



# Identification of the Bacterial Biosynthetic Gene Clusters of the Oral Microbiome Illuminates the Unexplored Social Language of Bacteria during Health and Disease

Gajender Aleti,<sup>a</sup> Jonathon L. Baker,<sup>a</sup> Xiaoyu Tang,<sup>a</sup> Ruth Alvarez,<sup>b</sup> Márcia Dinis,<sup>b</sup> Nini C. Tran,<sup>b</sup> Alexey V. Melnik,<sup>d</sup> Cuncong Zhong,<sup>c</sup> Madeleine Ernst,<sup>d</sup> Pieter C. Dorrestein,<sup>d</sup> Anna Edlund<sup>a</sup>

<sup>a</sup>Genomic Medicine Group, J. Craig Venter Institute, La Jolla, California, USA

<sup>b</sup>School of Dentistry, The University of California, Los Angeles, California, USA

<sup>c</sup>Department of Electric Engineering and Computer Science, The University of Kansas, Lawrence, Kansas, USA

<sup>d</sup>Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, The University of California San Diego, La Jolla, California, USA

**ABSTRACT** Small molecules are the primary communication media of the microbial world. Recent bioinformatic 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 more than 100 unclassified novel BGCs. The newly identified BGCs were cross-referenced against 254 metagenomes and metatranscriptomes derived from individuals either with good oral health or with dental caries or periodontitis. This analysis revealed 2,473 BGCs, which were differentially represented across the oral microbiomes associated with health versus disease. Coabundance network analysis identified numerous inverse correlations between BGCs and specific oral taxa. These correlations were present in healthy individuals but greatly reduced in individuals with 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

**Citation** Aleti G, Baker JL, Tang X, Alvarez R, Dinis M, Tran NC, Melnik AV, Zhong C, Ernst M, Dorrestein PC, Edlund A. 2019. Identification of the bacterial biosynthetic gene clusters of the oral microbiome illuminates the unexplored social language of bacteria during health and disease. *mBio* 10:e00321-19. <https://doi.org/10.1128/mBio.00321-19>.

**Invited Editor** Egija Zaura, Academic Center for Dentistry Amsterdam

**Editor** Nina R. Salama, Fred Hutchinson Cancer Research Center

**Copyright** © 2019 Aleti et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Anna Edlund, [aedlund@cvi.org](mailto:aedlund@cvi.org).

**Received** 5 February 2019

**Accepted** 12 March 2019

**Published** 16 April 2019

the chemical communication network of the healthy oral microbiome and how it devolves in the case of two prominent diseases.

**KEYWORDS** biosynthetic gene clusters, caries, genome mining, oral microbiome, periodontitis, small molecules

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 structurally diverse small molecules and that these small molecules serve as mediators in a variety of microbe-microbe and host-microbe interactions (2–4). These interactions 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 significant insight into the chemical communication network of the healthy oral microbiome and how it devolves in the case of dental caries and periodontitis.

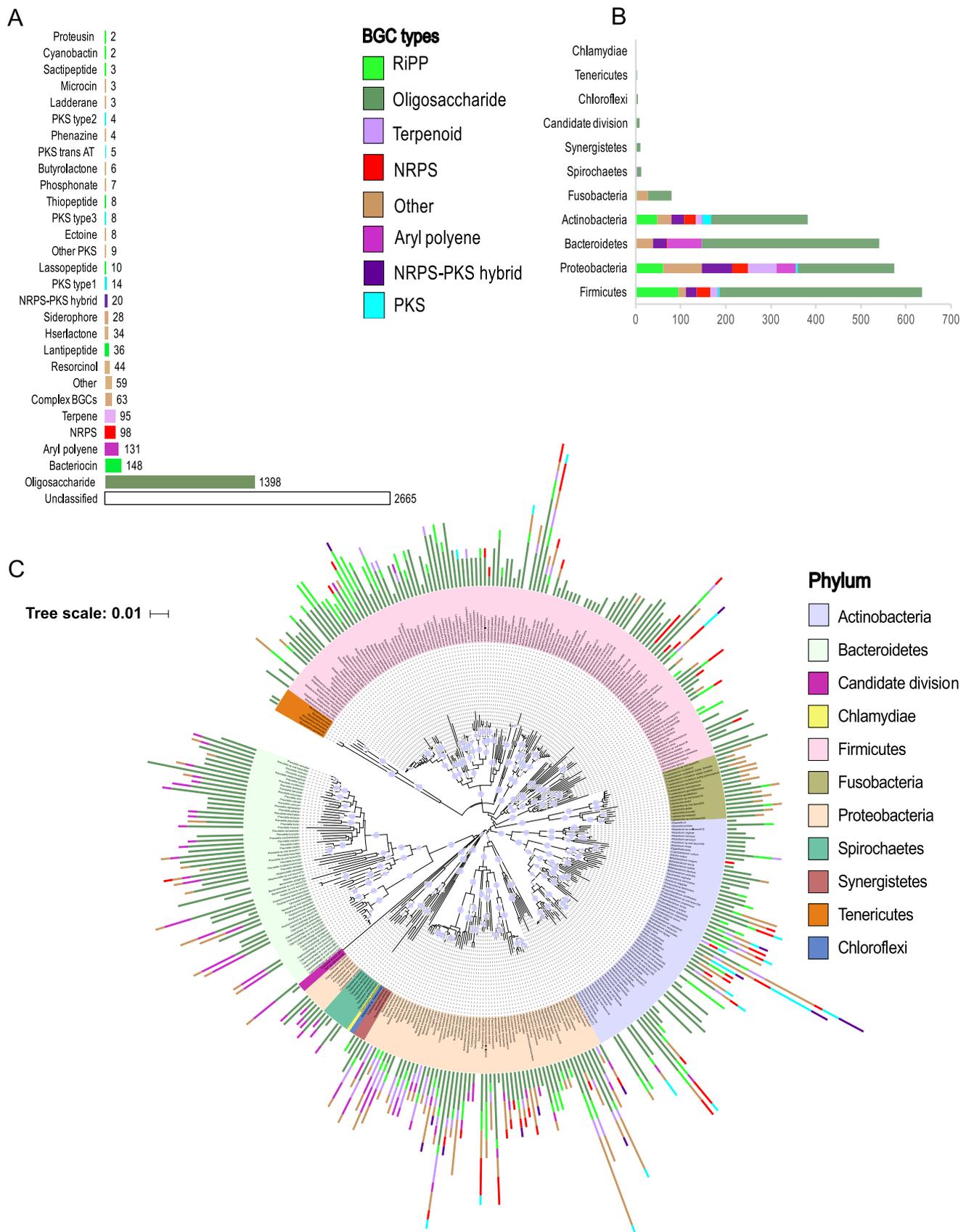
## RESULTS AND DISCUSSION

**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 (see Table S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). A few genomes representing transient exogenous bacteria (e.g., legume symbiont *Mesorhizobium loti*, pathogenic *Yersinia pestis*), which also have oral taxon IDs in the Human Oral Microbiome Database (HOMD) were included in our analysis to represent an additional source of unexplored BGC diversity. However, our sequence read mapping exercises (both with regards to metagenomic and metatranscriptomic reads) showed that most of these BGCs were of less importance, indicating that the majority were either of low abundance or absent in the oral cavity (Table S2 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). 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 bioinformatic study of 169 *Streptococcus 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 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). 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 posted at the above website) (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 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 *Fusobacteriaria*, with the highest number being identified in *Firmicutes* (Fig. 1B and C). However, the high abundance of *Firmicutes* BGCs may reflect that *Streptococcus* genomes are highly sequenced compared to genomes of other oral taxa which may have added bias to the analysis. The ecological roles of oligosaccharides are 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 posttranslationally 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 of genetic transformation in *Streptococcus* show that competence is tightly linked to bacteriocin production (39),



**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. (Continued on next page)

which suggests that these molecules also play important roles in the horizontal transfer of genes and ultimately in niche differentiation and population structure changes.

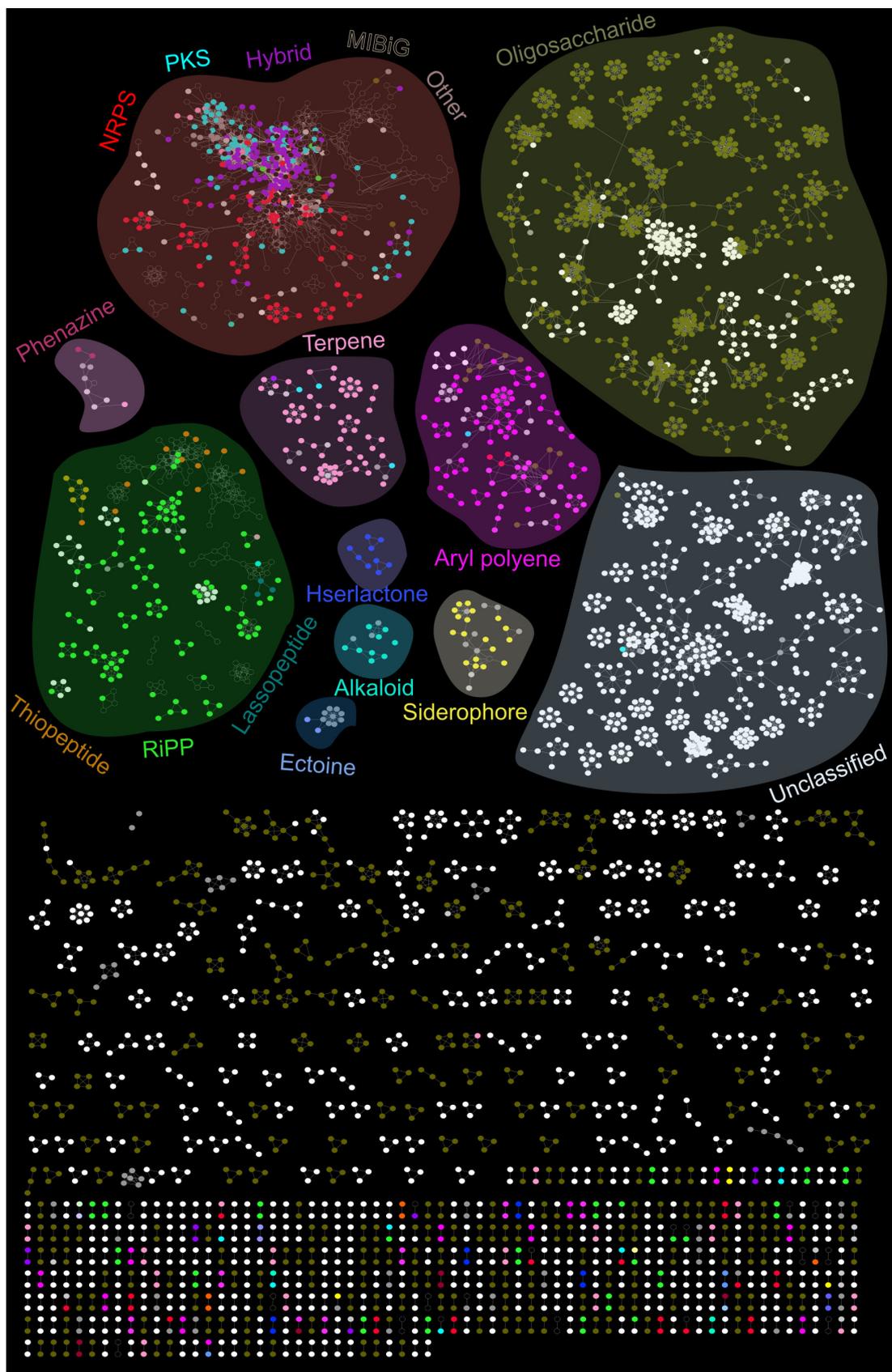
BGCs encoding aryl polyene-like molecules in several *Bacteroidetes* and *Proteobacteria* genomes were identified (131 BGCs or 5.8% of BGCs of a known class). Aryl polyenes are predicted to function as protective agents against oxidative stress (40). However, only a few candidates have been experimentally characterized, leaving this group of small molecules highly underexplored. A diversity of nonribosomal 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). In the same study, it was also observed that BGCs show biogeographic signatures within the oral cavity, which suggests that different social signaling skills are required to inhabit different ecological niches. NRPS and PKS 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 a 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 networks reveals distant and close homologies to 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 Text S1 in the supplemental material). 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 subnetwork with BGCs predicted to be of a known class, particularly the oligosaccharide, RiPP, and aryl polyene classes (Fig. 2). These data provide inferences as to the function of more than 100 previously unclassified novel BGCs.

The largest subnetwork, 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 at times categorized with primary metabolism, and not natural product-producing BGCs, as is the case in this study. The second-largest major subnetwork 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

#### FIG 1 Legend (Continued)

(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. The 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. Sixty-three (~3%) of such unresolved BGCs were encountered and categorized as the “complex” BGC type (for convenience, we combined these BGCs with BGC types “Other” for subsequent analysis).



**FIG 2** Similarity networking identified putative product classes for novel BGCs. Similarity networks between the BGCs identified in the oral cavity and the experimentally characterized reference BGCs obtained from the MIBiG repository are depicted. (Continued on next page)

Downloaded from <http://mbio.asm.org/> on April 16, 2019 by guest

repository (Table S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). 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 subnetworked 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; see also Table S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). Several BGCs annotated as saccharides, other, unclassified, PKS and NRPS BGC types grouped with aryl polyene BGCs, which may represent novel hybrid classes of BGCs. Other small subnetworks include biosynthesis of terpene phenazine, homoserine lactone, alkaloid, siderophore, and ectoine. These subnetworks 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 regard 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 future 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 (see Fig. S1 in the supplemental material). A total of 1,393 (~25%) BGCs did not appear to have significant homology to BGCs in MIBiG, based upon the E value (see Materials and Methods for details). Such a detailed annotation of BGCs harbored by the human oral microbiome has not been accomplished previously.

**Specific BGCs and bacterial taxa 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 were obtained from eight previous studies, which represented all publicly available metagenomes and metatranscriptomes associated with caries or periodontal disease compared to oral health at the time of this study (28 December 2017; Table S2 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). Due to the limited number of publicly available deep sequencing libraries of high quality that represent the same sampling location within the oral cavity (i.e., supra- or subgingival plaque and saliva), we included libraries with mixed sample origin to represent healthy and disease states, respectively (Table S2 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). This could have created discrepancies in our analysis, and differences that we present here may represent differences between sampling location and not between oral health states. In addition, DNAs from 47 saliva samples representing 23 children with caries and 24 healthy children were sequenced, and putative BGCs were identified. Nonsupervised exploratory ordination through

#### FIG 2 Legend (Continued)

Subnetworks 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 subclusters of unclassified and oligosaccharide BGCs with a minimum number of eight nodes are organized into given highlighted constellation.

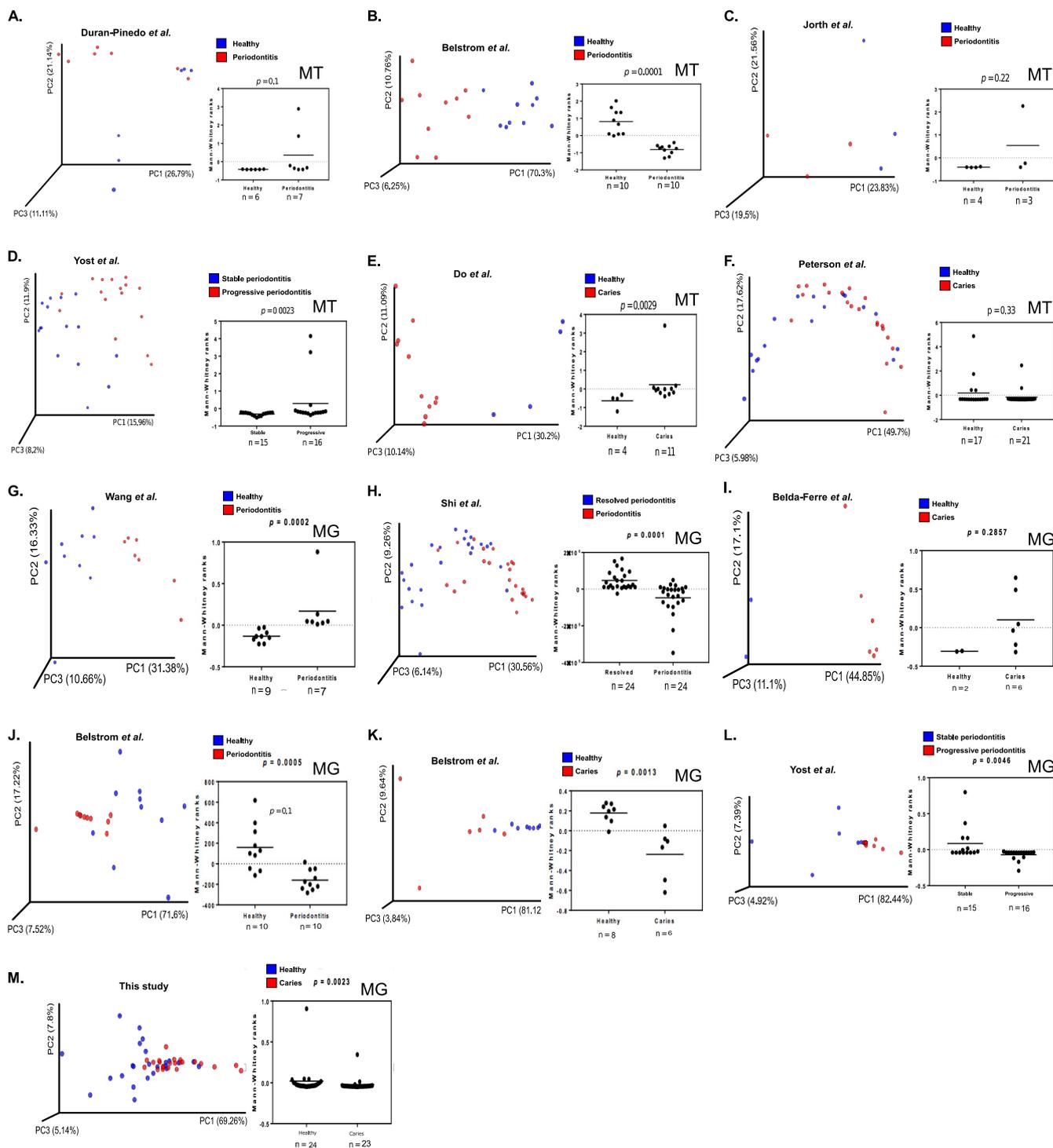
PCoA revealed significant differences in the representation of BGCs between healthy and diseased subjects in five of the six metatranscriptome studies and six of the seven metagenome studies investigated (Fig. 3). The 1,804 BGCs which were differentially represented in health versus disease in the metagenomes 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 types particularly included BGCs of the oligosaccharide, aryl polyene, terpene, bacteriocin, and NRPS classes (Fig. 4). 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. 4). Previous studies illustrated that aryl polyenes act as protective agents against oxidative stress and that terpenes function as anti-inflammatory agents (40). These findings are in line with previous findings that high levels of *Actinomyces* were associated with severe childhood caries (44). In the caries-associated samples, known cariogenic species belonging to the *Streptococcus*, *Veillonella*, and *Lactobacillus* genera (45) showed notable changes in bacteriocin and oligosaccharide BGC expression profiles (Fig. 4).

In periodontitis, a high number of differentially expressed BGCs (170 BGCs) were identified in community members belonging to the *Bacteroidetes* phylum. 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. 4). These results suggest that *Neisseria* spp. and its small molecules may play a role in periodontitis, which is worth further explorations in future studies. Homologs to specific BGC products in the MIBiG database which displayed differential expression in health and disease are detailed in Fig. S2. Analysis of the metagenomic studies yielded trends similar to those detailed above (Fig. 5).

Next, a subset of differentially represented BGCs, which showed high levels of expression in either healthy or diseased states, was examined to determine whether they commonly occur across studies. The results were visualized as a binary occurrence matrix (Fig. S3). In all studies analyzed, only a minor fraction of the differential features (<10 BGCs) were shared between any two studies. The reason for this discrepancy may be due to both biological and technical variations such as differences in anatomical features at the collection site of dental plaque and what time of the day saliva samples were collected. Additionally, sequencing platforms, methodological variations in sample collection, nucleic acid extraction techniques (with or without DNA or RNA/rRNA removal methods), and use of different library preparation methods (see Table S2 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1> and Fig. S4) may also play important roles. Given the complexity of metagenomes and metatranscriptomes, these factors can certainly influence the sequence composition and lead to the differences we observed here. It is also important to note that due to the fact that BGCs can be horizontally transferred between different bacterial species, mRNA read mapping to BGC databases does not provide definite taxonomic classification and predictions of species-level activities.

**Differential interactions between BGCs and oral taxa in caries and health.** To examine possible relationships between BGCs and bacterial taxa during health and disease, a focused comparative analysis of the shotgun metagenomics data obtained in this study from saliva samples from healthy children and children with caries (Table S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>) was performed. Children in the caries group had a minimum of two active caries lesions. No overall differences in diet or oral hygiene habits (brushing



**FIG 3** Principal coordinate analysis (PCoA) of BGC profiles representing oral bacterial communities in health and disease. (A) Metatranscriptomics (MT) of BGC profiles from healthy individuals versus individuals with periodontitis from Duran-Pinedo et al. (17). (B) Metatranscriptomics of healthy versus periodontitis BGC profiles from Belstrom et al. (18). (C) Metatranscriptomics of stable versus progressive periodontitis profiles from Jorth et al. (19). (D) Metatranscriptomics of healthy versus caries profiles from Yost et al. (20). (E) Metatranscriptomics of healthy versus caries profiles from Do et al. (21). (F) Metatranscriptomics of healthy versus caries profiles from Peterson et al. (22). (G) Metagenomics (MG) of BGC profiles from healthy individuals versus individuals with periodontitis from Wang et al. (23). (H) Metagenomics of resolved versus periodontitis profiles from Shi et al. (24). (I) Metagenomics of healthy versus caries profiles from Belda-Ferre et al. (25). (J) Metagenomics of healthy versus periodontitis profiles from Belstrom et al. (18). (K) Metagenomics of healthy versus caries profiles from Belstrom et al. (18). (L) Metagenomics of stable versus progressive periodontitis profiles from Yost et al. (20). (M) Metagenomics of healthy versus caries in the current study. PCoA analyses are based on differentially represented BGCs calculated by using the DESeq2 package with false discovery rate (FDR) correction of  $<0.05$ . Manhattan distances between samples were visualized using the first three principal coordinates, and significance was tested by applying Mann-Whitney ranks on the most variance captured by the first principal coordinate.

**TABLE 1** Differentially represented or expressed biosynthetic pathways in samples from shotgun metatranscriptomic and metagenomic libraries<sup>a</sup>

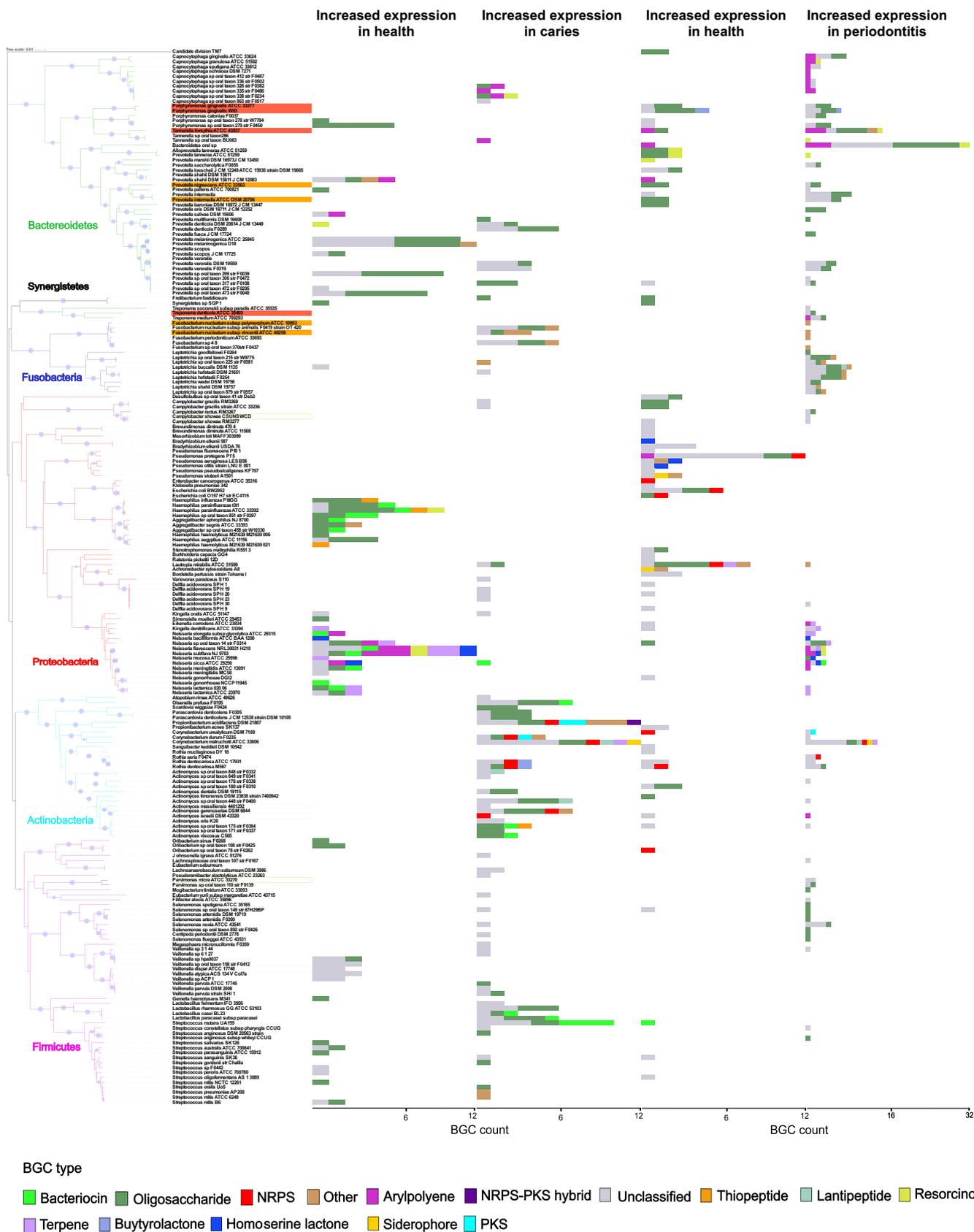
BGC type	Metatranscriptome (up/down) <sup>b</sup>		Metagenome (up/down) <sup>b</sup>	
	Caries	Periodontitis	Caries	Periodontitis
Aryl polyene	4/11	29/4	26/5	25/3
Bacteriocin	14/10	1/1	31/1	5/24
Butyrolactone	2/0	1/1	3/0	1/1
Homoserine lactone	0/4	3/4	10/0	1/1
Lantipeptide	7/0	4/0	11/2	1/6
Lasso peptide	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

<sup>a</sup>Number of differentially represented or expressed biosynthetic pathways in saliva, supra- and subgingival plaque samples, from shotgun metatranscriptomic and metagenomic libraries representing 231 subjects (oral healthy individuals [ $n = 110$ ] and individuals with dental caries [ $n = 77$ ] or periodontitis [ $n = 44$ ]). A total of 741 BGCs were differentially abundant in individuals with caries (515 enriched and 226 less abundant), and 1,063 BGCs were associated with periodontitis (670 enriched and 393 less abundant). A total of 355 BGCs were differentially expressed in caries (208 upexpressed and 147 downexpressed), while 421 BGCs were either up- or downexpressed in subjects with periodontitis (218 upexpressed and 203 downexpressed).

<sup>b</sup>Values in the table are the numbers of BGCs upexpressed/numbers of BGCs downexpressed in individuals with caries or periodontitis.

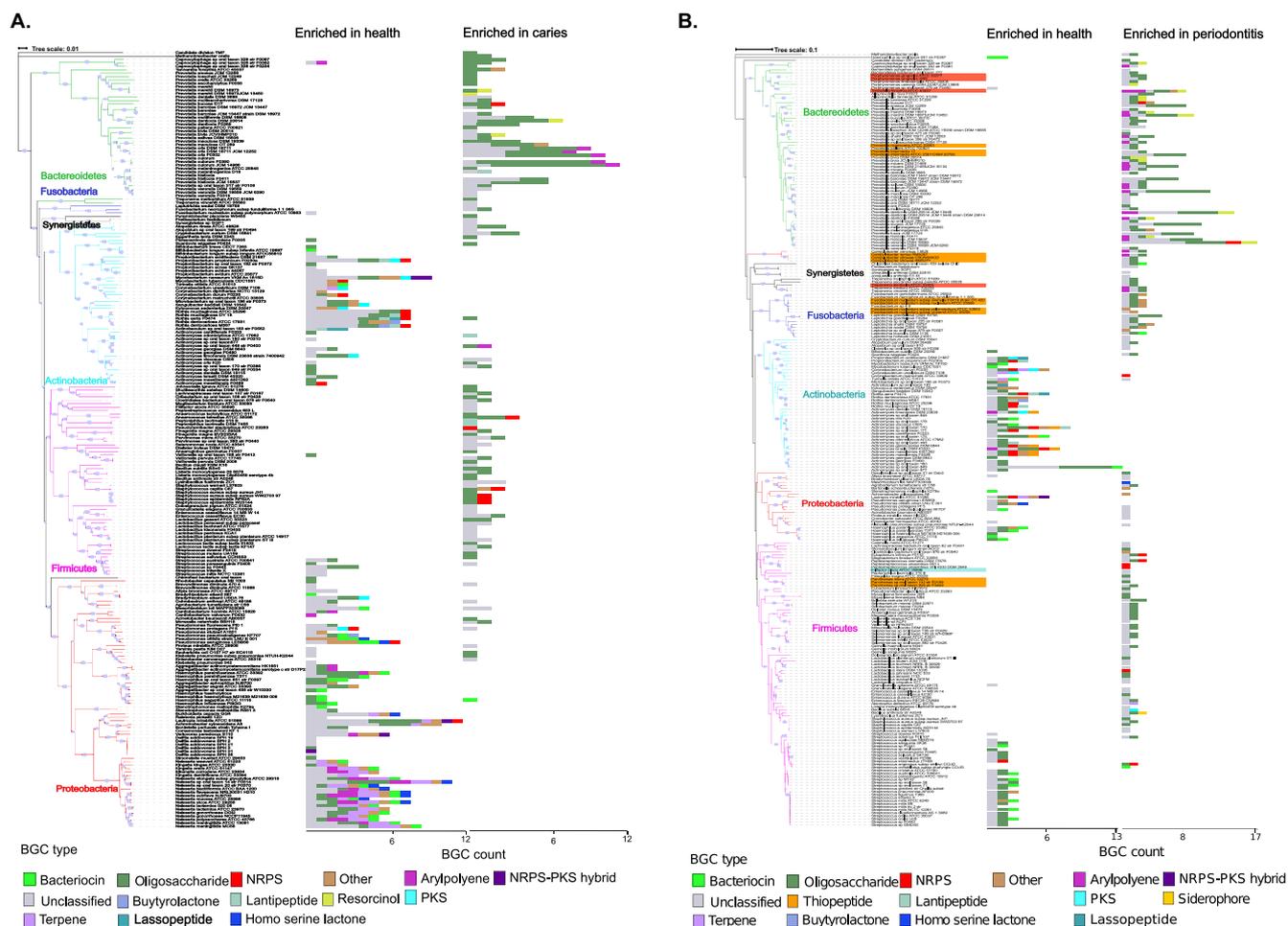
and flossing) were recorded between the groups. Also, gingiva index scores, which measure oral health were low for both groups (below 1.0) and not significantly different ( $P > 0.05$ , Table S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>), suggesting that both groups presented with good overall gingival health, except for cavities in the caries active group. 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 taxon-taxon 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 S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1> and Fig. S5 and S6). This suggests that antagonistic relationships, modulated through BGC-produced antimicrobial molecules, are highly significant to the ecology of the oral microbiome.

All BGCs that had significant correlations to oral taxa (a total 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. S1). In the oral microbiomes derived from healthy children, the interaction network was dominated by negative correlations between oral taxa and BGCs producing glycopeptidolipids, capsular polysaccharides, as well as a homolog of flexirubin (Fig. S5A and Table S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). 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. S5A and Table S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). The flexirubin homolog BGCs were encoded



**FIG 4** BGCs are differentially expressed in oral health and disease. Bar graphs illustrating phylogenetic distribution of biosynthetic pathways in oral health- and disease-associated oral microorganisms. Taxa with significant changes in BGC expression based on the analyzed metatranscriptomic data sets are shown in the phylogenetic tree on the left. Bacterial genomes associated with periodontal disease and that belong to the red and orange complexes are labeled in (Continued on next page)

Downloaded from <http://mbio.asm.org/> on April 16, 2019 by guest



**FIG 5** Overview of differentially represented biosynthetic pathways in oral bacterial genomes in oral health and disease states based on metagenomic data sets. Bacterial genomes associated with periodontal disease and that belong to the red and orange complexes are labeled in orange and red. (A) 740 differentially abundant BGCs in the salivary and supragingival microbiome in healthy individuals and individuals with caries, representing 69 subjects (healthy individuals [ $n = 34$ ]; individuals with caries [ $n = 35$ ]). (B) 1,065 differentially abundant BGCs in metagenomes obtained from saliva and subgingival plaque samples from individuals with periodontitis and healthy controls, representing 49 subjects (oral health [ $n = 25$ ]; periodontitis [ $n = 24$ ]). Taxonomic units with significant changes in BGC representation are shown in the phylogenetic tree. Bar graphs at the leaf tips are colored according to the BGC type and display numbers of BGCs expressed higher or lower. Noted that the x-axis scales are different for the left and right panels.

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 (positive correlation coefficients included here range between 0.8 and 1.0, and negative coefficients range between  $-0.8$  and 1) (Fig. S5A and Table S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). Additional negative correlation with *S. mutans* ranged between  $-0.5$  and  $-0.7$  (data not shown here). Homologs of the antibiotics bacillaene and pristinamycin (47, 48), harbored by genomes representing *Propionibacterium propionicum* F0230a and *Actinomyces timonensis* DSM 23838 displayed negative correlations with several pathogenic taxa: *Lactobacillus*, *Listeria*, *Lysinabacillus*, *Acinetobacter*, *Enterococcus*, *Neisseria*, *Staphylococcus*, *Kingella*, and *S. mutans* (49) (Fig. S5A). These associations are reminiscent of a previous study which observed similar macrolide-encoding BGCs widely distributed among oral bacterial genomes (2). These

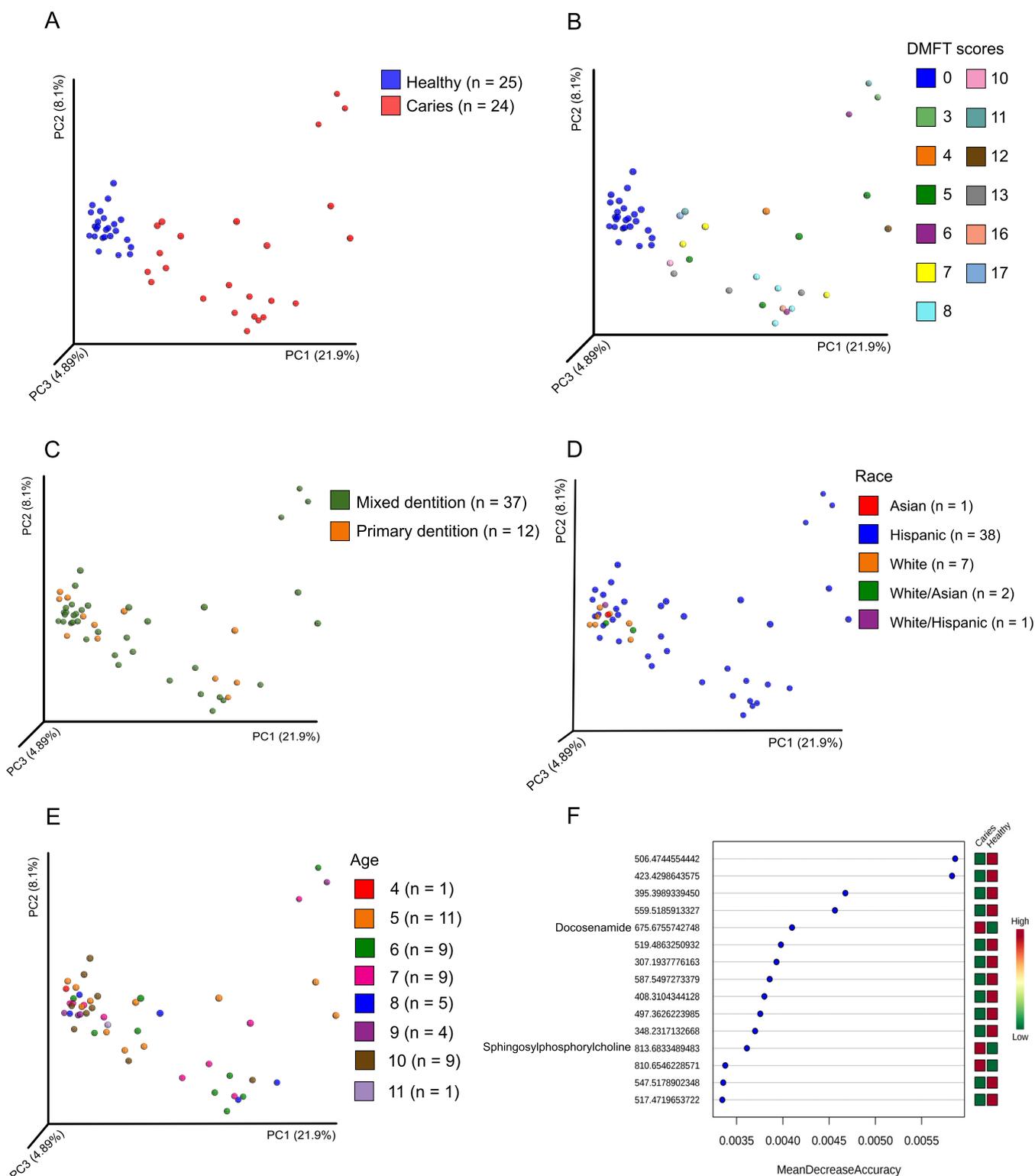
#### FIG 4 Legend (Continued)

orange and red. Bar graphs at the leaf tips display number of BGCs either up- or downexpressed and colored according to the BGC type. It should be noted that the x-axis scales are different in the 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  $< 0.05$ ).

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 pristinemycin-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 nonribosomal peptide pathway encoding the genotoxin colibactin (51). This BGC showed antagonistic associations with pathogenic genera: *Haemophilus*, *Aggregatibacter*, *Parascardovia*, *Capnocytophaga*, and *Streptococcus* (Table S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>).

Most intriguingly, the number of significant correlations between BGCs and microbial taxa was dramatically reduced in the samples derived from children with caries (Fig. S6A 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. S5B; Table S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). The nosiheptide-like BGC, encoded in the genome of *Corynebacterium matruchotii*, was the most predominant, with antagonistic interactions against ~90 taxa. These taxa included pathogens from the *Klebsiella*, *Helicobacter*, *Filifactor*, *Haemophilus*, *Enterococcus*, and *Fusobacterium* genera (Fig. S5B; Table S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). 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 suggests that during disease, the community lacks the ability to limit the abundance of this cariogenic pathogen. Taken together, these results show that in the oral microbiome, exclusion of particular taxa via antagonistic interactions, mediated by the products of BGCs, is widespread. Although such interactions were still present in the caries-associated oral microbiomes, there were fewer interactions. 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 from the differentially abundant BGCs that we identified in the analysis above (Fig. 3M), we performed untargeted liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis of the same saliva samples. 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. Fifty 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. S7). Substructures associated with macrolides, terpenoids, and macrolactams were also identified. At the chemical class level, there were no distinct relative abundance patterns between the oral health- and disease-associated samples. This is in stark contrast to PCoA analysis of the 1,369 MS features, which showed clear separation of samples between healthy and diseased states (Fig. 6A) rather than by DMFT scores (Fig. 6B), dentition stage (Fig. 6C), race (Fig. 6D), and age (Fig. 6E), in agreement with the BGC abundance profiles (Fig. 3M). By employing a random forest importance



**FIG 6** Analysis of small molecules shows clear differences between saliva samples derived from healthy children compared to children with caries. (A) Principal coordinate analysis (PCoA) of parent masses (*m/z*) derived from LC-MS/MS analysis of saliva. Manhattan distances between samples were visualized, and significance was tested by applying Mann-Whitney ranks on the principal coordinates ( $P < 0.0001$ ). (B) PCoA based on DMFT (decayed, missed, or filled teeth) scores ( $P > 0.05$ ). (C) PCoA based on primary and mixed dentition (primary and permanent teeth) ( $P > 0.05$ ). (D) PCoA based on race ( $P > 0.05$ ). (E) PCoA based on age ( $P > 0.05$ ). (F) Fifteen key metabolites responsible for the PCoA separation identified by a random forest importance model.

Downloaded from <http://mbio.asm.org/> on April 16, 2019 by guest

model, 15 key metabolites, which were distinct between healthy and disease states (Fig. 6F), were identified. Of these, 12 were significantly more abundant in healthy subjects, while 3 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 genomic and metabolomic approaches employed identify putative homologs and not exact matches. Thus, using current techniques and databases, it is not possible to definitively determine whether 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 were 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 extraoral) diseases and will guide development of improved therapeutic strategies to maintain oral health.

## MATERIALS AND METHODS

The ethics statement is provided in Text S1 (Supplementary Materials and Methods section) in the supplemental material.

**Study inclusion/exclusion criteria and collection of saliva.** Subjects were included in the study if the subject was 3 years old or older and in good general health according to a medical history and clinical judgment of the clinical investigator, subjects had at least 12 teeth. Subjects in the caries group had a supragingival (above the gum line) plaque score of 1 or greater on a scale of 0 to 3. Subjects were excluded from the study if they had generalized rampant dental caries, chronic systemic disease, or medical conditions that would influence the ability to participate in the proposed study (i.e., cancer treatment, HIV, rheumatic conditions, history of oral candidiasis). Subjects were also excluded if they had open sores or ulceration in the mouth, radiation therapy to the head and neck region of the body, significantly reduced saliva production or had been treated by anti-inflammatory or antimicrobial therapy. Ethnic origin was mixed for healthy subjects (Hispanic, Asian, Caucasian, Caucasian/Asian), while children with caries were of Hispanic origin. For the latter group, no other ethnic group enrolled despite several attempts to identify interested families/participants. Children with both primary and mixed dentition stages were included (caries group: 18 children with mixed dentition and 6 with primary dentition; healthy group: 19 children with mixed dentition and 6 with primary dentition). To further enable classification of health status (here caries and healthy), a comprehensive oral examination of each subject was performed by a single calibrated pediatric dental resident, using a standard dental mirror, illuminated by artificial light. The visual inspection was aided by tactile inspection with a community periodontal index (CPI) probe when necessary. The number of teeth present was recorded, and dental caries status was recorded using decayed (d), missing due to decay (m), or filled (f) teeth (t) in primary and permanent dentitions (dmft/DMFT), according to the criteria proposed by the World Health Organization in 1997. Duplicate examinations were performed on five randomly selected subjects to assess intraexaminer reliability. Subjects were dichotomized into two groups: caries free (dmft/DMFT = 0)

and caries active (subjects with  $\geq 2$  active dentin lesions). If the subject qualified for the study, (s)he was asked to brush their teeth after breakfast and abstain from eating and drinking for 2 h prior to saliva collection in the morning. According to a previously developed protocol (58), to ensure that the mouth was free of food or foreign substances, each volunteer rinsed their mouth three times with water 10 min prior to saliva collection. Following the same protocol, we collected approximately 2 ml saliva from each subject by having the subject spit directly in a sterile 15-ml Falcon tube over a 20-min period. Whole saliva was immediately transferred to sterile 2-ml cryovial tubes and snap-frozen for long term storage at  $-80^{\circ}\text{C}$ . For more information, see Text S1 (Supplementary Materials and Methods section) in the supplemental material.

**DNA extraction and metagenomic sequencing.** For detailed protocols of DNA extraction and metagenomic sequencing, see Text S1 (Supplementary Materials and Methods section).

**BGC identification and network analysis of known and putative oral BGCs.** A list of 1,362 described and curated human oral taxa (18 September 2017) was obtained from the Human Oral Microbiome Database (HOMD) (55). A few genomes belonging to transient exogenous bacterial taxa, which also have oral taxon IDs in HOMD were also added to this list (e.g., legume symbiont *Mesorhizobium loti*, pathogenic *Yersinia pestis*). In order to identify small-molecule- and secondary metabolite-encoding BGCs in the 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 (<https://www.ncbi.nlm.nih.gov/genome>), as well as from an in-house database (Table S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). These sequences 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 nonbiosynthetic genes commonly found in BGCs (2) (Table S1 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>) based on text within an attribute using advanced filter settings in CLC Workbench software v. 9. (CLCbio, Aarhus, Denmark). The resulting data set contained a total of 192,283 gene sequences from 4,915 BGCs and can be downloaded from the Massive repository (<ftp://massive.ucsd.edu/MSV000081832>) with the accession ID MSV000081832. For more information, see Text S1 (Supplementary Materials and Methods section).

**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 (excluding the fatty acid synthase encoding BGCs) derived from oral bacterial genomes with those encoding the biosynthetic pathways for known compounds, we performed multigene 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 Text S1 (Supplementary Materials and Methods section).

**16S rRNA gene (16S) phylogenetic analysis.** For a detailed protocol of 16S rRNA gene phylogenetic analysis, see Text S1 (Supplementary Materials and Methods section).

**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. (59), Do et al. (25), Peterson et al. (26), Yost et al. (27), Wang et al. (28), and Shi et al. (60), as well as our own study of metagenomes from saliva samples obtained from children with good dental health and children with dental caries was analyzed (sequence reads are accessible under BioProject accession no. *PRJNA1234*; Table S2 posted at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>). For a detailed protocol, see Text S1 (Supplementary Materials and Methods section).

**Differential abundance and expression analyses of BGCs.** We employed a systematic workflow for analyzing abundance and expression profiles of the BGCs. Using SRA toolkit utilities, reads were extracted from metatranscriptome and metagenome shotgun sequenced libraries available via NCBI. For a detailed protocol, see Text S1 (Supplementary Materials and Methods section).

**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 EMPPeror (61) tool. For a detailed protocol, see Text S1 (Supplementary Materials and Methods section).

**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 coabundance and coexclusion networks between species and corresponding BGCs. For a detailed protocol, see Text S1 (Supplementary Materials and Methods section).

**Experimental small molecule metabolite detection.** Approximately 150  $\mu\text{l}$  of saliva was lyophilized, and ethyl acetate was added to extract nonpolar molecules. Samples were then vortexed, centrifuged to remove the cell debris, and submitted to untargeted LC-MS/MS analysis. For a detailed protocol, see Text S1 (Supplementary Materials and Methods section).

**Mass spectral molecular networking.** LC-MS/MS spectra obtained from 49 saliva samples (25 healthy subjects and 24 subjects with caries) were preprocessed for feature extraction using MZmine2 (62) and submitted to mass spectral molecular networking through GNPS (43). For a detailed protocol, see Text S1.

**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 Text S1.

**Availability of data and material.** Sequence data have been submitted to NCBI under BioProject ID PRJNA478018 with SRA accession no. SRP151559. Mass spectral files, LC-MS/MS metadata file, and Nucleotide FASTA sequences of the oral biosynthetic gene cluster collection are accessible from the MassIVE repository (<ftp://massive.ucsd.edu/MSV000081832>) with the accession ID MSV000081832. Tables S1 and S2 are available at <https://massive.ucsd.edu/ProteoSAFe/status.jsp?task=1ed182b005a74b94ac730f769b2f37b1>.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mBio.00321-19>.

**TEXT S1**, DOCX file, 0.1 MB.

**FIG S1**, PDF file, 0.9 MB.

**FIG S2**, PDF file, 0.9 MB.

**FIG S3**, PDF file, 1.8 MB.

**FIG S4**, PDF file, 0.3 MB.

**FIG S5**, PDF file, 4.9 MB.

**FIG S6**, PDF file, 0.8 MB.

**FIG S7**, PDF file, 0.2 MB.

## ACKNOWLEDGMENTS

We 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.).

G.A., X.T., C.Z., and A.E. designed the experiments. G.A., A.V.M., R.A., M.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.

We declare that we have no competing interests.

## REFERENCES

- Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207–214. <https://doi.org/10.1038/nature11234>.
- Donia MS, Cimermancic P, Schulze CJ, Wieland Brown LC, Martin J, Mitreva M, Clardy J, Lington 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. <https://doi.org/10.1016/j.cell.2014.08.032>.
- Donia MS, Fischbach MA. 2015. Small molecules from the human microbiota. *Science* 349:1254766. <https://doi.org/10.1126/science.1254766>.
- 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, Lington RG, Fischbach MA. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158:412–421. <https://doi.org/10.1016/j.cell.2014.06.034>.
- Zipperer A, Konnerth MC, Laux C, Berscheid A, Janek D, Weidenmaier C, Burian M, Schilling NA, Slavetinsky C, Marschal M, Willmann M, Kalbacher H, Schitteck B, Brotz-Oesterhelt H, Grond S, Peschel A, Krüger B. 2016. Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* 535:511–516. <https://doi.org/10.1038/nature18634>.
- 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. <https://doi.org/10.1038/nature23874>.
- 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* 341:569–573. <https://doi.org/10.1126/science.1241165>.
- Jakubovics NS. 2015. Intermicrobial interactions as a driver for community composition and stratification of oral biofilms. *J Mol Biol* 427:3662–3675. <https://doi.org/10.1016/j.jmb.2015.09.022>.
- 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* 29:104–106. <https://doi.org/10.1264/jisme2.ME13151>.
- 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. <https://doi.org/10.1016/j.cell.2016.12.021>.
- Xia W. 2017. Competition for iron between host and pathogen: a structural case study on *Helicobacter pylori*. *Methods Mol Biol* 1535:65–75. [https://doi.org/10.1007/978-1-4939-6673-8\\_5](https://doi.org/10.1007/978-1-4939-6673-8_5).
- Rea K, Dinan TG, Cryan JF. 2016. The microbiome: a key regulator of stress and neuroinflammation. *Neurobiol Stress* 4:23–33. <https://doi.org/10.1016/j.ynstr.2016.03.001>.
- 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. <https://doi.org/10.1016/j.cmet.2014.10.016>.
- Mousa WK, Athar B, Merwin NJ, Magarvey NA. 2017. Antibiotics and specialized metabolites from the human microbiota. *Nat Prod Rep* 34:1302–1331. <https://doi.org/10.1039/c7np00021a>.
- Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A, Wade WG. 2010. The human oral microbiome. *J Bacteriol* 192:5002–5017. <https://doi.org/10.1128/JB.00542-10>.
- Costalonga M, Herzberg MC. 2014. The oral microbiome and the immunobiology of periodontal disease and caries. *Immunol Lett* 162:22–38. <https://doi.org/10.1016/j.imlet.2014.08.017>.

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. <https://doi.org/10.1038/sj.bdj.2016.865>.
18. Marsh PD. 2003. Are dental diseases examples of ecological catastrophes? *Microbiology* 149:279–294. <https://doi.org/10.1099/mic.0.26082-0>.
19. Bowen WH. 2002. Do we need to be concerned about dental caries in the coming millennium? *Crit Rev Oral Biol Med* 13:126–131. <https://doi.org/10.1177/154411130201300203>.
20. Dye BA, Thornton-Evans G. 2010. Trends in oral health by poverty status as measured by Healthy People 2010 objectives. *Public Health Rep* 125:817–830. <https://doi.org/10.1177/003335491012500609>.
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. <https://doi.org/10.1038/sj.bdj.2018.217>.
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* 8:1659–1672. <https://doi.org/10.1038/ismej.2014.23>.
23. Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, Romero H, Simón-Soro A, Pignatelli M, Mira A. 2012. The oral metagenome in health and disease. *ISME J* 6:46–56. <https://doi.org/10.1038/ismej.2011.85>.
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. <https://doi.org/10.1038/s41522-017-0031-4>.
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. <https://doi.org/10.3389/fcimb.2015.00025>.
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. <https://doi.org/10.3389/fcimb.2014.00108>.
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. <https://doi.org/10.1186/s13073-015-0153-3>.
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. <https://doi.org/10.1038/srep01843>.
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. <https://doi.org/10.1093/nar/gkx319>.
30. Medema MH, Takano E, Breitling R. 2013. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol Biol Evol* 30:1218–1223. <https://doi.org/10.1093/molbev/mst025>.
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. <https://doi.org/10.1038/srep37479>.
32. Doering TL. 2009. How sweet it is! Cell wall biogenesis and polysaccharide capsule formation in *Cryptococcus neoformans*. *Annu Rev Microbiol* 63:223–247. <https://doi.org/10.1146/annurev.micro.62.081307.162753>.
33. Krachler AM, Orth K. 2013. Targeting the bacteria-host interface: strategies in anti-adhesion therapy. *Virulence* 4:284–294. <https://doi.org/10.4161/viru.24606>.
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* 24:327–333.
35. Mazmanian SK, Liu CH, Zhanabos AO, Kasper DL. 2005. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* 122:107–118. <https://doi.org/10.1016/j.cell.2005.05.007>.
36. Mazmanian SK, Round JL, Kasper DL. 2008. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* 453:620–625. <https://doi.org/10.1038/nature07008>.
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* 187:3980–3989. <https://doi.org/10.1128/JB.187.12.3980-3989.2005>.
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* 265:11–17. <https://doi.org/10.1111/j.1574-6968.2006.00459.x>.
39. Shanker E, Federle MJ. 2017. Quorum sensing regulation of competence and bacteriocins in *Streptococcus pneumoniae* and *mutans*. *Genes (Basel)* 8:15. <https://doi.org/10.3390/genes8010015>.
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* 17:247–253. <https://doi.org/10.1002/cbic.201500474>.
41. Fischbach MA, Walsh CT. 2006. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* 106:3468–3496. <https://doi.org/10.1021/cr0503097>.
42. de las Heras B, Hortelano S. 2009. Molecular basis of the anti-inflammatory effects of terpenoids. *Inflamm Allergy Drug Targets* 8:28–39. <https://doi.org/10.2174/187152809787582534>.
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, Dürstherus S, Edwards DJ, Fewer DP, Garg N, Geiger C, Gomez-Escribano JP, Greule A, Hadjithomas M, Haines AS, Helfrich EJN, Hillwig ML, Ishida K, Jones AC, Jones CS, Jungmann K, Kegler C, Kim HU, Kötter P, Krug D, Masschelein J, Melnik AV, Mantovani SM, Monroe EA, Moore M, Moss N, Nützmann H-W, Pan G, Pati A, Petras D, Reen FJ, Rosconi F, Rui Z, Tian Z, Tobias NJ, Tsunematsu Y, Wiemann P, Wyckoff E, Yan X, Yim G, et al. 2015. Minimum information about a biosynthetic gene cluster. *Nat Chem Biol* 11:625–631. <https://doi.org/10.1038/nchembio.1890>.
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. <https://doi.org/10.1371/journal.pone.0122075>.
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* 46:1407–1417. <https://doi.org/10.1128/JCM.01410-07>.
46. Friedman J, Alm EJ. 2012. Inferring correlation networks from genomic survey data. *PLoS Comput Biol* 8:e1002687. <https://doi.org/10.1371/journal.pcbi.1002687>.
47. Muller S, Strack SN, Hoeffler BC, Straight PD, Kearns DB, Kirby JR. 2014. Bacillaene and sporulation protect *Bacillus subtilis* from predation by *Myxococcus xanthus*. *Appl Environ Microbiol* 80:5603–5610. <https://doi.org/10.1128/AEM.01621-14>.
48. Cooper EC, Curtis N, Cranswick N, Gwee A. 2014. Pristinamycin: old drug, new tricks?. *J Antimicrob Chemother* 69:2319–2325. <https://doi.org/10.1093/jac/dku167>.
49. Simon-Soro A, Mira A. 2015. Solving the etiology of dental caries. *Trends Microbiol* 23:76–82. <https://doi.org/10.1016/j.tim.2014.10.010>.
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* 44:1078–1080. <https://doi.org/10.1128/AAC.44.4.1078-1080.2000>.
51. Healy AR, Herzog SB. 2017. Molecular basis of gut microbiome-associated colorectal cancer: a synthetic perspective. *J Am Chem Soc* 139:14817–14824. <https://doi.org/10.1021/jacs.7b07807>.
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 W-T, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu C-C, Floros DJ, Gavilan RG, Kleigrew 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 C-C, Yang Y-L, Humpf H-U, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, et al. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 34:828–837. <https://doi.org/10.1038/nbt.3597>.
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 comprehen-

- sive, computable taxonomy. *J Cheminform* 8:61. <https://doi.org/10.1186/s13321-016-0174-y>.
54. Ruttkies C, Schymanski EL, Wolf S, Hollender J, Neumann S. 2016. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J Cheminform* 8:3. <https://doi.org/10.1186/s13321-016-0115-9>.
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. <https://doi.org/10.1093/database/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. <https://doi.org/10.1007/s11306-007-0082-2>.
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. <https://doi.org/10.1371/journal.pcbi.1006089>.
58. Henson BS, Wong DT. 2010. Collection, storage, and processing of saliva samples for downstream molecular applications. *Methods Mol Biol* 666: 21–30. [https://doi.org/10.1007/978-1-60761-820-1\\_2](https://doi.org/10.1007/978-1-60761-820-1_2).
59. 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. <https://doi.org/10.1128/mBio.01012-14>.
60. 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. <https://doi.org/10.1128/mBio.01926-14>.
61. Vazquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. 2013. EMPERor: a tool for visualizing high-throughput microbial community data. *Gigascience* 2:16. <https://doi.org/10.1186/2047-217X-2-16>.
62. 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. <https://doi.org/10.1186/1471-2105-11-395>.