AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Spectrum

| Human Microbiome | Research Article

Heterogeneous lineage-specific arginine deiminase expression within dental microbiome species

Allison E. Mann,[1] Brinta Chakraborty,[2] Lauren M. O'Connell,[1] Marcelle M. Nascimento,[3] Robert A. Burne,[2] Vincent P. Richards[1]

AUTHOR AFFILIATIONS See affiliation list on p. 23.

ABSTRACT Arginine catabolism by the bacterial arginine deiminase system (ADS) has anticariogenic properties through the production of ammonia, which modulates the pH of the oral environment. Given the potential protective capacity of the ADS pathway, the exploitation of ADS-competent oral microbes through pre- or probiotic applications is a promising therapeutic target to prevent tooth decay. To date, most investigations of the ADS in the oral cavity and its relation to caries have focused on indirect measures of activity or on specific bacterial groups, yet the pervasiveness and rate of expression of the ADS operon in diverse mixed microbial communities in oral health and disease remain an open question. Here, we use a multivariate approach, combining ultra-deep metatranscriptomic sequencing with paired metataxonomic and *in vitro* citrulline quantification to characterize the microbial community and ADS operon expression in healthy and late-stage cavitated teeth. While ADS activity is higher in healthy teeth, we identify multiple bacterial lineages with upregulated ADS activity on cavitated teeth that are distinct from those found on healthy teeth using both reference-based mapping and *de novo* assembly methods. Our dual metataxonomic and metatranscriptomic approach demonstrates the importance of species abundance for gene expression data interpretation and that patterns of differential expression can be skewed by low-abundance groups. Finally, we identify several potential candidate probiotic bacterial lineages within species that may be useful therapeutic targets for the prevention of tooth decay and propose that the development of a strain-specific, mixed-microbial probiotic may be a beneficial approach given the heterogeneity of taxa identified here across health groups.

IMPORTANCE Tooth decay is the most common preventable chronic disease, affecting more than two billion people globally. The development of caries on teeth is primarily a consequence of acid production by cariogenic bacteria that inhabit the plaque microbiome. Other bacterial strains in the oral cavity may suppress or prevent tooth decay by producing ammonia as a byproduct of the arginine deiminase metabolic pathway, increasing the pH of the plaque biofilm. While the benefits of arginine metabolism on oral health have been extensively documented in specific bacterial groups, the prevalence and consistency of arginine deiminase system (ADS) activity among oral bacteria in a community context remain an open question. In the current study, we use a multi-omics approach to document the pervasiveness of the expression of the ADS operon in both health and disease to better understand the conditions in which ADS activity may prevent tooth decay.

KEYWORDS metatranscriptomics, arginine deiminase system, probiotics, oral microbiome, caries

Arginine catabolism is a common metabolic pathway for anaerobic energy production used by a variety of organisms including bacteria, archaea, and microbial

Editor Paul A. Jensen, University of Michigan-Ann Arbor, Ann Arbor, Michigan, USA

Address correspondence to Robert A. Burne, rburne@dental.ufl.edu, or Vincent P. Richards, vpricha@clemson.edu.

Allison E. Mann and Brinta Chakraborty contributed equally to this article. Author order was decided empirically.

Marcelle M. Nascimento, Robert A. Burne, and Vincent P. Richards contributed equally to this article.

The authors declare no conflict of interest.

See the funding table on p. 23.

Received 4 April 2023
Accepted 6 February 2024
Published 27 February 2024

eukaryotes (1). Through a series of enzymatic modifications, arginine ingested from food, endogenously produced by the host, or synthesized by resident bacteria is broken down by the arginine deiminase system (ADS) to produce citrulline, ornithine, carbon dioxide, ammonia, and ATP (2–4). Along with acting as an important source of energy, carbon, and nitrogen production, the ADS pathway is protective against low pH environments through the production of ammonia (2, 5, 6). Furthermore, research has demonstrated that ADS activity affects the assembly and coaggregation of bacteria in the plaque biofilm matrix, prevents the colonization of the cariogenic bacterium *Streptococcus mutans*, and raises the pH of the oral environment, all of which may inhibit caries development (7–12). Moreover, the growth of bacteria in the presence of arginine *in vitro* or administration of arginine orally *in vivo* increases both ADS activity and the pH of the oral environment, and numerous clinical trials have documented a decrease in caries incidence when an arginine-supplemented toothpaste was used (6, 13–16). Given the capacity of ADS activity to buffer the effects of highly acidogenic microbes, ADS-competent bacteria that inhabit oral biofilms (i.e., dental plaque) have been extensively studied in the context of oral health and the prevention of tooth decay (6, 17–19). Of particular interest are the mechanisms of ADS pathway regulation in streptococci, many of which are important early colonizers of the oral cavity (20–23). Arginine and ADS-competent oral symbionts may therefore be candidate pre- and probiotic treatments, respectively, yet little is known about ADS activity among members of the oral cavity in a mixed microbial community context.

In the current study, we used a multivariate metatranscriptomic, metataxonomic, and *in vitro* assay approach to characterize the composition and function of the supragingival plaque microbial community in health and late-stage caries development. We found that while some *Streptococcus* sp. have higher expression of the ADS operon in health (i.e., *Streptococcus oralis* and *Streptococcus sanguinis*), others are more highly expressed in disease (e.g., *Streptococcus parasanguinis*). Importantly, the upregulation of genes involved in ADS activity for diseased teeth communities is limited to those with a low abundance of *Streptococcus mutans*, which may indicate an antagonistic relationship between these groups. Moreover, we find diverse non-*Streptococcus* bacteria with higher ADS operon expression in health when compared to disease. For instance, two poorly characterized members of the genus *Leptotrichia* have relatively high ADS expression in health and may represent an untapped source of genetically diverse potential probiotic strains. In addition, our results highlight the importance of community composition and within-species, lineage, or strain-level resolution in the interpretation of metatranscriptomic data from a mixed microbial community. We found that community composition is significantly correlated with observed differences in ADS gene expression profiles and that, within species, only some lineages (i) possess a full ADS operon, (ii) have relatively high community abundance, and (iii) have high ADS expression, all of which have important implications for the development of a probiotic. Moreover, comparing our reference database-mapping approach to that generated using a *de novo* assembly approach documents substantial genetic diversity within lineages that are not well represented in publicly available databases (e.g., Human Oral Microbiome Database, HOMD). Finally, we calculated coverage across the ADS operon in a subset of species and found that post-transcriptional regulatory activity may affect the efficacy of some strains to modulate the pH of the oral environment, a pattern which is substantiated by *de novo* assembly of full-length ADS transcripts in our data set. Assessment of the relative contributions of transcriptional versus post-transcriptional control would require extensive strain-level analyses. Finally, while the aim of this study is to characterize the microbial community and ADS activity in health and disease, the vast majority of observed functional variation between healthy teeth and those with carious lesions is driven by diverse and discordant gene expression among communities associated with severe tooth decay. The results of this study highlight the functional and taxonomic volatility in late-stage caries development, identify alternate microbial candidates for probiotic development in the context of the ADS pathway, and highlight the importance

of accounting for community composition when interpreting differential gene expression in mixed microbial contexts.

RESULTS

Increased microbiome diversity and citrulline production in health

We collected 60 supragingival plaque samples from 54 adults using a standard protocol (see Materials and Methods). Multiple teeth were sampled from a subset of individuals (n = 3) to document intraindividual functional variation. Each plaque sample was classified as either originating from a caries-free tooth (PF) or a tooth with an active, dentin cavitation or caries lesion (PD). In total, 27 plaque samples were collected from teeth with an active dentin cavity (PD) and 33 from a caries-free tooth (PF). Study participants included 35 females and 25 males ranging in age from 18 to 65 years old (average 31 ± 11 years).

First, to characterize the diversity of the oral microbiome and account for the differences in community composition driving gene expression, we used a high-resolution metataxonomic approach targeting a fragment of the bacterial *rpo*C gene to profile the community composition of each sample. We chose this approach as (i) it is more cost-effective than metagenomic sequencing, (ii) the *rpo*C gene is a single-copy core gene in bacteria and therefore should be a better representation of community membership as compared to other commonly used marker genes (e.g., 16S rRNA hypervariable regions), and finally (iii) the *rpo*C gene has high phylogenetic resolution at the species and strain level (24) (see Materials and Methods), which allows us to directly compare community composition to gene abundance data using a reference-based mapping approach. These data were compared to citrulline assay and metatranscriptomic results generated from the same samples to clarify the role of community composition and membership in ADS gene expression in health and disease using both biochemical and RNA sequencing techniques.

We found that the oral microbiome of healthy teeth (PF) is more diverse than those with caries (PD) as measured by Shannon diversity (Wilcoxon rank sum test: P = 0.003; Fig. 1a), and we found that citrulline production is significantly higher in PF as compared to PD samples (Wilcoxon rank sum test: P = 0.0004). Variation in citrulline production is significantly (albeit weakly) correlated with beta-diversity when controlling for tooth type (PD vs PF) (PERMANOVA: R^2 = 0.03; P = 0.03) (Fig. 1b). Importantly, however, not all PF samples have high citrulline production nor do all PD samples have low production. For example, a single PD sample (UF68PD) had unexpectedly high citrulline production (Fig. 1b), which may be the consequence of significantly higher ADS activity by *Streptococcus oralis* as detected by metatranscriptomic sequencing (see below) (Fig. 1c; Table S1). Specifically, although the proportion of *S. oralis* assigned to this sample is small in our paired metataxonomic data set (<1% of the total community), transcripts from *S. oralis* make up 29.56% of all ADS transcripts originating from UF68PD. The median proportion of transcripts assigned to this species in PF is 12.44% (±13.17%) while in PD it is 1.80% (±15.91%) making UF68PD more similar to other high-citrulline-producing PF samples in terms of the proportion of transcripts assigned to this species. This is in stark contrast to four low-citrulline-producing PF samples (UF33PF, UF36PF, UF40PFR, and UF48PF), where the proportion of transcripts assigned to *S. oralis* was less than 1% each. In addition to *S. oralis*, sample UF68PD also had a high number of ADS gene transcripts assigned to *Leptotrichia* sp. oral taxon (HMT) 215, *Kingella oralis*, and *Actinomyces naeslundii* (Table S2), which were also found to have higher ADS expression in PF as compared to PD.

Unequal expression of individual *arc*ABC genes

While arginine catabolism is a common metabolic pathway found across diverse bacterial groups, it is dispensable and the operon is absent or truncated in many oral bacteria, even among closely related phylogenetic groups (Table S3) (25). Moreover, the

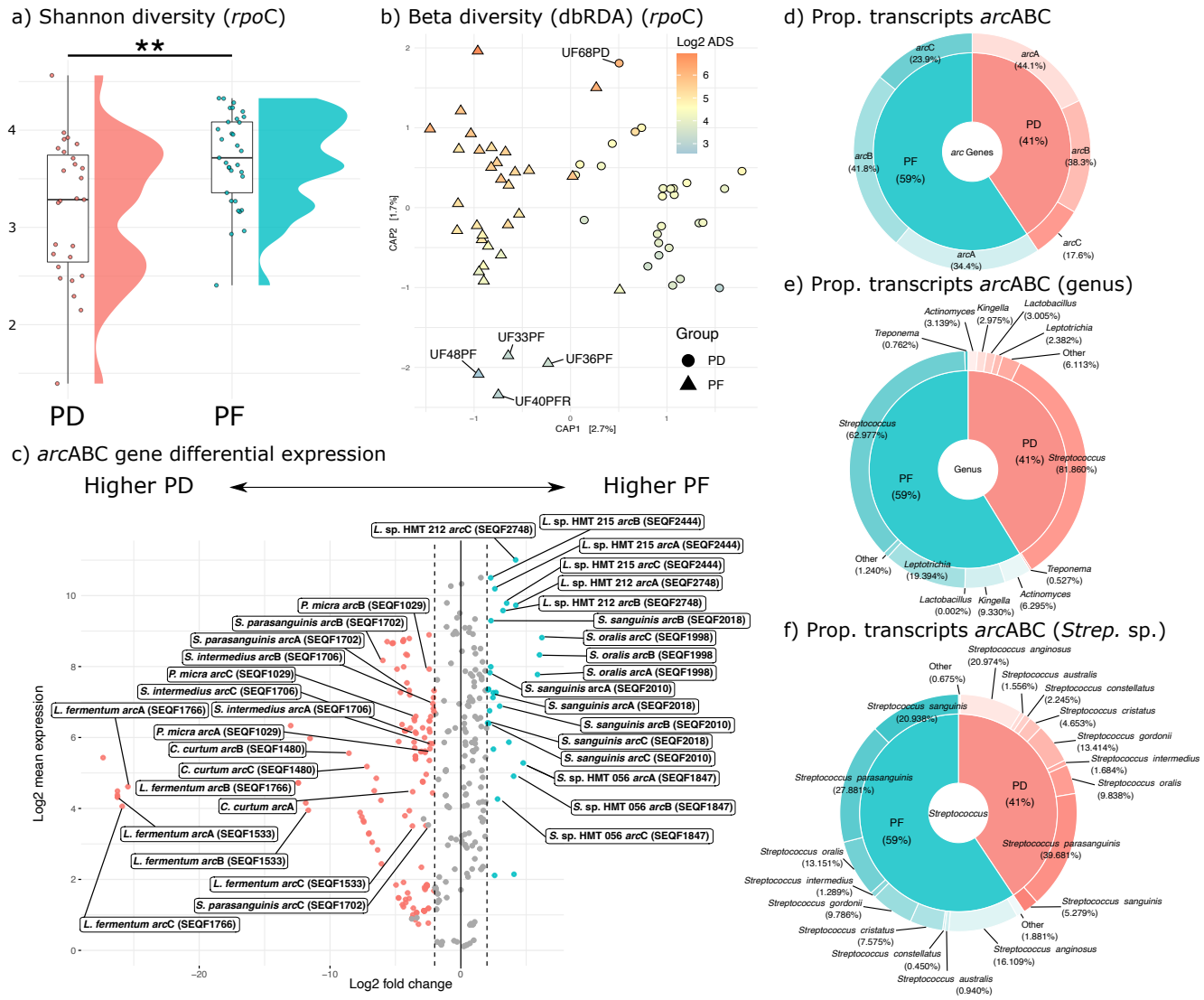

FIG 1 Significant differences in ADS activity are found when comparing PD and PF, which are predominantly driven by major differences in taxonomic composition. (a) Shannon diversity box and raincloud plots for PD and PF samples. (b) Beta-diversity colored by ADS expression levels as measured by our citrulline assay. Shapes represent the two groups. Samples with unexpectedly high or low citrulline production are labeled on the plot (see text). (c) Differential abundance volcano plot of metatranscriptome data. Colored points represent genes that were both statistically significant and passed log fold change abundance thresholds (see text). Labeled points represent all taxa upregulated for PF and six select taxa upregulated for PD (see text) that (i) had the highest adjusted *P*-value for either PD (pink) or PF (blue) and (ii) had all three *arc* genes marked as differentially abundant. One exception shown here is the *S. sanguinis* genes *arc*C1 and *arc*B, which, despite having a significantly higher expression of the *arc*A gene, did not pass thresholds. HOMD strain identifiers along with species identifications are shown in boxes. (d) Proportion of reads mapping to each of the three major *arc* genes in the ADS operon after manual curation for genes that were present in synteny within a contig in both PF and PD samples. (e) The proportion of RNA seq reads that mapped to *arc*ABC genes among different genera. (f) Proportion of reads that mapped to *arc*ABC across different *Streptococcus* species (strains collapsed to species).

gene content and organization of the ADS operon vary within and between ADS-competent bacteria (26–30). As such, we first identified ADS-competent oral taxa in the Human Oral Microbiome Database (31) through homologous gene clustering of three highly conserved *arc* genes that produce enzymes necessary for ADS activity (i) *arc*A (arginine deiminase EC: 3.5.3.6), (ii) *arc*B (ornithine carbamoyltransferase EC: 2.1.3.3), and (iii) *arc*C (carbamate kinase EC: 2.7.2.2). Only those bacteria with all three genes in synteny were included in differential gene expression analyses in our reference-based mapping approach (see Materials and Methods). Using this manually curated collection of ADS operons as a reference, we next compared the relative abundance of transcripts

assigned to individual *arc* genes across PF and PD samples. As only those bacteria with *arc*ABC organized in a syntenous operon were considered, we expected the proportion of transcripts that mapped to each gene within an operon to be approximately equivalent to one another. Interestingly, for both PF and PD samples, while the proportion of transcripts assigned to *arc*A and *arc*B were similar, the proportion of transcripts mapping to *arc*C tended to be lower. This is especially true of transcripts mapping to the *arc* genes in PD samples, where the proportion for *arc*A is nearly three times higher than *arc*C (Fig. 1d). This may be due to methodological biases (e.g., incomplete or scaffold-level genome assemblies used as reference) but also may be indicative of differences in operon usage or structure in ADS-competent bacteria that are potentially associated with health status. For example, specific genes within an operon may be regulated through mRNA degradation depending on the transcriptional needs of the cell (32–34), internal transcription promoter or termination sites, or through other forms of gene expression regulation (e.g., transcriptional attenuation through secondary DNA structures that prevent transcription of the full operon) (35). Notably, for this analysis, we only considered reference operons that were (i) complete (i.e., all three genes present) and (ii) present on the same contig (to reduce the impact of incomplete genomes), which suggests that these patterns are biological and not methodological in nature.

Of those transcripts that mapped to one of the three *arc* genes analyzed here, most were assigned to *Streptococcus,* but we also detected ADS *arc* gene expression for other genera including *Leptotrichia, Kingella, Lactobacillus, Actinomyces, Treponema,* and several low-frequency groups (Table S2; Fig. 1e). While both PD and PF samples had a high frequency of transcripts assigned to *Streptococcus parasanguinis* (PF: 27.88%, PD: 39.68%) and *Streptococcus anginosus* (PF: 16.11%, PD: 20.97%), PF samples had a comparatively higher proportion of transcripts assigned to *S. sanguinis* (PF: 20.94%, PD: 5.28%) and *S. oralis* (PF: 13.15%, PD: 9.84%) as compared to PD (Fig. 1f). To better understand how structure and regulation might impact the production and survival of transcripts, we mapped our quality-filtered metatranscriptomic data to the full ADS operon of six strains of interest (*Streptococcus sanguinis* SEQF2018, *S. parasanguinis* SEQF1919, *Streptococcus gordonii* SEQF1066, *S. oralis* SEQF1998, *Leptotrichia* oral taxon 212, and *Leptotrichia* oral taxon 215) and generated coverage plots (Fig. 2). We chose these specific taxa as they demonstrate the range of coverage patterns and intergenic region size variation we observed in the data set across different ADS-competent species. Operon coverage for *S. oralis* and both lineages of *Leptotrichia* exhibits a wave-like pattern, with inconsistent coverage levels across the full operon, while coverage for *S. sanguinis*, *S. parasanguinis*, and *S. gordonii* is wave-like in some regions but relatively consistent in others, particularly across *arc*B and to a lesser extent, *arc*A. Interestingly, we found that some of the most precipitous drops in coverage correspond to intergenic regions in the operon, the structure of which varies across the selected taxa presented here. For example, while *S. sanguinis, S. parasanguinis*, *S. oralis,* and *S. gordonii* have an intergenic region between *arc*A and *arc*B spanning approximately 51–85 base pairs long, both *Leptotrichia* lineages have an intergenic region between *arc*A and *arc*B spanning approximately 142–146 base pairs. Additionally, all *Streptococcus* sp. included in this analysis have an intergenic region ranging from between 86 (*S. parasanguinis*) and 313 base pairs (*S. oralis*) separating *arc*B and *arc*C. In contrast, both *Leptotrichia* sp. have only two or three base pairs separating the two genes. Importantly, the two most commonly observed species, *S. sanguinis* and *S. parasanguinis,* have markedly lower *arc*C coverage as compared to *arc*A and *arc*B, which may indicate differential regulation of this gene post-transcription. It is possible that drops in coverage for these two species correspond to hotspots or promoters in the operon for transcript degradation and intergenic regions may play a functional role in post-transcriptional regulation of the operon. Notably, the heterogeneity in operon structure and mRNA abundance observed in this study is consistent with the knowledge that ADS expression is known to be complex and responsive to a variety of environmental inputs [e.g., reference (36)]. A full table of *arc* operon coordinates and lengths of

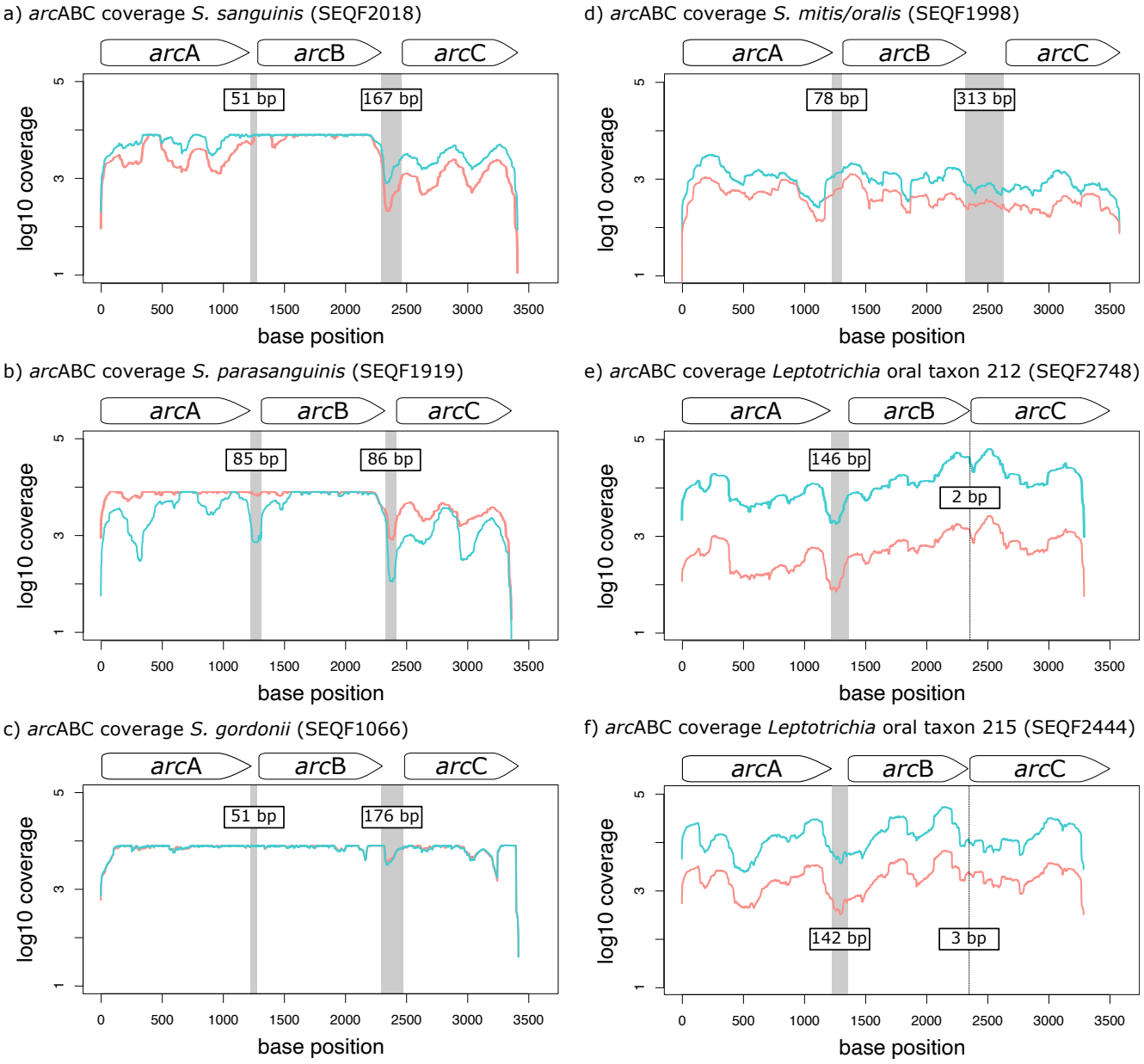

FIG 2 ADS operon structure and coverage vary across strains. (a) Log 10 coverage per base of the three *arc* genes in *Streptococcus sanguinis* (SEQF2018). (b) Log 10 coverage per base of the three *arc* genes in *Streptococcus parasanguinis* (SEQF1919). (c) Log 10 coverage per base of the three *arc* genes in *Streptococcus gordonii* (SEQF1066). (d) Log 10 coverage per base of the three *arc* genes in *Streptococcus mitis/oralis* (SEQF1998). (e) Log 10 coverage per base of the three *arc* genes in *Leptotrichia* sp. oral taxon 212 (SEQF2748). (f) Log 10 coverage per base of the three *arc* genes in *Leptotrichia* sp. oral taxon 215 (SEQF2444). Line colors in all coverage plots represent all reads mapped to health (blue) as compared to disease (red). Gray bars indicate intergenic regions. White internal labels on intergenic regions indicate the size of each region.

Streptococcus sp. can be found in Fig. 2 and the two *Leptotrichia* sp. analyzed here can be found in Table S4.

Strain-level differential expression of the ADS operon

Given the heterogeneity of ADS operon structure and presence or absence among closely related oral bacteria, we performed differential expression analyses at both the species level, defined as the total of all transcripts that aligned to a named oral species in the HOMD database, and at the strain level, defined as those transcripts that aligned

to a particular reference genome within the HOMD database or to a unique reference sequence generated using *de novo* assembly (see below). We take this tiered approach as previous research has demonstrated that ADS activity can vary significantly among closely related oral bacteria (28), which has implications for the development of probiotic treatments. Importantly, however, the HOMD database-based strain perspective has limitations. Specifically, a strain within the database used as a reference may not be present within a sample. Consequently, when a sequence read aligns to a specific genome within the database, we interpret this as the expression (RNA) or presence (DNA) of a strain from the same lineage as the strain in the database, and this is understood when we refer to strains throughout the manuscript.

First, we performed differential gene expression analysis to identify strains where all three ADS genes were significantly upregulated in health as compared to disease. Interestingly, only six bacterial strains met these criteria in health including *Streptococcus* sp. oral taxon 056 (SEQF1847), *S. oralis* (SEQF1998), *S. sanguinis* (SEQF2018 and SEQF2010), *Leptotrichia* sp. oral taxon 215 (SEQF2444), and *Leptotrichia* sp. oral taxon 212 (SEQF2748), while 13 were found upregulated in disease including *Parvimonas micra* (SEQF1029), *Cryptobacterium curtum* (SEQF1480), *Limosilactobacillus fermentum* (SEQF1766, SEQF1533, SEQF1964, SEQF2541, and SEQF2605), *S. parasanguinis* (SEQF1702, SEQF1706, SEQF1885, SEQF1919, SEQF2007, SEQF2222, SEQF2344, and SEQF2625), *Streptococcus intermedius, Actinomyces graevenitzii* (SEQF2031 and SEQF2450), *Lactobacillus crispatus* (SEQF2084 and SEQF2421), *Cutibacterium acnes* (SEQF2166), *Streptococcus constellatus* (SEQF2211, SEQF2589, SEQF2590, SEQF2591, and SEQF2643), *S. anginosus* (SEQF2334, SEQF2587, SEQF2588, SEQF2609), *Olsenella uli* (SEQF2631), *Olsenella* sp. oral taxon 807 (SEQF2751), and *Lactobacillus ultunensis* (SEQF2852) (Table S1; Fig. 1c). Genes of all six significant strains in PF and a subset of six significant strains in PD (as determined by the highest adjusted FDR score) are highlighted in Fig. 1c.

Species-level differential expression of the ADS operon

Next, as some species in our reference database have multiple representative strains (predominantly *Streptococcus*) and some strains were more highly expressed than others in the above analyses, we collapsed our results by species designation to compare to strain-level results (Table S5). While most differential expression results remained significant between groups after collapsing at the species level (n = 16), 11 were no longer identified as having higher or lower expression in either group (Tables S5 and S6), reflecting strain-level gene expression heterogeneity.

To compare our species and strain-level results further, we next built a maximum likelihood phylogeny using a concatenation of the ADS genes (*arc*A, *arc*B, and *arc*C) for all *Streptococcus* species detected in our metatranscriptomic data and identified strains within species that had higher or lower expression or abundance among health groups. For example, while there are a total of 16 representative genomes for *Streptococcus oralis* in HOMD (as of database version 9.15a) and we assigned transcripts to four of them (Fig. 3), only a single *S. oralis* strain had all three genes significantly upregulated in health (SEQF1998) and collapsing at the species level produced an insignificant adjusted (FDR) *P* value (0.8) (Tables S6 and S7). Similarly, while there are eight representative strains in HOMD for *S. intermedius* and all eight had mapped transcripts in our data set, only one had higher ADS gene expression in PD samples (SEQF1706) and the collapsed data were insignificant (FDR = 0.07). Importantly, unlike other *Streptococcus* groups where the average log fold change and base mean are relatively identical (indicating that most transcripts assigned to those lineages mapped equally well to other members of the clade), all *S. oralis* identified in the current data have a high average base mean, but only SEQF1998 had a high average log fold expression (Fig. 3).

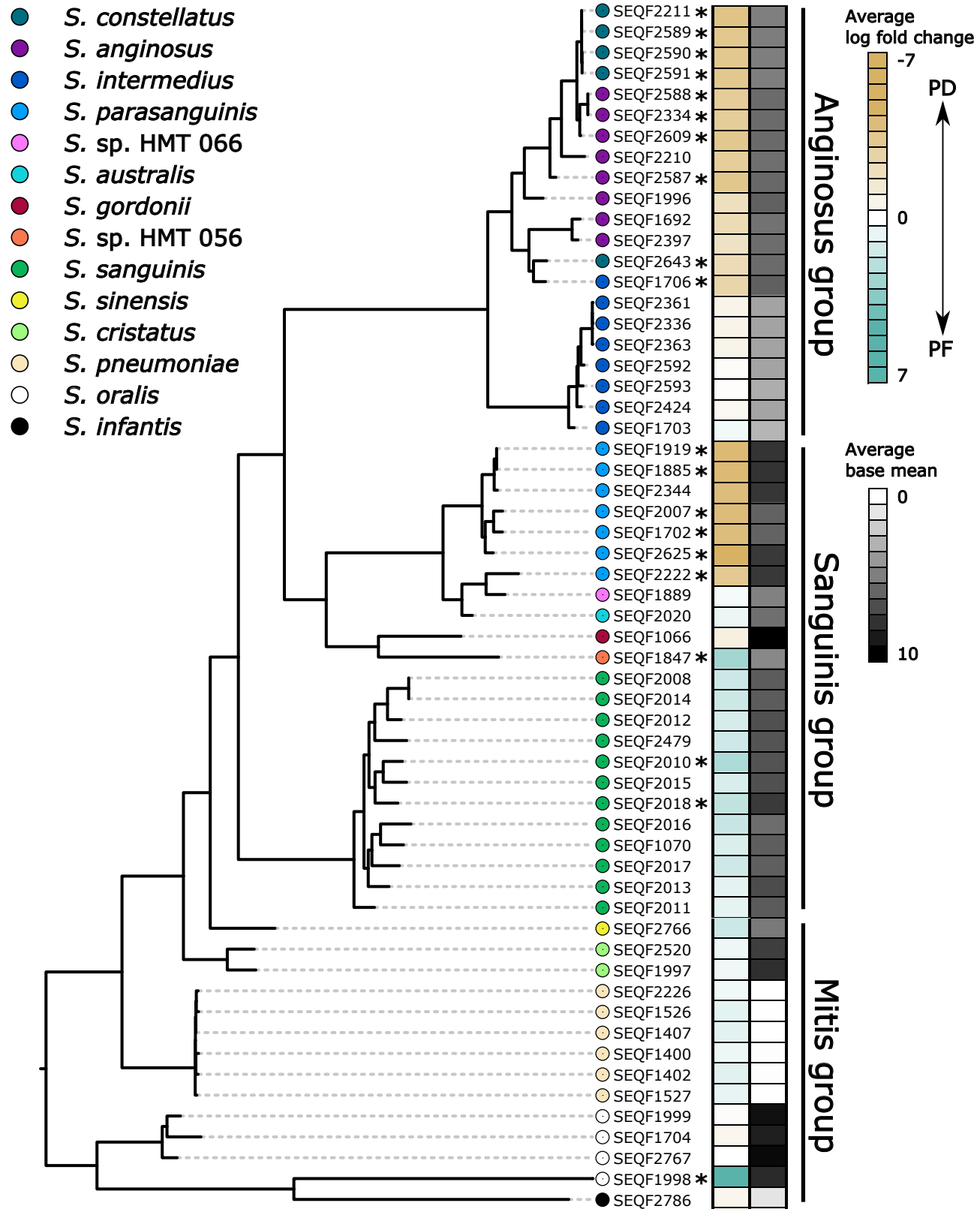

FIG 3 Differential expression and abundance of ADS operon vary by *Streptococcus* lineages. Maximum likelihood phylogeny for the *Streptococcus* sp. ADS operon (concatenated *arc*A, *arc*B, and *arc*C genes). Asterisk (*) indicates a lineage where all three *arc* genes were up- or downregulated as measured by DESeq2 analysis. Corresponding heatmap indicates the average log fold change and base mean per aligned lineage.

ADS transcript abundance and species abundance are correlated

Because we anticipate community structure to be an important confounding variable when comparing gene expression activity between PF and PD samples (Fig. 1b), we next calculated the correlation between log fold change of the abundance of ADS-competent species as measured by *rpo*C amplicon sequencing to the log fold change of gene expression for the ADS operon as measured by our reference-based mapping approach (Table S8). By inferring the taxonomic origin of individual transcripts and amplicon sequences with the same reference database, we can then directly compare patterns of community abundance and gene expression levels for specific bacterial species. In this way, we can better understand how community membership and abundance affect our interpretation of gene expression data. Log fold change for ADS expression was calculated at the species level by first summing all read counts per gene for each HOMD species and then averaged over the three *arc* genes to account for the differences in individual gene expression levels. We found that fold change of ADS expression and species abundance was highly correlated (R^2 = 0.62, P = 0.001), and in most cases, species with high differential expression levels are also found at higher abundance in their respective health category. For example, while transcripts assigned to the ADS pathway in *Streptococcus sanguinis* are more highly abundant among healthy samples, the relative abundance of amplicon sequence variants (ASVs) assigned to *S. sanguinis* is also higher in health, which suggests that the high abundance of transcripts assigned to the ADS pathway is linked to a higher abundance of this species in health as compared to disease. Only 48.15% of PD samples had at least one *rpo*C amplicon assigned to *S. sanguinis,* while the species was observed among all PF samples. Moreover, among those samples with detected levels of *S. sanguinis*, it was more highly abundant in PF as compared to PD (PF: 3.7% ± 4.5% relative abundance, PD: 0.5% ± 1.0% relative abundance). Similarly, while transcripts assigned to the ADS pathway in *S. parasanguinis* are more highly abundant in PD, the relative abundance of *S. parasanguinis* is also higher in PD as compared to PF when the amplicon data is considered (average 1.1% ± 2.3% of the total microbial community as compared to 0.4% ± 0.9% of the total community among PF) (Fig. 4a through c). This is despite the fact that there was an approximately equivalent proportion of PF and PD samples with *rpo*C reads assigned to *S. parasanguinis* (52.85% and 60.61%, respectively). Interestingly, we found a subset of ADS-competent bacteria that did not follow a simple pattern of higher community abundance coupled with higher ADS transcript expression. For example, *Bulleidia extructa, S. constellatus,* and *S. intermedius* are more highly abundant in PF as compared to PD samples but have higher ADS expression in PD. Conversely, *Cryptobacterium curtum, Parvimonas micra*, and *S. anginosus* have higher ADS expression in PD with very little difference in amplicon abundance between PD and PF samples. While these taxa have high ADS activity in disease that diverges substantially from amplicon sequence variant abundances, they are also relatively rare across individuals in this data set (Fig. 4a) and therefore may be less robust to different oral environments.

De novo assembly of transcripts provides higher ADS strain resolution

As described above, an ongoing challenge in microbial metatranscriptomics is that, lacking other evidence (e.g., metataxonomics or whole genome sequencing), the community of microorganisms across samples is largely unknown. Moreover, the composition of publicly available genomic databases tends to be heavily skewed toward model or pathogenic organisms. For example, of the 555 genomic sequences deposited in HOMD v9.15a for *Streptococcus*, 45% are the cariogenic pathogen *S. mutans*. Conversely, only 16 *Leptotrichia* genomes are available for reference within the database. As such, reference-based metatranscriptomic analysis underestimates true strain and species-level diversity in gene expression. Given the limitations of a reference-based mapping approach to characterize the full functional profile of a metatranscriptomic data set of unknown community composition, we next performed an additional *de novo*

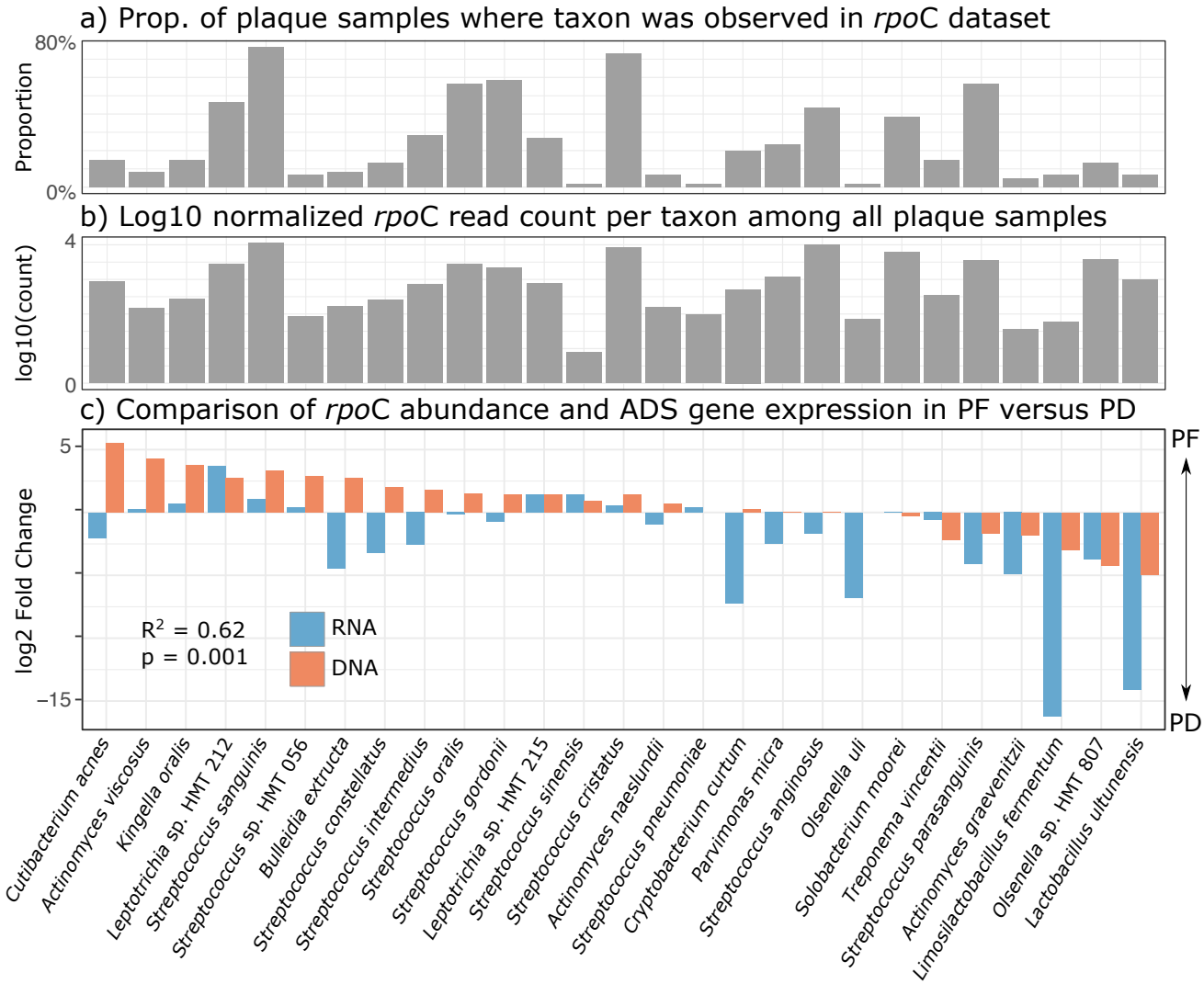

FIG 4 Species level ADS operon expression is significantly correlated to community composition as measured by *rpo*C amplicon sequencing. (a) The proportion of samples in the full amplicon data set where species listed in panel c were detected. (b) Log 10 normalized read count of amplicon sequences for each species corresponding to panel c. (c) Comparison of log2 fold changes for ADS-competent bacteria in the metataxonomic and metatranscriptomic data set. R^2 and alpha significance values were calculated using a Pearson correlation test.

assembly of our data set to document the diversity of ADS transcripts across all taxonomic groups.

On average, we generated 364,811 (±159,936) assembled transcripts per sample (Table S9), of which 1,723 were annotated with at least two of the *arc*ABC genes and 568 had all three genes annotated in synteny. Most transcripts were truncated, which supports our previous interpretation of post-transcriptional regulation of the operon using our reference-based mapping approach. Of those ADS transcripts with all three genes present, 132 had one or more genes upregulated in disease (PD) consisting mostly of transcripts assigned to *Streptococcus* sp. (35%) followed by *Lactobacillus ultunensis* (11%) and *Limosilactobacillus fermentum* (8%). Comparatively, 252 unique transcripts were upregulated in health (PF), 32% of which were assigned to *Leptotrichia* sp. oral taxon 212 followed by *Streptococcus* sp. (27%) and *Kingella oralis* (18%) (Table S10). The results of this analysis complement our reference-based approach and document the high within-strain diversity of ADS operon structure and activity in health and disease. For example, we found relatively consistent and high expression of ADS genes among the majority of transcripts assigned to *Leptotrichia* (Fig. 5; Table S11), while expression

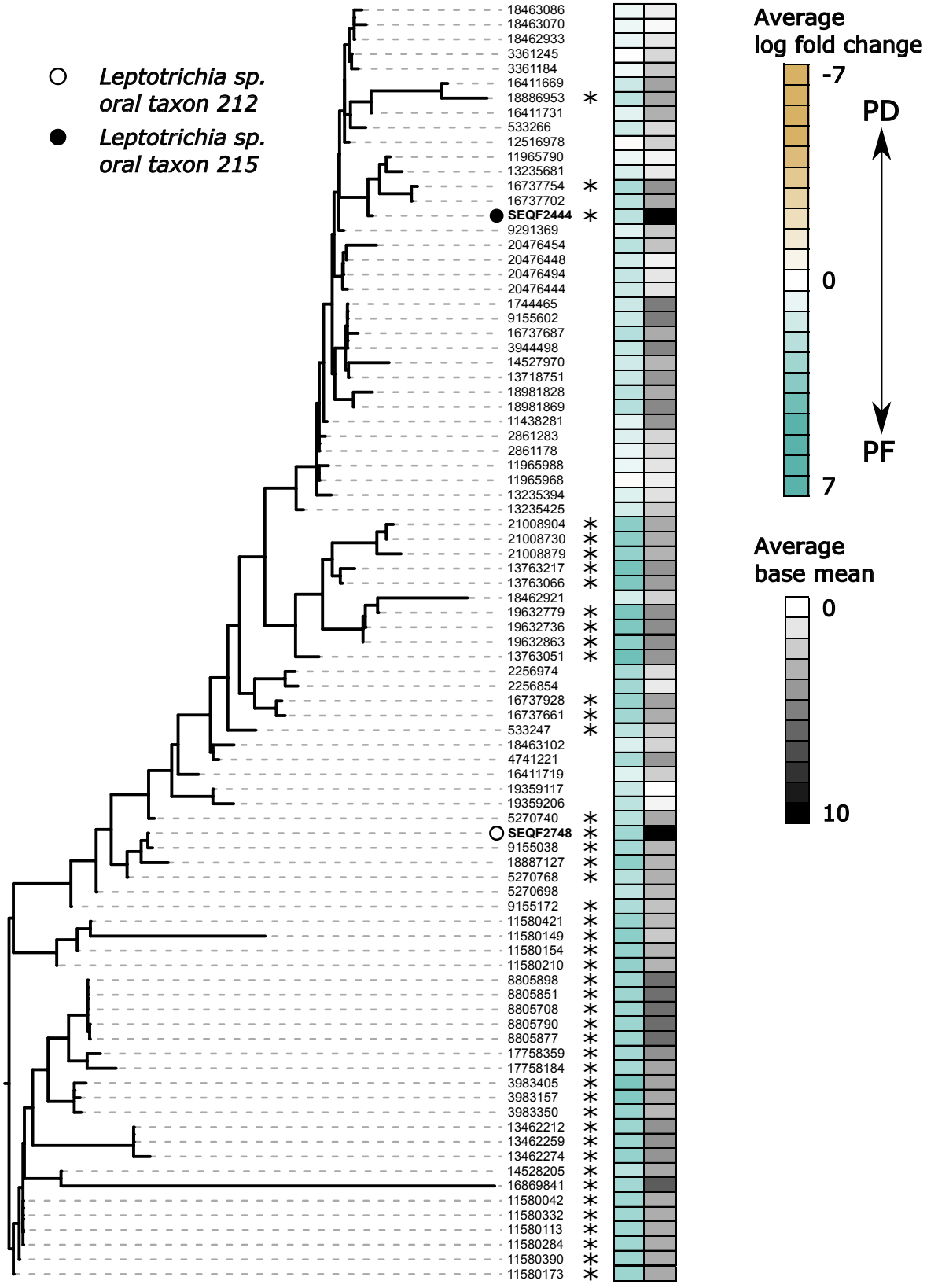

FIG 5 *De novo* assembly of ADS transcripts documents high operon diversity among oral *Leptotrichia*. Maximum likelihood tree of reference HOMD *Leptotrichia* (SEQF, bolded on tree) and *de novo*-assembled ADS transcripts assigned to these species (numbers only). Heatmap corresponds to the average log fold change over all three *arc* genes and the average base mean based on either HOMD reference genome mapping (SEQF lineages) or *de novo* assembly (numbered lineages). All lineages have relatively high gene expression in PF with many lineages having all three genes significantly upregulated (denoted by asterisks).

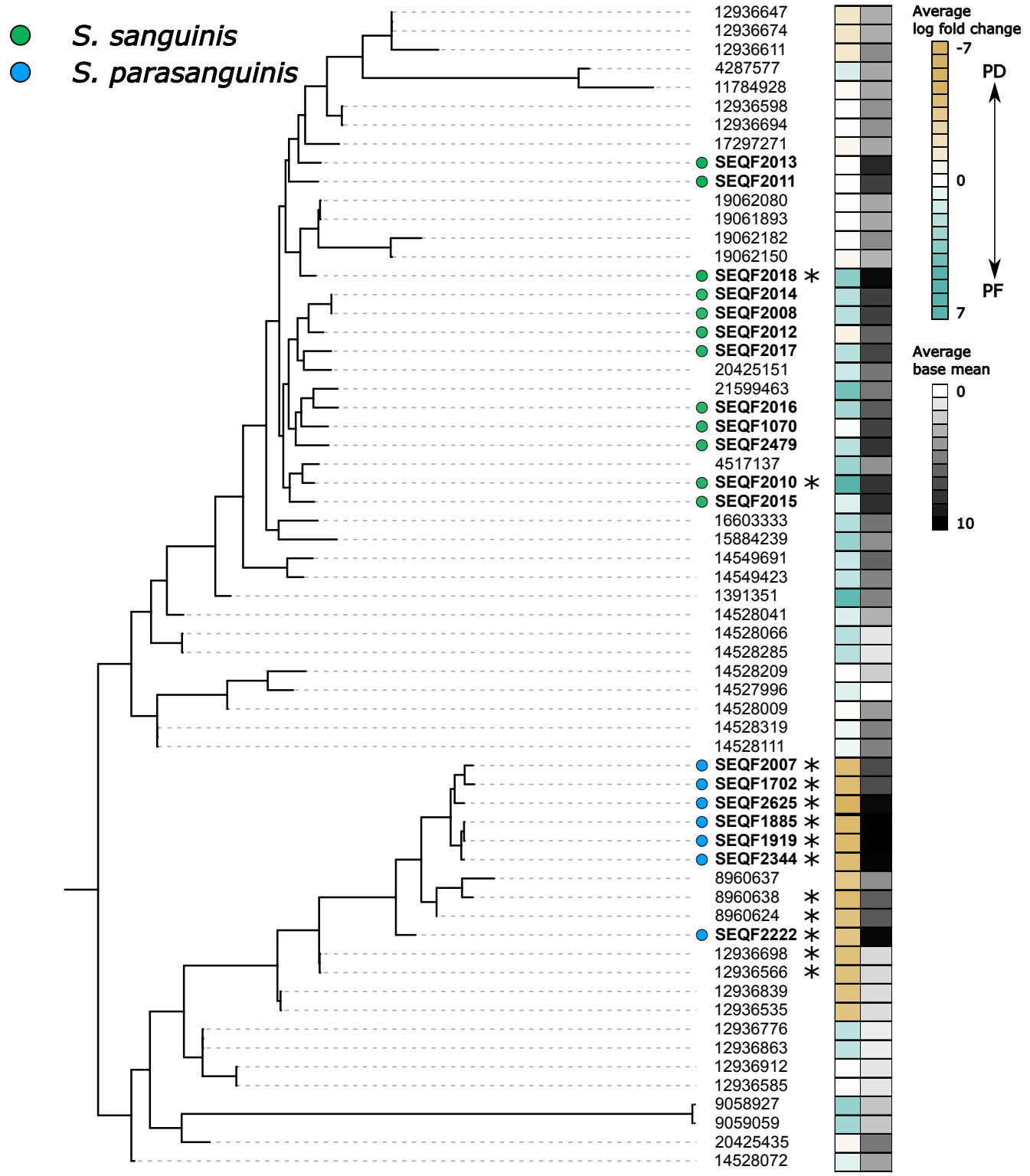

FIG 6 *De novo* assembly of ADS transcripts document variable expression in health and disease among two *Streptococcus* species. Maximum likelihood tree of reference HOMD *Streptococcus* found in a subset of Fig. 3 (*S. sanguinis* and *S. parasanguinis*; bolded SEQF annotations) and *de novo*-assembled ADS transcripts assigned to these species (numbers only). Heatmap corresponds to the average log fold change over all three *arc* genes and the average base mean based on either HOMD reference genome mapping (SEQF lineages) or *de novo* assembly (numbered lineages). Within species, there are variable expression levels among the two species in health and disease with only a subset having all three genes up- or downregulated (denoted by asterisks).

levels among *Streptococcus sanguinis* and *S. parasanguinis* are more variable (Fig. 6; Table S12). The results of this analysis demonstrate the diversity of the ADS operon among oral taxa, particularly in understudied groups like *Leptotrichia*.

Streptococcus mutans antagonism

Next, as ADS activity and ADS-competent bacteria have previously been shown to reduce the ability of *Streptococcus mutans* to proliferate within the plaque biofilm, we compared the proportions of ADS-competent bacteria and *S. mutans* in both PD and PF samples. We found that while PD samples as a whole had a higher prevalence across samples (77.78%) and overall community proportion (12.43% ± 18.47%) of *S. mutans* as compared to PF (prevalence: 46.88%, average proportion: 2.40% ± 8.44%), 55.56% had low (<5%) *S. mutans* abundance (Fig. 7a) and *S. mutans* abundance was not correlated with citrulline production (Fig. S1). Next, we performed selbal analysis (37) to identify groups of taxa that are most associated with citrulline production and found that the balance in the ratio between *Veillonella parvula* and *Corynebacterium matruchotii* have the highest association with the total amount of citrulline detected in our samples (Fig. 7b; $R^2 = 0.32$).

Finally, samples with low *S. mutans* (<10% of the total community) have upregulation of the ADS pathway for a variety of taxa including *Parvimonas micra, Actinomyces naeslundii, Bulleidia extructa, Fusobacterium* sp. HMT 370, *Lactobacillus ultunensis, Olsenella uli, Peptostreptococcaceae nodatum, Streptococcus constellatus*, and *Treponema*

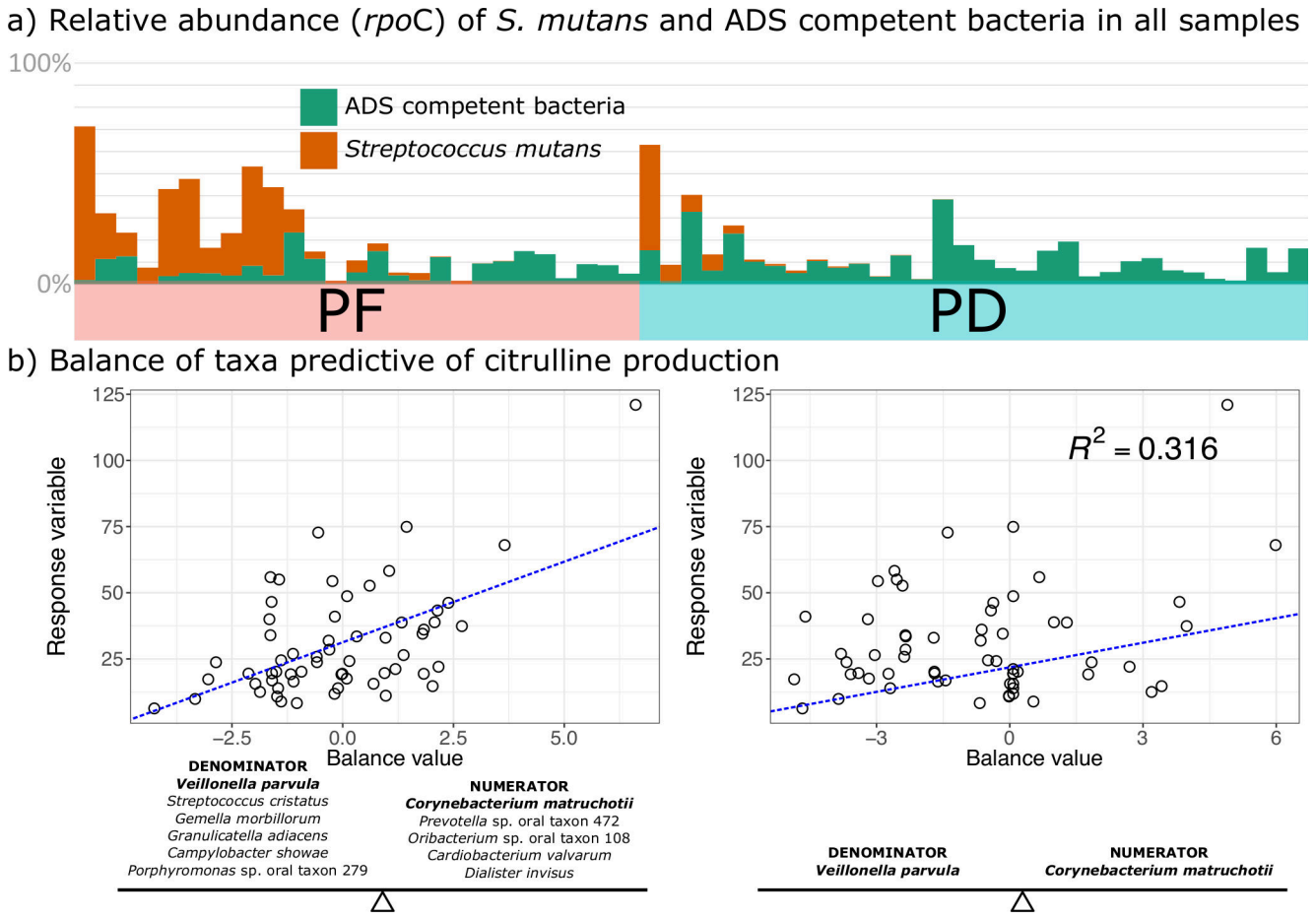

FIG 7 Evidence for *Streptococcus mutans* antagonism with ADS-competent bacteria. (a) Relative proportion of amplicon reads for ADS-competent bacteria identified in our metatranscriptomic data set as compared to *Streptococcus mutans* across all samples sequenced. (b) Balance of taxa most predictive of citrulline production. Scatter plot on the left illustrates the maximum number of taxa whose ratio is associated with the response variable. Plot on the right illustrates the two taxa whose ratio best predicts the maximum amount of variation in the response variable.

vincentii. We found no significant correlation between the relative abundance of these taxa and *S. mutans* in our ASV data set. Upregulation of the ADS pathway is disproportionately skewed by PD samples with low *S. mutans* prevalence, and only a single gene assigned to *Streptococcus infantis* (*arc*B) is upregulated in PD samples with high *S. mutans* (Fig. S2a; Table S13). Overall, samples with high *S. mutans* abundance also have a lower relative abundance of *S. oralis* and *S. sanguinis* with the exception of two PF samples (Fig. S3). Interestingly, both PF samples (UF17PF and UF55PF) with a relatively high abundance of *S. mutans* also have a large proportion of total ADS transcripts assigned to *S. sanguinis*. For example, 48% of *rpo*C amplicons derived from sample UF55PF were assigned to *S. mutans,* while 14% were assigned to *S. sanguinis*, yet 52% of all ADS transcripts originating from this sample were assigned to *S. sanguinis*.

Stress response and virulence genes are upregulated in disease

Finally, while this study focused on gene expression for a single metabolic pathway, arginine deiminase, we found 210,295 genes that had significantly different expressions between PD and PF samples, 63,268 (30%) of which are unknown hypothetical proteins. Much of the functional diversity at the community level is driven by PD samples, which exhibit low functional and community cohesion as compared to the PF samples (Fig. 8a). These differences are the result of a strong functional signal of cariogenic taxa [and particularly *Streptococcus mutans* and *Propionibacterium acidifaciens* (also known as

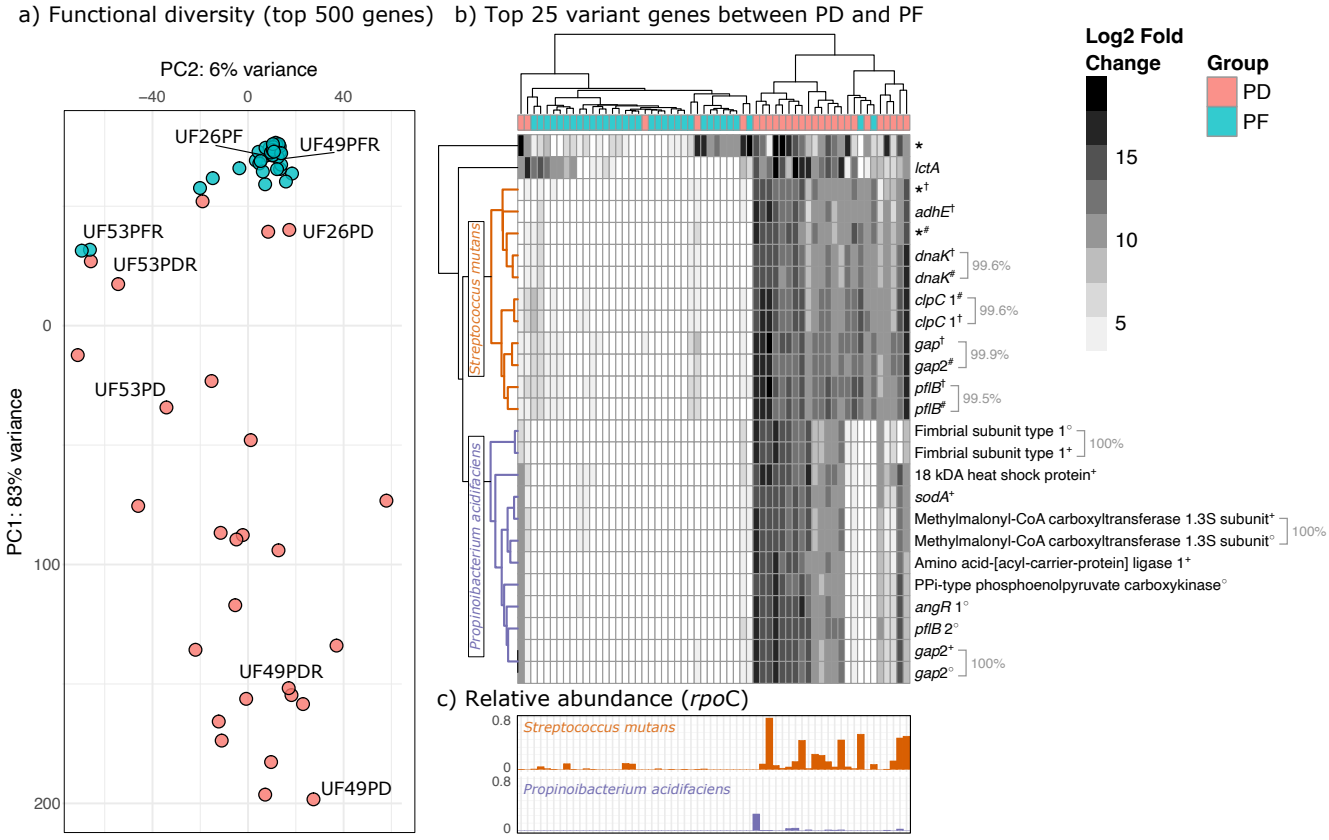

FIG 8 Community-level differential gene expression profiles are more individualistic in PD as compared to PF samples. (a) PCA plot showing the differences in functional profiles among samples using the top 500 varying genes between groups. Teeth sampled from the same individual are highlighted in the plot. (b) Hierarchically clustered heatmap of the top 25 genes (determined by highest variance) that differentiate between PD and PF samples. Genes marked by an asterisk are annotated as hypothetical proteins. Different lineages of *S. mutans* (SEQF1069†, SEQF1382#) and *P. acidifaciens* (SEQF1851+, SEQF2583○) are indicated by corresponding symbols. Identical genes detected in different lineages are indicated by brackets, and corresponding percentages indicate genetic identity between the two genes as determined by blastn alignment. (c) Relative abundance of *Streptococcus mutans* and *Propionibacterium acidifaciens* for each sample corresponding to heatmap dendrogram.

Acidipropionibacterium acidifaciens, Fig. 8b)]. Genes highly expressed in PD samples as compared to PF include a variety of stress response or virulence factors including those involved in biofilm formation, acid tolerance, the production of antimicrobial peptides, and response to environmental stressors. For example, *pfl*B (pyruvate format-lyase-encoding) genes are key factors in the colonization of *Salmonella* in the gut facilitated by host cell apoptosis (38), *clp*C encodes an ATPase involved in response to environmental stress and proteolytic activity (39, 40), *dna*K protects against a wide variety of environmental stressors and promotes biofilm growth and lactic acid fermentation at high temperatures in lactic acid bacteria (41–43), *adh*E promotes the production of mutacin by *S. mutans* that has antagonistic activity against a wide range of Gram-positive bacteria including other members of *Streptococcus* (44, 45), and genes encoding fimbriae (Fimbrial subunit type 1) are involved in biofilm development. Interestingly, the downregulation of genes encoding the major and minor fimbriae (*fim*A and *mfa*1) have previously been shown to be disrupted by ADS activity in the oral cavity (46).

We also find multiple *Streptococcus mutans* genes upregulated in the disease, which are key factors in carbohydrate metabolism, including *ccp*A, *cod*Y, *pac*L, *gtf*A, *gtf*C, *lev*D, and *fru*A (47, 48). Transcripts assigned to these genes are derived from two strains of *S. mutans:* UA159 and NN2025, which are also the dominant *S. mutans* strains found in the *rpo*C data set (44% and 52% of all *rpo*C amplicons assigned to *S. mutans*, respectively). Like our ADS pathway results, the abundance *of S. mutans* appears to be driving most of the expression-level data (Fig. 8c), where 78% of PD samples have *rpo*C reads assigned to *S. mutans* as compared to 48% of PF samples. Interestingly, though, *Propionibacterium acidifaciens* is found in very low abundance across the data set, even for those samples with high *P. acidifaciens* gene expression. About 59.26% of PD and 0% of PF samples had *rpo*C amplicons assigned to this species with an average abundance of 1.25% (± 4.29%) in PD. Similarly, on average 1,088 transcripts (±1,870) were assigned to *P. acidifaciens* in PD samples as compared to 0.10 (±0.09) in PF. Assuming this is not the result of the disproportionate representation of *S. mutans* as compared to *P. acidifaciens* in our database (247 vs 2 representative genomes), these results may indicate that this species may be an important driver of tooth decay, even at low frequency in the oral cavity. Importantly, two strains assigned to both *S. mutans* and *P. acidifaciens* are represented in Fig. 8b and contribute relatively equally to the variance between PF and PD as genes found in these strains are too similar or identical to one another to distinguish between the genome of origin (ranging from 99.5% to 100% genetic identity).

DISCUSSION

Manipulation of ADS activity in the oral cavity either through the administration of prebiotics in the form of arginine toothpaste or the development of ADS-competent probiotic strains (6, 10, 11, 13, 14, 16, 19, 49) are promising prophylactic treatments for the prevention of tooth decay, which remains a major global public health problem (50). The results of this study complement and build upon previous investigations of the significance of ADS activity in the oral cavity in preventing tooth decay by providing a multivariate analysis of differential gene expression of ADS-competent bacteria in the context of a mixed microbial ecosystem.

In agreement with previous studies (4, 6, 11), we found that while not completely lacking ADS activity, PD samples generally had lower citrulline production from arginine *in vitro* (Fig. 1b) and a lower proportion of ADS-competent bacteria compared to PF samples as measured by both amplicon frequency and transcript frequency (Fig. 5c and 1d). Interestingly, however, we detected 13 strains of bacteria with upregulation of the ADS operon in caries disease. Most strains identified as upregulated in PD make up a very small proportion of the total oral community (e.g., *Limosilactobacillus fermentum* and *Lactobacillus ultunensis*), which may indicate that to be effective in preventing tooth decay, the ADS operon must be expressed by a taxon or group of taxa that reach a certain abundance threshold. Alternatively, some taxa may have higher ADS activity in PD but not secrete sufficient levels of ammonia at an adequate rate into

the surrounding biofilm matrix to negate environmental acidification, independent of community abundance. Previous research has demonstrated that oral bacteria must be able to generate alkali products above a minimum threshold to prevent acidification of the environment by sugar metabolism (51) and caries development in experimental animals (52). Not all ammonia produced by alkali-generating metabolic pathways is rapidly released into the extracellular environment. Consider, for example, that while *Streptococcus mutans* lacks the ADS operon, it catabolizes agmatine to produce ammonia, CO_2, and ATP through the agmatine deiminase (AgDS) pathway, which closely resembles arginine catabolism through the ADS. AgDS contributes to acid tolerance and growth of *S. mutans* by ATP and intercellular ammonia production, the latter of which increases cytoplasmic pH, apparently without a substantial effect on the pH of the plaque biofilm (53). Moreover, as the ADS pathway is highly regulated, specific environmental conditions may be necessary to trigger higher ADS activity in some species that are not conducive to the prevention of tooth decay. For example, low pH increases the transcription of the ADS pathway in *Streptococcus gordonii* (36). It may be the case that the ADS pathway is upregulated at low pH in some taxa where cellular survival is possible but not the prevention of enamel demineralization. ADS activity can persist in some bacterial species (e.g., *Streptococcus*) at extremely low pH (5) and thus higher transcript expression of the ADS operon may continue even in highly acidic environments. As tooth demineralization is a consequence of not only environmental acidification but also the duration of a low pH environment in the plaque biofilm, potential preventative strains of bacteria should have higher ADS expression independent of environmental conditions excluding the availability of arginine itself (28). Finally, our ADS operon coverage analysis and *de novo* assembly of full ADS transcripts suggest that regions of the operon may undergo differential post-translational regulation, which may, in turn, lead to lower ammonia production in some strains or perhaps in certain environmental contexts (54). Further research on the regulatory mechanisms and metabolic behavior of the ADS-competent bacteria in PD samples detected here would clarify their efficacy in the prevention of tooth decay.

While we find ADS activity is upregulated in both health and disease, it is important to note that upregulation of the ADS pathway in PD samples was limited to those with low (<5%) *Streptococcus mutans* abundance, which may suggest that ADS activity prevents the proliferation of this cariogenic taxon in the oral cavity or, alternatively, that the presence of *S. mutans* inhibits ADS activity. Moreover, the oral environment across samples in both health and disease may be more or less conducive to ADS activity. For example, ADS activity is dependent on the availability and concentration of arginine as a secondary carbon source and can, in some taxa, be repressed by the oxygen tension in particular sites in the oral cavity (55, 56). Additionally, coverage plots and *de novo* assembly of ADS transcripts generated in the current study suggest the presence of gene-specific regulation of the ADS operon among certain groups, particularly of the *arc*C gene in *Streptococcus* sp. This post-transcriptional regulation may result from competition for substrates necessary for other metabolic pathways in certain environmental conditions. For example, the anaerobic growth of *Streptococcus thermophilus* is strongly dependent on carbamoylphosphate synthetase activity (CPS), which synthesizes carbamoyl phosphate from glutamine, CO_2, and two ATP molecules as a precursor for arginine and pyrimidine biosynthesis (57). The CPS synthetase domain is found in nearly all organisms (58), and its activity is strongly inhibited by arginine (3, 57, 59); thus, in anaerobic environments where arginine is depleted and CPS activity is high, *arc*C activity may be suppressed through selective mRNA degradation to prevent the catabolism of carbamoyl phosphate to ATP and NH_3, instead favoring its use as a substrate for arginine and pyrimidine production. Importantly, in taxa lacking CPS, the arginine deiminase pathway can replace CPS as the primary provider of carbamoyl phosphate for pyrimidine biosynthesis (60). Finally, the higher abundance of ADS-competent bacteria may lead to nutritional resource competition with *S. mutans* and other cariogenic bacteria that require arginine and arginine-containing peptides, all of which

may contribute to antagonism between *S. mutans* and ADS-competent bacteria in both PF and PD. Given that oral health exists on a continuum and tooth decay is a progressive disease, the relative contribution of the oral environment, microbial community structure, environmental secretion of alkaline products of metabolism, post-transcriptional mRNA regulation, and the presence or absence of specific strains in promoting health or disease likely fluctuates over time.

Despite this variability, we identified lineages for *Streptococcus oralis, S. sanguinis, Streptococcus* sp. HMT 056, and *Leptotrichia* sp. HMT 212 and 215, which may contain effective probiotic strains. Moreover, we detected taxa that may have important roles in the structural development of plaque, the inclusion of which may be necessary for probiotic success. For example, previous research identified *Corynebacterium matruchotii* as one of the taxa that increased with regular brushing with fluoride + arginine toothpaste while species of *Veillonella* decreased (15). In the current research, we find both *C. matruchotii* and *Veillonella parvula* to be associated with high or low citrulline production, respectively. In addition, *C. matruchotii* plays a significant role in biofilm formation, can induce the mineralization of plaque, and recently was shown to increase the growth rate of various *Streptococcus* sp. *in vitro* (61–63). It, therefore, may be a key structural taxon for the colonization and proliferation of ADS-competent bacteria in the oral cavity.

Importantly, the analysis of multiple *Streptococcus* genomes for each species highlighted that ADS upregulation is heterogeneous among lineages identified by our metatranscriptomics data set (e.g., *S. oralis* = 25%, *S. sanguinis* = 17%), that not all lineages annotated by HOMD possess the full ADS operon (e.g., *S. oralis* = 25%, *S. sanguinis* = 92%; Table S3), and that *de novo* assembly of the ADS operon can identify multiple distinct transcripts for the ADS operon assigned to a single species. Additionally, we found that relatively high proportions of ADS-competent bacteria do not equate to lower proportions of cariogenic bacteria in all cases. For example, two PF samples in our data set had both high frequency of *S. mutans* and *S. sanguinis* as measured by *rpo*C sequencing (Fig. S3). However, as these samples also had high ADS activity by *S. sanguinis,* one possible interpretation is that this activity has prevented the development of a cavity in these teeth, despite the presence of *S. mutans*. Alternatively, samples with a high relative frequency of *S. sanguinis* and *S. mutans* but low ADS activity for *S. sanguinis* (e.g., UF68PD) may reflect strain-level differences in ADS competency. Given that some lineages of *S. sanguinis* do not have a full ADS operon (Table S3), it may be that some strains are not capable of antagonizing *S. mutans* even at relatively high frequency. The importance of strain-level resolution in determining candidate probiotics to promote ADS activity in the oral cavity has previously been described among strains of closely related *Streptococcus* species (28). Our results confirm the importance of strain-level upregulation of ADS in the context of oral health for different *Streptococcus* species (e.g., *S. oralis, S. sanguinis,* and *Streptococcus* sp. HMT 056) and identify potential candidate species that are as of yet understudied for their role in the promotion of oral health. For example, we found two strains (lineages) of *Leptotrichia* (HMT 212 and HMT 215) that had high ADS activity in PF and comparatively high read abundance and prevalence across samples in the current data set. Moreover, *de novo* assembly of full ADS operons documents high levels of diversity within this group, which may contribute to health or disease. Conversely, our results support the suggestion that lineages of bacteria that are typically associated with health (e.g., non-mutans *Streptococcus* sp.) may be important contributors to the etiological development of caries (64, 65). Along with potential probiotic uses, identification of specific bacterial lineages consistently associated with health and with known beneficial properties in the oral cavity may be used as additional markers of risk assessment beyond the presence of cariogenic taxa, though further investigation into the prevalence and functional activity of the taxa identified here is necessary to clarify their role in maintaining favorable pH in the oral cavity.

Analysis and interpretation of metatranscriptomic data sets generated from mixed microbial communities are complicated by the differences in community composition

and membership (66, 67), particularly when the microbial community found in comparative groups is highly dissimilar. Thus, accounting for the differences in community composition between groups is critical (68, 69). Here, we compared log-fold changes in a parallel metataxonomic data set to investigate the impact of community composition and species abundance on our metatranscriptomics results. Our paired metatranscriptomic and metataxonomic results demonstrate that often higher gene expression is a consequence of differential community membership or abundance and not truly neutral differences in gene expression profiles. For example, all three *Streptococcus* species found here to have higher ADS activity on healthy teeth typically make up a much larger proportion of PF communities, and so more transcripts may be a simple function of more cells contributing to the effect. Conversely, taxa found at low frequency may be functionally more active in some circumstances (e.g., *Limosilactobacillus fermentum, Lactobacillus ultunensis,* and *Propionibacterium acidifaciens*). Given that, a good candidate probiotic will need not only to efficiently catabolize arginine to produce ammonia but also have the ability to effectively colonize the plaque biofilm, comparing genome and transcript frequency is an important consideration. Moreover, transcript abundance alone does not necessarily precisely correlate to actual *in vivo* biochemical activity, and a variety of confounding factors including strain-level translational efficiency and enzyme stability may impede our accurate interpretation of activity levels alone. For the purposes of this study, we focus on taxa with both high RNA transcript abundance and comparatively high ASV abundance as these taxa have (i) high ADS expression that is not directly linked to high relative abundance in a sample and (ii) are prevalent across multiple individuals (and thus may be more efficient colonizers of the plaque biofilm).

Finally, we find that while ASV diversity is lower in PD as compared to PF samples (Fig. 1a), functional diversity is much more dispersed as compared to healthy teeth, which tend to have a much more consistent functional profile (Fig. 6a). The main functional differences between PD and PF communities are largely driven by a variety of genes associated with virulence factors and carbohydrate metabolism in known cariogenic taxa, and these effects are largely individualistic in scope, which suggests the presence of high functional volatility during caries development. Interestingly, we detected two major taxa driving the majority of differentially abundant genes in PD vs PF, *Streptococcus mutans,* and *Propionibacterium acidifaciens,* even though *P. acidifaciens* occurs at low frequency for both groups in our corresponding amplicon data set. While this may be the result of taxonomic biases inherent in amplicon data sets, it may also demonstrate the important functional activity by relatively low-frequency taxa. Moreover, as *P. acidifaciens* biofilm formation is inhibited by *S. mutans* growth (70), a better understanding of the functional interaction between these species may elucidate their impact on the total ecology. The volatility of the functional profile of PD teeth, coupled with lowered diversity, suggests that candidate probiotics should not be limited to a single species.

Results of the current study highlight potential candidates for probiotic panels, including oral bacteria where pH modulation through the ADS pathway has previously been identified (e.g., *Streptococcus* sp.) (2), as well as those that are less well characterized (e.g., *Leptotrichia* sp.). In conclusion, while our multivariate approach substantiates the role of the ADS pathway in health and disease, it highlights the importance of accounting for taxonomic shifts in the analysis of functional differences in mixed microbial ecosystems such as dental plaque, which will impact the efficacy of any caries treatment, and suggests that probiotic development may benefit from further strain-level investigation of a wider assortment of species for ADS activity and pH modulation potential in mixed microbial communities.

MATERIALS AND METHODS

Samples and study design

A total of 54 adult individuals were recruited for this study. Carious lesions were diagnosed using the International Caries Detection and Assessment System II (ICDAS-II) visual criteria (71) and the lesion activity (72) scoring system by a calibrated examiner (M.M.N.) as previously described (11, 73). Sampling of site-specific plaque samples followed protocols previously published by this group (11). Briefly, study participants were instructed to refrain from oral hygiene procedures for 8 hours prior to plaque sampling, after which supragingival plaques were collected from individual teeth and classified as caries-free (PF) (ICDAS score = 0) or with an active dentin carious lesion (PD) (ICDAS score ≥ 4). PD samples were recovered from the internal surface of the lesion. Full metadata for each sample can be found in Table S14.

Citrulline assay

A total of 74 supragingival plaque samples were tested for their potential to generate citrulline from arginine via the arginine deiminase system using a protocol previously validated and published by our group (28). Briefly, plaque samples were dispersed using external sonication for two cycles of 30 seconds with cooling on ice during the interval. The cells were then harvested, washed, and resuspended in 10 mM Tris-Maleate buffer and further permeabilized with toluene-acetone (1:9) for the measurement of ADS activity. The total protein concentration of the plaque sample was also measured using a BCA protein estimation kit (Pierce, Waltham, MA, USA) with a known concentration of bovine serum albumin as a standard. The ADS activity levels in the plaques were normalized to protein content and represented as nanomoles of citrulline generated per milligram of protein.

DNA and RNA extraction

Plaque samples were stored in RNAlater (ThermoFisher Scientific, Invitrogen, Waltham, MA, USA) immediately after sampling. DNA and RNA were extracted simultaneously following the mirVana miRNA isolation kit for RNA protocol (Ambion, ThermoFisher Scientific, Waltham, MA, USA). We chose this isolation kit as it is designed to efficiently isolate small RNAs (<200 nt), as well as total RNA, which play a crucial role in regulating bacterial global gene expression and are routinely used for microbiome studies (74–76). Cells were washed and collected by centrifugation for 10 minutes at maximum speed. A quantity of 300 µL of mirVana kit lysis/binding buffer, 300 µL acid-phenol:chloroform, and 250 µL of 0.1 mm sterilized zirconia-silica beads (BioSpec Products, Bartlesville, OK, USA) were added to the samples. Samples were bead beaten twice for 30 seconds with a 15-second interval in between. Samples were centrifuged at maximum speed for 5 minutes at room temperature for clear separation of aqueous and organic phases. For RNA isolation, the aqueous phase was collected and processed according to the manufacturer's protocol. For DNA isolation, the organic phase from the mirVana mRNA isolation was stored at 4°C until use. An extraction buffer containing 0.1 M NaCl, 10 mM Tris-HCL (pH 8.0 ± 0.2), 1 mM EDTA, and 1% SDS was prepared. The final pH of the buffer was adjusted to pH 12 (±0.2) with 10 N NaOH. We next added 250 µL of the extraction buffer to the organic phase containing DNA followed by an incubation at 4°C. After centrifugation at 10,000 *g* for 20 minutes, the aqueous phase was recovered and immediately stored at −80°C for an hour after the addition of 1/15 vol of 7.5 M ammonium acetate and 100% ethanol. The samples were then centrifuged for 20 minutes at 10,000 *g,* and the supernatant was discarded. The pellets were air dried and washed with ethanol before being resuspended in 17 µL of molecular-grade water.

RNA library preparation and sequencing

Total RNA samples were processed twice with MICROBEnrich and MICROBExpress bacterial mRNA enrichment kits (Ambion/Life Technologies, Grand Island, NY, USA) to

remove eukaryotic mRNA and 16S/23S rRNAs, respectively. The purified mRNA was resuspended in 25 µL of nuclease-free water. The quality of enriched mRNA was analyzed using an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Next, we generated cDNA libraries from 100 ng of enriched mRNA using the NEBNext Ultra directional RNA library preparation kit and NEBNext dual index oligonucleotides for Illumina (New England BioLabs, Ipswich, MA, USA). To ensure proper enrichment of bacterial transcripts, we first sequenced the pooled library using a MiSeq Reagent Nano Kit v2 (300 cycles) at Clemson University. Deep sequencing of the library was next performed by the NextGen DNA sequencing core laboratory of the Interdisciplinary Center for Biotechnology Research at the University of Florida (Gainesville, FL, USA). Samples were sequenced on the NovaSeq 6000 platform on a single S4 flow cell using a paired-end 2 × 150 bp chemistry.

DNA library preparation and sequencing

Each plaque sample was amplified and built into Illumina sequencing libraries using a custom primer set targeting an approximately 478-bp fragment of the bacterial *rpo*C gene-rpoCF (5′-MAYGARAARMGNATGYTNCARGA-3′) and rpoCR (5′-GMCATYTGRTCNCCRTCRAA-3′) (77). We adopted this approach as recent research has demonstrated the inability of other common metabarcoding approaches (e.g., 16S rRNA hypervariable regions) to resolve common oral groups (i.e., *Streptococcus* sp.) (78), and further research identified the *rpo*C gene as a promising alternative marker (24). We note, however, that like other amplicon-based approaches, *rpo*C gene fragment metabarcoding is not free from bias [see reference (77)], and the primers published here have only been tested using oral samples. Each 25 µL PCR reaction consisted of 0.5 µL of the forward and reverse adapter, 10 µL of H_2O, and 10 µL of Platinum Hot Start PCR Master Mix (2×) (Invitrogen, Carlsbad, CA, USA). PCR thermocycler conditions were as follows: 94°C for 3 minutes followed by 41 cycles of 94°C for 45 seconds, 39.5°C for 1 min, and 72°C for 1 min and 30 seconds. A final elongation step was performed at 72°C for 10 minutes. Amplification of PCR products for each sample was confirmed by both gel electrophoresis and Qubit Fluorometer (Invitrogen, Carlsbad, CA, USA). We processed PCR blanks (molecular-grade water) in parallel to all samples. Prepared libraries were sequenced on an Illumina MiSeq using 2 × 300 paired-end chemistry.

Metatranscriptomic data processing

On average, we generated 326,151,580 (SD ± 102,407,864) raw sequencing reads per sample. We first assessed the quality of the raw reads using FastQC v0.11.9 (79) before quality filtering, trimming adapters, and removing poly-G tails using Cutadapt v3.3 (80). On average, 302,238,342 (SD ± 9,238,698) reads passed our quality filtering thresholds. Next, we performed a preliminary filter on the trimmed reads by first mapping to the SILVA v138 non-redundant 99% SSU and LSU 18S rRNA and 16S rRNA databases (81) and then to the human reference genome (GRCh38) using Bowtie2 v2.4.5 (82). On average 0.02% (SD ± 0.06%) of our quality-filtered reads per sample were mapped to the SILVA database and 3.04% (SD ± 4.43%) to the human genome. We removed any reads that mapped to either the human reference genome or SILVA ribosomal databases using a custom Python script before performing downstream analyses.

For taxonomic and functional assignment, all filtered reads were mapped against the Human Oral Microbiome Database (v9.15) (31) using STAR v2.7.5c (83) with alignIntroMax set to 1 to disable spliced alignments and outFilterMismatchNoverLmax set to 0.05 to increase mapping specificity. On average, 198,385,752 (SD ± 68,279,777) reads mapped to our database per sample. Gene counts were calculated from each sample using the featureCounts function as implemented in the Subread v2.0.1 package (84, 85). Given that our reference-based approach included a database of diverse microbial genomes with varying levels of genomic congruence, when running featureCounts, we allowed for multimapping genes (-M), only pairs with both reads aligned were counted (-B),

and paired reads that mapped to separate chromosomes (or contigs in the case of incomplete genome reconstructions) were not counted for downstream analyses (-C).

To complement our reference-based approach, we performed an additional *de novo* assembly on a sample-by-sample basis using Trinity (v2.15.1) (86). On average, we generated 364,811 transcripts per sample (±159,936). We next merged the assembled transcripts and performed a full length dereplication using vsearch (v2.14.1) (87). We annotated the final dereplicated database using prokka (v1.14.5) (88). Finally, we mapped our filtered reads to the *de novo* reference using STAR (v2.7.5c) (83) in an identical manner as described above. We identified significantly up- or downregulated genes using DESeq2 (89). We built maximum likelihood trees of our *de novo* ADS transcripts with references from HOMD using mafft (v7.453) (90) and RAxML (v8.2.12) (91), which were visualized in iTOL (v5) (92). As many transcripts were truncated or otherwise incomplete, we built phylogenetic trees with only those transcripts that had annotations for all three *arc* genes (*arc*ABC) and trimmed the final alignments using trimAl (v1.4.rev15) with a gap threshold of 0.5, minimum position overlap in a column of 0.5, and minimum percentage of "good positions" a sequence must retain for inclusion of 50 (93). Assembly statistics can be found in Table S8.

Metataxonomic data processing

On average, we generated 18,417 raw amplicon sequences per sample (±17,284), of which 11,517 (±10,347) passed our quality filtering thresholds. Rarefaction curves for all samples can be found in Fig. S4. Briefly, we trimmed primer and adapter sequences from our raw reads using Cutadapt (v4.0) (80), after which we quality filtered, trimmed, merged, removed chimeric sequences, and generated ASVs using DADA2 (v1.22.0) (94). In total, we generated 3,928 unique ASVs. We next assigned a taxonomy to each ASV using Kraken2 (v2.1.2) (95) with the Human Oral Microbiome Database whole-annotated genomes as reference (31). Low-frequency ASVs (prevalence threshold = 0.01) or those with no or kingdom-only taxonomic assignments (i.e., Unknown or Bacteria) were removed from downstream analyses leaving 3,114 ASVs in the data set. A full ASV table with annotated taxonomy can be found in Table S15. All amplicon sequence analyses were performed in the R v4.1.2 environment (96). We calculated alpha and beta-diversity metrics using the Phyloseq (v1.38.0) (97), vegan (v2.6-2) (98), and ggdist (v3.2.0) (99) R libraries. We generated capscale plots using Bray-Curtis dissimilarity and a distance-based redundancy analysis approach (100) to visualize the differences in beta-diversity across samples. Ordination of samples was constrained by tooth type (PD vs PF) and citrulline expression (log2 transformed) (Fig. 1b). Significance between groups was calculated on isometric log-ratio transformed data normalized using PhILR (v1.20.1) (101). Finally, we identified groups of taxa associated with citrulline production using selbal (v0.1.0) (37).

ADS operon identification

The ADS operon was identified in the HOMD database by first selecting all genes annotated as those that produced one of the three main enzymes involved in the pathway. While gene content varies within and among groups, the enzymes encoded by three genes (i) *arc*A (arginine deiminase EC: 3.5.3.6), (ii) *arc*B (ornithine carbamoyltransferase EC: 2.1.3.3), and (iii) *arc*C (carbamate kinase EC: 2.7.2.2) are necessary for ADS activity and are thus the focus of this manuscript (1, 36). Because gene annotations are rarely consistent across published data sets, we first performed homologous gene clustering using the HOMD database to verify *arc* gene identity using BLAST (102) and the Markov Clustering algorithm implemented in MCLblastline (103). In brief, we created a BLAST database of all protein sequences that had been annotated as one of the three ADS enzymes (i.e., by the EC number) from the HOMD database. We next mapped the full HOMD protein database against our reference ADS database using BLASTp (e value 1e-10), the output of which we used to generate homologous gene clusters using MCLblastline with an inflation parameter of 1.8, which previously had been identified

as well suited for high-throughput data (104) and, in practice, can identify homologous gene clusters across diverse groups (105). Genes that did not cluster with any other *arc* gene in the database (e.g., singleton clusters) and annotated genes that fell into clusters that included a mixture of genes annotated as *arc*A, *arc*B, and *arc*C were removed from downstream analyses. Using results from this analysis, we further identified species with a full ADS operon as those wherein all three genes (*arc*A, *arc*B, and *arc*C) were (i) present and (ii) in synteny with one another in the reference genome. Synteny was manually identified using locus tag assignment. Species, where synteny of the *arc* genes was interrupted by a single gene (e.g., hypothetical protein), were included in downstream analyses to account for operon variability. A full list of species that carry the ADS operon from this analysis can be found in Table S2.

Differential gene expression and species abundance analyses

We calculated differential gene expression using DESeq2 v1.34.0 (89) in R v4.1.2 (96). To minimize the effect of large fold changes between groups that were the result of low transcript count, we next ran lfcShrink on our DESeq2 results using the apeglm R package (106). Our threshold for differential expression between groups was an adjusted *P*-value (FDR) of less than or equal to 0.05 and a minimum fold change of four. Because we expect community composition and species presence to be meaningful confounding variables in comparing the estimates of gene expression across groups, we calculated the correlation between log fold change values in transcript abundance to *rpo*C gene abundance for a selection of ADS-competent oral taxa (Fig. 5c; Table S6).

Streptococcus ADS operon phylogeny

First, we concatenated the HOMD-annotated *arc*A, *arc*B, and *arc*C genes for each *Streptococcus* lineage in the HOMD database, which were validated by our homologous gene clustering into a single contiguous sequence using the concat function in SeqKit (v2.3.1). Concatenated sequences were then aligned using mafft (v7.505) (90) and built into a maximum likelihood phylogeny using RAxML (v8.2.12) (91). The phylogeny was midpoint rooted and visualized in FigTree (v1.4.4) (107). Collapsed log fold change and base mean were calculated by taking the average across all three *arc* genes for each *Streptococcus* lineage. Lineages that fell outside of an expected major species clade were checked using the BLAST web application (102) and the full NCBI NT database as a reference to ensure that it was annotated with the correct taxonomic assignment.

ADS operon coverage plots

Using sequence coordinates listed for each *arc* gene in the corresponding HOMD GFF files, we used the BEDTools (v2.27.1) (108) function getfasta to pull the *arc*ABC operon for each species of interest. Next, we mapped the full quality filtered metatranscriptomic sequencing data for each sample to the selected operon sequences using Bowtie2 (v2.4.5) (82). The depth of coverage was calculated using SAMtools (v1.12) depth (109) and plotted in R (v4.2.0) (96).

ACKNOWLEDGMENTS

We thank Alejandro Riveros Walker for comments on an early version of this manuscript and discussions on methodology.

This research was supported by the National Institutes of Health National Institute of Dental and Craniofacial Research (R01DE25832 to R.A.B.) and, in part, by computational resources provided by the Clemson University Genomics and Bioinformatics Facility, which receives support from two Institutional Development Awards (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant numbers P20GM146584 and P20GM139767.

AUTHOR AFFILIATIONS

[1]Department of Biological Sciences, Clemson University, Clemson, South Carolina, USA
[2]Department of Oral Biology, College of Dentistry, University of Florida, Gainesville, Florida, USA
[3]Division of Operative Dentistry, Department of Restorative Dental Sciences, College of Dentistry, University of Florida, Gainesville, Florida, USA

AUTHOR ORCIDs

Allison E. Mann http://orcid.org/0000-0001-7170-6017
Robert A. Burne http://orcid.org/0000-0002-4234-0316
Vincent P. Richards http://orcid.org/0000-0002-9309-6595

FUNDING

Funder	Grant(s)	Author(s)
HHS | NIH | National Institute of Dental and Craniofacial Research (NIDCR)	R01DE25832	Robert A. Burne

AUTHOR CONTRIBUTIONS

Allison E. Mann, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review and editing | Brinta Chakraborty, Data curation, Methodology, Writing – original draft, Writing – review and editing | Lauren M. O'Connell, Investigation, Methodology, Writing – review and editing | Marcelle M. Nascimento, Investigation, Writing – review and editing | Robert A. Burne, Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review and editing | Vincent P. Richards, Conceptualization, Methodology, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

Conda environments and processing scripts are provided at https://github.com/aemann01/ads_plaque and archived under the DOI: 10.5281/zenodo.7760990. Raw metatranscriptomic data have been deposited in the European Nucleotide Archive under the accession number PRJEB60355 and rpoC amplicon data under the accession number PRJEB60856.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Figures S1 to S4 (Spectrum01445-23-S0001.docx). All supplemental figures.
Supplemental Tables S1 to S15 (Spectrum01445-23-S0002.xlsx). All supplemental tables.

REFERENCES

1. Novák L, Zubáčová Z, Karnkowska A, Kolisko M, Hroudová M, Stairs CW, Simpson AGB, Keeling PJ, Roger AJ, Čepička I, Hampl V. 2016. Arginine deiminase pathway enzymes: evolutionary history in metamonads and other eukaryotes. BMC Evol Biol 16:197. https://doi.org/10.1186/s12862-016-0771-4
2. Burne RA, Marquis RE. 2000. Alkali production by oral bacteria and protection against dental caries. FEMS Microbiol Lett 193:1–6. https://doi.org/10.1111/j.1574-6968.2000.tb09393.x
3. Cunin R, Glansdorff N, Piérard A, Stalon V. 1986. Biosynthesis and metabolism of arginine in bacteria. Microbial Rev 50:314–352. https://doi.org/10.1128/mr.50.3.314-352.1986
4. Nascimento MM, Alvarez AJ, Huang X, Hanway S, Perry S, Luce A, Richards VP, Burne RA. 2019. Arginine metabolism in supragingival oral biofilms as a potential predictor of caries risk. JDR Clin Trans Res 4:262–270. https://doi.org/10.1177/2380084419834234
5. Marquis RE, Bender GR, Murray DR, Wong A. 1987. Arginine deiminase system and bacterial adaptation to acid environments. Appl Environ Microbiol 53:198–200. https://doi.org/10.1128/aem.53.1.198-200.1987

6. Nascimento MM, Browngardt C, Xiaohui X, Klepac-Ceraj V, Paster BJ, Burne RA. 2014. The effect of arginine on oral biofilm communities. Mol Oral Microbiol 29:45–54. https://doi.org/10.1111/omi.12044
7. Chakraborty B, Burne RA. 2017. Effects of arginine on *Streptococcus mutans* growth, virulence gene expression, and stress tolerance. Appl Environ Microbiol 83:e00496-17. https://doi.org/10.1128/AEM.00496-17
8. He J, Hwang G, Liu Y, Gao L, Kilpatrick-Liverman L, Santarpia P, Zhou X, Koo H. 2016. L-arginine modifies the exopolysaccharide matrix and thwarts *Streptococcus mutans* outgrowth within mixed-species oral biofilms. J Bacteriol 198:2651–2661. https://doi.org/10.1128/JB.00021-16
9. Kamaguch A, Nakayama K, Ohyama T, Watanabe T, Okamoto M, Baba H. 2001. Coaggregation of *Porphyromonas gingivalis* and *Prevotella intermedia*. Microbiol Immunol 45:649–656. https://doi.org/10.1111/j.1348-0421.2001.tb01298.x
10. Nascimento MM, Gordan VV, Garvan CW, Browngardt CM, Burne RA. 2009. Correlations of oral bacterial arginine and urea catabolism with caries experience. Oral Microbiol Immunol 24:89–95. https://doi.org/10.1111/j.1399-302X.2008.00477.x
11. Nascimento MM, Liu Y, Kalra R, Perry S, Adewumi A, Xu X, Primosch RE, Burne RA. 2013. Oral arginine metabolism may decrease the risk for dental caries in children. J Dent Res 92:604–608. https://doi.org/10.1177/0022034513487907
12. Sharma S, Lavender S, Woo J, Guo L, Shi W, Kilpatrick-Liverman L, Gimzewski JKY. 2014. Nanoscale characterization of effect of L-arginine on *Streptococcus mutans* biofilm adhesion by atomic force microscopy. Microbiology (Reading) 160:1466–1473. https://doi.org/10.1099/mic.0.075267-0
13. Acevedo AM, Machado C, Rivera LE, Wolff M, Kleinberg I. 2005. The inhibitory effect of an arginine bicarbonate/calcium carbonate (Cavistat) -containing dentifrice on the development of dental caries in Venezuelan school children. J Clin Dent 16:63–70.
14. Acevedo AM, Montero M, Rojas-Sanchez F, Machado C, Rivera LE, Wolff M, Kleinberg I. 2008. Clinical evaluation of the ability of Cavistat in a mint confection to inhibit the development of dental caries in children. J Clin Dent 19:1–8.
15. Carda-Diéguez M, Moazzez R, Mira A. 2022. Functional changes in the oral microbiome after use of fluoride and arginine containing dentifrices: a metagenomic and metatranscriptomic study. Microbiome 10:159. https://doi.org/10.1186/s40168-022-01338-4
16. Wolff MS, Schenkel AB. 2018. The Anticaries efficacy of a 1.5% arginine and fluoride toothpaste. Adv Dent Res 29:93–97. https://doi.org/10.1177/0022034517735298
17. Casiano-Colón A, Marquis RE. 1988. Role of the arginine deiminase system in protecting oral bacteria and an enzymatic basis for acid tolerance. Appl Environ Microbiol 54:1318–1324. https://doi.org/10.1128/aem.54.6.1318-1324.1988
18. Nascimento MM, Burne RA. 2014. Caries prevention by arginine metabolism in oral biofilms: translating science into clinical success. Curr Oral Health Rep 1:79–85. https://doi.org/10.1007/s40496-013-0007-2
19. Nascimento MM. 2018. Potential uses of arginine in dentistry. Adv Dent Res 29:98–103. https://doi.org/10.1177/0022034517735294
20. Curran TM, Ma Y, Rutherford GC, Marquis RE. 1998. Turning on and turning off the arginine deiminase system in oral streptococci. Can J Microbiol 44:1078–1085. https://doi.org/10.1139/cjm-44-11-1078
21. Ferro KJ, Bender GR, Marquis RE. 1983. Coordinately repressible arginine deiminase system in *Streptococcus sanguis*. Curr Microbio 9:145–149. https://doi.org/10.1007/BF01567287
22. Poolman B, Driessen AJ, Konings WN. 1987. Regulation of arginine-ornithine exchange and the arginine deiminase pathway in *Streptococcus lactis*. J Bacteriol 169:5597–5604. https://doi.org/10.1128/jb.169.12.5597-5604.1987
23. Quivey RG, Kuhnert WL, Hahn K. 2001. Genetics of acid adaptation in oral streptococci. Crit Rev Oral Biol Med 12:301–314. https://doi.org/10.1177/10454411010120040201
24. Hassler HB, Probert B, Moore C, Lawson E, Jackson RW, Russell BT, Richards VP. 2022. Phylogenies of the 16S rRNA gene and its hypervariable regions lack concordance with core genome phylogenies. Microbiome 10:104. https://doi.org/10.1186/s40168-022-01295-y
25. Tian J, Utter DR, Cen L, Dong P-T, Shi W, Bor B, Qin M, McLean JS, He X. 2022. Acquisition of the arginine deiminase system benefits epiparasitic *Saccharibacteria* and their host bacteria in a mammalian niche environment. Proc Natl Acad Sci USA 119:e2114909119. https://doi.org/10.1073/pnas.2114909119
26. Dong Y, Chen Y-YM, Snyder JA, Burne RA. 2002. Isolation and molecular analysis of the gene cluster for the arginine deiminase system from *Streptococcus gordonii DI1*. Appl Environ Microbiol 68:5549–5553. https://doi.org/10.1128/AEM.68.11.5549-5553.2002
27. Noens EEE, Lolkema JS. 2017. Convergent evolution of the arginine deiminase pathway: the ArcD and ArcE arginine/ornithine exchangers. Microbiologyopen 6:e00412. https://doi.org/10.1002/mbo3.412
28. Velsko IM, Chakraborty B, Nascimento MM, Burne RA, Richards VP. 2018. Species designations belie phenotypic and genotypic heterogeneity in oral *streptococci* 3:14. mSystems 3:e00158-18. https://doi.org/10.1128/mSystems.00158-18
29. Zúñiga M, Champomier-Verges M, Zagorec M, Pérez-Martínez G. 1998. Structural and functional analysis of the gene cluster encoding the enzymes of the arginine deiminase pathway of *Lactobacillus* sake. J Bacteriol 180:4154–4159. https://doi.org/10.1128/JB.180.16.4154-4159.1998
30. Zúñiga M, Miralles M del C, Pérez-Martínez G. 2002. The product of arcR, the sixth gene of the arc operon of *Lactobacillus sakei*, is essential for expression of the arginine deiminase pathway. Appl Environ Microbiol 68:6051–6058. https://doi.org/10.1128/AEM.68.12.6051-6058.2002
31. Chen T, Yu W-H, 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
32. Belasco JG, Beatty JT, Adams CW, von Gabain A, Cohen SN. 1985. Differential expression of photosynthesis genes in *R. capsulata* results from segmental differences in stability within the polycistronic rxcA transcript. Cell 40:171–181. https://doi.org/10.1016/0092--8674(85)90320-4
33. Dar D, Sorek R. 2018. Extensive reshaping of bacterial operons by programmed mRNA decay. PLoS Genet 14:e1007354. https://doi.org/10.1371/journal.pgen.1007354
34. Newbury SF, Smith NH, Higgins CF. 1987. Differential mRNA stability controls relative gene expression within a polycistronic operon. Cell 51:1131–1143. https://doi.org/10.1016/0092-8674(87)90599-x
35. Güell M, Yus E, Lluch-Senar M, Serrano L. 2011. Bacterial transcriptomics: what is beyond the RNA horiz-ome?. Nat Rev Microbiol 9:658–669. https://doi.org/10.1038/nrmicro2620
36. Liu Y, Dong Y, Chen Y-YM, Burne RA. 2008. Environmental and growth phase regulation of the *Streptococcus gordonii* arginine deiminase genes. Appl Environ Microbiol 74:5023–5030. https://doi.org/10.1128/AEM.00556-08
37. Rivera-Pinto J, Egozcue JJ, Pawlowsky-Glahn V, Paredes R, Noguera-Julian M, Calle ML. 2018. Balances: a new perspective for microbiome analysis. mSystems 3:e00053-18. https://doi.org/10.1128/mSystems.00053-18
38. Anderson CJ, Medina CB, Barron BJ, Karvelyte L, Aaes TL, Lambertz I, Perry JSA, Mehrotra P, Gonçalves A, Lemeire K, Blancke G, Andries V, Ghazavi F, Martens A, van Loo G, Vereecke L, Vandenabeele P, Ravichandran KS. 2021. Microbes exploit death-induced nutrient release by gut epithelial cells. Nature 596:262–267. https://doi.org/10.1038/s41586-021-03785-9
39. Biswas S, Dhaked HPS, Keightley A, Biswas I. 2021. Involvement of ClpE ATPase in physiology of *Streptococcus mutans*. Microbiol Spectr 9:e0163021. https://doi.org/10.1128/Spectrum.01630-21
40. Miethke M, Hecker M, Gerth U. 2006. Involvement of *Bacillus subtilis* ClpE in CtsR degradation and protein quality control. J Bacteriol 188:4610–4619. https://doi.org/10.1128/JB.00287-06
41. Al-MahinASugimoto S, Higashi C, Matsumoto S, Sonomoto K. 2010. Improvement of multiple-stress tolerance and lactic acid production in *Lactococcus lactis* NZ9000 under conditions of thermal stress by heterologous expression of *Escherichia coli* DnaK. Appl Environ Microbiol 76:4277–4285. https://doi.org/10.1128/AEM.02878-09

42. Jayaraman GC, Burne RA. 1995. Dnak expression in response to heat shock of *Streptococcus mutans*. FEMS Microbiol Lett 131:255–261. https://doi.org/10.1111/j.1574-6968.1995.tb07785.x
43. Lemos JA, Luzardo Y, Burne RA. 2007. Physiologic effects of forced down-regulation of dnaK and groEL expression in *Streptococcus mutans*. J Bacteriol 189:1582–1588. https://doi.org/10.1128/JB.01655-06
44. Merritt J, Qi F. 2012. The mutacins of *Streptococcus mutans*: regulation and ecology. Mol Oral Microbiol 27:57–69. https://doi.org/10.1111/j.2041-1014.2011.00634.x
45. Tsang P, Merritt J, Shi W, Qi F. 2006. IrvA-dependent and IrvA-independent pathways for mutacin gene regulation in *Streptococcus mutans*. FEMS Microbiol Lett 261:231–234. https://doi.org/10.1111/j.1574-6968.2006.00351.x
46. Cugini C, Stephens DN, Nguyen D, Kantarci A, Davey ME. 2013. Arginine deiminase inhibits *Porphyromonas gingivalis* surface attachment. Microbiology (Reading) 159:275–285. https://doi.org/10.1099/mic.0.062695-0
47. Abranches J, Nascimento MM, Zeng L, Browngardt CM, Wen ZT, Rivera MF, Burne RA. 2008. CcpA regulates central metabolism and virulence gene expression in *Streptococcus mutans*. J Bacteriol 190:2340–2349. https://doi.org/10.1128/JB.01237-07
48. Moye ZD, Zeng L, Burne RA. 2014. Fueling the caries process: carbohydrate metabolism and gene regulation by *Streptococcus mutans*. J Oral Microbiol 6:10. https://doi.org/10.3402/jom.v6.24878
49. Nascimento MM, Alvarez AJ, Huang X, Browngardt C, Jenkins R, Sinhoreti MC, Ribeiro APD, Dilbone DA, Richards VP, Garrett TJ, Burne RA. 2019. Metabolic profile of supragingival plaque exposed to arginine and fluoride. J Dent Res 98:1245–1252. https://doi.org/10.1177/0022034519869906
50. Petersen PE, Bourgeois D, Ogawa H, Estupinan-Day S, Ndiaye C. 2005. The global burden of oral diseases and risks to oral health. Bull World Health Organ 83:661–669.
51. Clancy A, Burne RA. 1997. Construction and characterization of a recombinant ureolytic *Streptococcus mutans* and its use to demonstrate the relationship of urease activity to pH modulating capacity. FEMS Microbiol Lett 151:205–211. https://doi.org/10.1111/j.1574-6968.1997.tb12571.x
52. Clancy KA, Pearson S, Bowen WH, Burne RA. 2000. Characterization of recombinant, ureolytic *Streptococcus mutans* demonstrates an inverse relationship between dental plaque ureolytic capacity and cariogenicity. Infect Immun 68:2621–2629. https://doi.org/10.1128/IAI.68.5.2621-2629.2000
53. Griswold AR, Jameson-Lee M, Burne RA. 2006. Regulation and physiologic significance of the agmatine deiminase system of *Streptococcus mutans* UA159. J Bacteriol 188:834–841. https://doi.org/10.1128/JB.188.3.834-841.2006
54. Hui MP, Foley PL, Belasco JG. 2014. Messenger RNA degradation in bacterial cells. Annu Rev Genet 48:537–559. https://doi.org/10.1146/annurev-genet-120213-092340
55. Dong Y, Chen Y-YM, Burne RA. 2004. Control of expression of the arginine deiminase operon of *Streptococcus gordonii* by CcpA and Flp. J Bacteriol 186:2511–2514. https://doi.org/10.1128/JB.186.8.2511-2514.2004
56. Zeng L, Dong Y, Burne RA. 2006. Characterization of cis-acting sites controlling arginine deiminase gene expression in *Streptococcus gordonii*. J Bacteriol 188:941–949. https://doi.org/10.1128/JB.188.3.941-949.2006
57. Arioli S, Monnet C, Guglielmetti S, Mora D. 2009. Carbamoylphosphate synthetase activity is essential for the optimal growth of *Streptococcus thermophilus* in milk. J Appl Microbiol 107:348–354. https://doi.org/10.1111/j.1365-2672.2009.04213.x
58. Lawson FS, Charlebois RL, Dillon JA. 1996. Phylogenetic analysis of carbamoylphosphate synthetase genes: complex evolutionary history includes an internal duplication within a gene which can root the tree of life. Mol Biol Evol 13:970–977. https://doi.org/10.1093/oxfordjournals.molbev.a025665
59. Charlier D, Glansdorff N. 2004. Biosynthesis of arginine and polyamines. EcoSal Plus 1. https://doi.org/10.1128/ecosalplus.3.6.1.10
60. Nicoloff H, Hubert J-C, Bringel F. 2001. Carbamoyl-phosphate synthetases (CPS) in lactic acid bacteria and other gram-positive bacteria. Lait 81:151–159. https://doi.org/10.1051/lait:2001119
61. Almeida E. 2022. Investigating the role *Carynebacterium matruchoti* has within supragingival plaque spatial organization and its response to oral streptococci and saliva. University of Rhode Island, Kingston, RI.
62. Li Q, Zhou F, Su Z, Li Y, Li J. 2022. *Corynebacterium matruchotii*: a confirmed calcifying bacterium with a potentially important role in the supragingival plaque. Front Microbiol 13:940643. https://doi.org/10.3389/fmicb.2022.940643
63. Morillo-Lopez V, Sjaarda A, Islam I, Borisy GG, Mark Welch JL. 2022. Corncob structures in dental plaque reveal microhabitat taxon specificity. Microbiome 10:145. https://doi.org/10.1186/s40168-022-01323-x
64. Sansone C, Van Houte J, Joshipura K, Kent R, Margolis HC. 1993. The association of mutans streptococci and non-mutans streptococci capable of acidogenesis at a low pH with dental caries on enamel and root surfaces. J Dent Res 72:508–516. https://doi.org/10.1177/00220345930720020701
65. van Ruyven FOJ, Lingström P, van Houte J, Kent R. 2000. "Relationship among mutans streptococci, "low-pH" bacteria, and lodophilic polysaccharide-producing bacteria in dental plaque and early enamel caries in humans". J Dent Res 79:778–784. https://doi.org/10.1177/00220345000790021201
66. Baker JL, Morton JT, Dinis M, Alvarez R, Tran NC, Knight R, Edlund A. 2021. Deep metagenomics examines the oral microbiome during dental caries, revealing novel taxa and co-occurrences with host molecules. Genome Res 31:64–74. https://doi.org/10.1101/gr.265645.120
67. Gross EL, Beall CJ, Kutsch SR, Firestone ND, Leys EJ, Griffen AL. 2012. Beyond *Streptococcus mutans*: dental caries onset linked to multiple species by 16S rRNA community analysis. PLoS ONE 7:e47722. https://doi.org/10.1371/journal.pone.0047722
68. Klingenberg H, Meinicke P. 2017. How to normalize metatranscriptomic count data for differential expression analysis. PeerJ 5:e3859. https://doi.org/10.7717/peerj.3859
69. Zhang Y, Thompson KN, Huttenhower C, Franzosa EA. 2021. Statistical approaches for differential expression analysis in metatranscriptomics. Bioinformatics 37:i34–i41. https://doi.org/10.1093/bioinformatics/btab327
70. Obata J, Fujishima K, Nagata E, Oho T. 2019. Pathogenic mechanisms of cariogenic propionibacterium acidifaciens. Arch Oral Biol 105:46–51. https://doi.org/10.1016/j.archoralbio.2019.06.005
71. Ekstrand KR, Martignon S, Ricketts DJN, Qvist V. 2007. Detection and activity assessment of primary coronal caries lesions: a methodologic study. Oper Dent 32:225–235. https://doi.org/10.2341/06-63
72. Braga MM, Mendes FM, Martignon S, Ricketts DNJ, Ekstrand KR. 2009. *In vitro* comparison of nyvad's system and ICDAS-II with lesion activity assessment for evaluation of severity and activity of occlusal caries lesions in primary teeth. Caries Res 43:405–412. https://doi.org/10.1159/000239755
73. Richards VP, Alvarez AJ, Luce AR, Bedenbaugh M, Mitchell ML, Burne RA, Nascimento MM. 2017. Microbiomes of site-specific dental plaques from children with different caries status. Infect Immun 85:e00106-17. https://doi.org/10.1128/IAI.00106-17
74. Azpiroz MA, Orguilia L, Palacio MI, Malpartida A, Mayol S, Mor G, Gutiérrez G. 2021. Potential biomarkers of infertility associated with microbiome imbalances. Am J Reprod Immunol 86:e13438. https://doi.org/10.1111/aji.13438
75. Tong F, Tang G, Wang X. 2023. Characteristics of human and microbiome RNA profiles in saliva. RNA Biol 20:398–408. https://doi.org/10.1080/15476286.2023.2229596
76. Yost S, Stashenko P, Choi Y, Kukuruzinska M, Genco CA, Salama A, Weinberg EO, Kramer CD, Frias-Lopez J. 2018. Increased virulence of the oral microbiome in oral squamous cell carcinoma revealed by metatranscriptome analyses. 4. Int J Oral Sci 10:32. https://doi.org/10.1038/s41368-018-0037-7
77. Mann AE, O'Connell LM, Osagie E, Akhigbe P, Obuekwe O, Omoigberale A, Kelly C, Coker MO, Richards VP, DOMHaIN Study Team. 2023. Impact of HIV on the oral microbiome of children living in sub-Saharan Africa, determined by using an rpoC gene fragment metataxonomic

approach. Microbiol Spectr 11:e0087123. https://doi.org/10.1128/spectrum.00871-23
78. O'Connell LM, Blouin T, Soule A, Burne RA, Nascimento MM, Richards VP. 2022. Optimization and evaluation of the 30s-S11 rRNA gene for taxonomic profiling of oral streptococci. Appl Environ Microbiol 88:e0045322. https://doi.org/10.1128/aem.00453-22
79. Andrews S. 2019. FastQc a quality control tool for high throughput sequence data (0.11.9)
80. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10. https://doi.org/10.14806/ej.17.1.200
81. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The silva ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–6. https://doi.org/10.1093/nar/gks1219
82. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923
83. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635
84. Liao Y, Smyth GK, Shi W. 2013. The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res 41:e108. https://doi.org/10.1093/nar/gkt214
85. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930. https://doi.org/10.1093/bioinformatics/btt656
86. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-seq data. Nat Biotechnol 29:644–652.
87. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584. https://doi.org/10.7717/peerj.2584
88. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
89. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with Deseq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8
90. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436
91. Stamatakis A. 2014. RaxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033
92. Letunic I, Bork P. 2021. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301
93. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348
94. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from illumina amplicon data. 7. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869
95. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biol 20:257. https://doi.org/10.1186/s13059-019-1891-0
96. R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
97. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8:e61217. https://doi.org/10.1371/journal.pone.0061217
98. Oksanen J, Blanchet FG, Friendly M, Kiindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2019. vegan: community ecology package (R package version 2.5-6)
99. Kay M. 2022. ggdist: visualizations of distributions and uncertainty in the grammar of graphics. Open Science Framework. https://doi.org/10.31219/osf.io/2gsz6
100. Legendre P, Anderson MJ. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecol Monograp 69:1–24. https://doi.org/10.1890/0012--9615(1999)069[0001:DBRATM]2.0.CO;2
101. Silverman JD, Washburne AD, Mukherjee S, David LA. 2017. A phylogenetic transform enhances analysis of compositional microbiota data. Elife 6:e21887. https://doi.org/10.7554/eLife.21887
102. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
103. Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 30:1575–1584. https://doi.org/10.1093/nar/30.7.1575
104. Brohée S, van Helden J. 2006. Evaluation of clustering algorithms for protein-protein interaction networks. BMC Bioinformat 7:488. https://doi.org/10.1186/1471-2105-7-488
105. Townsend JP, Hassler HB, Wang Z, Miura S, Singh J, Kumar S, Ruddle NH, Galvani AP, Dornburg A. 2021. The durability of immunity against reinfection by SARS-CoV-2: a comparative evolutionary study. Lancet Microbe 2:e666–e675. https://doi.org/10.1016/S2666-5247(21)00219-6
106. Zhu A, Ibrahim JG, Love MI. 2019. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics 35:2084–2092. https://doi.org/10.1093/bioinformatics/bty895
107. Rambaut A. 2018. Figtree. http://tree.bio.ed.ac.uk/software/figtree/.
108. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. https://doi.org/10.1093/bioinformatics/btq033
109. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and samtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352