Identification of the bacterial biosynthetic gene clusters of the oral microbiome illuminates the unexplored social language of bacteria during health and disease

Gajender Aleti[1], Jonathon L. Baker[1], Xiaoyu Tang[1], Ruth Alvarez[2], Márcia Dinis[2], Nini C. Tran[2], Alexey V. Melnik[4], Cuncong Zhong[3], Madeleine Ernst[4], Pieter C. Dorrestein[4], Anna Edlund[1#]

[1] *Genomic Medicine Group, J. Craig Venter Institute, La Jolla, California, USA.*

[2] *School of Dentistry, The University of California, Los Angeles, USA.*

[3] *Department of Electric Engineering and Computer Science, The University of Kansas, Lawrence, Kansas, USA.*

[4] *Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, The University of California San Diego, La Jolla, California, USA.*

Running Title: Biosynthetic pathways of the human oral microbiome

[#] Address correspondence to Anna Edlund, mail: aedlund@jcvi.org

ABSTRACT

Small molecules are the primary communication media of the microbial world. Recent bioinformatics studies, exploring the biosynthetic gene clusters (BGCs) which produce many small molecules, have highlighted the incredible biochemical potential of the signaling molecules encoded by the human microbiome. Thus far, most research efforts have focused on understanding the social language of the gut microbiome, leaving crucial signaling molecules produced by oral bacteria, and their connection to health versus disease, in need of investigation. In this study, a total of 4,915 BGCs were identified across 461 genomes representing a broad taxonomic diversity of oral bacteria. Sequence similarity networking provided a putative product class for over 100 unclassified novel BGCs. The newly identified BGCs were cross-referenced against 254 metagenomes and metatranscriptomes derived from individuals with either good oral health, dental caries, or periodontitis. This analysis revealed 2,473 BGCs, which were differentially represented across the oral microbiomes associated with health versus disease. Co-abundance network analysis identified numerous inverse correlations between BGCs and specific oral taxa. These correlations were present in health, but greatly reduced in dental caries, which may suggest a defect in colonization resistance. Finally, corroborating mass spectrometry identified several compounds with homology to products of the predicted BGC classes. Together, these findings greatly expand the number of known biosynthetic pathways present in the oral microbiome and provide an atlas for experimental characterization of these abundant, yet poorly understood, molecules and socio-chemical relationships, which impact the development of caries and periodontitis, two of the world's most common chronic diseases.

IMPORTANCE

The healthy oral microbiome is symbiotic with the human host, importantly providing colonization resistance against potential pathogens. Dental caries and periodontitis are two of the world's most common and costly chronic infectious diseases, and are caused by a localized dysbiosis of the oral microbiome. Bacterially produced small molecules, often encoded by BGCs, are the primary communication media of bacterial communities, and play a crucial, yet largely unknown, role in the transition from health to dysbiosis. This study provides a comprehensive mapping of the BGC repertoire of the human oral microbiome and identifies major differences in health compared to disease. Furthermore, BGC representation and expression is linked to the abundance of particular oral bacterial taxa in health versus dental caries and periodontitis. Overall, this study provides a significant insight into the chemical communication network of the healthy oral microbiome, and how it devolves in the case of two prominent diseases.

KEY WORDS: oral microbiome, small-molecules, biosynthetic gene cluster, caries, periodontitis

INTRODUCTION

The human body is inhabited by rich and diverse bacterial communities , which are intimately linked to the health of the human host (1). Small molecules, which are often encoded by biosynthetic gene clusters (BGCs), are the primary means of communication in this microbial world. Recent studies suggest that the human microbiota has the potential to synthesize a myriad of exquisite small molecules, and that these small molecules serve as mediators in a variety of microbe-microbe and host-microbe interactions (2-4). These include: antibacterial activity (5), bacterial signaling (6), immune modulation (7), biofilm formation (8, 9), host colonization (10), nutrient-scavenging (11) and stress protection (12). Disruption of the finely-tuned equilibrium of the bacterial ecosystems in the human microbiome, referred to as dysbiosis, is associated with a plethora of diseases. While the mechanistic underpinnings of a shift to a dysbiotic community remain poorly understood, there is little doubt that signaling via the small molecules produced by microbial BGCs plays a critical role in the transition to dysbiosis, and associated pathogenesis (13, 14).

The human oral cavity contains an assortment of ecological niches, and as such, harbors one of the most diverse microbial populations in the human body (1, 15). Dental caries and periodontitis are two of the most common and costly chronic conditions afflicting humans, and are the result of localized dysbiosis in the oral cavity (16-20). Unlike the rest of the human digestive tract, the oral cavity is consistently exposed to the exterior environment. Therefore, an indispensable portion of the first line of defense against invading pathogens is the colonization resistance provided by a healthy oral microbiome. Indeed, dysbiosis of the oral microbiome is not only directly linked to oral diseases, but is also implicated in system-wide health (21), stressing the urgent need to unravel the underlying factors that shape and maintain a healthy human oral microbiome.

Elucidating the transmissions relayed by oral bacterial small molecules could lead to a deeper understanding of key ecological factors that set the stage for oral community succession, in health and pathogenesis. A large and growing body of literature suggests that the microbial composition and metabolic potential of the saliva and dental plaque varies significantly in healthy versus disease states (22-28). Therefore, we hypothesize that the abundance and expression of BGCs, which produce small molecules, may drive crucial bacterial interactions which contribute to health or disease. To explore this further, the biosynthetic capacity of 461 well-annotated oral bacterial genomes was investigated, and an enormous diversity of BGCs was revealed. In addition, sequence reads from 294 publicly available metagenomes and metatranscriptomes, which were associated with health, dental caries, or periodontitis, were mapped to these novel oral BGCs. This analysis identified 2,473 biosynthetic pathways which were differentially represented in health versus disease. In addition, the BGC content in salivary metagenomes obtained from 24 healthy children and 23 children with dental caries was analyzed. A Bayesian network approach was employed to identify both positive and inverse correlations between BGCs and bacterial taxa, which revealed differentially abundant signaling networks and species in health compared to dental caries. Overall, this study provides a significant insight into the chemical communication network of the healthy oral microbiome, and how it devolves in the case of dental caries and periodontitis.

RESULTS AND DISCUSSION

The human oral microbiome encodes thousands of diverse BGCs from an array of species. To explore the metabolic capacity of the human oral microbiome in-depth, a comprehensive pipeline for mining bacterial genomes was established, utilizing antiSMASH infrastructure v4 (accessible at https://antismash.secondarymetabolites.org/) (29), including MultiGeneBlast (30). An oral bacterial genome sequence database was assembled to include a total of 461 well-curated and annotated bacterial genomes, representing 113 unique bacterial genera and 298 taxonomically unique species, as well as 72 taxa unclassified at the species level (Table S1). Genomes were selected based on their completeness and level of annotation. A single genome sequence for each bacterial species was included to circumvent the overrepresentation of BGCs from bacteria with a high number of genome representatives. Indeed, in a previous bioinformatics study of 169 *S. mutans* genomes, ~1,000 putative BGCs were identified, revealing an incredible potential to produce small molecules within one bacterial species (31). Therefore, it should be noted that the estimated BGC diversity reported here is likely underestimated. Clearly, strain-level diversity is important to explore in future studies. However, this will require extensive genome sequencing, since to-date most oral bacterial species lack multiple reference genomes. By applying the genome-mining pipeline described above, a total of 4,915 BGCs of known and unknown types were identified (Table S1). BGCs annotated as fatty acid synthases, which are often involved in primary metabolism, were excluded. Approximately 50% of the identified BGCs were of an unknown class, congruent with the observations of other efforts to identify BGCs (Table S1)(2). The remaining 50% of BGCs (2,250) shared sequence similarities with an extensive range of previously characterized BGC classes, which is likely reflective of the high taxonomic diversity observed within the oral cavity as compared to many other body sites (1) (Fig. 1A).

Of the BGCs of a known class, a substantial fraction (1,398 BGCs, 62%) were annotated as oligosaccharides, making it the most abundant class of BGCs in the oral cavity. Oligosaccharide pathways are widely distributed across bacterial phyla and are predominant in Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria and Fusobacteria, with the highest number being identified in Firmicutes (Fig. 1B and C). Their ecological roles are largely underexplored, but studies show important functions such as capsule formation in virulence development (32) and attachment to surfaces, including neighboring bacterial species and host cells (33). Furthermore, diffusible oligosaccharides are known to display antibacterial activities (34), for example a previous study showed that polysaccharide A from the human gut bacterium *Bacteroides fragilis* can modulate the gut mucosal immune response (35, 36).

Another highly represented BGC class was ribosomally synthesized and post-translationally modified peptides (RiPPs), for which 209 BGCs (9.3% of BGCs of a known class) were identified. RiPPs include molecules such as bacteriocins, lantipeptides, sactipeptides, cyanobactins, and proteusins (denoted as fluorescent green in Fig 1). Of these RiPP types, bacteriocin-encoding BGCs were the most abundant as they contributed ~75% of the total RiPP diversity. Interestingly, although bacteriocin producing-BGCs were abundant in the oral microbiome overall, they were depleted in all Bacteroidetes genomes (Fig. 1C). The role of RiPPs, such as the bacteriocins, demands further exploration, as they exhibit antagonistic activities against other microbes sharing the same ecological niche, and influence competition for persistence between commensals and pathogens (37, 38). Furthermore, multiple studies genetic transformation in *Streptococcus* show that competence is tightly linked to bacteriocin production (39), which suggests that these molecules also play important roles in the horizontal transfer of genes and ultimately in niche differentiation and population structure changes.

BGCs encoding aryl polyene-like molecules in several Bacteroidetes and Proteobacteria genomes were identified (131 BGCs or 5.8% of BGCs of a known class). Aryl polyenes are predicted to function as protective agents against oxidative stress (40). However, only a few candidates have been experimentally characterized, leaving this group of small molecules highly underexplored. A diversity of non-ribosomal peptide synthetases (NRPSs), polyketide synthase (PKS), and NRPS-PKS hybrid BGCs (ranging between 0.9% and 4.4% of BGCs of a known class) were identified, in line with a prior study, which classified BGCs in the human microbiome in multiple body habitats (2). These compound classes are known for their antimicrobial activities and were previously characterized as possessing various nutrient-scavenging, immunosuppressant, surfactant, and cytotoxic properties (41). BGCs of the terpene class were also identified (95 BGCs, 4.2% of BGCs of a known class). This diverse group of small molecules may also be of ecological and medicinal interest since their activities have been reported as both anti-inflammatory and antimicrobial (42). The class 'other' encompasses BGCs that fall outside the known categories of antiSMASH-annotation, includes rare classes found in only few species, and constituted 9.4% of the total BGCs identified (Fig. 1A). Taken together, these results show that the oral microbiome encodes a vast and highly diverse array of small molecules that have largely unexplored, yet likely pivotal, roles in ecology and health.

Sequence similarity networking reveals unexplored BGC diversity, even in well-studied classes of BGCs. In order to assess the evolutionary relationships between conserved domains in the proteins encoded by BGCs, as well as to group BGCs of similar putative function to evaluate novelty, a sequence similarity network approach was applied (see File S1). Briefly, the BGCs that were identified from the bacterial genomes using antiSMASH were aligned to the MIBiG repository (43) of 1,409 experimentally validated reference BGCs using the BiG-SCAPE algorithm (https://git.wageningenur.nl/medema-group/BiG-SCAPE). The resulting network comprised 4,242 nodes and 19,847 connecting edges revealing both close and distant homology to characterized biosynthetic pathways (Fig. 2). Notably, a significant fraction of the previously unclassified BGCs did sub-network with BGCs predicted to be of a known class, particularly the oligosaccharide, RiPP and aryl polyene classes (Fig. 2). This data provides inferences as to the function of over 100 previously unclassified novel BGCs.

The largest sub-network, comprised of mainly oligosaccharide-encoding BGCs, showed no significant homology with any experimentally validated BGCs in the MIBiG repository (Fig. 2). This may be due to the fact that oligosaccharide-producing BGCs are in at times categorized with primary metabolism, and not natural product-producing BGCs, as is the case in this study. The second-largest major sub-network was comprised of primarily unclassified BGCs. These may encompass distinct chemical scaffolds, and may represent a rich source of novel BGC types. The NRPS, PKS, NRPS-PKS hybrids, and a few terpene BGCs, grouped together forming a subnetwork implying a set of common core domains involved in these biosynthetic assembly lines, as described previously (4, 41). The majority of NRPS, PKS, NRPS-PKS hybrids, and RiPPs (in particular thiopeptides and lantipeptides) showed strong associations with MIBiG reference BGC sequences. It should be noted that these are the most prevalent classes in the MIBiG repository (Table S2). Currently, only four experimentally characterized aryl polyene BGCs exist in the MIBiG database, therefore it was not surprising that none of the nodes in the aryl polyene cluster sub-networked with MIBiG reference BGCs. Given that aryl polyenes are thought to be the most abundant BGC class in the human microbiome (4), this indicates that this class of molecules is severely understudied (Fig. 1 and Table S2). Several BGCs annotated as saccharides, other, unclassified, PKS and NRPS BGC types grouped with aryl polyene BGCs, which may represent novel hybrid classes of BGC. Other small sub-networks include biosynthesis of terpene phenazine, homoserine lactone, alkaloid, siderophore, and ectoine. These sub-networks did not associate with MIBiG reference BGCs, indicating that they also await experimental validation. Our implemented analysis approach, using the MIBiG/BiG-SCAPE pipeline, is powerful with regards to predicting the functions of novel BGCs. The annotations we generated here provide deeper insights of which BGCs and compound classes are most likely to be identified in futures studies, due to knowledge of their closest neighbor's biochemical properties. The BGCs remaining with completely unknown functions represent exciting future challenges, which could be addressed by generating large-insert BGC expression libraries.

While antiSMASH and network analysis were employed for broad classification of BGCs into known classes, MultiGeneBlast was also utilized at the level of the entire gene cluster to further annotate BGCs in-depth and identify homologs against the MIBiG repository (30). Using this approach, the 4,915 BGCs were classified into four major categories based upon the level of homology to known experimentally validated BGCs in the MIBiG repository. This categorization resulted in 1,146 (20%) BGCs closely homologous, 848 (15%) BGCs moderately homologous and 2,221 (40%) BGCs distantly homologous to well-characterized BGCs (Fig. S2). 1,393 (~25%) BGCs did not appear to have significant homology to BGCs in MIBiG, based upon the E-value (see Methods section for details). Such a detailed annotation of BGCs harbored by the human oral microbiome has not been accomplished previously.

Specific BGCs are associated with periodontitis and dental caries. We next systematically examined the differential representation of bacterial BGCs in saliva and dental plaque across 294 human subjects with good oral health, dental caries, or periodontitis. The data from 247 subjects was obtained from eight previous studies, which represented all publicly available metagenomes and metatranscriptomes associated with caries or periodontal disease, compared to health, at the time of this study (Table S3). In addition, DNA from 47 saliva samples representing 23 children with caries and 24 healthy children was sequenced and putative BGCs were identified (see Fig S3 for workflow). Non-supervised exploratory ordination through PCoA revealed significant differences in the representation of BGCs between healthy and diseased subjects in five of the six metatranscriptome studies and six of the seven metagenome studies investigated (Fig.

S4). The 1,804 BGCs which were differentially represented in health versus disease in the metagenomes and metatranscriptomes are summarized in Table 1.

The BGCs associated with disease in the metatranscriptome studies were related to the synthesis of a broad range of small molecule types. These particularly included BGCs of the oligosaccharide, aryl polyene, terpene, bacteriocin and NRPS classes (Fig. 3). BGCs encoding PKS, NRPS, and bacteriocins from *Actinomyces*, *Rothia* and *Corynebacterium* had increased expression in subjects with caries, while BGCs encoding terpenes and aryl polyenes from *Neisseria spp.* and Proteobacteria had increased expression in healthy subjects (Fig. 3). Previous studies illustrated that aryl polyenes act as protective agents against oxidative stress, and that terpenes function as anti-inflammatory agents (40). Interestingly, high levels of *Actinomyces* were previously associated with severe early childhood caries (44). In the caries associated samples, known caries-associated species belonging to the *Streptococcus, Veillonella*, and *Lactobacillus* genera (45) showed notable changes in bacteriocins and oligosaccharides BGC expression profiles (Fig. 3).

In periodontitis, a high number of differentially expressed BGCs (170 BGCs) were identified in community members belonging to the Bacteroidetes phylum. Interestingly, several BGCs encoded by periodontal pathogens of the red and orange complexes (e.g. *Porphyromonas gingivalis*) were differentially expressed in health compared to periodontal disease. Known red complex species had increased expression of BGCs belonging to the aryl polyene, oligosaccharide, homoserine lactone and resorcinol classes in diseased states. *Neisseria spp*. also showed interesting signatures, such as increased expression of BGCs belonging to the terpene, resorcinol, bacteriocin, and homoserine lactone classes (Fig. 3). Homologs to specific BGC products in the MIBiG database which displayed differential expression in health and disease are detailed in Figure S5. Analysis of the metagenomic studies yielded similar trends to those detailed above (Figs. S6 and Fig. S7).

Next, a subset of differentially represented BGCs, which showed high expression in either healthy or diseased states, was examined to determine if they commonly occur across studies. The results were visualized as a binary occurrence matrix (Fig. S8). In all studies analyzed, only a minor fraction of the differential features (< 10 BGCs) were shared between any two studies. Besides high inter- and intra-individual variations in the microbial composition, the significant study-to-study variation can likely be attributed to differences in sequencing platforms (Table S3). This factor may have influenced the sequence composition and sequencing depth, particularly considering the metagenome and metatranscriptome complexity (Fig. S9 and Table S4). Based on the above comparisons, the authors suggest that differences between sequencing and computational platforms (e.g. alignment parameters and sequence read filtering) must be considered, and that future efforts to obtain high-quality, deep-coverage sequencing data will help alleviate the study-to-study noise observed here.

Correlations between BGCs and oral taxa are depleted in dental caries as compared to health. To examine the relationship between BGCs and bacterial taxa during health and disease, a focused comparative analysis of the shotgun metagenomics data obtained in this study from healthy children and children with caries was performed. Interactions between BGCs and microbial taxa were examined by employing cooccurrence network analysis using the SparCC algorithm, which has the benefit of limiting the number of spurious correlations identified due to species data being compositional (46). While positive correlations were more evident among taxa-taxa relationships, (i.e. different taxa benefit from one another's presence), almost all significant correlations that were identified between specific BGCs and taxa were negative (Table S5 and Figs. S10 and S11). This suggests that antagonistic relationships, modulated through BGC-produced antimicrobial molecules, are highly significant to the ecology of the oral microbiome.

All BGCs which had significant correlations to oral taxa (a total number of 36) were annotated as close homologs to previously characterized BGCs belonging to the PKS, NRPS, NRPS-PKS hybrid, oligosaccharide and aryl polyene classes (Fig. S10). In the oral microbiomes derived from healthy children, the interaction network was dominated by negative correlations between oral taxa and BGCs producing glycopeptidolipids, capsular polylsaccharides, as well as a homolog of flexirubin (Fig. S11A). The glycopeptidolipids were encoded by the opportunistic pathogens *Kytococcus sedentarius* and *Mycobacterium neoaurum,* and were primarily shown to vary inversely with the oral taxa *Lactobacillus*, *Prevotella*, *Capnocytophaga* and *Enterococcus* (Fig. S11A and Table S5). The flexirubin homolog BGCs were encoded by *Actinomyces massiliensis* and *Prevotella oralis,* and displayed antagonistic activity against 122 taxa, including *Streptococcus mutans,* historically considered the primary etiologic species of dental caries (Fig. S11A and Table S5). Homologs of the antibiotics bacillaene and pristinamycin (47, 48), harbored by genomes of the health-associated species *Propionibacterium propionicum* F0230a and *Actinomyces timonensis* DSM 23838 (Fig. S11A, Table S5), displayed negative correlations with several pathogenic taxa: *Lactobacillus*, *Listeria*, *Lysinabacillus*, *Acinetobacter*, *Enterococcus*, *Neisseria*, *Staphylococcus*, *Kingella* and *S. mutans* (49) (Table S5). These associations are reminiscent of a previous study which observed similar macrolide-encoding BGCs widely distributed amongst oral bacterial genomes (2). These macrolide structures were also reported to inhibit the growth of cariogenic Streptococci (50). This collective evidence indicates that the isolation and characterization of bacillaene- and pristinamycin-like molecules in future studies may be key to understanding important health-protective mechanisms in the oral cavity. Finally, *P. propionicum* F0230a encoded a BGC with high sequence homology to a non-ribosomal peptide pathway encoding the genotoxin, colibactin (51). This BGC showed antagonistic associations with pathogenic genera: *Haemophilus*, *Aggregatibacter*, *Parascardovia*, *Capnocytophaga* and *Streptococcus*.

Most intriguingly, the number of significant correlations between BGCs and microbial taxa was dramatically reduced in the samples derived from children with caries (Fig. S10A to C). This may indicate that in the oral cavities exhibiting disease, the well-documented colonization resistance of the oral microbiome may be impaired. Of the few significant correlations between BGCs and taxa within the interaction network of the caries-associated microbiome, the vast majority involved BGCs encoding RiPPs with close homology to nosiheptide and hygromycin BGCs (Fig. S11B, Table S5). The nosiheptide-like BGC, encoded in the genome of *C. matruchotii,* was the most predominant, with antagonistic interactions against ~90 taxa. These included pathogens from the *Klebsiella*, *Helicobacter*, *Filifactor*, *Haemophilus*, *Enterococcus*, *Fusobacterium* genera (Fig. S11B, Table S5). The hygromycin-like BGC from *P. propionicum* negatively correlated with several pathogens belonging to the genera *Lactobacillus*, *Neisseria*, *Klebsiella*, *Anaerococcus* and *Pseudoramibacter*. Interestingly, there were no significant correlations between *S. mutans* and BGCs in the caries-associated oral microbiomes, which may indicate that during disease, the community lacks the ability to limit the abundance of this keystone pathogen. Taken together, these results suggest that in the oral microbiome, exclusion of particular taxa via antagonistic interactions, mediated by the products of BGCs, is widespread (Table S5). Although such interactions were still present in the caries-associated oral microbiomes, they were much fewer in number. This underscores the importance of ecology, and the role of BGC-produced small molecules, in the balance between health and disease.

Homologs of BGC-produced small molecules are present in oral metabolomes associated with caries and health. To validate the production of small molecules by differentially abundant BGCs, untargeted liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis of saliva samples was performed. Utilizing the Global Natural Products Social Molecular Networking (GNPS) (52) analysis platform, a mass spectral molecular network consisting of 1,369 mass spectral features grouped into 69 molecular families (two or more connected components of a graph) was obtained. 50 matches were acquired between the query MS/MS spectra and characterized reference spectra from GNPS. To further enhance mass spectrometry annotations and to link annotations to known chemical structures encoded by BGCs, major chemical classes were putatively identified by integrating mass spectral molecular networking with *in silico* annotations and automated chemical classification approaches (53-55). This allowed identification of approximately 38% of the nodes in the mass spectral molecular network at the chemical class level. The most predominant chemical classes within the network were carboxylic acids and derivatives, prenol lipids, fatty acyls, and flavonoids (Fig. S12). Substructures associated with macrolides, terpenoids, and macrolactams were also identified. At the chemical class level, distinct relative abundance patterns between the health and disease-associated samples could be observed for carboxylic acids and derivatives. The PCoA analysis of the 1,369 unidentified MS features, showed clear separation of samples between healthy and diseased states (Fig. S13A), in agreement with the BGC abundance profiles (Fig. S4M). By employing a random forest importance model, 15 key metabolites, which were distinct between healthy and disease states (Fig. S13B), were identified. 12 of the 15 key metabolites were significantly more abundant in healthy subjects, while three were more abundant in the subjects with dental caries. Out of the three key metabolites that were significantly more abundant in the diseased subjects, two matches were obtained to lipid compounds from GNPS reference spectra resulting in a level-2 metabolite identification (56). These matches were N-Nervonoyl-D-erythro-sphingophosphorylcholine and 13-Docosenamide. These molecules are likely to originate from the human host and warrant further investigation.

Using the *in silico* Network Annotation Propagation tool (NAP) (57), putative structural matches were obtained for 6 out of the 12 key metabolites that were more abundant in the healthy subjects, including terpenoids, phenylpropanoids as well as fatty alcohols. It should be noted however, that one of the limitations of *in silico* annotation is the uncertainty around the correct structure among the predicted candidate structures. Results should therefore be interpreted with care, and an accurate prediction of the putative identity would require follow-up investigations, which is outside the scope of the present study. It should be also noted that both the genomics and metabolomics approaches employed identify putative homologs and not exact matches. Thus, using current techniques and databases, is it not possible to definitively determine if the small molecules identified by LC-MS/MS were in-fact produced by the specific BGCs predicted by antiSMASH. However, the LC-MS/MS analyses largely support the results of the genomic analyses by detecting classes of small molecules and homologs which were similar to those discovered by the complementary BGC genomics analyses.

Concluding remarks. This study significantly expands the number of identified BGCs encoded by bacteria of the human oral microbiome and designates putative products to many novel clusters. Representation and expression of the newly identified BGCs, as well as their relationship to the abundance of oral bacterial taxa was examined during health, dental caries, and periodontitis, revealing significant differences in microbial social ecology and communication among the three host outcomes. This work provides an atlas for further examination and experimental validation of the identified socio-chemical relationships and their role in the pathogenesis of dental caries and periodontal disease. A deeper elucidation of the social activities of the microbes residing in the oral cavity will significantly improve our understanding of the pathogenesis of oral (and extra-oral) diseases and will guide development of improved therapeutic strategies to maintain health.

MATERIALS AND METHODS

The ethics statement is provided in the Supplementary Materials and Methods section of Supplementary Material File S1**.**

Study inclusion/exclusion criteria and collection of saliva. Approximately 2 ml saliva was collected by spitting method in a 15 ml Falcon tube over a 20 min period. Whole saliva was immediately transferred to sterile 2 ml cryovial tubes and centrifuged at 6000 x g for 5 minutes to remove eukaryotic cells and solid debris. Supernatants were collected, mixed with glycerol (20%), and snap-frozen for long term storage at -80˚C. For detailed protocol, see Supplementary Materials and Methods.

DNA extraction and metagenomics sequencing. For a detailed protocol, see Supplementary Materials and Methods.

BGC identification and network analysis of known and putative oral BGCs. A list of 1,362 described and curated human oral taxa (18th September 2017) was obtained from HOMD, Human Oral Microbiome Database (55). In order to identify small molecule and secondary metabolite-encoding BGCs in genomes of bacterial taxa representative of a broad oral bacterial diversity, 461 complete and high-quality draft genomic sequences, annotated as dynamic and static, were obtained from the National Center of Biotechnology Information genome database (http://www.ncbi.nlm.nih.gov/genome), as well as from an in-house database (Table S1). These were concatenated into a major query-database and fed to antiSMASH, (Antibiotics & Secondary Metabolite Analysis Shell, version 4.0) (29). Multiple nucleotide FASTA sequences from BGCs were constructed. We excluded a list of 320 previously described non-biosynthetic genes commonly found in BGCs (2) (Table S6) based on text within an attribute using advanced filter settings in CLC Workbench software v. 9. (CLCbio, Aahus, Denmark). The resulting dataset contained a total of 192,283 gene sequences from 4,915 BGCs and can be downloaded from the MassIVE repository (https://massive.ucsd.edu/) with the accession ID MSV000081832. For more information, see Supplementary Materials and Methods.

Comparison of BGCs with known biosynthetic pathways. A reference MIBiG database comprising multiple amino acid sequences for each BGC was constructed using MultiGeneBlast (30). To further compare BGCs derived (excluding the fatty acid synthase encoding BGCs) from oral bacterial genomes with those encoding the biosynthetic pathways for known compounds, we performed multi-gene homology searches using complete gene cluster sequences against the MIBiG database by using the stand-alone version of MultiGeneBlast (http://multigeneblast.sourceforge.net/) algorithm with default settings. Subsequently, for each queried BGC, we extracted information from the top hit (with the highest cumulative BLAST bit score) from an output of multiple BLAST hits using an in-house python script. For additional information, see Supplementary Materials and Methods.

16S rRNA gene (16S) phylogenetic analysis. For a detailed protocol, see Supplementary Materials and Methods.

Metagenomic and metatranscriptomic data collection. Shotgun metatranscriptomic and metagenomic sequencing data published previously by Duran-Pinedo et al. (22), Belda-Ferre et al. (23), Belstrøm et al. (24), Jorth et al. (58), Do et al. (25), Peterson et al. (26), Yost et al. (27), Wang et al. (28), and Shi et al. (59), as well as our own study of metagenomes from saliva obtained from children with good dental health, or children with dental caries was analyzed (sequence reads are accessible under BioProject PRJNA1234. Table S3). For detailed protocol, see Supplementary Materials and Methods.

Differential abundance and expression analyses of BGCs. We employed a systematic workflow for analyzing abundance and expression profiles of the BGCs (see Fig. S3). Using SRA toolkit utilities, reads were extracted from metatranscriptome and metagenome shotgun sequenced libraries available via NCBI. For a detailed protocol, see Supplementary Materials and Methods.

Principal Coordinate analysis. The differences between samples from healthy versus diseased individuals was investigated by applying Principal Coordinates Analysis (PCoA) on Manhattan distances generated on the DESeq2 normalized count file using the EMPeror (60) tool. For a detailed protocol, see Supplementary Materials and Methods.

Correlation network analysis. The correlation network was constructed using the SparCC algorithm (46) python package (available at https://bitbucket.org/yonatanf/sparcc) to represent both co-abundance and co-exclusion networks between species and corresponding BGCs. For a detailed protocol, see Supplementary Materials and Methods.

Experimental small molecule metabolites detection. Approximately 150µl of saliva was lyophilized and ethyl acetate was added to extract non-polar molecules. Samples were then vortexed, centrifuged to remove the cell debris and submitted to untargetd LC-MS/MS analysis. For a detailed protocol, see Supplementary Materials and Methods.

Mass spectral molecular networking. LC-MS/MS spectra were preprocessed for feature extraction using MZmine2 (61) and submitted to mass spectral molecular networking through GNPS (43). For a detailed protocol, see Supporting Information.

Putative chemical structure annotation. To putatively annotate chemical structures in our mass spectral molecular networks, we performed *in silico* structure annotation through Network Annotation Propagation (NAP) (57) both for [M+H]+ and [M+Na]+ adducts. For a detailed protocol, see Supporting Information.

Acknowledgements

The authors thank Alexander Richter for providing custom python scripts to support the data analysis and thank Ricardo R. da Silva for supporting mass spectral analysis. This research was funded by NIH/NIDCR grants R00-DE024543-3 (A.E.), K99-DE024543 (A.E.), and F32-DE026947 (J.L.B.).

Author contributions

G.A., X.T., C.Z. and A.E. designed the experiments, G.A., A.V.M., R.A., M.B.D., N.C.T. collected and processed the samples, G.A., X.T., C.Z. A.V.M. M.E. and A.E. analyzed the data, and G.A., J.L.B and A.E. wrote the manuscript. All authors helped edit the manuscript. All authors read and approved the final manuscript.

Additional information

Correspondence and requests for materials should be addressed to A.E.

Competing interests

The authors declare that they have no competing interests.

Availability of data and material

Sequence data has been submitted to NCBI under BioProject ID PRJNA478018 with SRA accession SRP151559. Mass spectral files, LCMS/MS metadata file, Nucleotide FASTA sequences of the oral biosynthetic gene cluster collection are accessible from the MassIVE repository (https://massive.ucsd.edu/) with the accession ID MSV000081832.

REFERENCES

1. Human Microbiome Project C. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207-14.
2. Donia MS, Cimermancic P, Schulze CJ, Wieland Brown LC, Martin J, Mitreva M, Clardy J, Linington RG, Fischbach MA. 2014. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. Cell 158:1402-1414.
3. Donia MS, Fischbach MA. 2015. HUMAN MICROBIOTA. Small molecules from the human microbiota. Science 349:1254766.
4. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, Birren BW, Takano E, Sali A, Linington RG, Fischbach MA. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158:412-421.
5. Zipperer A, Konnerth MC, Laux C, Berscheid A, Janek D, Weidenmaier C, Burian M, Schilling NA, Slavetinsky C, Marschal M, Willmann M, Kalbacher H, Schittek B, Brotz-Oesterhelt H, Grond S, Peschel A, Krismer B. 2016. Human commensals producing a novel antibiotic impair pathogen colonization. Nature 535:511-6.
6. Cohen LJ, Esterhazy D, Kim SH, Lemetre C, Aguilar RR, Gordon EA, Pickard AJ, Cross JR, Emiliano AB, Han SM, Chu J, Vila-Farres X, Kaplitt J, Rogoz A, Calle PY, Hunter C, Bitok JK, Brady SF. 2017. Commensal bacteria make GPCR ligands that mimic human signalling molecules. Nature 549:48-53.
7. Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly YM, Glickman JN, Garrett WS. 2013. The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. Science 341:569-73.
8. Jakubovics NS. 2015. Intermicrobial Interactions as a Driver for Community Composition and Stratification of Oral Biofilms. J Mol Biol 427:3662-75.
9. Ono K, Oka R, Toyofuku M, Sakaguchi A, Hamada M, Yoshida S, Nomura N. 2014. cAMP signaling affects irreversible attachment during biofilm formation by Pseudomonas aeruginosa PAO1. Microbes Environ 29:104-6.
10. Guo CJ, Chang FY, Wyche TP, Backus KM, Acker TM, Funabashi M, Taketani M, Donia MS, Nayfach S, Pollard KS, Craik CS, Cravatt BF, Clardy J, Voigt CA, Fischbach MA. 2017. Discovery of Reactive Microbiota-Derived Metabolites that Inhibit Host Proteases. Cell 168:517-526 e18.
11. Xia W. 2017. Competition for Iron Between Host and Pathogen: A Structural Case Study on Helicobacter pylori. Methods Mol Biol 1535:65-75.
12. Rea K, Dinan TG, Cryan JF. 2016. The microbiome: A key regulator of stress and neuroinflammation. Neurobiol Stress 4:23-33.
13. Sharon G, Garg N, Debelius J, Knight R, Dorrestein PC, Mazmanian SK. 2014. Specialized metabolites from the microbiome in health and disease. Cell Metab 20:719-730.
14. Mousa WK, Athar B, Merwin NJ, Magarvey NA. 2017. Antibiotics and specialized metabolites from the human microbiota. Nat Prod Rep 34:1302-1331.
15. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A, Wade WG. 2010. The human oral microbiome. J Bacteriol 192:5002-17.

16. Costalonga M, Herzberg MC. 2014. The oral microbiome and the immunobiology of periodontal disease and caries. Immunol Lett 162:22-38.

17. Kilian M, Chapple IL, Hannig M, Marsh PD, Meuric V, Pedersen AM, Tonetti MS, Wade WG, Zaura E. 2016. The oral microbiome - an update for oral healthcare professionals. Br Dent J 221:657-666.

18. Marsh PD. 2003. Are dental diseases examples of ecological catastrophes? Microbiology 149:279-94.

19. Bowen WH. 2002. Do we need to be concerned about dental caries in the coming millennium? Crit Rev Oral Biol Med 13:126-31.

20. Dye BA, Thornton-Evans G. 2010. Trends in oral health by poverty status as measured by Healthy People 2010 objectives. Public Health Rep 125:817-30.

21. Jia G, Zhi A, Lai PFH, Wang G, Xia Y, Xiong Z, Zhang H, Che N, Ai L. 2018. The oral microbiota - a mechanistic role for systemic diseases. Br Dent J 224:447-455.

22. Duran-Pinedo AE, Chen T, Teles R, Starr JR, Wang X, Krishnan K, Frias-Lopez J. 2014. Community-wide transcriptome of the oral microbiome in subjects with and without periodontitis. ISME J 8:1659-72.

23. Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, Romero H, Simon-Soro A, Pignatelli M, Mira A. 2012. The oral metagenome in health and disease. ISME J 6:46-56.

24. Belstrom D, Constancias F, Liu Y, Yang L, Drautz-Moses DI, Schuster SC, Kohli GS, Jakobsen TH, Holmstrup P, Givskov M. 2017. Metagenomic and metatranscriptomic analysis of saliva reveals disease-associated microbiota in patients with periodontitis and dental caries. NPJ Biofilms Microbiomes 3:23.

25. Do T, Sheehy EC, Mulli T, Hughes F, Beighton D. 2015. Transcriptomic analysis of three Veillonella spp. present in carious dentine and in the saliva of caries-free individuals. Front Cell Infect Microbiol 5:25.

26. Peterson SN, Meissner T, Su AI, Snesrud E, Ong AC, Schork NJ, Bretz WA. 2014. Functional expression of dental plaque microbiota. Front Cell Infect Microbiol 4:108.

27. Yost S, Duran-Pinedo AE, Teles R, Krishnan K, Frias-Lopez J. 2015. Functional signatures of oral dysbiosis during periodontitis progression revealed by microbial metatranscriptome analysis. Genome Med 7:27.

28. Wang J, Qi J, Zhao H, He S, Zhang Y, Wei S, Zhao F. 2013. Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. Sci Rep 3:1843.

29. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res 45:W36-W41.

30. Medema MH, Takano E, Breitling R. 2013. Detecting sequence homology at the gene cluster level with MultiGeneBlast. Mol Biol Evol 30:1218-23.

31. Liu L, Hao T, Xie Z, Horsman GP, Chen Y. 2016. Genome mining unveils widespread natural product biosynthetic capacity in human oral microbe Streptococcus mutans. Sci Rep 6:37479.

32. Doering TL. 2009. How sweet it is! Cell wall biogenesis and polysaccharide capsule formation in Cryptococcus neoformans. Annu Rev Microbiol 63:223-47.

33. Krachler AM, Orth K. 2013. Targeting the bacteria-host interface: strategies in anti-adhesion therapy. Virulence 4:284-94.
34. Lee BC, Kim MS, Choi SH, Kim KY, Kim TS. 2009. In vitro and in vivo antimicrobial activity of water-soluble chitosan oligosaccharides against Vibrio vulnificus. Int J Mol Med 24:327-33.
35. Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL. 2005. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. Cell 122:107-18.
36. Mazmanian SK, Round JL, Kasper DL. 2008. A microbial symbiosis factor prevents intestinal inflammatory disease. Nature 453:620-5.
37. van der Ploeg JR. 2005. Regulation of bacteriocin production in Streptococcus mutans by the quorum-sensing system required for development of genetic competence. J Bacteriol 187:3980-9.
38. Kreth J, Merritt J, Zhu L, Shi W, Qi F. 2006. Cell density- and ComE-dependent expression of a group of mutacin and mutacin-like genes in Streptococcus mutans. FEMS Microbiol Lett 265:11-7.
39. Shanker E, Federle MJ. 2017. Quorum Sensing Regulation of Competence and Bacteriocins in Streptococcus pneumoniae and mutans. Genes (Basel) 8.
40. Schoner TA, Gassel S, Osawa A, Tobias NJ, Okuno Y, Sakakibara Y, Shindo K, Sandmann G, Bode HB. 2016. Aryl Polyenes, a Highly Abundant Class of Bacterial Natural Products, Are Functionally Related to Antioxidative Carotenoids. Chembiochem 17:247-53.
41. Fischbach MA, Walsh CT. 2006. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. Chem Rev 106:3468-96.
42. de las Heras B, Hortelano S. 2009. Molecular basis of the anti-inflammatory effects of terpenoids. Inflamm Allergy Drug Targets 8:28-39.
43. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, Cruz-Morales P, Duddela S, Dusterhus S, Edwards DJ, Fewer DP, Garg N, Geiger C, Gomez-Escribano JP, Greule A, Hadjithomas M, Haines AS, Helfrich EJ, Hillwig ML, Ishida K, Jones AC, Jones CS, Jungmann K, Kegler C, Kim HU, Kotter P, Krug D, Masschelein J, Melnik AV, Mantovani SM, Monroe EA, Moore M, Moss N, Nutzmann HW, Pan G, Pati A, Petras D, Reen FJ, Rosconi F, Rui Z, Tian Z, Tobias NJ, Tsunematsu Y, Wiemann P, Wyckoff E, Yan X, et al. 2015. Minimum Information about a Biosynthetic Gene cluster. Nat Chem Biol 11:625-31.
44. Ma C, Chen F, Zhang Y, Sun X, Tong P, Si Y, Zheng S. 2015. Comparison of oral microbial profiles between children with severe early childhood caries and caries-free children using the human oral microbe identification microarray. PLoS One 10:e0122075.
45. Aas JA, Griffen AL, Dardis SR, Lee AM, Olsen I, Dewhirst FE, Leys EJ, Paster BJ. 2008. Bacteria of dental caries in primary and permanent teeth in children and young adults. J Clin Microbiol 46:1407-17.
46. Friedman J, Alm EJ. 2012. Inferring correlation networks from genomic survey data. PLoS Comput Biol 8:e1002687.

47. Muller S, Strack SN, Hoefler BC, Straight PD, Kearns DB, Kirby JR. 2014. Bacillaene and sporulation protect Bacillus subtilis from predation by Myxococcus xanthus. Appl Environ Microbiol 80:5603-10.
48. Cooper EC, Curtis N, Cranswick N, Gwee A. 2014. Pristinamycin: old drug, new tricks? J Antimicrob Chemother 69:2319-25.
49. Simon-Soro A, Mira A. 2015. Solving the etiology of dental caries. Trends Microbiol 23:76-82.
50. Ono T, Shiota S, Hirota K, Nemoto K, Tsuchiya T, Miyake Y. 2000. Susceptibilities of oral and nasal isolates of Streptococcus mitis and Streptococcus oralis to macrolides and PCR detection of resistance genes. Antimicrob Agents Chemother 44:1078-80.
51. Healy AR, Herzon SB. 2017. Molecular Basis of Gut Microbiome-Associated Colorectal Cancer: A Synthetic Perspective. J Am Chem Soc 139:14817-14824.
52. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crusemann M, Boudreau PD, Esquenazi E, Sandoval-Calderon M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, et al. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol 34:828-837.
53. Djoumbou Feunang Y, Eisner R, Knox C, Chepelev L, Hastings J, Owen G, Fahy E, Steinbeck C, Subramanian S, Bolton E, Greiner R, Wishart DS. 2016. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. J Cheminform 8:61.
54. Ruttkies C, Schymanski EL, Wolf S, Hollender J, Neumann S. 2016. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J Cheminform 8:3.
55. Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. 2010. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. Database (Oxford) 2010:baq013.
56. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR. 2007. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 3:211-221.
57. da Silva RR, Wang M, Nothias LF, van der Hooft JJJ, Caraballo-Rodriguez AM, Fox E, Balunas MJ, Klassen JL, Lopes NP, Dorrestein PC. 2018. Propagating annotations of molecular networks using in silico fragmentation. PLoS Comput Biol 14:e1006089.
58. Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M. 2014. Metatranscriptomics of the human oral microbiome during health and disease. MBio 5:e01012-14.
59. Shi B, Chang M, Martin J, Mitreva M, Lux R, Klokkevold P, Sodergren E, Weinstock GM, Haake SK, Li H. 2015. Dynamic changes in the subgingival microbiome and their potential for diagnosis and prognosis of periodontitis. MBio 6:e01926-14.

60. Vazquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. 2013. EMPeror: a tool for visualizing high-throughput microbial community data. Gigascience 2:16.

61. Pluskal T, Castillo S, Villar-Briones A, Oresic M. 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11:395.

Table legends

Table 1 Number of differentially represented or expressed biosynthetic pathways in saliva, supra- and sub-gingival plaque samples, from shotgun metatranscriptomics and metagenomics libraries representing 231 subjects (oral health: n = 110, dental caries: n = 77, periodontitis: n = 44). 741 BGCs were differentially abundant in caries (515 enriched, and 226 less abundant) and 1,063 BGCs were periodontitis associated (670 enriched, and 393 less abundant). 355 BGCs were differentially expressed in caries (208 up-regulated, and 147 down-regulated), while 421 BGCs were either up- or down-regulated in subjects with periodontitis (218 up-regulated, and 203 down-regulated).

	Metatranscriptome (up/down)		**Metagenome (up/down)**	
BGC type	**Caries**	**Periodontitis**	**Caries**	**Periodontitis**
Aryl polyene	4/11	29/4	26/5	25/3
Bacteriocin	14/10	1/1	31/1	5/24
Butyrolactone	2/0	1/1	3/0	1/1
Homoserine lactone	0/4	3/4	10/0	1/1
Lantipeptide	7/0	4/0	11/2	1/6
Lassopeptide	0	0	0	0/3
NRPS	7/0	2/8	11/7	10/14
NRPS-PKS hybrid	1/0	0/1	5/0	0/1
Oligosaccharide	66/68	100/40	116/118	287/132
Other	7/0	2/11	36/3	16/15
Phenazine	0	0/1	2/0	0/1
PKS	3/0	1/0	8/0	1/4
Proteusin	0	0	1/0	1/0
Resorcinol	1/4	3/8	4/4	19/0
Sactipeptide	0	0	0	1/0
Siderophore	1/0	1/2	0	1/0
Terpene	1/10	9/1	42/0	1/4
Thiopeptide	1/0	0	0	0/7
Unclassified	93/50	62/121	209/86	300/177

Figure legends

FIG 1: The oral microbiome contains a massive diversity of BGCs encoded by a multitude of taxa. (A) Bar graph illustrating the most common BGC subtypes identified in this study. Bars are colored according to higher level BGC class. **(B)** Bar graph illustrating the distribution of eight major classes of BGCs by phyla. **(C)** Phylogenetic tree based on 16S rRNA gene sequences showing the distribution of BGCs encoded by oral bacteria. Nodes with bootstrap values higher than 80% are displayed in the tree. Numbers of BGC types identified within each genome are shown in the bar graph and colored by BGC type. Leaf labels are colored by phyla. antiSMASH often identifies BGCs that encompass multiple gene clusters of different types fused into a single large gene cluster. 63 (~3%) of such unresolved BGCs and were encountered, and were categorized as the 'complex' BGC type (For convenience, we combined these BGCs with BGC types 'Other' for subsequent analysis). Distribution of BGCs is presented in more detail in Fig. S1.

FIG 2: Similarity networking identified putative product classes for novel BGCs. Similarity network between the BGCs identified in the oral cavity and the experimentally characterized reference BGCs obtained from the MIBiG repository. Sub-networks representing major BGC classes, as determined by antiSMASH and BiG-SCAPE, are highlighted with different background colors to visualize BGCs as constellations within the biosynthetic landscape. Nodes (small circles) represent amino acid sequences of BGC domains and are colored by BGC class. Unfilled nodes represent reference BGCs from the MIBiG repository. Edges drawn between the nodes correspond to pairwise distances, computed by BiG-SCAPE as the weighted combination of the Jaccard, adjacency and

domain sequence similarity indices. For increased simplicity, only sub-clusters of unclassified and oligosaccharide BGCs with a minimum number of eight nodes are organized into given highlighted constellation.

FIG 3: BGCs are differentially expressed in health and disease. Bar graphs illustrating phylogenetic distribution of biosynthetic pathways in health- and disease-associated oral microorganisms. Species with significant changes in BGC expression based on the analyzed metatranscriptomic data sets are shown in the phylogenetic tree on the left. Bar graphs at the leaf tips display number of BGCs either over or under expressed and colored according to the BGC type. It should be noted that the x-axis scales are different in left and right panels. Significant differences in the expression of BGCs were determined based on negative binomial distribution model using DESeq2 with FDR correction (p-value < 0.05).

Figure 1

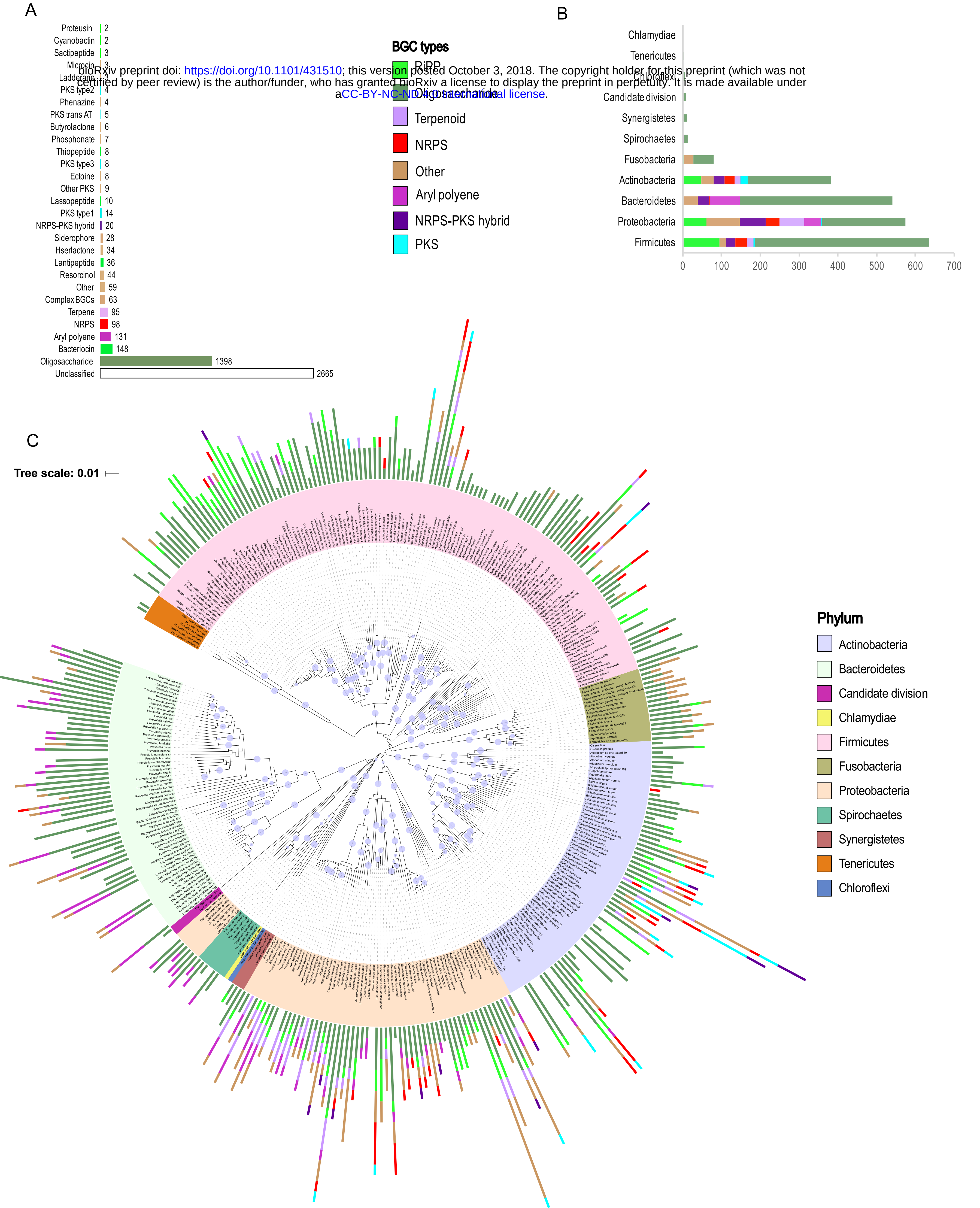

A
Proteusin 2
Cyanobactin 2
Sactipeptide 3
Microcin 3
Ladderane 3
PKS type2 4
Phenazine 4
PKS trans AT 5
Butyrolactone 6
Phosphonate 7
Thiopeptide 8
PKS type3 8
Ectoine 8
Other PKS 9
Lassopeptide 10
PKS type1 14
NRPS-PKS hybrid 20
Siderophore 28
Hserlactone 34
Lantipeptide 36
Resorcinol 44
Other 59
Complex BGCs 63
Terpene 95
NRPS 98
Aryl polyene 131
Bacteriocin 148
Oligosaccharide 1398
Unclassified 2665
BGC types
RiPP
Oligosaccharide
Terpenoid
NRPS
Other
Aryl polyene
NRPS-PKS hybrid
PKS
B
Chlamydiae
Tenericutes
Chloroflexi
Candidate division
Synergistetes
Spirochaetes
Fusobacteria
Actinobacteria
Bacteroidetes
Proteobacteria
Firmicutes
0
100
200
300
400
500
600
700
C
Tree scale: 0.01
Phylum
Actinobacteria
Bacteroidetes
Candidate division
Chlamydiae
Firmicutes
Fusobacteria
Proteobacteria
Spirochaetes
Synergistetes
Tenericutes
Chloroflexi

Figure 2

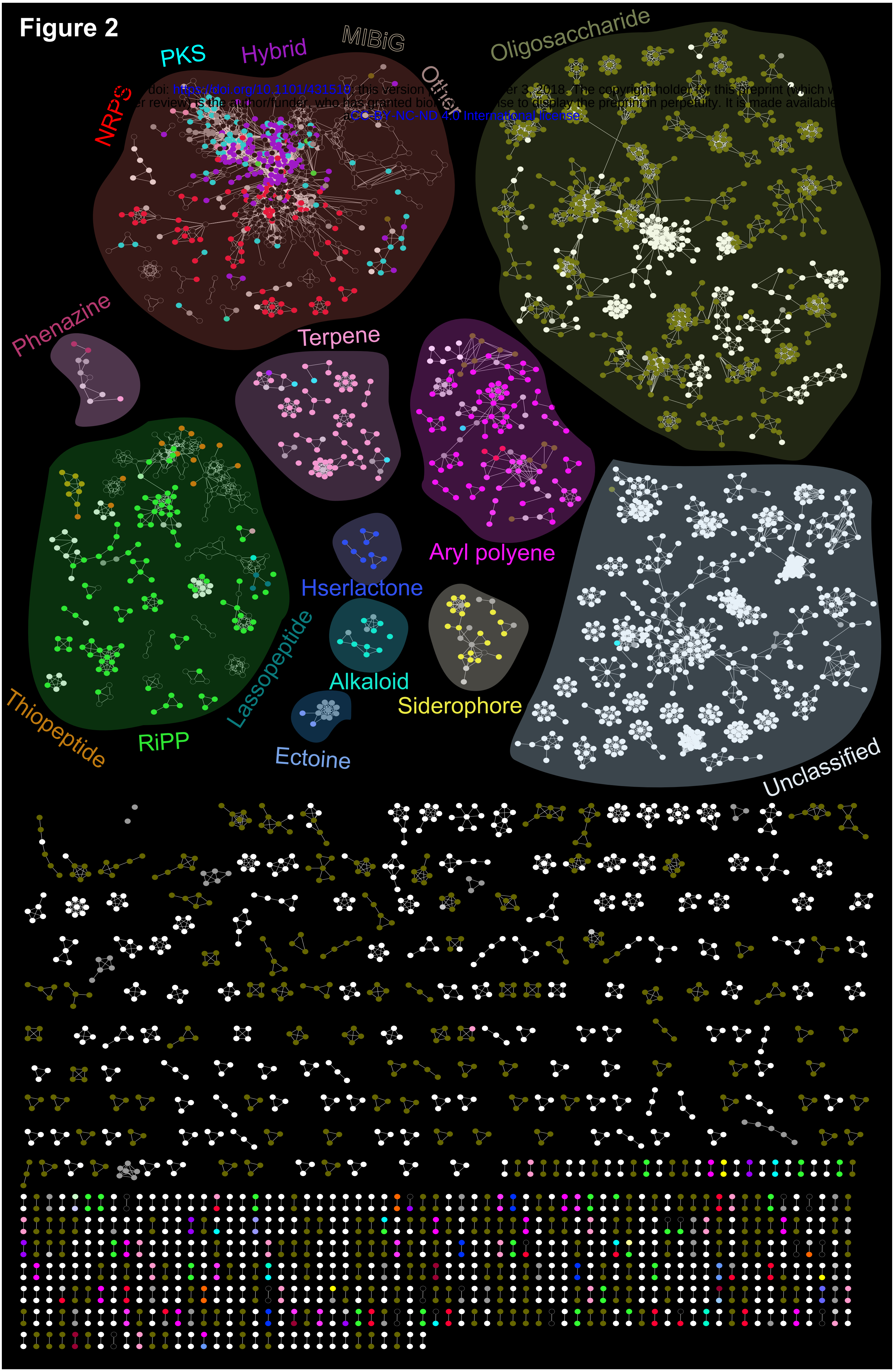

PKS
Hybrid
MIBiG
Oligosaccharide
NRPS
Other
Phenazine
Terpene
Aryl polyene
Hserlactone
Lassopeptide
Alkaloid
Siderophore
Thiopeptide
RiPP
Ectoine
Unclassified

Figure 3

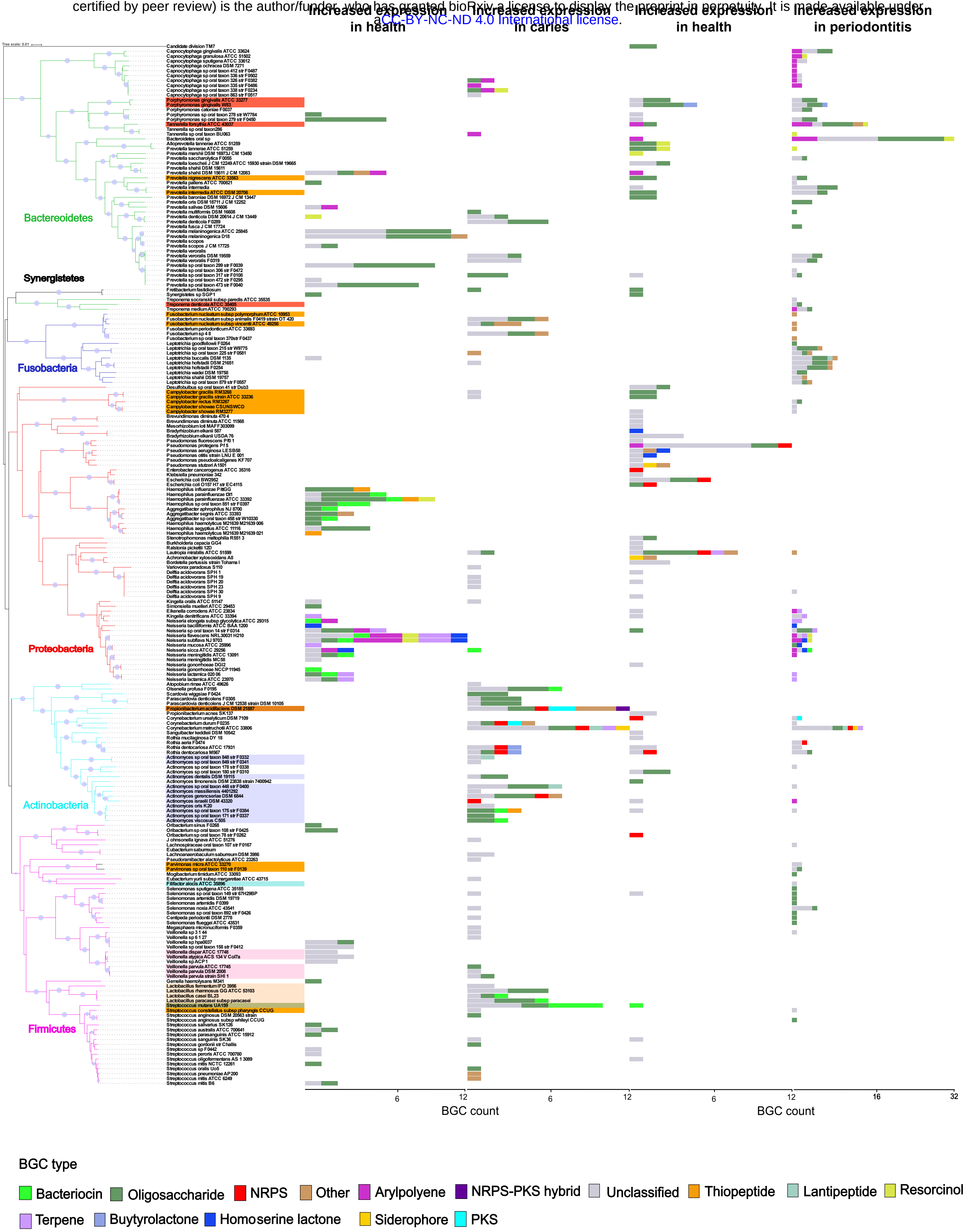
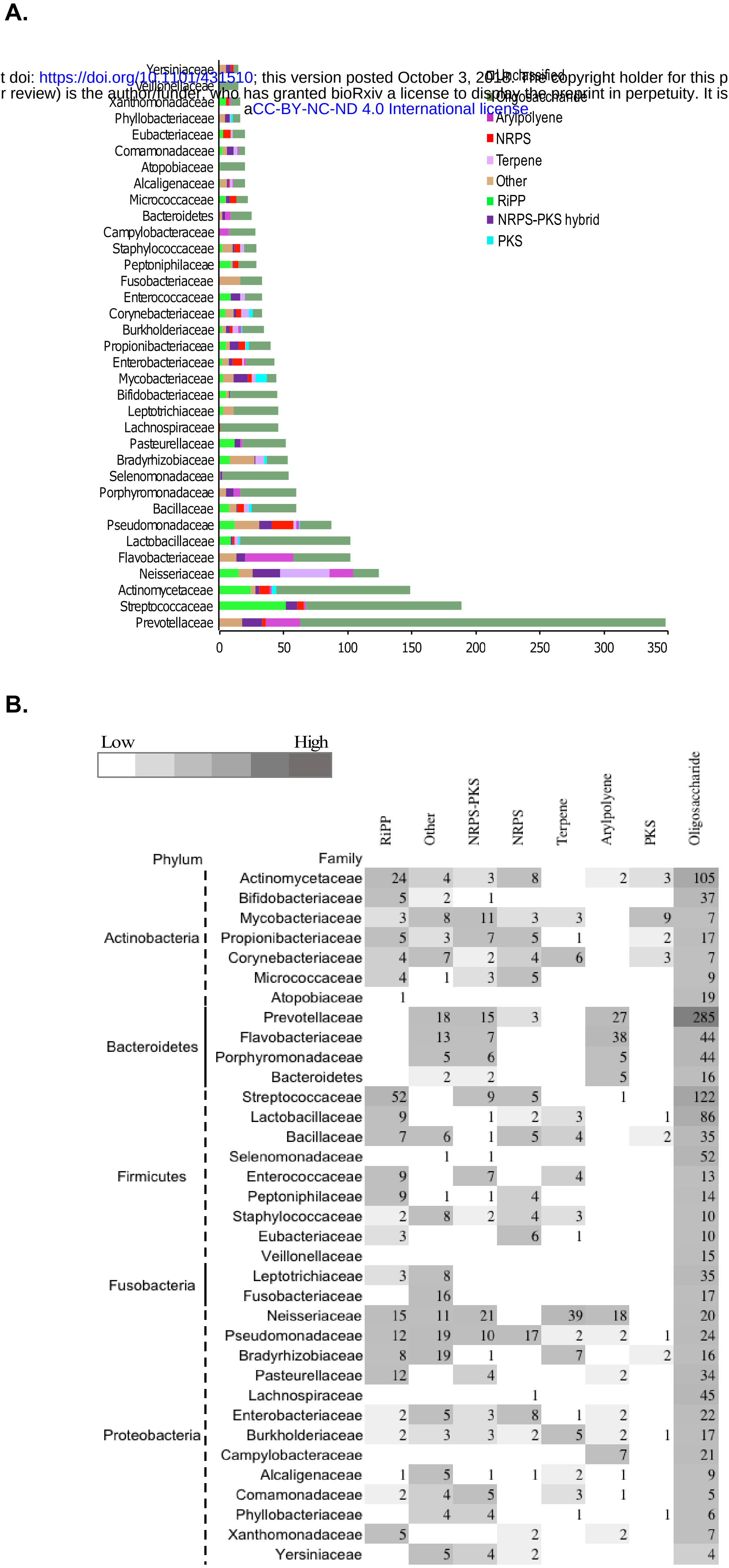
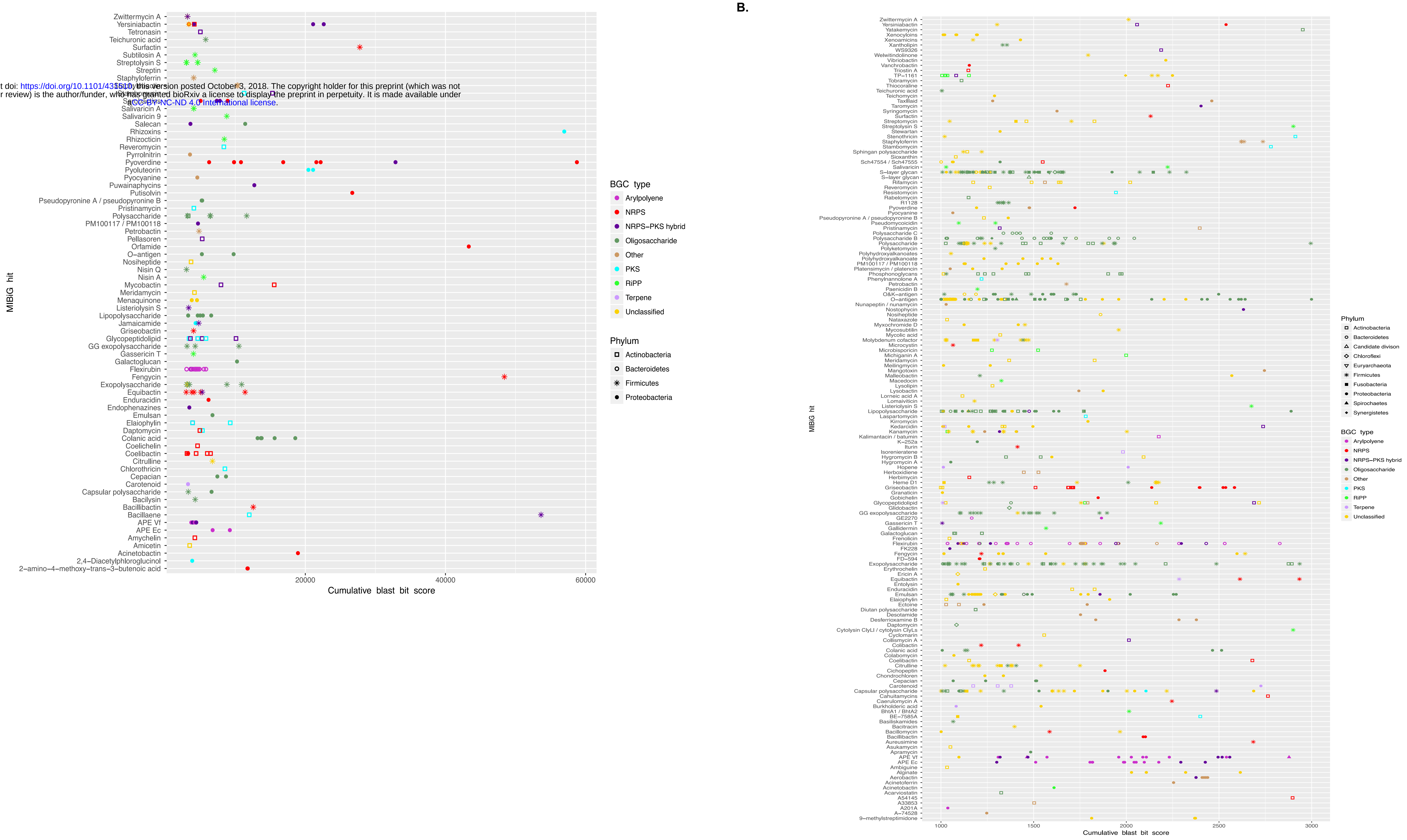

Increased expression in health
Increased expression in caries
Increased expression in health
Increased expression in periodontitis
Tree scale: 0.01
Bactereoidetes
Synergistetes
Fusobacteria
Proteobacteria
Actinobacteria
Firmicutes
Candidate division TM7
Capnocytophaga gingivalis ATCC 33624
Capnocytophaga granulosa ATCC 51502
Capnocytophaga sputigena ATCC 33612
Capnocytophaga ochracea DSM 7271
Capnocytophaga sp oral taxon 412 str F0487
Capnocytophaga sp oral taxon 336 str F0502
Capnocytophaga sp oral taxon 326 str F0382
Capnocytophaga sp oral taxon 335 str F0486
Capnocytophaga sp oral taxon 338 str F0234
Capnocytophaga sp oral taxon 863 str F0517
Porphyromonas gingivalis ATCC 33277
Porphyromonas gingivalis W83
Porphyromonas catoniae F0037
Porphyromonas sp oral taxon 278 str W7784
Porphyromonas sp oral taxon 279 str F0450
Tannerella forsythia ATCC 43037
Tannerella sp oral taxon286
Tannerella sp oral taxon BU063
Bacteroidetes oral sp
Alloprevotella tannerae ATCC 51259
Prevotella tannerae ATCC 51259
Prevotella marshii DSM 16973J CM 13450
Prevotella saccharolytica F0055
Prevotella loescheii J CM 12249 ATCC 15930 strain DSM 19665
Prevotella shahii DSM 15611
Prevotella shahii DSM 15611 J CM 12083
Prevotella nigrescens ATCC 33563
Prevotella pallens ATCC 700821
Prevotella intermedia
Prevotella intermedia ATCC DSM 20706
Prevotella baroniae DSM 16972 J CM 13447
Prevotella oris DSM 18711 J CM 12252
Prevotella salivae DSM 15606
Prevotella multiformis DSM 16608
Prevotella denticola DSM 20614 J CM 13449
Prevotella denticola F0289
Prevotella fusca J CM 17724
Prevotella melaninogenica ATCC 25845
Prevotella melaninogenica D18
Prevotella scopos
Prevotella scopos J CM 17725
Prevotella veroralis
Prevotella veroralis DSM 19559
Prevotella veroralis F0319
Prevotella sp oral taxon 299 str F0039
Prevotella sp oral taxon 306 str F0472
Prevotella sp oral taxon 317 str F0108
Prevotella sp oral taxon 472 str F0295
Prevotella sp oral taxon 473 str F0040
Fretibacterium fastidiosum
Synergistetes sp SGP1
Treponema socranskii subsp paredis ATCC 35535
Treponema denticola ATCC 35405
Treponema medium ATCC 700293
Fusobacterium nucleatum subsp polymorphum ATCC 10953
Fusobacterium nucleatum subsp animalis F0419 strain OT 420
Fusobacterium nucleatum subsp vincentii ATCC 49256
Fusobacterium periodonticum ATCC 33693
Fusobacterium sp 4 8
Fusobacterium sp oral taxon 370str F0437
Leptotrichia goodfellowii F0264
Leptotrichia sp oral taxon 215 str W9775
Leptotrichia sp oral taxon 225 str F0581
Leptotrichia buccalis DSM 1135
Leptotrichia hofstadii DSM 21651
Leptotrichia hofstadii F0254
Leptotrichia wadei DSM 19758
Leptotrichia shahii DSM 19757
Leptotrichia sp oral taxon 879 str F0557
Desulfobulbus sp oral taxon 41 str Dsb3
Campylobacter gracilis RM3268
Campylobacter gracilis strain ATCC 33236
Campylobacter rectus RM3267
Campylobacter showae CSUNSWCD
Campylobacter showae RM3277
Brevundimonas diminuta 470 4
Brevundimonas diminuta ATCC 11568
Mesorhizobium loti MAFF303099
Bradyrhizobium elkanii 587
Bradyrhizobium elkanii USDA 76
Pseudomonas fluorescens Pf0 1
Pseudomonas protegens Pf 5
Pseudomonas aeruginosa LESB58
Pseudomonas otitis strain LNU E 001
Pseudomonas pseudoalcaligenes KF707
Pseudomonas stutzeri A1501
Enterobacter cancerogenus ATCC 35316
Klebsiella pneumoniae 342
Escherichia coli BW2952
Escherichia coli O157 H7 str EC4115
Haemophilus influenzae PittGG
Haemophilus parainfluenzae T3T1
Haemophilus parainfluenzae ATCC 33392
Haemophilus sp oral taxon 851 str F0397
Aggregatibacter aphrophilus NJ 8700
Aggregatibacter segnis ATCC 33393
Aggregatibacter sp oral taxon 458 str W10330
Haemophilus haemolyticus M21639 006
Haemophilus aegyptius ATCC 11116
Haemophilus haemolyticus M21639 021
Stenotrophomonas maltophilia R551 3
Burkholderia cepacia GG4
Ralstonia pickettii 12D
Lautropia mirabilis ATCC 51599
Achromobacter xylosoxidans A8
Bordetella pertussis strain Tohama I
Variovorax paradoxus S110
Delftia acidovorans SPH 1
Delftia acidovorans SPH 19
Delftia acidovorans SPH 20
Delftia acidovorans SPH 23
Delftia acidovorans SPH 30
Delftia acidovorans SPH 9
Kingella oralis ATCC 51147
Simonsiella muelleri ATCC 29453
Eikenella corrodens ATCC 23834
Kingella denitrificans ATCC 33394
Neisseria elongata subsp glycolytica ATCC 29315
Neisseria bacilliformis ATCC BAA 1200
Neisseria sp oral taxon 14 str F0314
Neisseria flavescens NRL30031 H210
Neisseria subflava NJ 9703
Neisseria mucosa ATCC 25996
Neisseria sicca ATCC 29256
Neisseria meningitidis ATCC 13091
Neisseria meningitidis MC58
Neisseria gonorrhoeae DGI2
Neisseria gonorrhoeae NCCP 11945
Neisseria lactamica 020 06
Neisseria lactamica ATCC 23970
Atopobium rimae ATCC 49626
Olsenella profusa F0195
Scardovia wiggsiae F0424
Parascardovia denticolens F0305
Parascardovia denticolens J CM 12538 strain DSM 10105
Propionibacterium acidifaciens DSM 21887
Propionibacterium acnes SK137
Corynebacterium urealyticum DSM 7109
Corynebacterium durum F0235
Corynebacterium matruchotii ATCC 33806
Sanguibacter keddieii DSM 10542
Rothia mucilaginosa DY 18
Rothia aeria F0474
Rothia dentocariosa ATCC 17931
Rothia dentocariosa M567
Actinomyces sp oral taxon 848 str F0332
Actinomyces sp oral taxon 849 str F0341
Actinomyces sp oral taxon 178 str F0338
Actinomyces sp oral taxon 180 str F0310
Actinomyces dentalis DSM 19115
Actinomyces timonensis DSM 23838 strain 7400942
Actinomyces sp oral taxon 448 str F0400
Actinomyces massiliensis 4401292
Actinomyces gerencseriae DSM 6844
Actinomyces israelii DSM 43320
Actinomyces oris K20
Actinomyces sp oral taxon 175 str F0384
Actinomyces sp oral taxon 171 str F0337
Actinomyces viscosus C505
Oribacterium sinus F0268
Oribacterium sp oral taxon 108 str F0425
Oribacterium sp oral taxon 78 str F0262
J ohnsonella ignava ATCC 51276
Lachnospiraceae oral taxon 107 str F0167
Eubacterium saburreum
Lachnoanaerobaculum saburreum DSM 3986
Pseudoramibacter alactolyticus ATCC 23263
Parvimonas micra ATCC 33270
Parvimonas sp oral taxon 110 str F0139
Mogibacterium timidum ATCC 33093
Eubacterium yurii subsp margaretiae ATCC 43715
Filifactor alocis ATCC 35896
Selenomonas sputigena ATCC 35185
Selenomonas sp oral taxon 149 str 67H29BP
Selenomonas artemidis DSM 19719
Selenomonas artemidis F0399
Selenomonas noxia ATCC 43541
Selenomonas sp oral taxon 892 str F0426
Centipeda periodontii DSM 2778
Selenomonas flueggei ATCC 43531
Megasphaera micronuciformis F0359
Veillonella sp 3 1 44
Veillonella sp 6 1 27
Veillonella sp hpa0037
Veillonella sp oral taxon 158 str F0412
Veillonella dispar ATCC 17748
Veillonella atypica ACS 134 V Col7a
Veillonella sp ACP1
Veillonella parvula ATCC 17745
Veillonella parvula DSM 2008
Veillonella parvula strain SHI 1
Gemella haemolysans M341
Lactobacillus fermentum IFO 3956
Lactobacillus rhamnosus GG ATCC 53103
Lactobacillus casei BL23
Lactobacillus paracasei subsp paracasei
Streptococcus mutans UA159
Streptococcus constellatus subsp pharyngis CCUG
Streptococcus anginosus DSM 20563 strain
Streptococcus anginosus subsp whileyi CCUG
Streptococcus salivarius SK126
Streptococcus australis ATCC 700641
Streptococcus parasanguinis ATCC 15912
Streptococcus sanguinis SK36
Streptococcus gordonii str Challis
Streptococcus sp F0442
Streptococcus peroris ATCC 700780
Streptococcus oligofermentans AS 1 3089
Streptococcus mitis NCTC 12261
Streptococcus oralis Uo5
Streptococcus pneumoniae AP200
Streptococcus mitis ATCC 6249
Streptococcus mitis B6
6
12
6
12
6
12
16
32
BGC count
BGC count
BGC type
🟩 Bacteriocin
🟩 Oligosaccharide
🟥 NRPS
🟫 Other
🟪 Arylpolyene
🟪 NRPS-PKS hybrid
⬜ Unclassified
🟧 Thiopeptide
🟩 Lantipeptide
🟨 Resorcinol
🟪 Terpene
🟦 Buytyrolactone
🟦 Homoserine lactone
🟨 Siderophore
🟦 PKS

Figure S1



Phylum	Family	RiPP	Other	NRPS-PKS	NRPS	Terpene	Arylpolyene	PKS	Oligosaccharide
Actinobacteria	Actinomycetaceae	24	4	3	8		2	3	105
	Bifidobacteriaceae	5	2	1					37
	Mycobacteriaceae	3	8	11	3	3		9	7
	Propionibacteriaceae	5	3	7	5	1		2	17
	Corynebacteriaceae	4	7	2	4	6		3	7
	Micrococcaceae	4	1	3	5				9
	Atopobiaceae	1							19
Bacteroidetes	Prevotellaceae		18	15	3		27		285
	Flavobacteriaceae		13	7			38		44
	Porphyromonadaceae		5	6			5		44
	Bacteroidetes		2	2			5		16
Firmicutes	Streptococcaceae	52		9	5		1		122
	Lactobacillaceae	9		1	2	3		1	86
	Bacillaceae	7	6	1	5	4		2	35
	Selenomonadaceae		1	1					52
	Enterococcaceae	9		7		4			13
	Peptoniphilaceae	9	1	1	4				14
	Staphylococcaceae	2	8	2	4	3			10
	Eubacteriaceae	3			6	1			10
	Veillonellaceae								15
Fusobacteria	Leptotrichiaceae	3	8						35
	Fusobacteriaceae		16						17
Proteobacteria	Neisseriaceae	15	11	21		39	18		20
	Pseudomonadaceae	12	19	10	17	2	2	1	24
	Bradyrhizobiaceae	8	19	1		7		2	16
	Pasteurellaceae	12		4			2		34
	Lachnospiraceae				1				45
	Enterobacteriaceae	2	5	3	8	1	2		22
	Burkholderiaceae	2	3	3	2	5	2	1	17
	Campylobacteraceae						7		21
	Alcaligenaceae	1	5	1	1	2	1		9
	Comamonadaceae	2	4	5		3	1		5
	Phyllobacteriaceae		4	4		1		1	6
	Xanthomonadaceae	5			2		2		7
	Yersiniaceae		5	4	2				4

Figure S2



A.
MIBiG hit
Zwittermycin A
Yersiniabactin
Tetronasin
Teichuronic acid
Surfactin
Subtilosin A
Streptolysin S
Streptin
Staphyloferrin
Salivaricin A
Salivaricin 9
Salecan
Rhizoxins
Rhizocticin
Reveromycin
Pyrrolnitrin
Pyoverdine
Pyoluteorin
Pyocyanine
Puwainaphycins
Putisolvin
Pseudopyronine A / pseudopyronine B
Pristinamycin
Polysaccharide
PM100117 / PM100118
Petrobactin
Pellasoren
Orfamide
O−antigen
Nosiheptide
Nisin Q
Nisin A
Mycobactin
Meridamycin
Menaquinone
Listeriolysin S
Lipopolysaccharide
Jamaicamide
Griseobactin
Glycopeptidolipid
GG exopolysaccharide
Gassericin T
Galactoglucan
Flexirubin
Fengycin
Exopolysaccharide
Equibactin
Enduracidin
Endophenazines
Emulsan
Elaiophylin
Daptomycin
Colanic acid
Coelichelin
Coelibactin
Citrulline
Chlorothricin
Cepacian
Carotenoid
Capsular polysaccharide
Bacilysin
Bacillibactin
Bacillaene
APE Vf
APE Ec
Amychelin
Amicetin
Acinetobactin
2,4−Diacetylphloroglucinol
2−amino−4−methoxy−trans−3−butenoic acid
20000
40000
60000
Cumulative blast bit score
BGC type
Arylpolyene
NRPS
NRPS−PKS hybrid
Oligosaccharide
Other
PKS
RiPP
Terpene
Unclassified
Phylum
Actinobacteria
Bacteroidetes
Firmicutes
Proteobacteria
B.
MIBiG hit
Zwittermycin A
Yersiniabactin
Yatakemycin
Xenocoloins
Xenoamicins
Xantholipin
WS9326
Welwitindolinone
Vibriobactin
Vanchrobactin
Triostin A
TP−1161
Tobramycin
Thiocoraline
Teichuronic acid
Teichomycin
Taxllaid
Taromycin
Syringomycin
Surfactin
Streptomycin
Streptolysin S
Stewartan
Stenothricin
Staphyloferrin
Stambomycin
Sphingan polysaccharide
Sioxanthin
Sch47554 / Sch47555
Salivaricin
S−layer glycan
S−layer glycan
Rifamycin
Reveromycin
Resistomycin
Rabelomycin
R1128
Pyoverdine
Pyocyanine
Pseudopyronine A / pseudopyronine B
Pseudomycoicidin
Pristinamycin
Polysaccharide C
Polysaccharide B
Polysaccharide
Polyketomycin
Polyhydroxyalkanoates
Polyhydroxyalkanoate
PM100117 / PM100118
Platensimycin / platencin
Phosphonoglycans
Phenylnannolone A
Petrobactin
Paenicidin B
O&K−antigen
O−antigen
Nunapeptin / nunamycin
Nostophycin
Nosiheptide
Natasazole
Myxochromide D
Mycosubtilin
Mycolic acid
Molybdenum cofactor
Microcystin
Microbisporicin
Michiganin A
Meridamycin
Meilingmycin
Mangotoxin
Malleobactin
Macedocin
Lysolipin
Lysobactin
Lorneic acid A
Lomaiviticin
Listeriolysin S
Lipopolysaccharide
Laspartomycin
Kirromycin
Kedarcidin
Kanamycin
Kalimantacin / batumin
K−252a
Iturin
Isorenieratene
Hygromycin B
Hygromycin A
Hopene
Herboxidiene
Herbimycin
Heme D1
Griseobactin
Granaticin
Gobichelin
Glycopeptidolipid
Glidobactin
GG exopolysaccharide
GE2270
Gassericin T
Gallidermin
Galactoglucan
Frenolicin
Flexirubin
FK228
Fengycin
FD−594
Exopolysaccharide
Erythrochelin
Ericin A
Equibactin
Entolysin
Enduracidin
Emulsan
Elaiophylin
Ectoine
Diutan polysaccharide
Desotamide
Desferrioxamine B
Daptomycin
Cytolysin CylLl / cytolysin CylLs
Cyclomarin
Collismycin A
Colibactin
Colanic acid
Colabomycin
Coelibactin
Citrulline
Cichopeptin
Chondrochloren
Cepacian
Carotenoid
Capsular polysaccharide
Cahuitamycins
Caerulomycin A
Burkholderic acid
BhtA1 / BhtA2
BE−7585A
Basiliskamides
Bacitracin
Bacillomycin
Bacillibactin
Aureusimine
Asukamycin
Apramycin
APE Vf
APE Ec
Ambiguine
Alginate
Aerobactin
Acinetoferrin
Acinetobactin
Acarviostatin
A54145
A33853
A201A
A−74528
9−methylstreptimidone
1000
1500
2000
2500
3000
Cumulative blast bit score
Phylum
Actinobacteria
Bacteroidetes
Candidate divison
Chloroflexi
Euryarchaeota
Firmicutes
Fusobacteria
Proteobacteria
Spirochaetes
Synergistetes
BGC type
Arylpolyene
NRPS
NRPS−PKS hybrid
Oligosaccharide
Other
PKS
RiPP
Terpene
Unclassified

Figure S3

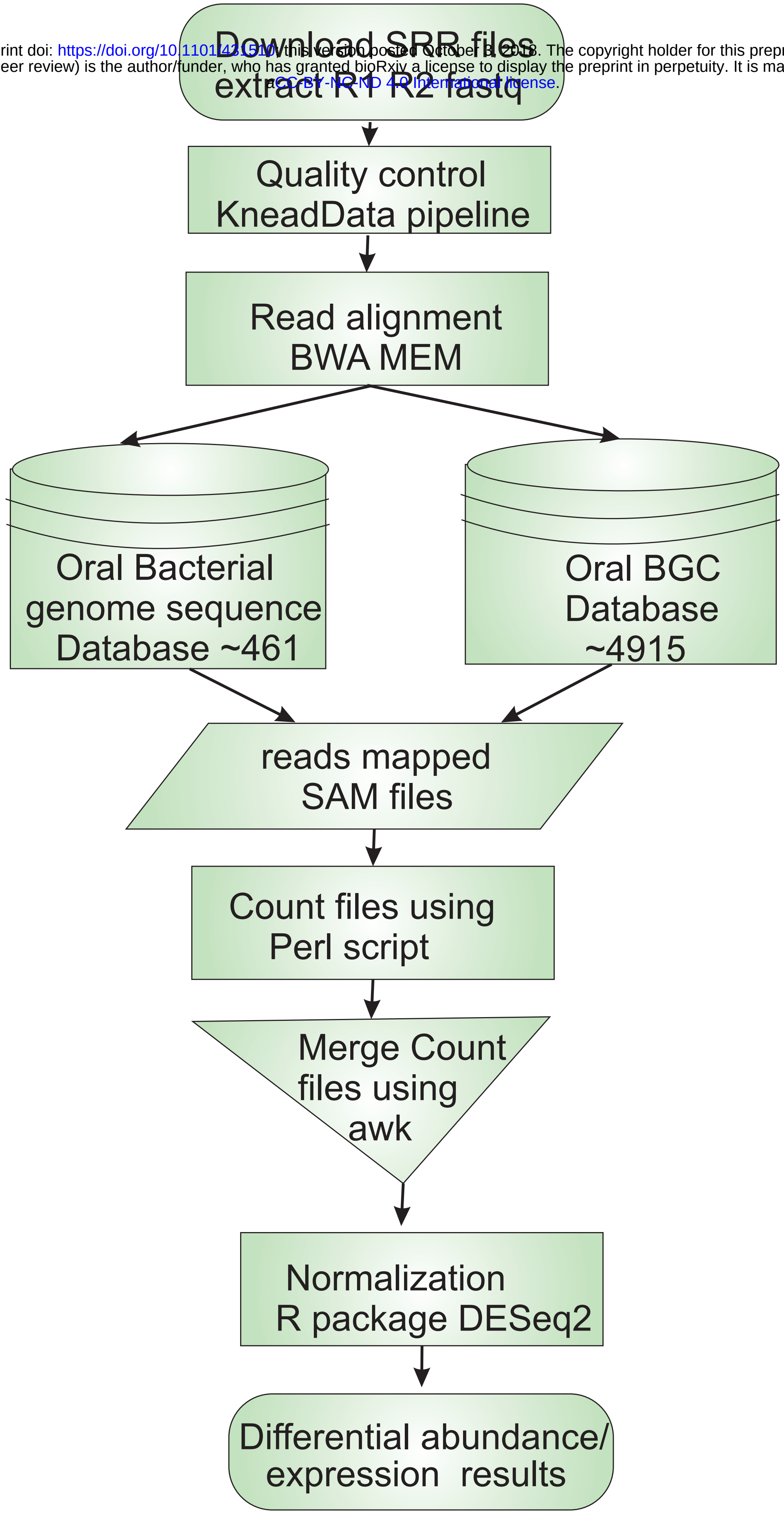

Download SRR files
extract R1 R2 fastq
Quality control
KneadData pipeline
Read alignment
BWA MEM
Oral Bacterial
genome sequence
Database ~461
Oral BGC
Database
~4915
reads mapped
SAM files
Count files using
Perl script
Merge Count
files using
awk
Normalization
R package DESeq2
Differential abundance/
expression results

Figure S4

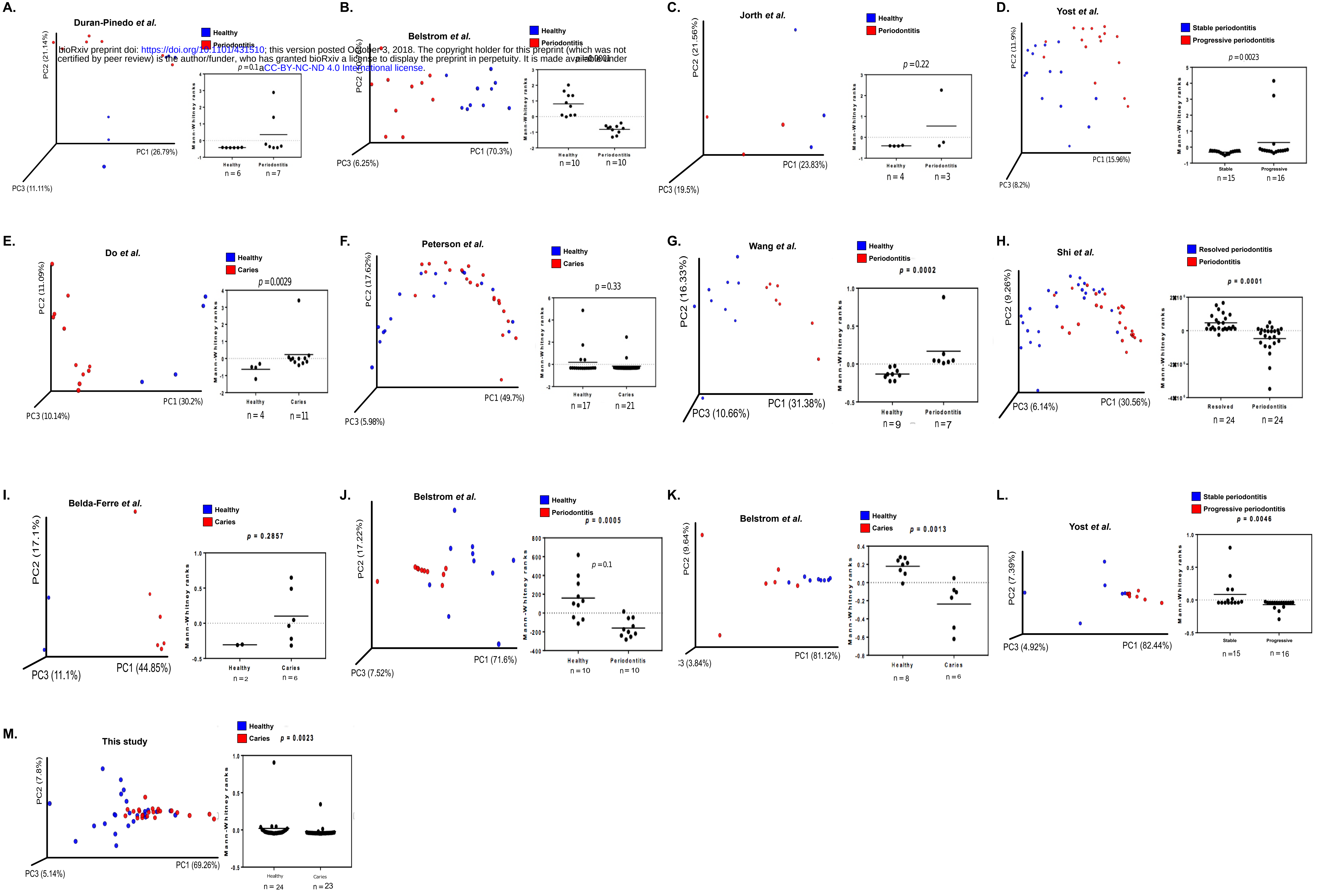

A.
Duran-Pinedo et al.
Healthy
Periodontitis
p = 0.1
PC2 (21.14%)
PC1 (26.79%)
PC3 (11.11%)
Mann-Whitney ranks
Healthy
Periodontitis
n = 6
n = 7
B.
Belstrom et al.
Healthy
Periodontitis
p = 0.0001
PC1 (70.3%)
PC3 (6.25%)
n = 10
n = 10
C.
Jorth et al.
Healthy
Periodontitis
p = 0.22
PC2 (21.56%)
PC1 (23.83%)
PC3 (19.5%)
n = 4
n = 3
D.
Yost et al.
Stable periodontitis
Progressive periodontitis
p = 0.0023
PC2 (11.9%)
PC1 (15.96%)
PC3 (8.2%)
Stable
Progressive
n = 15
n = 16
E.
Do et al.
Healthy
Caries
p = 0.0029
PC2 (11.09%)
PC1 (30.2%)
PC3 (10.14%)
n = 4
n = 11
F.
Peterson et al.
Healthy
Caries
p = 0.33
PC2 (17.62%)
PC1 (49.7%)
PC3 (5.98%)
n = 17
n = 21
G.
Wang et al.
Healthy
Periodontitis
p = 0.0002
PC2 (16.33%)
PC1 (31.38%)
PC3 (10.66%)
n = 9
n = 7
H.
Shi et al.
Resolved periodontitis
Periodontitis
p = 0.0001
PC2 (9.26%)
PC1 (30.56%)
PC3 (6.14%)
Resolved
Periodontitis
n = 24
n = 24
I.
Belda-Ferre et al.
Healthy
Caries
p = 0.2857
PC2 (17.1%)
PC1 (44.85%)
PC3 (11.1%)
n = 2
n = 6
J.
Belstrom et al.
Healthy
Periodontitis
p = 0.0005
p = 0.1
PC2 (17.22%)
PC1 (71.6%)
PC3 (7.52%)
n = 10
n = 10
K.
Belstrom et al.
Healthy
Caries
p = 0.0013
PC2 (9.64%)
PC1 (81.12%)
PC3 (3.84%)
n = 8
n = 6
L.
Yost et al.
Stable periodontitis
Progressive periodontitis
p = 0.0046
PC2 (7.39%)
PC1 (82.44%)
PC3 (4.92%)
n = 15
n = 16
M.
This study
Healthy
Caries
p = 0.0023
PC2 (7.8%)
PC1 (69.26%)
PC3 (5.14%)
n = 24
n = 23

Figure S5

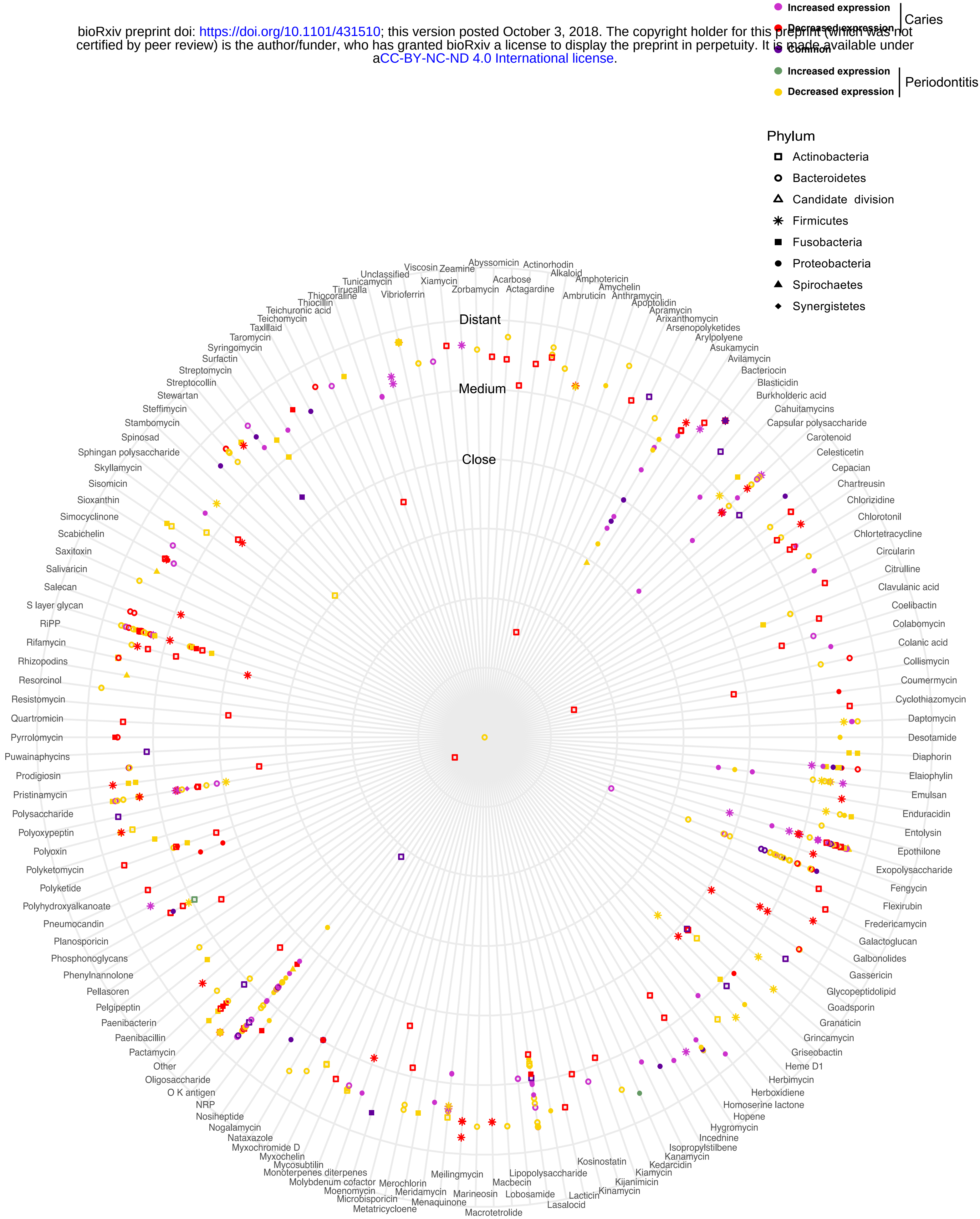

● Increased expression
● Decreased expression
● Common
Caries
● Increased expression
● Decreased expression
Periodontitis
Phylum
□ Actinobacteria
○ Bacteroidetes
△ Candidate division
✳ Firmicutes
■ Fusobacteria
● Proteobacteria
▲ Spirochaetes
◆ Synergistetes
Distant
Medium
Close
Abyssomicin
Actinorhodin
Acarbose
Alkaloid
Actagardine
Amphotericin
Amychelin
Ambruticin
Anthramycin
Apoptolidin
Apramycin
Arixanthomycin
Arsenopolyketides
Arylpolyene
Asukamycin
Avilamycin
Bacteriocin
Blasticidin
Burkholderic acid
Cahuitamycins
Capsular polysaccharide
Carotenoid
Celesticetin
Cepacian
Chartreusin
Chlorizidine
Chlorotonil
Chlortetracycline
Circularin
Citrulline
Clavulanic acid
Coelibactin
Colabomycin
Colanic acid
Collismycin
Coumermycin
Cyclothiazomycin
Daptomycin
Desotamide
Diaphorin
Elaiophylin
Emulsan
Enduracidin
Entolysin
Epothilone
Exopolysaccharide
Fengycin
Flexirubin
Fredericamycin
Galactoglucan
Galbonolides
Gassericin
Glycopeptidolipid
Goadsporin
Granaticin
Grincamycin
Griseobactin
Heme D1
Herbimycin
Herboxidiene
Homoserine lactone
Hopene
Hygromycin
Incednine
Isopropylstilbene
Kanamycin
Kedarcidin
Kiamycin
Kijanimicin
Kinamycin
Kosinostatin
Lacticin
Lasalocid
Lipopolysaccharide
Lobosamide
Macbecin
Marineosin
Macrotetrolide
Menaquinone
Meilingmycin
Meridamycin
Merochlorin
Metatricycloene
Microbisporicin
Moenomycin
Molybdenum cofactor
Monoterpenes diterpenes
Mycosubtilin
Myxochelin
Myxochromide D
Nataxazole
Nogalamycin
Nosiheptide
NRP
O K antigen
Oligosaccharide
Other
Pactamycin
Paenibacillin
Paenibacterin
Pelgipeptin
Pellasoren
Phenylnannolone
Phosphonoglycans
Planosporicin
Pneumocandin
Polyhydroxyalkanoate
Polyketide
Polyketomycin
Polyoxin
Polyoxypeptin
Polysaccharide
Pristinamycin
Prodigiosin
Puwainaphycins
Pyrrolomycin
Quartromicin
Resistomycin
Resorcinol
Rhizopodins
Rifamycin
RiPP
S layer glycan
Salecan
Salivaricin
Saxitoxin
Scabichelin
Simocyclinone
Sioxanthin
Sisomicin
Skyllamycin
Sphingan polysaccharide
Spinosad
Stambomycin
Steffimycin
Stewartan
Streptocollin
Streptomycin
Surfactin
Syringomycin
Taromycin
Taxlllaid
Teichomycin
Teichuronic acid
Thiocillin
Thiocoraline
Tirucalla
Tunicamycin
Unclassified
Vibrioferrin
Viscosin
Xiamycin
Zeamine
Zorbamycin

Figure S6

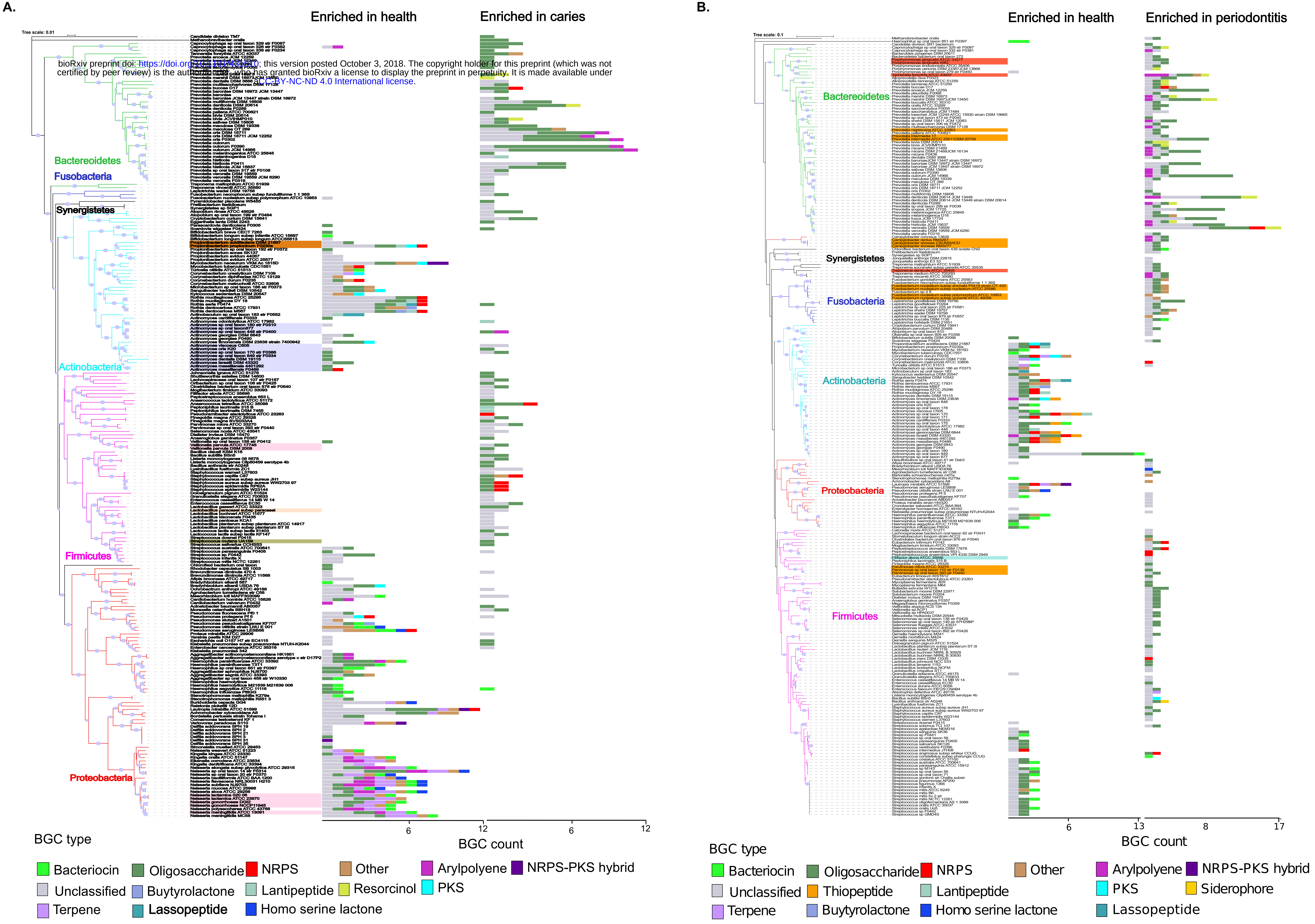

A.
Enriched in health
Enriched in caries
Tree scale: 0.01
Candidate division TM7
Methanobrevibacter oralis
Capnocytophaga sp oral taxon 329 str F0087
Capnocytophaga sp oral taxon 326 str F0382
Capnocytophaga sp oral taxon 338 str F0234
Tannerella forsythia ATCC 43037
Prevotella enoeca JCM 12259
Bactereoidetes
Fusobacteria
Synergistetes
Actinobacteria
Firmicutes
Proteobacteria
0
6
12
6
12
BGC type
BGC count
🟩 Bacteriocin
🟩 Oligosaccharide
🟥 NRPS
🟫 Other
🟪 Arylpolyene
🟪 NRPS-PKS hybrid
⬜ Unclassified
🟦 Buytyrolactone
🟩 Lantipeptide
🟨 Resorcinol
🟦 PKS
🟪 Terpene
🟦 Lassopeptide
🟦 Homo serine lactone
B.
Enriched in health
Enriched in periodontitis
Tree scale: 0.1
Bactereoidetes
Synergistetes
Fusobacteria
Actinobacteria
Proteobacteria
Firmicutes
6
13
8
17
BGC type
BGC count
🟩 Bacteriocin
🟩 Oligosaccharide
🟥 NRPS
🟫 Other
🟪 Arylpolyene
🟪 NRPS-PKS hybrid
⬜ Unclassified
🟧 Thiopeptide
🟩 Lantipeptide
🟦 PKS
🟨 Siderophore
🟪 Terpene
🟦 Buytyrolactone
🟦 Homo serine lactone
🟦 Lassopeptide

Figure S7

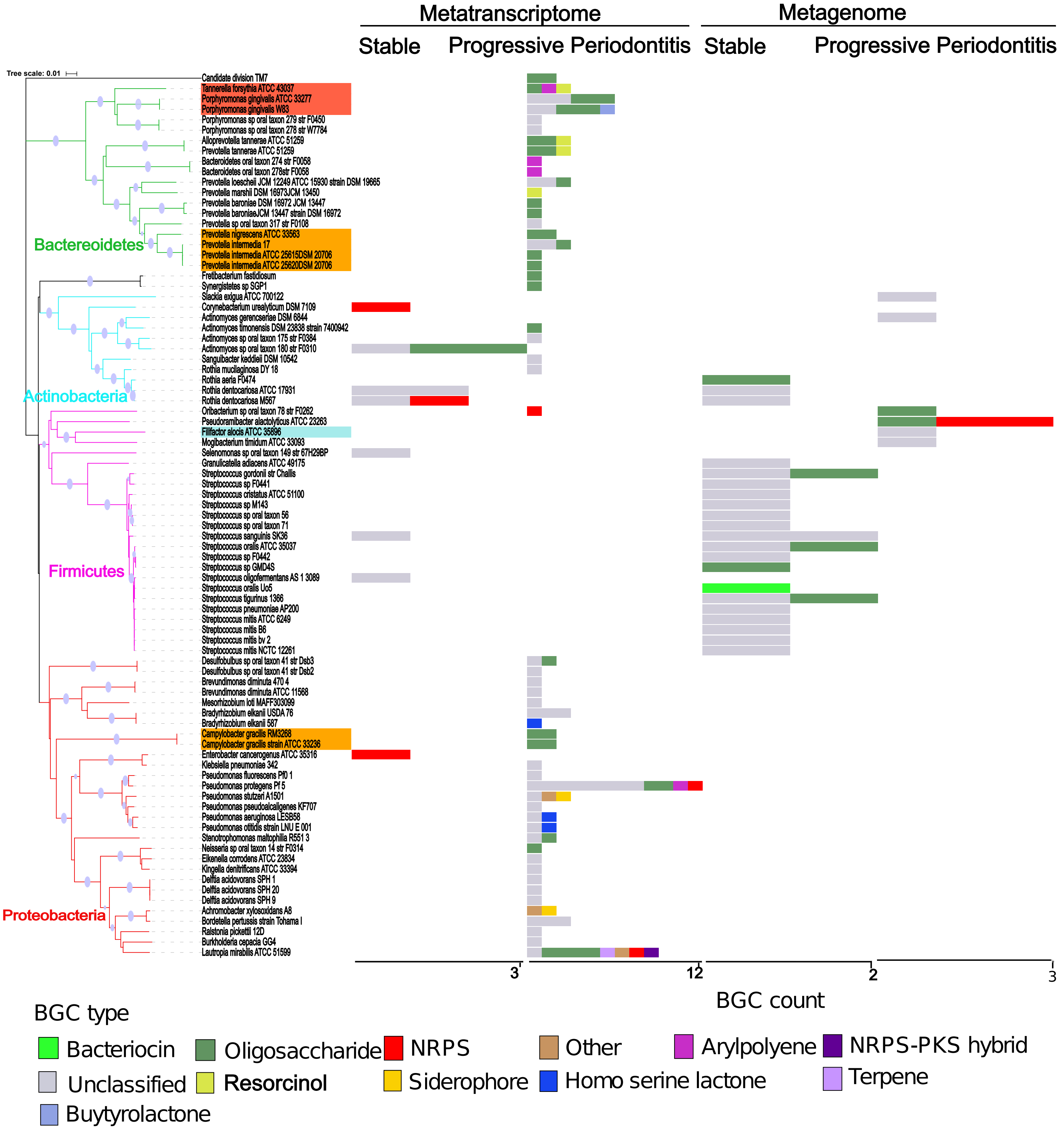

Metatranscriptome
Metagenome
Stable
Progressive Periodontitis
Stable
Progressive Periodontitis
Tree scale: 0.01
Bactereoidetes
Actinobacteria
Firmicutes
Proteobacteria
Candidate division TM7
Tannerella forsythia ATCC 43037
Porphyromonas gingivalis ATCC 33277
Porphyromonas gingivalis W83
Porphyromonas sp oral taxon 279 str F0450
Porphyromonas sp oral taxon 278 str W7784
Alloprevotella tannerae ATCC 51259
Prevotella tannerae ATCC 51259
Bacteroidetes oral taxon 274 str F0058
Bacteroidetes oral taxon 278str F0058
Prevotella loescheii JCM 12249 ATCC 15930 strain DSM 19665
Prevotella marshii DSM 16973JCM 13450
Prevotella baroniae DSM 16972 JCM 13447
Prevotella baroniaeJCM 13447 strain DSM 16972
Prevotella sp oral taxon 317 str F0108
Prevotella nigrescens ATCC 33563
Prevotella intermedia 17
Prevotella intermedia ATCC 25615DSM 20706
Prevotella intermedia ATCC 25620DSM 20706
Fretibacterium fastidiosum
Synergistetes sp SGP1
Slackia exigua ATCC 700122
Corynebacterium urealyticum DSM 7109
Actinomyces gerencseriae DSM 6844
Actinomyces timonensis DSM 23838 strain 7400942
Actinomyces sp oral taxon 175 str F0384
Actinomyces sp oral taxon 180 str F0310
Sanguibacter keddieii DSM 10542
Rothia mucilaginosa DY 18
Rothia aeria F0474
Rothia dentocariosa ATCC 17931
Rothia dentocariosa M567
Oribacterium sp oral taxon 78 str F0262
Pseudoramibacter alactolyticus ATCC 23263
Filifactor alocis ATCC 35896
Mogibacterium timidum ATCC 33093
Selenomonas sp oral taxon 149 str 67H29BP
Granulicatella adiacens ATCC 49175
Streptococcus gordonii str Challis
Streptococcus sp F0441
Streptococcus cristatus ATCC 51100
Streptococcus sp M143
Streptococcus sp oral taxon 56
Streptococcus sp oral taxon 71
Streptococcus sanguinis SK36
Streptococcus oralis ATCC 35037
Streptococcus sp F0442
Streptococcus sp GMD4S
Streptococcus oligofermentans AS 1 3089
Streptococcus oralis Uo5
Streptococcus tigurinus 1366
Streptococcus pneumoniae AP200
Streptococcus mitis ATCC 6249
Streptococcus mitis B6
Streptococcus mitis bv 2
Streptococcus mitis NCTC 12261
Desulfobulbus sp oral taxon 41 str Dsb3
Desulfobulbus sp oral taxon 41 str Dsb2
Brevundimonas diminuta 470 4
Brevundimonas diminuta ATCC 11568
Mesorhizobium loti MAFF303099
Bradyrhizobium elkanii USDA 76
Bradyrhizobium elkanii 587
Campylobacter gracilis RM3268
Campylobacter gracilis strain ATCC 33236
Enterobacter cancerogenus ATCC 35316
Klebsiella pneumoniae 342
Pseudomonas fluorescens Pf0 1
Pseudomonas protegens Pf 5
Pseudomonas stutzeri A1501
Pseudomonas pseudoalcaligenes KF707
Pseudomonas aeruginosa LESB58
Pseudomonas otitidis strain LNU E 001
Stenotrophomonas maltophilia R551 3
Neisseria sp oral taxon 14 str F0314
Eikenella corrodens ATCC 23834
Kingella denitrificans ATCC 33394
Delftia acidovorans SPH 1
Delftia acidovorans SPH 20
Delftia acidovorans SPH 9
Achromobacter xylosoxidans A8
Bordetella pertussis strain Tohama I
Ralstonia pickettii 12D
Burkholderia cepacia GG4
Lautropia mirabilis ATCC 51599
3
12
2
3
BGC count
BGC type
Bacteriocin
Oligosaccharide
NRPS
Other
Arylpolyene
NRPS-PKS hybrid
Unclassified
Resorcinol
Siderophore
Homo serine lactone
Terpene
Buytyrolactone

Figure S8

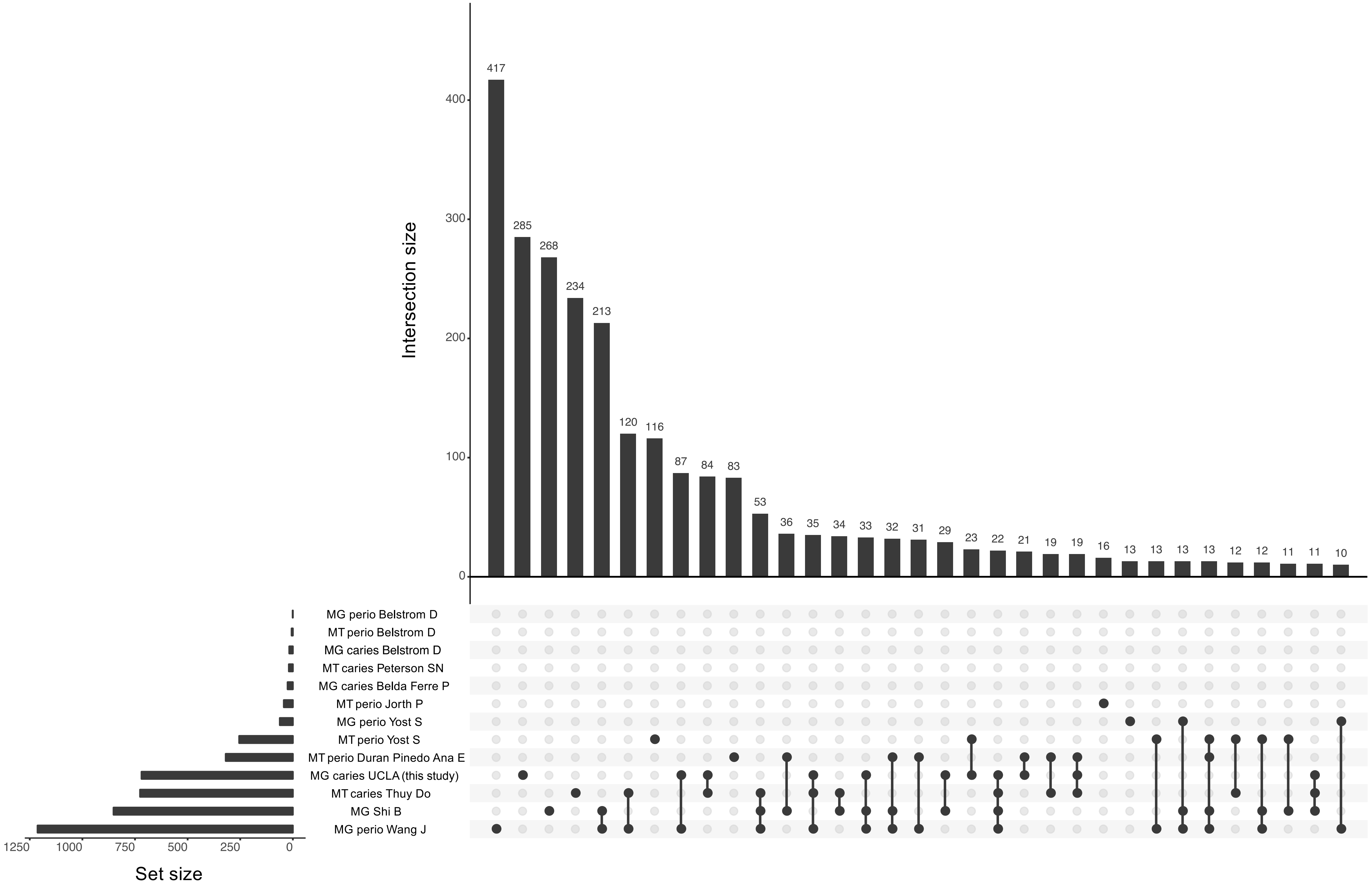

Intersection size
0
100
200
300
400
417
285
268
234
213
120
116
87
84
83
53
36
35
34
33
32
31
29
23
22
21
19
19
16
13
13
13
13
12
12
11
11
10
MG perio Belstrom D
MT perio Belstrom D
MG caries Belstrom D
MT caries Peterson SN
MG caries Belda Ferre P
MT perio Jorth P
MG perio Yost S
MT perio Yost S
MT perio Duran Pinedo Ana E
MG caries UCLA (this study)
MT caries Thuy Do
MG Shi B
MG perio Wang J
1250
1000
750
500
250
0
Set size

Figure S9

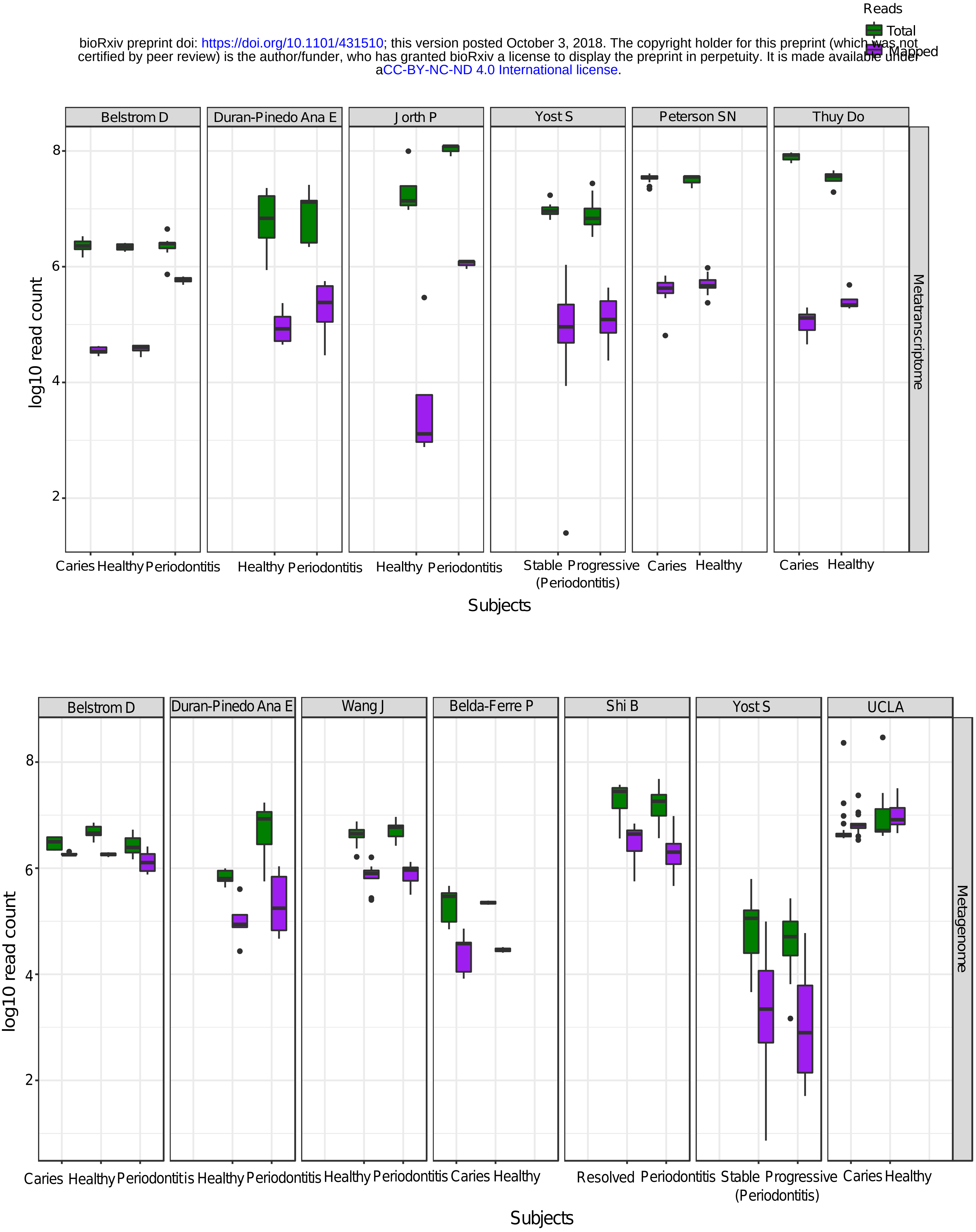

Reads
🟩 Total
🟪 Mapped
Belstrom D
Duran-Pinedo Ana E
J orth P
Yost S
Peterson SN
Thuy Do
Metatranscriptome
log10 read count
8
6
4
2
Caries Healthy Periodontitis
Healthy Periodontitis
Healthy Periodontitis
Stable Progressive (Periodontitis)
Caries Healthy
Caries Healthy
Subjects
Belstrom D
Duran-Pinedo Ana E
Wang J
Belda-Ferre P
Shi B
Yost S
UCLA
Metagenome
log10 read count
8
6
4
2
Caries Healthy Periodontitis
Healthy Periodontitis
Healthy Periodontitis
Caries Healthy
Resolved Periodontitis
Stable Progressive (Periodontitis)
Caries Healthy
Subjects

Figure S10

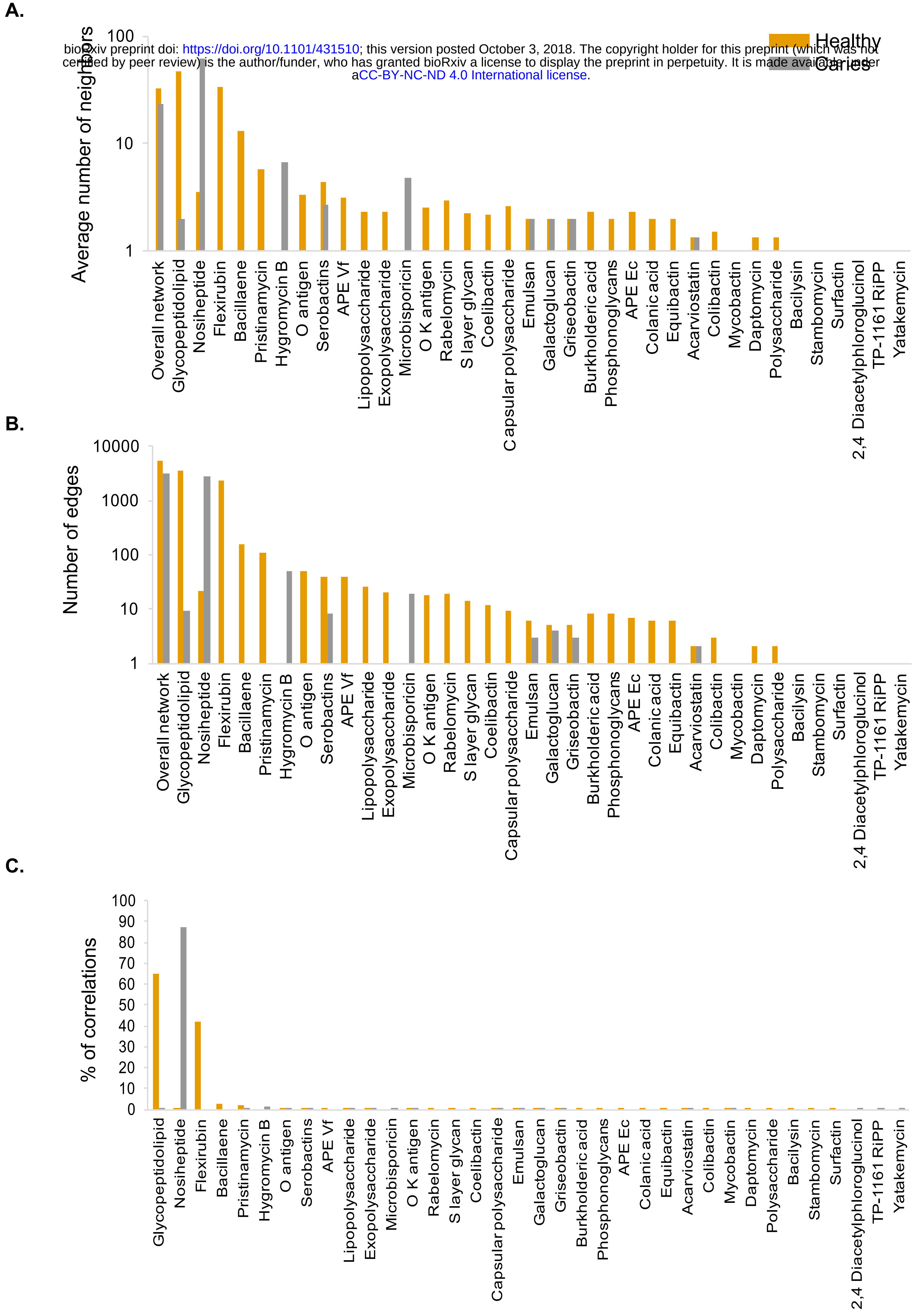

A.
Average number of neighbors
100
10
1
Healthy
Caries
Overall network
Glycopeptidolipid
Nosiheptide
Flexirubin
Bacillaene
Pristinamycin
Hygromycin B
O antigen
Serobactins
APE Vf
Lipopolysaccharide
Exopolysaccharide
Microbisporicin
O K antigen
Rabelomycin
S layer glycan
Coelibactin
Capsular polysaccharide
Emulsan
Galactoglucan
Griseobactin
Burkholderic acid
Phosphonoglycans
APE Ec
Colanic acid
Equibactin
Acarviostatin
Colibactin
Mycobactin
Daptomycin
Polysaccharide
Bacilysin
Stambomycin
Surfactin
2,4 Diacetylphloroglucinol
TP-1161 RiPP
Yatakemycin
B.
Number of edges
10000
1000
100
10
1
Overall network
Glycopeptidolipid
Nosiheptide
Flexirubin
Bacillaene
Pristinamycin
Hygromycin B
O antigen
Serobactins
APE Vf
Lipopolysaccharide
Exopolysaccharide
Microbisporicin
O K antigen
Rabelomycin
S layer glycan
Coelibactin
Capsular polysaccharide
Emulsan
Galactoglucan
Griseobactin
Burkholderic acid
Phosphonoglycans
APE Ec
Colanic acid
Equibactin
Acarviostatin
Colibactin
Mycobactin
Daptomycin
Polysaccharide
Bacilysin
Stambomycin
Surfactin
2,4 Diacetylphloroglucinol
TP-1161 RiPP
Yatakemycin
C.
% of correlations
100
90
80
70
60
50
40
30
20
10
0
Glycopeptidolipid
Nosiheptide
Flexirubin
Bacillaene
Pristinamycin
Hygromycin B
O antigen
Serobactins
APE Vf
Lipopolysaccharide
Exopolysaccharide
Microbisporicin
O K antigen
Rabelomycin
S layer glycan
Coelibactin
Capsular polysaccharide
Emulsan
Galactoglucan
Griseobactin
Burkholderic acid
Phosphonoglycans
APE Ec
Colanic acid
Equibactin
Acarviostatin
Colibactin
Mycobactin
Daptomycin
Polysaccharide
Bacilysin
Stambomycin
Surfactin
2,4 Diacetylphloroglucinol
TP-1161 RiPP
Yatakemycin

Figure S11

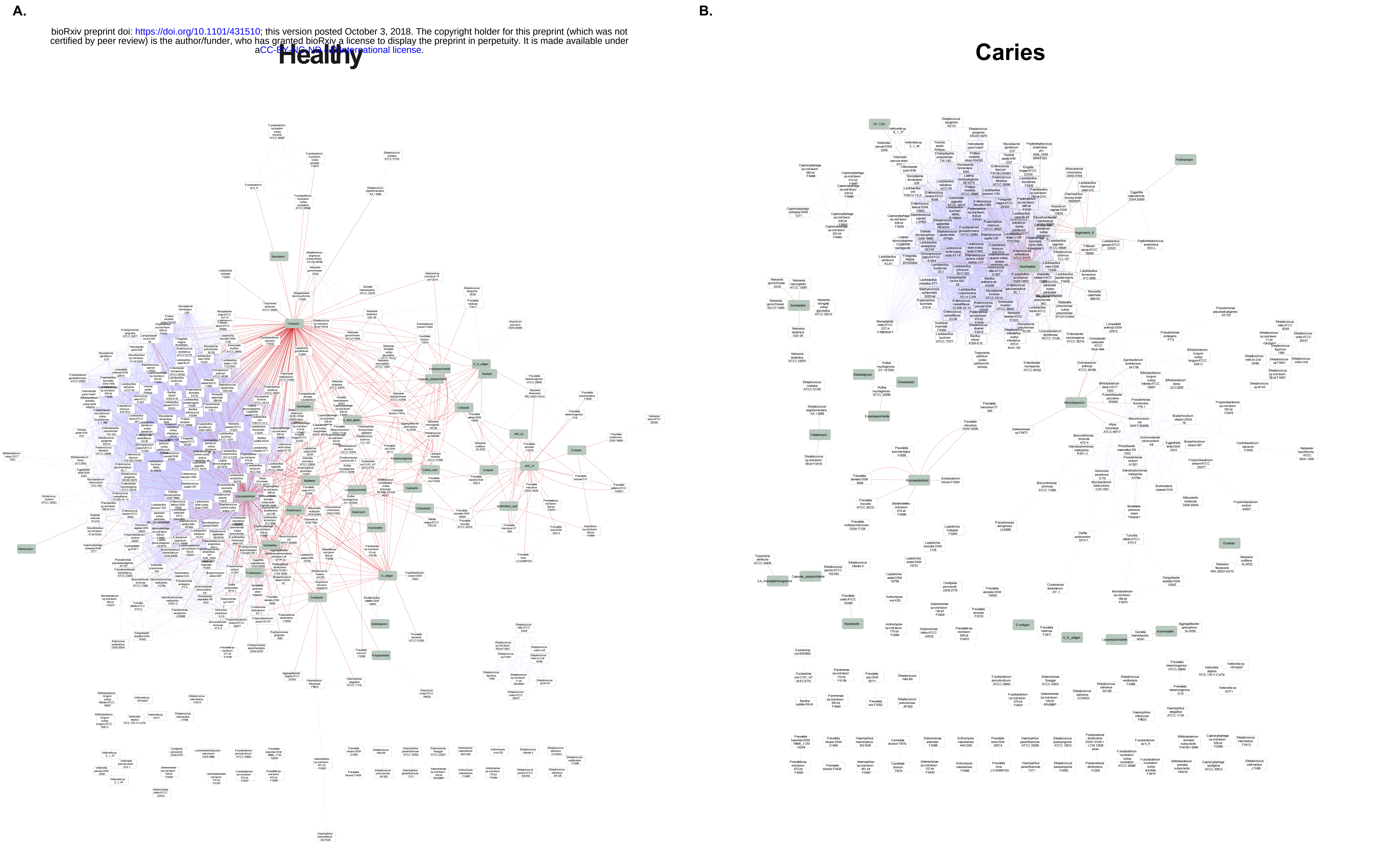


Figure S12

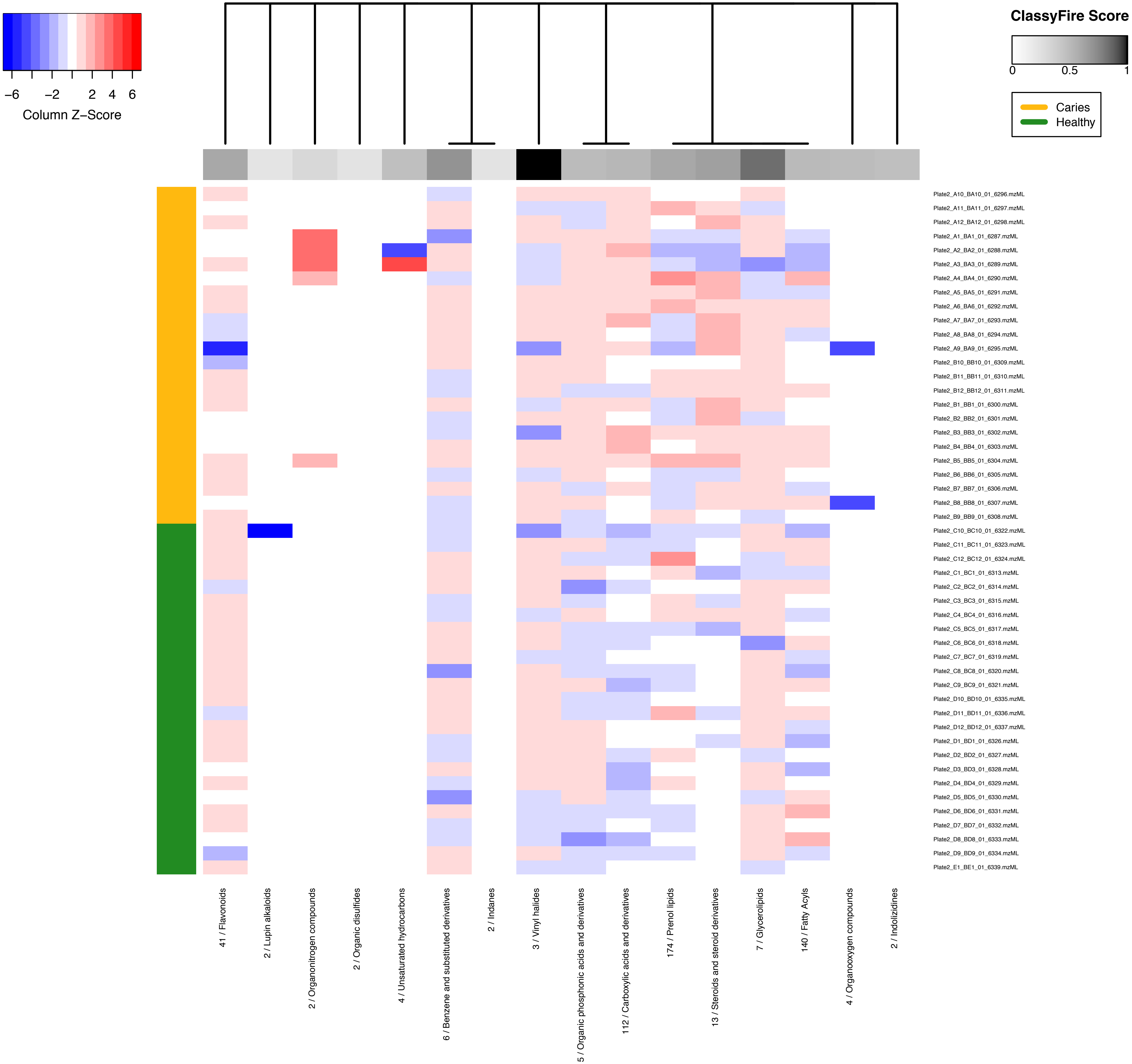
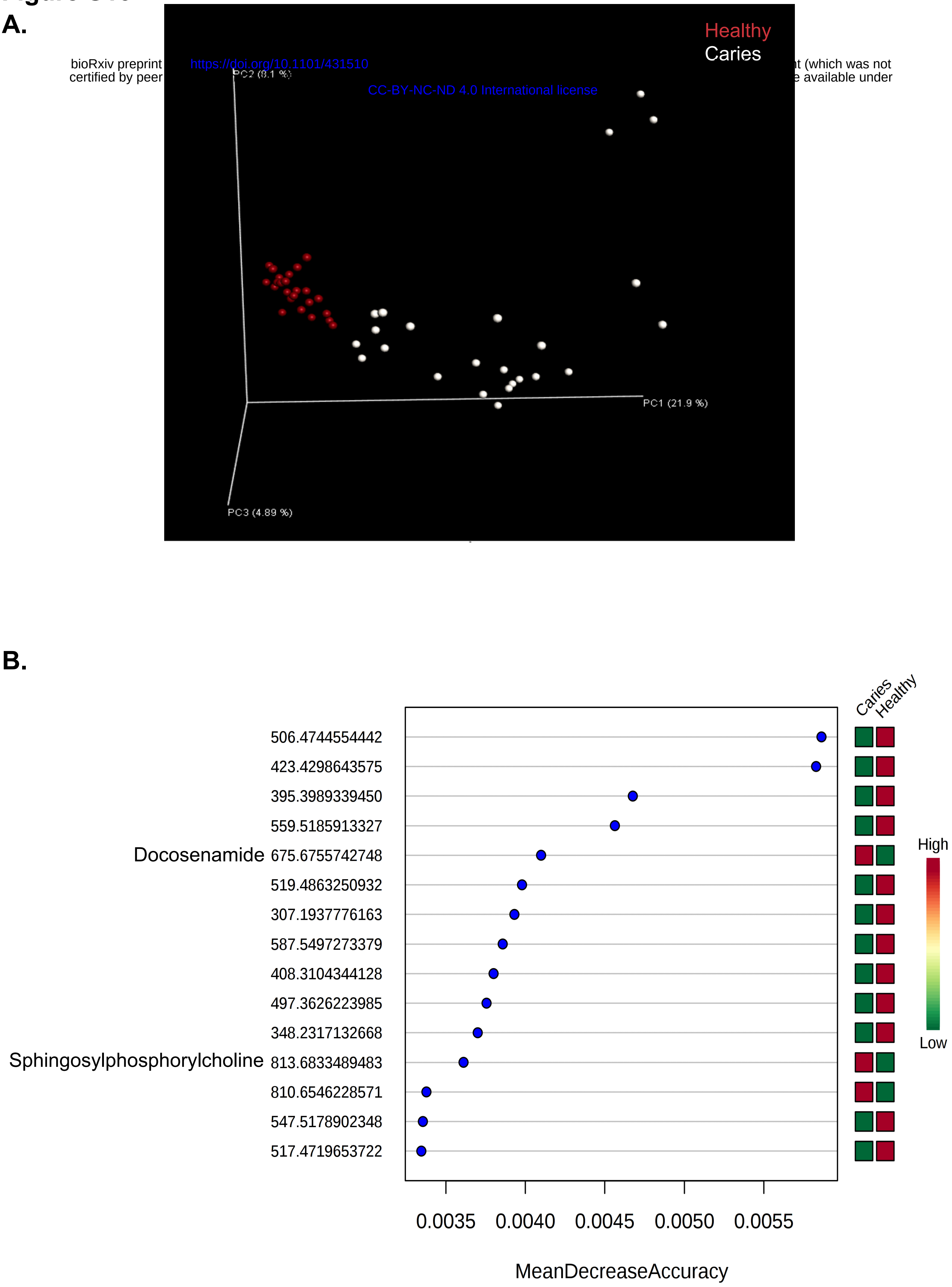

−6
−2
2
4
6
Column Z−Score
ClassyFire Score
0
0.5
1
Caries
Healthy
Plate2_A10_BA10_01_6296.mzML
Plate2_A11_BA11_01_6297.mzML
Plate2_A12_BA12_01_6298.mzML
Plate2_A1_BA1_01_6287.mzML
Plate2_A2_BA2_01_6288.mzML
Plate2_A3_BA3_01_6289.mzML
Plate2_A4_BA4_01_6290.mzML
Plate2_A5_BA5_01_6291.mzML
Plate2_A6_BA6_01_6292.mzML
Plate2_A7_BA7_01_6293.mzML
Plate2_A8_BA8_01_6294.mzML
Plate2_A9_BA9_01_6295.mzML
Plate2_B10_BB10_01_6309.mzML
Plate2_B11_BB11_01_6310.mzML
Plate2_B12_BB12_01_6311.mzML
Plate2_B1_BB1_01_6300.mzML
Plate2_B2_BB2_01_6301.mzML
Plate2_B3_BB3_01_6302.mzML
Plate2_B4_BB4_01_6303.mzML
Plate2_B5_BB5_01_6304.mzML
Plate2_B6_BB6_01_6305.mzML
Plate2_B7_BB7_01_6306.mzML
Plate2_B8_BB8_01_6307.mzML
Plate2_B9_BB9_01_6308.mzML
Plate2_C10_BC10_01_6322.mzML
Plate2_C11_BC11_01_6323.mzML
Plate2_C12_BC12_01_6324.mzML
Plate2_C1_BC1_01_6313.mzML
Plate2_C2_BC2_01_6314.mzML
Plate2_C3_BC3_01_6315.mzML
Plate2_C4_BC4_01_6316.mzML
Plate2_C5_BC5_01_6317.mzML
Plate2_C6_BC6_01_6318.mzML
Plate2_C7_BC7_01_6319.mzML
Plate2_C8_BC8_01_6320.mzML
Plate2_C9_BC9_01_6321.mzML
Plate2_D10_BD10_01_6335.mzML
Plate2_D11_BD11_01_6336.mzML
Plate2_D12_BD12_01_6337.mzML
Plate2_D1_BD1_01_6326.mzML
Plate2_D2_BD2_01_6327.mzML
Plate2_D3_BD3_01_6328.mzML
Plate2_D4_BD4_01_6329.mzML
Plate2_D5_BD5_01_6330.mzML
Plate2_D6_BD6_01_6331.mzML
Plate2_D7_BD7_01_6332.mzML
Plate2_D8_BD8_01_6333.mzML
Plate2_D9_BD9_01_6334.mzML
Plate2_E1_BE1_01_6339.mzML
41 / Flavonoids
2 / Lupin alkaloids
2 / Organonitrogen compounds
2 / Organic disulfides
4 / Unsaturated hydrocarbons
6 / Benzene and substituted derivatives
2 / Indanes
3 / Vinyl halides
5 / Organic phosphonic acids and derivatives
112 / Carboxylic acids and derivatives
174 / Prenol lipids
13 / Steroids and steroid derivatives
7 / Glycerolipids
140 / Fatty Acyls
4 / Organooxygen compounds
2 / Indolizidines

Figure S13


A.
Healthy
Caries
PC2 (8.1 %)
PC1 (21.9 %)
PC3 (4.89 %)
B.
Caries
Healthy
506.4744554442
423.4298643575
395.3989339450
559.5185913327
Docosenamide 675.6755742748
519.4863250932
307.1937776163
587.5497273379
408.3104344128
497.3626223985
348.2317132668
Sphingosylphosphorylcholine 813.6833489483
810.6546228571
547.5178902348
517.4719653722
0.0035
0.0040
0.0045
0.0050
0.0055
MeanDecreaseAccuracy
High
Low