01012-14 (2014).
50. Slots, J. Periodontal herpesviruses: prevalence, pathogenicity, systemic risk. *Periodontol 2000* **69**, 28–45, doi: 10.1111/prd.12085 (2015).
51. Hall, R. J. *et al.* Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *Journal of virological methods* **195**, 194–204, doi: 10.1016/j.jviromet.2013.08.035 (2014).
52. Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214, doi: 10.1038/nature11234 (2012).
53. Blew, R. D. On the Definition of Ecosystem. *Bulletin of the Ecological Society of America* **77**, 171–173, doi: 10.2307/20168066 (1996).
54. Loesche, W. J. *et al.* Relationship between oxygen tension and subgingival bacterial flora in untreated human periodontal pockets. *Infect Immun* **42**, 659–667 (1983).
55. Vianna, M. E. *et al.* Quantitative analysis of three hydrogenotrophic microbial groups, methanogenic archaea, sulfate-reducing bacteria, and acetogenic bacteria, within plaque biofilms associated with human periodontal disease. *J Bacteriol* **190**, 3779–3785, doi: 10.1128/JB.01861-07 (2008).
56. Eckburg, P. B., Lepp, P. W. & Relman, D. A. Archaea and their potential role in human disease. *Infect Immun* **71**, 591–596 (2003).
57. Loesche, W. J. & Grossman, N. S. Periodontal disease as a specific, albeit chronic, infection: diagnosis and treatment. *Clin Microbiol Rev* **14**, 727–752, table of contents, doi: 10.1128/CMR.14.4.727-752.2001 (2001).
58. Niederman, R., Zhang, J. & Kashket, S. Short-chain carboxylic-acid-stimulated, PMN-mediated gingival inflammation. *Crit Rev Oral Biol Med* **8**, 269–290 (1997).
59. Jiao, Y., Hasegawa, M. & Inohara, N. The Role of Oral Pathobionts in Dysbiosis during Periodontitis Development. *Journal of Dental Research* **93**, 539–546, doi: 10.1177/0022034514528212 (2014).
60. Hajishengallis, G. & Lamont, R. J. Beyond the red complex and into more complexity: the polymicrobial synergy and dysbiosis (PSD) model of periodontal disease etiology. *Molecular oral microbiology* **27**, 409–419, doi: 10.1111/j.2041-1014.2012.00663.x (2012).
61. Emingil, G., Atilla, G., Baskesen, A. & Berdeli, A. Gingival crevicular fluid EMAP-II, MIP-1alpha and MIP-1beta levels of patients with periodontal disease. *J Clin Periodontol* **32**, 880–885 (2005).
62. Haffajee, A. D. *et al.* Clinical and microbiological features of subjects with adult periodontitis who responded poorly to scaling and root planing. *J Clin Periodontol* **24**, 767–776 (1997).
63. Listgarten, M. A. *et al.* Incidence of periodontitis recurrence in treated patients with and without cultivable Actinobacillus actinomycetemcomitans, Prevotella intermedia, and Porphyromonas gingivalis: a prospective study. *J Periodontol* **62**, 377–386 (1991).
64. Wennstrom, J. L., Dahlen, G., Svensson, J. & Nyman, S. Actinobacillus actinomycetemcomitans, Bacteroides gingivalis and Bacteroides intermedius: predictors of attachment loss? *Oral Microbiol Immunol* **2**, 158–162 (1987).
65. Di Murro, C. *et al.* Occurrence of Porphyromonas gingivalis, Bacteroides forsythus, and Treponema denticola in periodontally healthy and diseased subjects as determined by an ELISA technique. *J Periodontol* **68**, 18–23 (1997).
66. MacFarlane, T. W. *et al.* Longitudinal study of untreated periodontitis (II). Microbiological findings. *J Clin Periodontol* **15**, 331–337 (1988).
67. Perinetti, G. *et al.* Longitudinal monitoring of subgingival colonization by Actinobacillus actinomycetemcomitans, and crevicular alkaline phosphatase and aspartate aminotransferase activities around orthodontically treated teeth. *J Clin Periodontol* **31**, 60–67 (2004).

## Author Contributions

Shareef M. Dabdoub: Study design, Data analysis, manuscript preparation. Sukirth M. Ganesan: Study design, Sample preparation, data analysis, manuscript preparation. Purnima S. Kumar: Study design, subject recruitment and clinical evaluation, data analysis, manuscript preparation.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Dabdoub, S. M. *et al.* Comparative metagenomics reveals taxonomically idiosyncratic yet functionally congruent communities in periodontitis. *Sci. Rep.* **6**, 38993; doi: 10.1038/srep38993 (2016).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016