Supragingival Plaque Microbiome Ecology and Functional Potential in the Context of Health and Disease

Josh L. Espinoza,a Derek M. Harkins,b Manolito Torralba,c Andres Gomez,c,⁎ Sarah K. Highlander,c,⁎ Marcus B. Jones,d Pamela Leong,a Richard Saffery,a Michelle Bockmann,f Claire Kuelbs,c Jason M. Inman,c Toby Hughes,f Jeffrey M. Craig,g Karen E. Nelson,b,c Chris L. Duponta

aDepartment of Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, California, USA
bDepartments of Human Biology and Genomic Medicine, J. Craig Venter Institute, Rockville, Maryland, USA
cDepartments of Human Biology and Genomic Medicine, J. Craig Venter Institute, La Jolla, California, USA
dHuman Longevity Institute, La Jolla, California, USA
eMurdoch Children’s Research Institute and Department of Pediatrics, University of Melbourne, Royal Children’s Hospital, Parkville, VIC, Australia
fSchool of Dentistry, The University of Adelaide, Adelaide, SA, Australia
gCentre for Molecular and Medical Research, School of Medicine, Deakin University, Geelong, VIC, Australia

ABSTRACT To address the question of how microbial diversity and function in the oral cavities of children relates to caries diagnosis, we surveyed the supragingival plaque biofilm microbiome in 44 juvenile twin pairs. Using shotgun sequencing, we constructed a genome encyclopedia describing the core supragingival plaque microbiome. Caries phenotypes contained statistically significant enrichments in specific genome abundances and distinct community composition profiles, including strain-level changes. Metabolic pathways that are statistically associated with caries include several sugar-associated phosphotransferase systems, antimicrobial resistance, and metal transport. Numerous closely related previously uncharacterized microbes had substantial variation in central metabolism, including the loss of biosynthetic pathways resulting in auxotrophy, changing the ecological role. We also describe the first complete Gracilibacteria genomes from the human microbiome. Caries is a microbial community metabolic disorder that cannot be described by a single etiology, and our results provide the information needed for next-generation diagnostic tools and therapeutics for caries.

IMPORTANCE Oral health has substantial economic importance, with over $100 billion spent on dental care in the United States annually. The microbiome plays a critical role in oral health, yet remains poorly classified. To address the question of how microbial diversity and function in the oral cavities of children relate to caries diagnosis, we surveyed the supragingival plaque biofilm microbiome in 44 juvenile twin pairs. Using shotgun sequencing, we constructed a genome encyclopedia describing the core supragingival plaque microbiome. This unveiled several new previously uncharacterized but ubiquitous microbial lineages in the oral microbiome. Caries is a microbial community metabolic disorder that cannot be described by a single etiology, and our results provide the information needed for next-generation diagnostic tools and therapeutics for caries.

KEYWORDS Streptococcus, metabolism, metagenomics, microbial ecology, oral microbiology

The oral microbiome is a critical component of human health. There are an estimated 2.4 billion cases of untreated tooth decay worldwide, making the study of carious lesions, and their associated microbiota, a topic of utmost importance from the interest

Received 31 July 2018 Accepted 17 October 2018 Published 27 November 2018

Citation Espinoza JL, Harkins DM, Torralba M, Gomez A, Highlander SK, Jones MB, Leong P, Saffery R, Bockmann M, Kuelbs C, Inman JM, Hughes T, Craig JM, Nelson KE, Dupont CL. 2018. Supragingival plaque microbiome ecology and functional potential in the context of health and disease. mBio 9:e01631-18. https://doi.org/10.1128/mBio.01631-18.

Editor David A. Relman, VA Palo Alto Health Care System
Copyright © 2018 Espinoza et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.
Address correspondence to Chris L. Dupont, cdupont@jcvi.org.
* Present address: Andres Gomez, Department of Animal Science, University of Minnesota, St. Paul, Minnesota, USA; Sarah Highlander, Translational Genomics Research Institute, Flagstaff, Arizona, USA.
of public health (1). Oral disease poses a considerable socioeconomic burden within the United States; the national dental care expenditure exceeded $113 billion in 2014 (2). Despite the widespread impact of oral diseases, much of the oral microbiome is poorly characterized; of approximately 700 species identified (3), 30% have yet to be cultivated (4). On average, less than 60% of oral metagenomic reads can be mapped to reference databases at species-level specificity (5). The characterization of functional potential plasticity, (intra/inter)organismal interactions, and responses to environmental stimuli in the oral microbiomes are essential in interpreting the complex narratives that orchestrate host phenotypes.

Oral microbiomes are important not only in the immediate environment of the oral cavity, but also systemically as well. For instance, although dental caries, the most common chronic disease in children (6), is of a multifactorial nature, it usually occurs when sugar is metabolized by acidogenic plaque-associated bacteria in biofilms, resulting in increased acidity and dental demineralization, a condition exacerbated by frequent sugar intake (7). Similarly, in periodontitis, a bacterial community biofilm (plaque) elicits local and system inflammatory responses in the host, leading to the destruction of periodontal tissue, pocket formation, and tooth loss (8, 9). Likewise, viruses and fungi in oral tissues can trigger gingival lesions associated with herpes and candidiasis (10) in addition to cancerous tissue in oral cancer (11). The connections between oral microbes and health extend beyond the oral cavity as cardiometabolic, respiratory, and immunological disorders and some gastrointestinal cancers are thought to have associations with oral microbes (12–15). For example, the enterosalivary nitrate cycle originates from dissimilatory nitrite reduction in the oral cavity but influences nitric oxide production throughout the body (16). Consequently, unraveling the forces that shape the oral microbiome is crucial for the understanding of both oral and broader systemic health.

*Streptococcus mutans* has been the focal point of research involving cariogenesis and the associated dogma since the species discovery in 1924 (17). However, prior to the characterization of *S. mutans*, dental caries etiology was believed to be community driven rather than the product of a single organism (18). There is no question that *S. mutans* can be a driving factor in cariogenesis; *S. mutans* colonized onto sterilized teeth form caries (17). However, the source of the precept transpired from an era inherently subject to culturing bias. Mounting evidence exists supporting the claim that *S. mutans* is not the sole pathogenic agent responsible for carious lesions (19–25) with characterization of oral biofilms by metabolic activity and taxonomic composition rather than only listing dominant species (26). A recent large-scale 16S rRNA gene survey of juvenile twins by our team found that while there are heritable microbes in supragingival plaque, the taxa associated with the presence of caries are environmentally acquired and seem to displace the heritable taxa with age (27). Another surprising facet of this recent study was the lack of a statistically robust association between the relative abundance of *S. mutans* and the presence of carious lesions.

The nature of the methods used in that recent 16S rRNA amplicon study addressed questions related to community composition but did provide the means to examine links between microbiome phylogenetic diversity, functional potential, and host phenotype. Here, we characterize the microbiome in supragingival biofilm swabs of monozygotic (MZ) and dizygotic (DZ) twin pairs, including children both with and without dental caries, using shotgun metagenomic sequencing, coassembly, and binning techniques.

**RESULTS**

**Data overview.** The supragingival microbiomes of 44 twin pairs were sampled via tooth swab, with the twin pairs being chosen based on 16S rRNA data (see “Study design” in Materials and Methods). Of the 88 samples, 50 (56.8%) were positive for caries. Shotgun sequencing of whole-community DNA was conducted to produce an average of 3 million fragments of non-human DNA per sample. Human reads, accounting for ~20% of the total, were removed, and all data were subjected to metagenomics...
coassembly using metaSPAdes and subsequent quality control with QUAST (28). The contigs were annotated for phylogenetic and functional content and then binned by k-mer utilization and across sample representation into genome bins. Thirty-four of these passed quality control and were graduated to metagenome assembled genomes (MAGs) (Table 1; see Table S5 in the supplemental material; see also Fig. 5). MAGs are sets of assembled contigs determined to likely originate from the same organism based on basic nucleotide usage (e.g., k-mer) or coverage across different sample sets; there are established methods for determining if these are pure (29) or complete (30). These are likely to not capture genomic “island regions” that are specific to individual samples but can be considered a consensus core genome for a specific species (31). Contigs annotated as Streptococcus interfered with the k-mer binning but were also easily identified; therefore, we removed these contigs to facilitate the binning of other genomes. The Streptococcus population was examined using the pangenomic analysis package MIDAS (5). Reads from each subject were mapped back to the genome bins to create an assembly-linked matrix of relative abundance, phylogenetically characterized genome origin, genome bin functional contents, variance component estimation, and the phenotypic state of the host (caries or not).

Core supragingival microbiome community, phenotype-specific taxonomic enrichment, and cooccurrence. We observed a genome-resolved microbiome dominated by Streptococcus and Neisseria across the cohort with relatively little variability (Fig. 1). Many of the taxa previously associated with supragingival biofilms were found (4), but we report the first binned genomes for Gracilbacteria, which have only been previously observed in the oral cavity using 16S rRNA screens. Three genomes for the TM7 lineage (32–34) were also recovered with substantial differences in genome architecture. Each of the 88 subjects contains the same genomes; these genomes comprise a core supragingival plaque microbiome that has relatively minor but statistically significant changes in proportional abundance. Sequences and genome bins for Archaea, Eukaryota (other than the host), or viruses were not observed. The core supragingival plaque microbiome defined here certainly includes far more than the 34 genome bins and 119 Streptococcus strains recovered here, though they are by nature numerically rare members of the community. These rare organisms could be later captured using deeper sequencing and a combination of reference-based and assembly-centric binning in an iterative manner.

The presence of caries is associated with statistically significant enrichments in the relative abundance of specific taxa when comparing the caries-negative and caries-positive cohorts (Fig. 1B). Namely, the abundance profiles of Streptococcus, Catonella morbi, and Granulicatella elegans were enriched in caries-positive subjects, while Tannerella forsythia, Gracilbacteria, Capnocytophaga gingivalis, Bacteroides sp. oral taxon 274, and Campylobacter gracilis were more abundant in the healthy cohort (Fig. 1A; Table S2). Community composition profiles were different according to caries progression; Alloprevotella rava, Porphyromonas gingivalis, Neisseria oralis, and Bacteroides oral taxon 274 were significantly enriched in subjects with carious enamel only (Fig. 1A). The Streptococcus community across all microbiomes contained 119 strain genomes detected at high resolution (Fig. 2). Within this mixed Streptococcus community, 86% is composed of S. sanguinis, S. mitis, S. oralis, and an unnamed Streptococcus sp. As with the bulk community, we found statistically significant strain abundance variations associated with the presence of caries in S. parasanguinis, S. australis, S. salivarius, S. intermedius, S. constellatus, and S. vestibularis; this enrichment was specific to rare strains that make up less than 5% of the Streptococcus (Fig. 2A and B; Table S2). During the global progression of carious lesions from enamel to dentin, we observed a modest statistical enrichment (P < 0.05) in S. parasanguinis and S. cristatus (Table S2). However, these are general trends considering phenotype subsets for a single time point per individual; this data set does not support a longitudinal analysis, which will be the topic of future studies.

To investigate the role of community alpha diversity in caries, we compared Shannon entropy measures for the healthy and diseased cohorts separately in both the
### TABLE 1 Microbial community description

| Taxonomy          | Genus                        | Species                  | Group | Completeness | Contamination | CheckM estimation | Genome | Variance component estimation | No. of ORFs |
|-------------------|------------------------------|--------------------------|-------|--------------|---------------|-------------------|--------|-------------------------------|------------|
|                   | Actinomycetaeae              | Actinomyces               | Group I.A | 4.44         | 0             | 0.002546        | 74.289 | 0.27712                       | 1,799,133  |
|                   | Actinomycetaeae              | Actinomyces               | Group I.B | 0            | 0             | 0.002165        | 78.45  | 0.18824                       | 1,541,794  |
|                   | Prevotella                   | Alloprevotella rava      | Group II.A | 5.33         | 0             | 0.001236        | 89.42  | 0.36261                       | 2,086,006  |
|                   | Prevotella                   | Alloprevotella rava      | Group II.B | 44.75        | 0             | 0.002046        | 89.51  | 0.46292                       | 3,203,447  |
|                   | Unclassified Bacteroidetes   | Bacteroidetes oral taxon 274 |           | 40.66        | 0             | 0.002016        | 96.655 | 0.35578                       | 2,486,193  |
|                   | Campylobacteraceae           | Campylobacter              | Group I   | 0            | 0             | 0.001236        | 49.92  | 0.47692                       | 2,010,123  |
|                   | Campylobacteraceae           | Campylobacter              | Group II   | 55.88        | 0             | 0.001236        | 85.89  | 0.31143                       | 3,163,899  |
|                   | Unclassified Saccharibacteria| "Candidatus Saccharimonas" <i>Saccharibacteria</i> | Group II.A | 9.95         | 0             | 0.001236        | 74.84  | 0.14287                       | 596.323    |
|                   | Unclassified Saccharibacteria| "Candidatus Saccharimonas" <i>Saccharibacteria</i> | Group II.B | 11.41        | 0             | 0.001236        | 65.91  | 0.2286                         | 1,065,796  |
|                   | Flavobacteriaceae            | Capnocytophaga             | Group I   | 207.14       | 0             | 0.001236        | 85.34  | 0.47343                       | 7,550,959  |
|                   | Flavobacteriaceae            | Capnocytophaga             | Group II   | 49.06        | 0             | 0.001236        | 73.46  | 0.82298                       | 3,863,326  |
|                   | Lachnospiraceae              | Catonella                 | Group I   | 28.29        | 0             | 0.001236        | 72.41  | 0.32865                       | 2,490,472  |
|                   | Lachnospiraceae              | Catonella                 | Group II   | 4.5          | 0             | 0.001236        | 89.34  | 0.37219                       | 2,342,592  |
|                   | Unclassified Gracilibacteria  | Unclassified Gracilibacteria | Group I.A | 56.66        | 0             | 0.001236        | 79.31  | 0.79204                       | 1,663,606  |
|                   | Unclassified Gracilibacteria  | Unclassified Gracilibacteria | Group I.B | 88.17        | 0             | 0.001236        | 93.1   | 0.82443                       | 1,991,479  |
|                   | Carnobacteriaceae            | Granulicatella            | Group I   | 49.96        | 0             | 0.001236        | 86.21  | 0.62874                       | 2,163,485  |
|                   | Neisseriaceae                | Kingella                  | Group I   | 3.45         | 0             | 0.001236        | 79.95  | 0.6778                         | 1,689,409  |
|                   | Neisseriaceae                | Kingella                  | Group II   | 5.17         | 0             | 0.001236        | 66.22  | 0.78043                       | 1,716,807  |
|                   | Lachnospiraceae              | Lachnospiraceae bacterium oral taxon 500 | Group I   | 243.39        | 0             | 0.001236        | 95.69  | 0.35816                       | 3,580,1036 |
|                   | Burkholderiaceae             | Lautropia                 | Group I   | 22.57        | 0             | 0.001236        | 93.97  | 0.51076                       | 4,240,867  |
|                   | Leptotrichiaceae             | Leptotrichia              | Group I   | 7.12         | 0             | 0.001236        | 68.13  | 0.39736                       | 1,129,771  |
|                   | Moraxellaceae                | Moraxella                 | Group I   | 16.09        | 0             | 0.001236        | 100    | 0.33461                       | 2,551,229  |
|                   | Neisseriaceae                | Neisseria                 | Group I   | 7.68         | 0             | 0.001236        | 84.5   | 0.3099                        | 7,906,650  |
|                   | Neisseriaceae                | Neisseria                 | Group II   | 1.72         | 0             | 0.001236        | 84.5   | 0.3099                        | 7,906,650  |
|                   | Porphyromonadaceae           | Porphyromonas              | Group I   | 0            | 0             | 0.001236        | 79.31  | 0.5141                        | 1,549,578  |
|                   | Porphyromonadaceae           | Porphyromonas              | Group II   | 1.03         | 0             | 0.001236        | 85.11  | 0.51461                       | 2,201,265  |
|                   | Prevotella                   | Prevotella                | Group I   | 25.24        | 0             | 0.001236        | 69.83  | 0.41557                       | 2,593,194  |
|                   | Prevotella                   | Prevotella                | Group II   | 20.69        | 0             | 0.001236        | 69.51  | 0.40583                       | 4,133,493  |
|                   | Prevotella                   | Prevotella                | Group V.A  | 21.07        | 0             | 0.001236        | 69.69  | 0.33461                       | 2,023,841  |
|                   | Staphylococcaceae            | Staphylococcus             | Group I   | 12.07        | 0             | 0.001236        | 79     | 0.39938                       | 1,754,497  |
|                   | Streptococcaceae             | Streptococcus              | Group I   | 288.36       | 0             | 0.001236        | 99.22  | 0.3098                        | 1,243,852  |
|                   | Tannierellaceae              | Tannierella               | Group I   | 1.72         | 0             | 0.001236        | 80.09  | 0.46838                       | 2,246,045  |
|                   | Veillonellaceae              | Veillonella               | Group I   | 91.07        | 0             | 0.001236        | 91.07  | 0.387835111                  | 5,471      |

*Recovered draft genomes with taxonomy, identifier mapping, CheckM statistics, variance component estimation (i.e., ACE model), and genome statistics.*
core supragingival and strain-level Streptococcus communities. At the bulk community scale, healthy individuals (mode $/H_11005^3.793$) had statistically greater taxonomic diversity than the diseased subset (mode $/H_11005^2.796$) as shown in Fig. 1C. Adversely, the Streptococcus strain-level analysis revealed a statistical enrichment of diversity within the diseased cohort (mode $/H_11005^5.452$) compared to the healthy subset (mode $/H_11005^5.177$) as shown in Fig. 2D. An examination of a genome cooccurrence network revealed a distinct topology illustrated in Fig. 3. For each genome, we also computed PageRank centrality, a variant of eigenvector centrality (35), which allows us to measure the influence of bacterial nodes within our cooccurrence networks. A select few MAGs form a cooccurrence cluster (Fig. 3; cluster 3, 4; $n = 6$) containing the 6 most highly connected taxa, 3 of which are statistically enriched in the diseased cohort while 2 are enriched in the healthy cohort. This microbial clique contains the highest-ranking PageRank centrality MAGs in the network and can be cleanly divided into a caries-affiliated subset and a subset enriched in bacteria associated with healthy phenotypes. The Streptococcus strain-level cooccurrence network topology varied greatly between caries-positive and caries-negative individuals with red indicating enrichment of taxa in the caries-positive cohort. (Top) Caries-positive versus caries-negative individuals with red indicating enrichment of taxa in the caries-positive cohort. (Bottom) Subjects with caries that has progressed to the dentin layer versus enamel-only caries with blue denoting taxa enriched in the enamel. (B) Relative abundance of TPM values for each genus of de novo assembly grouped by caries-positive and caries-negative cohorts. Actinomyces $= 0.36$, Alloprevotella $= 0.0515$, Campylobacter $= 0.0124$, "Candidatus Saccharimonas" $= 0.107$, Capnocytophaga $= 0.000875$, Cardiobacterium $= 0.0621$, Catonella $= 0.000264$, Corynebacterium $= 0.0552$, Granulicatella $= 0.00255$, Kingella $= 0.252$, Lachnoclostridium $= 0.0791$, Lautropia $= 0.0829$, Moraxella $= 0.215$, Neisseria $= 0.455$, Porphyromonas $= 0.136$, Prevotella $= 0.12$, Riemenella $= 0.0621$, Streptococcus $= 0.000901$, Tannerella $= 0.0061$, unclassified Bacteroidetes $= 0.0111$, unclassified Gracilibacteria $= 0.0316$, Veillonella $= 0.471$. (C) Kernel density estimation of Shannon entropy alpha diversity distributions for caries-positive (red) and caries-negative (gray) subjects calculated from normalized core supragingival community composition. Vertical lines indicate the mode of the kernel density estimate distributions for each cohort. Statistical significance ($P = 0.0109$).
negative cohorts (Fig. 2C). Specifically, we observed that healthy subjects harbored a Streptococcus community influenced by a few key strains with high PageRank measures. In contrast to the healthy cohort, caries-affected individuals harbored a much more dispersed strain-level Streptococcus cooccurrence network.

Phylogeny-linked functional profiles across caries phenotypes. To determine if caries presence is associated with trends in functional potential of the supragingival microbiome, we tested for functional enrichment in both a taxonomically anchored and unanchored fashion (Fig. 4). The “unanchored approach” simply tests whether KEGG modules are statistically enriched in the taxa statistically significant to caries status. The “anchored approach” combines the functional potential of a MAG, represented by a KEGG module completion ratio (MCR), with its normalized abundance values collapsing into a single phylogenomically binned functional potential (PBFP) profile. PBFP encodes information regarding genome size, community proportions, and functional potential for each sample which can be used in parallel with subject-specific metadata for machine learning and statistical methods. The overlap of these two methods identified 37 KEGG metabolic modules positively associated with caries-positive microbiomes (Fig. 4 and Table 2). Numerous phosphotransferase sugar uptake systems were more abundant in the caries-positive communities, including systems for the uptake of glucose (M00809, M00265, and M00266), galactitol (M00279), lactose (M00281), maltose (M00266), alpha glucoside (M00268), cellobiose (M00275), and N-acetylgalactosamine (M00277). Also positively associated with diseased host phenotypes was an increased abundance of numerous two-component histidine kinase-response regulator systems, including AgrC-AgrA (exoprotein synthesis) (M00495), BceS-BceR (bacitracin transport) (M00492), LiaS-LiaR (cell wall stress response) (M00481), and LytS-LytR (M00492) along
with those putatively associated with virulence regulation such as SaeS-SaeR (staphylococcal virulence regulation) (M00468), Ihk-Irr (M00719), and ArlS-ArlR (M00716). Note that while these two-component systems have been characterized in specific lineages (e.g., Staphylococcus aureus), the KEGG modules detect their orthologs across diverse lineages. Modules associated with xenobiotic efflux pump transporters were enriched, including BceAB transporter (bacitracin resistance) (M00738), multidrug resistance EfrAB (M00706), multidrug resistance MdiAB/SmdAB (M00707), and tetracycline resistance TetAB(46) (M00635) along with antibiotic resistance, including cationic antimicrobial peptide (CAMP) resistance (M00725) and nisin resistance (M00754). Furthermore, numerous trace metal transport systems were enriched, including those for manganese/zinc (M00791, M00792), nickel (M00245, M00246), and cobalt (M00245).

**Metabolic tradeoffs between closely related species.** The k-mer-based genome binning unveiled several closely related pairs or triplets of genomes (Fig. 5A). These include Actinomycetes (group I), Alloprevotella (group II), “Candidatus Saccharimonas” TM7 (group III), Gracilibacteria (group IV), and Prevotella (group V) lineages (Table 1). This allowed us to examine the extent to which metabolic potential varies for ubiquitous microbial taxa in the human supragingival microbiome, two of which are from recently identified bacterial phyla (i.e., TM7 and Gracilibacteria). As shown in Fig. 5B, substantial variations in metabolic potential distinguish the different MAGs with emphasis on the Alloprevotella group. Similarly, the MAGs also display different abun-
| KEGG module ID | Metabolic modules significant to host phenotype<sup>a</sup> |
|---------------|-----------------------------------------------------|
| M00495       | Two-component regulatory system                      |
| M00716       | Two-component regulatory system                      |
| M00550       | Cofactor and vitamin biosynthesis                     |
| M00738       | Drug efflux transporter/pump                         |
| M00469       | Two-component regulatory system                      |
| M00170       | Carbon fixation                                      |
| M00925       | Terpenoid backbone biosynthesis                      |
| M00245       | Metallic cation, iron siderophore, and vitamin B<sub>2</sub> transport system |
| M00455       | Histidine metabolism                                 |
| M00375       | Carbon fixation                                      |
| M00719       | Two-component regulatory system                      |
| M00481       | Two-component regulatory system                      |
| M00492       | Two-component regulatory system                      |
| M00791       | Metallic cation, iron siderophore, and vitamin B<sub>2</sub> transport system |
| M00792       | Metallic cation, iron siderophore, and vitamin B<sub>2</sub> transport system |
| M00706       | Drug efflux transporter/pump                         |
| M00707       | Drug efflux transporter/pump                         |
| M00298       | ABC-2-type and other transport systems               |
| M00205       | Saccharide, polyol, and lipid transport systems       |
| M00246       | Metallic cation, iron siderophore, and vitamin B<sub>2</sub> transport system |
| M00754       | Drug resistance                                      |
| M00277       | Phosphotransferase system (PTS)                      |
| M00268       | Phosphotransferase system (PTS)                      |
| M00275       | Phosphotransferase system (PTS)                      |
| M00279       | Phosphotransferase system (PTS)                      |
| M00287       | Phosphotransferase system (PTS)                      |
| M00809       | Phosphotransferase system (PTS)                      |
| M00265       | Phosphotransferase system (PTS)                      |
| M00281       | Phosphotransferase system (PTS)                      |
| M00266       | Phosphotransferase system (PTS)                      |
| M00603       | Saccharide, polyol, and lipid transport system        |
| M00589       | Phosphate and amino acid transport system            |
| M00468       | Two-component regulatory system                      |
| M00683       | Central carbohydrate metabolism                      |
| M00635       | Drug efflux transporter/pump                         |
| M00159       | ATP synthesis                                        |

<sup>a</sup>KEGG modules significant in both anchored and unanchored approaches with hierarchical descriptions.
FIG 5 Metagenome assembled genomes recovered in assembly. Multiple MAGs of individual species and their functional differences. (A) t-SNE embeddings of center-log-ratio-transformed 5-mer profiles for each contig. Gold contigs have higher C and E coefficients in the ACE model distinguishing environmental acquisition while teal contigs indicate high A coefficients of heritable organisms. (B) Differences in functional potential of A. rava strains where one has a high heritability coefficient while the other has high environmentally associated coefficients according to the ACE model.
dances across subjects, cooccurrence patterns relative to other MAGs, and, in one case, heritability (Fig. 1A, 5A, and 6A).

The two *Alloprevotella* genomes recovered (denoted as MAGs II.A and II.B in Fig. 5A) contain one representative that has a high genetic component and another that is environmentally acquired, differing in the potential for nitrate reduction, biosynthesis of sulfur-containing amino acids, PRPP, and sugar nucleotides. *Alloprevotella* MAG II.B contains complete metabolic pathways for phosphoribosyl pyrophosphate (PRPP) biosynthesis from ribose-5-phosphate (M00005) and cysteine biosynthesis from serine (M00021); these metabolic modules are completely absent in the *Alloprevotella* MAG.
II.A. Similarly, the potential for dissimilatory nitrate reduction (M00530) is lacking in the heritable Alloprevotella MAG II.A but is present in MAG II.B.

Three MAGs for the recently discovered candidate phylum TM7 (32–34) were recovered, with each of them showing high heritability estimates. The recent cultivation of TM7 from the oral microbiome revealed a parasitic nature with Actinobacteria as hosts (34). With our recovered TM7 genomes, we observe very little cooccurrence patterns between TM7 MAG III.C and any other taxa. Curiously, TM7 MAGs III.A and III.B have an inverse relationship, in terms of cooccurrence profiles ($\rho = -0.34$), suggesting different hosts or functional niches. TM7 MAG III.B is positively correlated with taxa associated with the caries-positive subjects—Streptococcus, C. morbi, and G. elegans—while MAG III.A is negatively associated with these taxa. TM7 MAG III.A contains complete pathways for dTDP-\(L\)-rhamnose biosynthesis (M00793), F-type ATPase (M00157), putative polar amino acid transport system (M00236), energy-coupling factor transport system (M00582), and SenX3-RegX3 (phosphate starvation response) two-component regulatory system (M00443) that are entirely absent in TM7 MAG III.B. TM7 MAGs III.A and III.B provide unique functional capabilities to the entire community, containing components for F420 biosynthesis (M00378) and SasA-RpaAB (circadian timing-mediating) two-component regulatory system (M00467), respectively.

Two MAGs with phylogenetic affinity to Actinomyces have been recovered in our analysis. Actinomyces MAG I.A exclusively has all of the components for the rhamnose transport system (M00220) while Actinomyces MAG I.B exclusively has components for the manganese transport system (M00316) compared to the rest of the community. Both Actinomyces MAG I.A and MAG I.B are the only genomes to contain components for the DevS-DevR (redox response) two-component regulatory system (M00482).

Two Prevotella sp. MAGs most similar to the Human Microbiome Project (HMP) oral taxon 472 (36) have been recovered in our analysis. Both Prevotella MAGs uniquely contain complete pathways for CAM (crassulacean acid metabolism) (M00169), pyruvate oxidation via the conversion of pyruvate to acetyl-CoA (M00307), PRPP biosynthesis (M00005), beta-oxidation, acyl-CoA synthesis (M00086), adenine ribonucleotide biosynthesis via IMP-to-adenosine (di/tri)phosphate conversion (M00049), guanine ribonucleotide biosynthesis via IMP-to-guanosine (di/tri)phosphate conversion (M00050), coenzyme A biosynthesis from pantothenate conversion (M00120), cell division transport system (M00256), and putative ABC transport system (M00258).

Two MAGs for Gracilibacteria were recovered. Mapping of recent human oral WGS samples from the extended Human Microbiome Project (37) to these genomes show that they are also found in North American subjects (Table S4). Metabolic reconstructions describe the capability for anoxic fermentation of glucose to acetate, as well as auxotrophies for vitamin B12 and a preponderance of type II and IV secretion systems. Both Gracilibacteria cooccurrence profiles show negative relationships with Veillonella sp. oral taxon 780 while showing positive relationships with Capnocytophaga gingivalis, suggesting competitive and commensal relationships, respectively. Gracilibacteria MAG IV.B is the only organism in our recovered oral community that has components for the BasS-BasR (antimicrobial peptide resistance) two-component regulatory system (M00451) and the cationic antimicrobial peptide (CAMP) resistance amBCADTEF operon (M00721) while MAG IV.A exclusively contains components for the NrsS-NrsR (nickel tolerance) two-component regulatory system (M00464). Gracilibacteria MAG IV.A contains complete pathways for phosphate acetyltransferase-acetate kinase (M000579), the conversion of acetyl-CoA into acetate, while being completely absent in MAG IV.B. Gracilibacteria MAG IV.B contains complete components of the ABC-2 type transport system (M00254), which is completely absent in MAG IV.A.

**DISCUSSION**

**Community-scale profiles for caries.** In this study, we took a hybrid approach of utilizing read mapping for taxa with excellent genome representation (e.g., Streptococcus) while generating new reference genomes *de novo* for less well represented
organisms. These genome bins also allow for improved databases for future read mapping analyses and indeed identified taxa present in previous oral microbiome studies (37) that were overlooked. Our results suggest that caries status can be accurately associated with and, potentially, diagnosed by profiling specific taxa beyond Streptococcus mutans. We observed a loss of community diversity within the diseased subjects (Fig. 1C), a finding previously reported in 16S rRNA amplicon studies (21), with an increase in Streptococcus strain diversity (Fig. 2D). The observed differences in the overall community structure, as seen from our cohort study, suggest that profiling the abundance of multiple taxa may present an opportunity for caries diagnosis and preventative methods. This finding is also functionally relevant as many organisms, other than Streptococcus, statistically enriched within the caries-positive subjects are also both anaerobic glucose fermenters and acidogenic, a trend extending to strain-level Streptococcus populations as well. In the healthy cohort, the Streptococcus community exhibits less variation in terms of abundance and cooccurrence at the strain level. This is not the case in the diseased cohort, in which we observe higher Shannon entropy (Fig. 2). In an independent analysis, unsupervised clustering of diseased and healthy cohorts showed that ecological states in healthy subjects consistently have a Streptococcus community abundance that is either lower than or comparable to the rest of the healthy cohort (≤0.27), with the exception of cluster 7 (0.44), while being much more unpredictable in the diseased cohort (see Fig. S3 in the supplemental material). This phenomenon may be the result of grouping carious lesions of various degree into a single classification. However, this finding also supports the fact that caries is a dynamic and progressive disease whose rate of progression can change dramatically over short periods of time. These results suggest that caries onset cannot be described by a single bacterium but should be described as the perturbation of an entire ecosystem, consistent with hypotheses sourcing the disease from metabolic and community-driven origins.

The dynamics of the Streptococcus community, at both the genus and strain levels, explain the observation of low connectivity in the diseased cooccurrence network, supporting our hypothesis that caries onset is a community perturbation. The highly connected cluster in the microbial community cooccurrence network (Fig. 2C) can be interpreted as influential organisms that drive the dynamics of other taxa in the community. The aforementioned high-PageRank microbial clique can be subdivided into a subset enriched in diseased subjects (Fig. 3, cluster 3) and a subset enriched in healthy individuals (Fig. 3, cluster 4). It is possible that disease state is influenced by a balance between these microbes, though empirical in situ studies are required to move past speculation.

Caries is a community-scale metabolic disorder. Our metabolic and community composition analyses challenge a single-organism etiology for caries and coincide with previously published ecological perspectives (19, 23, 25, 26, 38, 39). These authors suggested that the changes in functional activities of the microbiome were the major cause of caries while simultaneously suggesting regime shifts (e.g., see reference 40) in community composition. They posited that increases in sugar catabolism potential are the key markers of caries progression; here we observed enrichments in nearly a dozen phosphotransferase sugar uptake systems. A novel insight was the enrichment for diverse sugar uptake pathways instead of just those associated with glucose, notably, glucose, galactitol, lactose, maltose, alpha-glucoside, cellobiose, and N-acetylgalactosamine phosphotransferase uptake pathways were all enriched. Takahashi and Nyvad (7) proposed an “extended ecological caries hypothesis,” suggesting that community metabolism regulates adaptation and selection; once acid production starts to proceed, it provides evolutionary selection for adaptations to variable pH. Analogous to the expansion of the S. mutans-centric paradigm to the entire community, the structural potential of the community for sugar catabolism diversifies greatly in caries-positive states. This suggests that a healthy phenotype has self-stabilizing functional potential; numerous sugar compounds are not easily taken up by the community, preventing the
associated pH decrease. In contrast, the progression to a caries state increases the diversity of sugar compounds available to the community catabolic network, thereby facilitating a pH decrease from an increased proportion of dietary input. This likely leads to increased pH fluctuations in both frequency, magnitude, and, potentially, duration. This is entirely consistent with the hypothesis that caries dysbiosis results from a community-scale metabolic shift and that progression has a feed-forward evolutionary adaptation pressure. At this point, it is difficult to identify the originating environmental pressure that might positively select for catabolism of diverse sugars, though host diet is a possible explanation.

The increase in sugar uptake potential is unsurprising but also increases confidence in the other trends that have not been described before. Within caries-positive microbiomes, we observed a notable enrichment of numerous two-component histidine-kinase response-regulator pairs, which are responsible for transcriptional changes in response to environmental stimuli. It is likely that this represents a community adaptation to increased variations in biofilm pH due to increased sugar catabolism, as well as a more dynamic community-scale microbial interaction network. Pathways for antibiotic resistance were also enriched in the communities of caries-positive subjects. To our knowledge this has not been previously reported. However, this could be an artifact of database bias, considering that carious lesions are associated with enrichment of *Streptococcus*, which has an abundance of well-characterized antibiotic resistance gene systems. An alternative explanation leverages the observation that carious states are characterized by inherently chaotic community profiles. Specifically, healthy caries-free states are associated with a relatively stable *Streptococcus* community, whereas the caries-positive *Streptococcus* communities are far more dynamic and diverse (Fig. 2D). It is possible that increased pH temporal gradients associated with caries select for antibiotic resistance due to increased antagonistic microbial interactions. The possibility that antibiotics included in toothpaste influence the prevalence of antibiotic resistance pathways could not be tested here.

The most surprising and difficult-to-explain enrichment in community functional potential associated with caries is the gain of metal transport pathways involved in the efflux of Zn or Mn, Ni, and Co. This is not likely due to a host response, as this normally involves host acquisition of Fe, Mn, and Zn, along with Cu efflux (41–43). We propose that the degradation of enamel, and subsequently dentin, results in the release of locally toxic levels of trace metals for the biofilm constituents, thereby providing an environmental selection pressure for the detoxification of these elements. This parallels geomicrobiology surveys where microbial activity degrades physical substrates like rocks and inorganic minerals (44). It also follows that other governing principles based on geomicrobiology systems may apply to the human supragingival microbiomes.

**Functional plasticity within poorly described core supragingival microbiome lineages.** We often ascribe functional potential to species based on type strains. To some extent, this is sensible as function and phylogeny are interlinked, and this has resulted in computational tools that associate phylogenetic loci, such as 16S rRNA, with functional potential (45). However, the genome plasticity makes it difficult to extrapolate function with confidence in some cases. For example, it has been shown that some strains of *S. mutans* can be less acidogenic than other streptococcal species (46), which could affect the organism’s pathogenicity in the context of cariogenesis. This phenomenon may serve to explain why we did not witness an enrichment of *S. mutans* in the diseased microbiota, an observation noticed in previously published studies (47–49). Our genome-centric approach revealed rather dramatic differences in functional potential between MAGs in the core supragingival microbiome. These include changes in vitamin and amino acid auxotrophy, environmental sensing, and nitrate reduction. Closely related genomes also exhibited differences in patterns of abundance and cooccurrence across the 88 subjects. The differences in functional potential and genome autecology suggest these taxonomically similar microbes have different metabolic roles in the supragingival microenvironment as mentioned in previous microbiome studies from the gut.
consortia (50) and in vitro biofilm cultures (51). Many of the newly described core supragingival microbiome MAGs greatly increase our genomic knowledge about previously poorly described lineages. In particular, to date only two *Alloprevotella* genomes have been published, only one TM7 oral microbial genome has been described, and this is the first description of oral *Gracilibacteria*. In all cases, multiple genomes with different autecology and metabolic capabilities were recovered.

**Alloprevotella.** The type strain, *Alloprevotella rava*, historically referred to as *Bacteroides melaninogenicus*, is an anaerobic fermenter producing low levels of acetic acid and high levels of succinic acid as fermentation end products while being weakly to moderately saccharolytic (52). Only two reference genomes are available for *Alloprevotella*, including *A. rava* and *Alloprevotella* sp. oral taxon 302 (37). As mentioned above, our analysis recovered two MAGs of *A. rava*, one of which has a high genetic component (MAG II.A; A/H11005 0.64) and the other of which is environmentally acquired (MAG II.B), according to the ACE model. The environmentally acquired *Alloprevotella* MAG II.B has reduced potential for cysteine biosynthesis, specifically the conversion of methionine to cysteine (M00609), and appears to be a methionine auxotroph; the potential for methionine biosynthesis from aspartate (M00017) and the methionine salvage pathway (M00034) are both higher in MAG II.A (Fig. 5B). Furthermore, *Alloprevotella* MAG II.A is completely lacking all the components for cysteine biosynthesis from serine, but can synthesize cysteine from methionine. The amino acid auxotrophies can be satisfied by the amino acids found in saliva, but the divergence in cysteine biosynthesis pathways is striking. A similar scenario exists for cationic antimicrobial peptide (CAMP) resistance: *Alloprevotella* MAG II.A contains more envelope protein folding and degradation factors (e.g., DegP and DsbA) while phosphoethanolamine transferase EptB is found in the environmentally acquired MAG II.B. Interestingly, the heritable strain contains more components for two-component regulatory systems (M00504, M00481, and M00459), transport systems (M00256 and M00188), and nisin resistance (M00754) than environmentally acquired MAG II.B. A final key metabolic acquisition novel to *Alloprevotella* MAG II.B relative to both cultivated and uncultivated *Alloprevotella* is the presence of components necessary for dissimilatory nitrate reduction. The oral production of nitrite from dietary or saliva-derived nitrate is the first step in the enterosalivary nitrate circulation (16), and this is the first implication of *Alloprevotella* in this cycle. The environmentally acquired *Alloprevotella* MAG II.B has a larger genome with a slightly higher CheckM-calculated contamination than MAG I.A, suggesting that there may be several *Alloprevotella rava* strains that are acquired by the environment with similar k-mer content. Interestingly, *Alloprevotella* MAG II.B is also enriched in enamel caries compared to caries that have progressed to the dentin layer, indicating that the environmentally acquired strains may play a role in the onset of carious lesions. Longitudinal metagenomic studies would be necessary to determine if the environmentally acquired *Alloprevotella rava* displaces the heritable strain with age and diet.

**TM7.** TM7 has been found in a variety of environments using cultivation-independent methods such as 16S rRNA sequencing. The first described genomes were assembled from wastewater reactor (53) and groundwater aquifer metagenomes, while the first cultivated strain was derived from an oral microbiome (34). A comparison of those three genomes revealed highly conserved genomic content and, generally, low percent GC (34). The recovery of three TM7 MAGs (named oral_TM7_JCVI III.A, III.B, and III.C) greatly adds to our knowledge about this unique lineage. Even the most coarse-grained comparative genomics analysis, that is, average GC content, revealed that MAGs III.A, III.B, and III.C are quite divergent (Fig. 6). TM7 MAG III.B is most similar to the cultivated strains at 35% GC, while MAG III.A is 46% GC and MAG III.C is 55% GC. TM7 MAG III.A is almost certainly a pangenome of multiple TM7 strains of relatively intermediate GC content based on the bimodal GC profile and the CheckM contamination of 159.28, though they were clearly separated in 5-mer space (Fig. 5). TM7 MAG III.B has
an inverse cooccurrence pattern with MAG III.A ($\rho = -0.34$) and no cooccurrence relation with MAG III.C ($\rho = 0.03$) in this study (Fig. 6A), strongly suggesting they have different ecological roles. Altogether, these TM7 genomes likely each represent a new family within the TM7 phylum.

The cultivated TM7 is an obligate epibiont parasite of oral Actinomyces (34). However, we did not observe strong cooccurrence patterns between the TM7 MAGs and any of the Actinobacteria, though it is not known if these relationships should have a strong cooccurrence or a lagged predator-prey cycle. The TM7 MAG III.B genome contains a near-complete pentose phosphate pathway, which indicates that this MAG has gained at least one form of energy-generating sugar metabolism relative to the other TM7. The TM7 MAG III.A genome uniquely contains the pathway for the synthesis of cofactor F420, though none of the components for methanogenesis. Instead, it is likely a cofactor in a flavin oxidoreductase of unknown function (54) and possibly involved in nitrosative (55) or oxidative stress (56). Given the likely production of nitrite by Alloprevotella and other members of the supragingival biofilm via dissimilatory nitrate reduction, the ability to detoxify nitrite has a protective role for both TM7 MAG III.A and its putative host.

Gracilibacteria. The presence of Gracilibacteria in the human oral microbiome was first reported in 2014 through 16S rRNA screening, with two 16S clades from human oral or skin samples (33). The first genome for Gracilibacteria was a single amplified genome from a deep-sea hydrothermal environment (57), and here we describe two Gracilibacteria MAGs that, analogous to TM7, are differentiated by fundamental genomic characteristics. Gracilibacteria MAG IV.A has a percent GC of around 37% while MAG IV.B averages 25% GC, making it one of the lowest-GC genomes reported (Fig. 6B). Both Gracilibacteria MAGs use the alternative opal stop codon. The Gracilibacteria MAGs detected here are most phylogenetically similar to those recovered from the East Central Atlantic hydrothermal chimney (57, 58). Both of our recovered Gracilibacteria MAGs appear to be anaerobic glucose fermenters producing acetate as a product, while containing numerous putative auxotrophies for vitamins and amino acids (see Table S5 in the supplemental material). Each genome bin contains at least two secretion systems (type II and IV) and CRISPR/CAS9 systems. Based on the small genomes, low GC content, and limited metabolic potential, we propose that these organisms are likely intracellular, or epibiont, parasites of other bacteria analogous to TM7. The two oral Gracilibacteria differ moderately with regard to functional potential and genome autecology. Gracilibacteria MAG IV.B contains an enrichment in functional potential for cationic antimicrobial peptide (CAMP) resistance (M00721, M00722, and M00728) and two-component pathways suggesting more metabolic interactions than MAG IV.A. Gracilibacteria MAG IV.B also uniquely contains an ABC-2-type polysaccharide or drug efflux system. Gracilibacteria MAG IV.A is also significantly enriched in healthy individuals, suggesting that this MAG may be beneficial for maintaining stable microbial solution states with regard to caries status.

Gracilibacteria MAGs IV.A and IV.B strongly cooccur with one another, suggesting a commensal or synergistic relationship. Gracilibacteria MAG IV.A has negative cooccurrence with Veillonella sp. oral taxon 780 and Actinomyces MAG I.B ($\rho = -0.549$ and $-0.497$, respectively) with a positive cooccurrence with Capnocytophaga gingivalis and Prevotella MAG V.B ($\rho = 0.467$ and 0.422, respectively). Gracilibacteria MAG IV.B has negative cooccurrence with Prevotella oulorum and Veillonella sp. oral taxon 780 ($\rho = -0.589$ and $-0.516$, respectively) with a positive cooccurrence with Cardiobacterium hominis and Capnocytophaga gingivalis ($\rho = 0.486$ and 0.462, respectively). Veillonella plays a role in the anaerobic fermentation of lactate to propionate and acetate by the methylmalonyl-CoA pathway. As mentioned above, Capnocytophaga gingivalis is found along the Corynebacterium base of the biofilm but in high density within the annulus due to a high demand for carbon dioxide (59). This may suggest that Gracilibacteria resides in the annulus of the biofilm, moderately interacting with Corynebacterium in regions where carbon dioxide is accessible. The enrichment in cooccurrence of Gracili-
bacteria MAG IV.B with Neisseria oralis compared to MAG IV.A \((\text{rho} = 0.314 \text{ and } 0.122, \text{ respectively})\) may suggest that MAG IV.B colonizes the biofilm periphery in microaerophilic environments.

**Conclusions.** Collectively, we have provided a holistic overview of the juvenile supragingival plaque microbiome. Many of the same genomes were found across all 88 subjects, describing a core microbiome. With the presence of caries, the abundance of constituents of the core microbiome changes to a variety of ecological states where the community networks are perturbed. The phenotypes of health and carious lesions should be characterized not only by the abundance of taxa but also by the functional potential of the community. Sugar-fermenting bacteria are always enriched in carious states, as are the abundance and diversity of sugar uptake pathways. This strongly supports the idea that caries phenotype is a community metabolism disorder. Carious lesions are also accompanied with an enrichment in environmental sensing and antibiotic resistance, suggesting that acid production provides a selection pressure. Finally, we provide genomic information about the metabolic diversity of several organisms represented by poorly described microbial lineages.

**MATERIALS AND METHODS**

**Study design.** Our objective was to compare the metagenomic signatures, in dental plaque, of children with and without dental caries. Our hypotheses were that (i) there would be a measurable difference in species and diversity between the two groups and (ii) some species present in plaque samples associated with caries will have been previously implicated in cariogenesis. Dental plaque samples were collected from participants of the University of Adelaide Craniofacial Biology Research Group (CBRG) and the Murdoch Children’s Research Institute (MCRI’s) Peri/Postnatal Epigenetic Twins Study (PETS) (60). The PETS \((n = 193)\) and CBRG \((n = 292)\) cohorts were composed of twins 5 to 11 years old. Twins from the state of Victoria, Australia (PETS), were recruited during the gestation period. All contactable twins from the PETS cohort will be eligible for participation. Inclusions were those twins whose parent consented to this particular wave of the study and who were recruited into the study during the gestation period. Ethics approval was attained from the University of Adelaide (Adelaide, Australia) Human Research Ethics Committee, The Royal Children’s Hospital Melbourne (Parkville, Australia) Human Research Ethics Committee, and the J. Craig Venter Institute Institutional Review Board.

Dental plaque samples were obtained at the commencement of a dental examination. The participants had not brushed their teeth the night preceding the plaque collection and on the day of collection. Additional data were collected from three separate questionnaires completed by the parents during the period from consent to prior to the dental examination being undertaken. The combined questionnaires consisted of 132 questions regarding oral health, dietary patterns and general health and development. Fifteen of these were extracted for use by JCVI for this analysis.

The entire dentition of each participant was assessed using the International Caries Detection and Assessment System (ICDAS II) (61). The ICDAS II is used to assess and define dental caries at the initial and early enamel lesion stages through to dentinal and final stages of the disease. Examiners were experienced clinicians who had undergone rigorous calibration and were routinely recalibrated across measurement sites to minimize error. Caries experience in each participant was initially reduced to a whole-mouth score, and three classifications were utilized: no evidence of current or previous caries experience, evidence of current caries affecting the enamel layer only on one or more tooth surfaces, and evidence of previous or current caries experience that has progressed through the enamel layer to involve the dentin on one or more tooth surfaces (including restorations or tooth extractions due to caries). For the purpose of this analysis, we classified disease phenotypes in twins as presence of caries in enamel or dentin.

The number of pairs selected for metagenomic sequencing was constrained by budget; thus, a subset of the broader clinical cohort was subsampled. Twin pairs were selected for sequencing manually by examining ordination plots from the broader 16S rRNA gene sequencing study (27) and then selecting (i) twins of the same phenotype that were closely related and (ii) twins discordant for caries that were divergent in ordination space.

**Sample collection, DNA extraction, library prep, and sequencing.** Plaque sample collection and DNA extraction were conducted as specified in reference 27. Libraries were prepared using the NEBNext Illumina DNA library preparation kit according to manufacturer’s specifications (New England Biolabs, Ipswich, MA). Metagenomic libraries were sequenced using the Illumina NextSeq 500 High-Output kit for 300 cycles following standard manufacturer’s specifications (Illumina Inc., La Jolla, CA). Subjects without caries in enamel or dentin are referred to as healthy while subjects with caries in either enamel or dentin are referred to as diseased unless otherwise noted.

**Metagenomic coassembly.** KneadData (62) was used for quality trimming on the raw sequencing reads, as well as screening out host-associated reads. Assembly was performed using SPAdes genome assembler v3.9.0 (metaSPAdes mode) (63) with the memory limit set to 1,024 GB. Reads were pooled, and initial assembly attempts resulted in exceeding memory limits. Therefore, each library was subsampled randomly to 25% of the total for the final coassembly. Quality-trimmed reads from all samples were
mapped back to the final metagenomics coassembly to generate abundance profiles across all 88 subjects.

**Genome binning.** The t-Distributed Stochastic Neighborhood Embedding algorithm (64, 65) applied to center-log-transformed k-mer profiles (k = 5) implemented in VizBin (66) raises a memory error when computing metagenomics assemblies of this scale. To address this issue, we developed a semisupervised iterative linear algebra technique to extract metagenome assembled genomes (MAGs) from the deluge of contigs in the de novo assembly. The pipeline is as follows: (step 1) use a conservative contig size threshold of 2,500 nucleotides for the initial binning; (step 2) eigendecomposition to calculate a representative vector in the direction of greatest variance (i.e., the 1st principal component [PC1]) of (2a) the coverage profiles and (2b) the k-mer profiles for each visually identified bin (66) of larger contigs; (step 3) subset the remaining contigs between the bandwidth of 300 and 2,500 nucleotides while computing the Pearson correlation between each bin’s PC1 and each contig in the smaller-contig-length subset; (step 4) extract the contigs (n = 500 contigs per bin) with the highest Pearson correlation to each bin’s PC1 from steps 2a and 2b; (step 5) merge the results of step 4 with the coarse bins from step 1 and recalculate the embeddings to generate finer-scale bins; and (step 6) iterate step 5 until convergence to yield the finalized draft genome bin (MAG). The quality of each of the finalized draft genome bins, or MAGs, was assessed for completeness, contamination, and strain-level heterogeneity using CheckM (v1.07) (29). Streptococcal annotated contigs were set aside during the binning process due to their promiscuous k-mer usage between strains.

**Annotation.** Annotation methods were as described in reference 67 with slight modifications. Open reading frames (ORFs) were called with FragGeneScan (v1.16), except the *Gracilibacteria* bins, which were called with Prodigal (v2.6.3), using the Candidate Division SR1 and *Gracilibacteria* genetic code (trans_ table = 25) (68, 69). For ORF annotations, domains were characterized using a custom compilation database with HMMER (v3.1b2) and functionality was further assessed via best-hit BLAST results and manual curation (70, 71). Contig and draft genome bin taxonomic assignments were determined by calculating a running sum of percent amino acid identities of the ORFs for each bin grouped by their respective taxonomy identifier from PhyloDB (https://github.com/allenlab/PhyloDB). PhyloDB is a comprehensive database of existing NCBI genomes, JGI-only single amplified genomes, and the MMETSP (72). The maximum weighted taxon at the species level was used to classify each MAG. The phylogenetic tree traversal was conducted using ete3’s NCBI taxonomy object, implemented in Python, to extract labeled hierarchies from taxonomic identifiers (73).

**Microbial community composition.** Metagenomic reads were mapped to contigs using CLC with a minimum spatial coverage of 50% and a minimum percent identity of 85%. The resulting count tables were transformed by adapting the transcript per million (TPM) (74) calculation for use on contigs to incorporate length, coverage, and relative abundance into the measure, an essential normalization for metagenome assemblies due to the inherently wide distribution of contig lengths. Summations of TPM values grouped by bin assignment were used as abundance values for draft genome composition and downstream analysis unless otherwise noted.

**Strain-resolved Streptococcus abundances** were calculated using the Metagenomic Intra-Species Diversity Analysis System (MIDAS) (5) with the subsampled subject-specific reads. MIDAS subprograms “genes” and “species” were run with default settings of 75% alignment coverage, 94% percent identity, and mean quality greater than 20. Streptococcus strains were parsed from the subject-specific count matrices to build a master count matrix containing all Streptococcus strains with respect to each individual twin.

**Statistical analyses.** Pairwise log_{2} fold change (logFC) profiles were calculated between groups by combinatorically computing the logFC between each nonredundant pair of the comparison groups and taking the mean of the distributions. All P values were calculated using the Mann-Whitney U test unless specifically noted otherwise. The statistical significance values were set at an inclusive threshold of 0.05 for determining enrichment in the following contexts: (context 1) MAGs via microbial abundance between diseased and healthy phenotypes, (context 2) functional module via phylogenomically binned functional potential (PBFP) profiles (see below) between diseased and healthy phenotypes, and (context 3) functional modules via PBFP profiles that were statistically significant between the groups identified in context 1. Spearman’s correlation coefficients are denoted as rho unless otherwise noted.

**Cooccurrence network analysis.** Pairwise Spearman’s correlations were used to robustly measure monotonic relationships between microbial abundance profiles, calculated for the MAGs and Streptococcus strains separately. MAG abundance profiles were standardized by TPM normalization, log transformation, and z-score normalization for each subject. In the Streptococcus strain-level cooccurrence analysis, strains were dropped from the calculation if they were not present in at least 10% of the samples. To account for domain errors during log transformation, due to zero values in the Streptococcus strain-level counts matrix, a pseudocount of 1e-4 was added to the entire dataframe. The WGCNA R package was used to calculate the adjacency and topological overlap measures (TOM) of the weighted cooccurrence network (75, 76). The TOM similarity matrix was used to construct the fully connected undirected NetworkX graph structure (77). The dendrogram visualizations were constructed from the dissimilarity representation of the TOM matrix (i.e., 1 − TOM_{minmax}) using ward linkage in SciPy (v1.0). PageRank centrality values (35) were computed on the fully connected weighted networks using the implementation provided in NetworkX. PageRank centrality is a variant of eigenvector centrality which allows us to measure the influence of bacterial nodes within our cooccurrence networks. We implemented PageRank centrality because it can be applied to fully connected weighted networks while many other, more common measures (e.g., Katz centrality) are not applicable in this setting. All analyses were
conducted in Python (v3.6.4), and figures were generated using Matplotlib (v2.0.2) unless otherwise noted (78).

**Variance component estimation.** Assessing the additive genetic and environmental factors driving MAG abundance as determined by the ACE model (79), controlling for sex, age, and health phenotype, was described previously in reference 27. The ACE model assumes that the variability of a given attribute is explained by additive (A) genetic effects, the shared/common (C) environment, and nonshared/unique environmental (E) factors. To these ends, the MAG abundance data were standardized by the following procedure: (i) calculating proportions of each bin (i.e., relative abundance of summed TPM values per bin); (ii) log transformation of normalized microbial abundances; and (iii) z-score normalization for each draft genome bin. The ACE model for variant component estimation was implemented using the mets package in R (80).

**Functional potential profiling.** To determine the functional components of MAG, we translated ORFs for each bin to build a putative proteome. The University of Kyoto’s Metabolic And Physiological potential Evaluator (MAPLE v2.3.0) (81, 82) was used to compute the Module Completion Ratios (MCRs) representing KEGG pathways, complexes, functions, and specific signatures. The MCR is calculated using a Boolean algebra-like equation previously described in reference 82. In order to identify latent interactions between these KEGG modules, MAGs, and specific subjects, we developed a metric that incorporates genome coverage, proportion, and MCRs referred to from this point forward as Phylogenomically Binned Functional Potential (PBFP). With this measure, we were able to identify differences in functional modules within the context of subject metadata (e.g., caries status). Subject-specific PBFP profiles were computed by summing matrix $D$ ($n$ = subjects, $m$ = contigs) across the contig axis with respect to draft genome bin assignment to produce matrix $X$ ($n$ = subjects, $p$ = MAGs). Matrix multiplication was computed for the MCRs in matrix $C$ ($p$ = MAGs, $q$ = modules) and the abundance measures in $X$ to yield a transformed matrix $A$ ($n$ = subjects, $q$ = modules) with subject-specific PBFP profiles for subsequent analysis.

**Data availability.** New methods are available at https://github.com/jolespin/supragingival_plaque_microbiome. All reads and assemblies are available in BioProject PRJNA383868. The overall coassembly is available through biosample SAMN10133834. The genome bines are available through biosample accession numbers SAMN10134551 to SAMN10134584. The individual reads for each library are available through accession numbers SRR6865436 to SRR6865523.

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at https://doi.org/10.1128/mBio.01631-18.

**FIG S1,** EPS file, 1.1 MB.  
**FIG S2,** EPS file, 1.5 MB.  
**FIG S3,** EPS file, 1.3 MB.  
**TABLE S1,** TXT file, 0.02 MB.  
**TABLE S2,** TXT file, 0.01 MB.  
**TABLE S3,** TXT file, 0.02 MB.  
**TABLE S4,** TXT file, 0.01 MB.  
**TABLE S5,** TXT file, 0.1 MB.

**ACKNOWLEDGMENTS**

We thank all twins and their families; Grant Townsend and Nicky Kilpatrick for their dental expertise; Tina Vaiano, Jane Loke, Anna Czajko, Blessy Manil, Chrissie Robinson, Mihiir Silva, and Supriya Raj for their expertise and assistance with collection of data and samples; and Anna Edlund for constructive comments on an early draft of the manuscript.

The research in this publication was supported by the National Institute of Dental and Craniofacial Research of the National Institutes of Health under award number R01DE019665. PETS was supported by grants from the Australian National Health and Medical Research Council (grant numbers 437015 and 607358 to J.M.C. and R.S.), the Bonnie Babes Foundation (grant number BBF20704 to J.M.C.), the Financial Markets Foundation for Children (grant no. 032-2007 to J.M.C.), and the Victorian Government’s Operational Infrastructure Support Program. CBRG was supported by grants from the Australian National Health and Medical Research Council (grant numbers 349448 and 1006294 to T.H.) and the Financial Markets Foundation for Children (grant no. 223-2009 to T.H.).

J.L.E. conducted the majority of the data analysis with contributions from C.L.D., J.M.I., and D.M.H. M.T. and C.K. conducted sample extractions and oversaw sequencing. A.G., S.K.H., and M.B.J. were involved in project coordination during the sample gathering phase. P.L., R.S., M.B., T.H., and J.M.C. oversaw clinical sampling. K.E.N. attained
funding and guided the research. C.L.D. directed the data analysis and interpretation. C.L.D. and J.L.E. wrote the paper with contributions from all authors. The authors declare no competing interests.

REFERENCES

1. Marquesen W, Kassebaum NJ, Bernabe E, Flaxman A, Naghavi M, Lopez A, Murray CJ. 2013. Global burden of oral conditions in 1990-2010: a systematic analysis. J Dent Res 92:592–597. https://doi.org/10.1177/0022034513490168.

2. Wall TA, Vujicic M. 2015. U.S. dental spending continues to be flat. Health Policy Institute research brief. American Dental Association, Chicago, IL.

3. Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE, 2005. Defining the normal bacterial flora of the oral cavity. J Clin Microbiol 43:5721–5732. https://doi.org/10.1128/JCM.43.11.5721-5732.2005.

4. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu W-H, Lakshmanan A, Wade WG. 2010. The human oral microbiome. J Bacteriol 192:5002–5017. https://doi.org/10.1128/JB.00542-10.

5. S. 2010. Oral health: the silent epidemic. Public Health Rep 125:158–159. https://doi.org/10.1177/003335491012500202.

6. Takahashi N, Nyvad B. 2011. The role of bacteria in the caries process: ecological perspectives. J Dent Res 90:294–303. https://doi.org/10.1177/0022034510376902.

7. Loesche WI. 1996. Microbiology of dental decay and periodontal disease. Chapter 99. In Baron S (ed), Medical microbiology, 4th ed. University of Texas Medical Branch at Galveston, Galveston, TX.

8. Mayer IC. 1999. Development of a classification system for periodontal diseases and conditions. Ann Periodontol 4:1–6. https://doi.org/10.2002/annals.1999.4.1.1.

9. Holmstrup P. 1999. Non- plaque-induced gingival lesions. Ann Periodontol 4:29–31. https://doi.org/10.2002/annals.1999.4.1.2.

10. Schmidt BL, Kuczynski J, Bhatthacharya A, Huey B, Corby PM, Queiroz EL, Nightingale K, Kerr AR, DeLacure MD, Veeramachaneni R, Olsen RB, Albertson DG. 2014. Changes in abundance of oral microbiota associated with oral cancer. PLoS One 9:e98741. https://doi.org/10.1371/journal.pone.0098741.

11. Beck JD, Slade G, Oftedal S. 2000. Oral disease, cardiovascular disease and systemic inflammation. Periodontol 2000 23:110–120. https://doi.org/10.1046/j.1600-0757.2000.230111.x.

12. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. 2013. Fusobacterium nucleatum promotes colorectal cancerogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. Cell Host Microbe 13(Suppl 4):3–10. https://doi.org/10.1016/j.chom.2014.10.012.

13. Seymour GJ, Ford PJ, Cullinan MP, Leishman S, Yamazaki K. 2007. Relationship between periodontal infections and systemic disease. Clin Microbiol Infect 13(Suppl 4):3–10. https://doi.org/10.1016/j.cmi.2006.10.013.

14. Clarke JK. 1924. On the bacterial factor in the aetiology of dental caries. Br J Exp Pathol 5:141–147.

15. Goodby KW. 1908. Mythology of the mouth. Longmans, Green, and Co, London, United Kingdom.

16. Lundberg JO, Gladwin MT, Ahsaluvala A, Benjamin N, Bryan NS, Butler A, Gabrales P, Fago A, Feilich M, Ford PC, Freeman BA, Frennaux M, Friedman J, Kelm M, Kevil CG, Kim-Shapiro DB, Kozlov A, Lieberman RS, Nealson K, Friedman R, Venter JC. 2012. Cultivable anaerobic microbiota of severe early childhood caries. J Clin Microbiol 48:4121–4128. https://doi.org/10.1128/JCM.01232-10.

17. Hirosue H, Hirosue K, Isogai E, Mima H, Ueda I. 1993. Close association between Streptococcus sobrinus in the saliva of young children and smooth-surface caries increment. Caries Res 27:292–297. https://doi.org/10.1159/000261553.

18. Kleinberg I. 2002. A mixed-bacteria ecological approach to understanding the role of the oral bacteria in dental caries causation: an alternative to Streptococcus mutans and the specific- plaque hypothesis. Crit Rev Oral Biol Med 13:108–125. https://doi.org/10.1177/104307660201300202.

19. Nissen H, Saveliev V, Yashih J, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

20. Marsh PD. 2018. In sickness and in health—what does the oral microbiome mean to us? An ecological perspective. Adv Dent Res 29:60–65. https://doi.org/10.1022/204517735295.

21. Gromez A, Espinoza JL, Harksins DM, Leong P, Saffery R, Bockmann M, Torralba M, Kuebels P, Kodukula R, Inman J, Hughes T, Craig JM, Highlander SK, Jones MB, Duport CL, Nelson KE. 2017. Host genetic control of the oral microbiome in health and disease. Cell Host Microbe 22:269–278.e3. https://doi.org/10.1016/j.chom.2017.08.013.

22. Lewin K, Leong P, Liu J, Mckinney J, Williams C, Shearer J, Benitez J. 2013. QIIME: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

23. Banks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Biol 25:1043–1055. https://doi.org/10.1186/s13059-015-0927-9.

24. Duan J, Russo DC, Landmesser MJ, Richter RA, Valas R, Novotny M, Greenbaum JS, Selengut JD, Ha H, Halpern AL, Lasken RS, Nealsom K, Friedman R, Venter JC. 2012. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. ISME J 6:1186–1199. https://doi.org/10.1038/ismej.2011.189.

25. Russo DC, Martyne AC, Duport CL, Halpern AL, Venter JC. 2010. Characterization of Prochlorococcus clades from iron-depleted oceanic regions. Proc Natl Acad Sci U S A 107:16184–16189. https://doi.org/10.1073/pnas.1009151107.

26. Brining MM, Lepp PW, Ouvrerc C, Armitage GC, Elmran DA. 2003. Prevalence of bacteria of division TM7 in human subgingival plaque and their association with disease. Appl Environ Microbiol 69:1687–1694. https://doi.org/10.1128/AEM.69.6.1687-1694.2003.

27. Camanocha A, Dewhirst FE. 2014. Host-associated bacterial taxa from Chlorobi, Chloroflexi, GN02, Synergistetes, SR1, TM7, and WPS-2 phyla/ candidate divisions. J Oral Microbiol 6:25468. https://doi.org/10.3402/jom.v6i25468.

28. He X, McLean JS, Edlund A, Yooseph S, Hall AP, Shi W, Dorrestein PC, Esquenazi E, Hunter RC, Cheng G, Nelson KE, Lux R, Shi W. 2015. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc Natl Acad Sci U S A 112:244–249. https://doi.org/10.1073/pnas.1419038112.

29. Page L, Brim S, Motwani R, Winograd T. 1999. The PageRank citation ranking; bringing order to the web. Stanford InfoLab, Stanford University, Stanford, CA.

30. Human Microbiome Jumpstart Reference Strains Consortium. 2010. A catalog of reference genomes from the human microbiome. Science 328:994–999. https://doi.org/10.1126/science.1193605.

31. Lloyd-Price J, Maturrana A, Rahnavard G, Crabtree J, Orvis J, Hall AB, Brady A, Creasy HH, McCracken C, Giglio MG, McDonald D, Franzosa EA, Knight R, White O, Huttenhower C. 2017. Strains, functions and dynamics in the
expanded Human Microbiome Project. Nature 550:61. https://doi.org/10.1038/s41586-018-03998-3.

38. Marsh PD. 1991. The significance of maintaining the stability of the natural microflora of the mouth. Br Dent J 171:174–177. https://doi.org/10.1038/sj.bdj.4807647.

39. Marsh PD. 1994. Microbial ecology of dental plaque and its significance in health and disease. Adv Dent Res 8:263–271. https://doi.org/10.1177/0895934494008002001.

40. Relman DA, Schneider JA, Costea PI, Coelho LP, Sunagawa S, Munch R, Huerta-Cepas J, Forslund K, Pedroni DA, Selengut JD, Haft DH. 2010. Unexpected abundance of coenzyme F420-dependent enzymes in Mycobacterium tuberculosis and other mycobacteria. Microbiome 3:1. https://doi.org/10.1186/2049-267X-3-1.

41. Curzon MEJ, Crocker DC. 1978. Relationships of trace elements in human tooth enamel to dental caries. Arch Oral Biol 23:647–653. https://doi.org/10.1016/0003-9969(78)90189-9.

42. Ghalimi E, Elmar H, Marelli B, Nazhat SN, Asgharian M, Vahli H, Tamifini M. 2013. Trace elements can influence the physical properties of tooth enamel. SpringerPlus 2:499. https://doi.org/10.1186/2193-1801-2-499.

43. He M, Lu H, Luc R, Ren T. 2016. Determining trace metal elements in the tooth enamel from Hui and Han ethnic groups in China using microwave digestion and inductively coupled plasma mass spectrometry (ICP-MS). Microchem J 127:142–144. https://doi.org/10.1016/j.microc.2016.02.009.

44. Gadd GM. 2010. Metals, minerals and microbes: geobiomicrobiology and bioremediation. Microbiology 156:609–643. https://doi.org/10.1099/mic.0.034714-0.

45. Langille MGJ, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Fierer N, Knights D. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31:814–821. https://doi.org/10.1038/nbt.2676.

46. de Soet JJ, Nyvad B, Kilian M. 2000. Strain-related acid production by streptococci and the clinical severity of root caries lesions. Oral Microbiol Immunol 15:212–217. https://doi.org/10.1034/j.1399-302X.2000.00470.x.

47. Trinchieri G, Banchereau J. 1995. Alloprevotella rava gen. nov., sp. nov., isolated from the human oral mucosa. Mol Microbiol 13:960. https://doi.org/10.1111/j.1365-2958.1994.tb04371.x.

48. Selengut JD, Haft DH. 2010. Unexpected abundance of coenzyme F420-dependent enzymes in Mycobacterium tuberculosis and other mycobacteria. Microbiome 3:1. https://doi.org/10.1186/2049-267X-3-1.

49. Hyatt D, Larsen J, Lopez R, M ears SE, Pettersson L, Persson AJ, Ronning E, Yooseph S, Rognes T, Quince C. 2016. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 17:119. https://doi.org/10.1186/s12859-016-1199-8.

50. Costea PI, Coelho LP, Sunagawa S, Munch R, Huerta-Cepas J, Forslund K, Pedroni DA, Selengut JD, Haft DH. 2010. Unexpected abundance of coenzyme F420-dependent enzymes in Mycobacterium tuberculosis and other mycobacteria. Microbiome 3:1. https://doi.org/10.1186/2049-267X-3-1.

51. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Manjunatha UH, Manjunatha UH. 2013. A novel F(420)-dependent enzyme in Mycobacterium tuberculosis and other mycobacteria. ISME J 9:2605–2619. https://doi.org/10.1038/ismej.2015.72.

52. Trinchieri G, Banchereau J. 1995. Alloprevotella rava gen. nov., sp. nov., isolated from the human oral mucosa. Mol Microbiol 13:960. https://doi.org/10.1111/j.1365-2958.1994.tb04371.x.

53. Manjunatha UH, Manjunatha UH. 2013. A novel F(420) -dependent enzymes in Mycobacterium tuberculosis and other mycobacteria. ISME J 9:2605–2619. https://doi.org/10.1038/ismej.2015.72.

54. Hyatt D, Larsen J, Lopez R, Mears SE, Pettersson L, Persson AJ, Ronning E, Yooseph S, Rognes T, Quince C. 2016. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 17:119. https://doi.org/10.1186/s12859-016-1199-8.

55. Van der Maaten L. 2013. Barnes-Hut-SNE. In Proceedings of the International Conference on Learning Representations.

56. Muzzupappa G, Stojanov P, Myllyla M, Salmi T, Paasonen E, Mencinger R, Ervik M, Friedberg G, Isaksson P, Kalliokoski J, Koskiniemi L, Lindqvist B, lobas E, Mikkola H, Nissinen M, Nuutila J, Peltomäki P, Rannikko S, Riordan J, Ronkkonen T, Sainio O, Sahlgren C, Salomaa K, Schuppler T, Serimaa T, Sillanaukee P, Sunna M, Tanavdikar S, Voutilainen H, von Ahnen K, Wotton D, Ylihannu J. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol 12:e1001889. https://doi.org/10.1371/journal.pbio.1001889.
73. Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol Biol Evol 33:1635–1638. https://doi.org/10.1093/molbev/msw046.

74. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesiak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. 2016. A survey of best practices for RNA-seq data analysis. Genome Biol 17:13. https://doi.org/10.1186/s13059-016-0881-8.

75. Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559. https://doi.org/10.1186/1471-2105-9-559.

76. Yip AM, Horvath S. 2007. Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics 8:22. https://doi.org/10.1186/1471-2105-8-22.

77. Hagberg AA, Schult DA, Swart PJ. 2008. Exploring network structure, dynamics, and function using NetworkX, p 11–15. In Varoquaux G, Vaught T, Millman J (ed), Proceedings of the 7th Python in Science Conference (SciPy2008).

78. Hunter JD. 2007. Matplotlib: a 2D graphics environment. Comput Sci Eng 9:90. https://doi.org/10.1109/MCSE.2007.55.

79. Eaves LJ, Last KA, Young PA, Martin NG. 1978. Model-fitting approaches to the analysis of human behavior. Heredity 41:249–320. https://doi.org/10.1038/hdy.1978.101.

80. Holst KK, Scheike T. 2017. mets: analysis of multivariate event times. https://www.r-pkg.org/pkg/mets.

81. Takami H. 2014. New method for comparative functional genomics and metagenomics using KEGG module, p 1–15. Springer, New York, NY.

82. Takami H, Taniguchi T, Moriya Y, Kuwahara T, Kanehisa M, Goto S. 2012. Evaluation method for the potential functionome harbored in the genome and metagenome. BMC Genomics 13:699. https://doi.org/10.1186/1471-2164-13-699.