Small RNA Transcriptome of the Oral Microbiome during Periodontitis Progression

Ana E. Duran-Pinedo, Susan Yost, Jorge Frias-Lopez
Forsyth Institute, Cambridge, Massachusetts, USA; Harvard School of Dental Medicine, Boston, Massachusetts, USA

The oral microbiome is one of the most complex microbial communities in the human body, and due to circumstances not completely understood, the healthy microbial community becomes dysbiotic, giving rise to periodontitis, a polymicrobial inflammatory disease. We previously reported the results of community-wide gene expression changes in the oral microbiome during periodontitis progression and identified signatures associated with increasing severity of the disease. Small noncoding RNAs (sRNAs) are key players in posttranscriptional regulation, especially in fast-changing environments such as the oral cavity. Here, we expanded our analysis to the study of the sRNA metatranscriptome during periodontitis progression on the same samples for which mRNA expression changes were analyzed. We observed differential expression of 12,097 sRNAs, identifying a total of 20 Rfam sRNA families as being overrepresented in progression and 23 at baseline. Gene ontology activities regulated by the differentially expressed (DE) sRNAs included amino acid metabolism, ethanolamine catabolism, signal recognition particle-dependent cotranslational protein targeting to membrane, intron splicing, carbohydrate metabolism, control of plasmid copy number, and response to stress. In integrating patterns of expression of protein coding transcripts and sRNAs, we found that functional activities of genes that correlated positively with profiles of expression of DE sRNAs were involved in pathogenesis, proteolysis, ferrous iron transport, and oligopeptide transport. These findings represent the first integrated sequencing analysis of the community-wide sRNA transcriptome of the oral microbiome during periodontitis progression and show that sRNAs are key regulatory elements of the dysbiotic process leading to disease.

Bacterial small noncoding RNAs (sRNAs) encompass a large and diverse group of RNA molecules that do not result in the translation of a protein product. They show high heterogeneity in size and structure and many are used in regulatory roles or other functional capacities upon transcription. sRNAs are important in bacteria because they can be used to adapt rapidly to changing environmental conditions, especially in an environment such as the oral cavity, which is exposed daily to regular changes in amount and quality of nutrients available for use by the oral biofilms. Most of these sRNAs are encoded in the 5’ and 3’ untranslated regions, as well as in intergenic regions (IGRs) of the genome (1).

sRNAs display a wide variety of mechanisms of action. sRNAs repress translation of mRNA by attaching to the ribosome binding site (RBS) competing with the ribosome and leading to the rapid degradation of the mRNA (2). Other noncanonical mechanisms of translation repression have been also described, such as binding cis-acting antisense RNA to a ribosome standby site upstream of the RBS (3). In addition, sRNAs can also activate the translations of mRNAs (4, 5) or can modulate gene expression by varying the level of transcript stability (6). sRNA can also modulate protein activity by mimicking the structures of other nucleic acids and sequestering proteins that otherwise will act on their real target (7).

Notwithstanding the importance of sRNAs in the regulation of bacterial metabolism, identifying them and predicting their targets is still labor-intensive and often combines an initial bioinformatic prediction with a later experimental confirmation. Computational identification of sRNAs has predicted the existence of large numbers of these elements as well as their putative targets, but their role should be ultimately confirmed in vivo. Most approaches for the identification of sRNAs in silico are based on structure similarity or comparative genomics searching for IGR homology in closely related species (8–10). Nonetheless, approaches based on noncomparative algorithms have also been used to identify sRNAs in silico. Several groups have searched for promoters and Rho-independent terminators in IGRs as a way of identifying potential sRNAs (11, 12). Similarly, there has been an interest in developing bioinformatic tools to predict sRNA targets most of them based on RNA-RNA interaction algorithms (9, 13).

Several experimental approaches have been used to identify bacterial sRNAs (14, 15). Methods such as direct labeling of RNA and sequencing or collecting sRNA genes by shotgun cloning of their cDNA are laborious and expensive, especially if a large number of elements are analyzed. Global detection of sRNAs has been facilitated by the development of microarrays and next-generation sequencing (NGS) techniques. NGS techniques have been applied to analyze complete transcriptomes of bacteria under various conditions, which also led to the discovery of numerous novel sRNA transcripts in different bacteria, such as Salmonella spp. (16), Vibrio cholerae (17), or Bacillus subtilis (18).

Since the majority of studies on the identification and charac-

Received 28 May 2015 Accepted 12 July 2015
Accepted manuscript posted online 17 July 2015

Editor: C. A. Elkins
Address correspondence to Jorge Frias-Lopez, jfrias@forsyth.org.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.01782-15.

Copyright © 2015, American Society for Microbiology. All Rights Reserved. doi:10.1128/AEM.01782-15
terization of microbial sRNAs have been performed on model organisms (19, 20), we have a limited knowledge of the role and ecological relevance of sRNAs in complex microbial communities in the environment. In a metatranscriptomic study of the ocean water, Shi et al. analyzed the distribution of sRNAs at different ocean depths and found new groups of previously unknown putative sRNAs. Differences in putative sRNA distributions indicated potential roles of these sRNAs in niche adaptation (21).

In a previous study, we analyzed the profiles of expression of periodontitis and health for encoding protein regions of a oral microbiome (22). Periodontitis is an inflammatory disease mediated by the presence of a polymicrobial biofilm. The components of the oral biofilm must adapt to constant and rapid changes in their environment. Thus, daily ingestion of food represents a periodic and drastic change in the environmental conditions under which these organisms live. Therefore, sRNAs may play an important role in adapting the metabolism of the biofilm to these changing conditions.

In the present study, we analyzed the profiles of expression of sRNAs encoded in the IGRs of the genomes of members of the oral microbiome in health and disease. We identified the sRNAs by mapping transcripts to an IGR database generated from the genomes used in the analysis. To this end, we compared expression patterns of stable and progressing sites from eight individuals with periodontal disease.

**MATERIALS AND METHODS**

**Study design, subject population, and sample collection.** Power calculation to determine the sample size was performed as described elsewhere (23). The subjects in the present study were recruited as part of a multicenter clinical trial to determine biomarkers of periodontal disease progression (ClinicalTrials.gov ID NCT01489839) and are the same as in a paper by Yost et al. (23), in which are described criteria to select them and sampling methods (Table 1).

**NGS.** The sequences used for analysis were generated as described by Yost et al. (23) and are accessible in the Human Oral Microbiome Database at ftp://www.homd.org/publication_data/20141024/RNA/. Detailed protocols for community RNA extraction, RNA amplification, and Illumina Sequencing are described in the same publication.

**Computational identification of sRNAs.** To identify putative sRNAs in silico, we generated an IGR database that was used to map our transcriptomic results. Genomes of *Archaea* and bacteria and their associated information were downloaded from the HOMD database server (http://www.homd.org/), the Pathosystems Resource Integration Center (PATRIC) ftp server (www.patricbrc.org/portal/portal/patric/Home) (24), and the J. Craig Venter Institute (www.jcvi.org). A total of 524 genomes from 312 species of bacteria and two genomes from one archaeal species were used in the analysis (23). Intergenic regions were obtained using a Perl script that uses put and genomic fasta files to identify them. Only IGRs longer than 50 bp with <10% Ns in the sequence were considered to be included in the database. In addition, we included tRNAs obtained from the PATRIC annotation in the database.

**Short-read sequence alignment analysis.** Low-quality sequences were removed from the query files. Fast clipper and fastq quality filter from the Fastx-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) were used to remove short sequences with a quality score of >20 in >80% of the sequence. Cleaned files were then aligned against the bacterial/archaeal database using bowtie2. We generated a *.gff* file to map hits to different regions in the genomes of our database. Read counts from the SAM files were obtained using bedtools multicov from bedtools (25).

**Phylogenetic analysis of the sRNA metatranscriptome.** Counts from the sRNA libraries were used to determine their phylogenetic composition. We created a *.gff* file containing information on whole genomes that was used to assign hits to genomes. Estimated counts were normalized by genome size and frequency and then log, transformed before final analysis. To identify significant differences between communities under the different conditions studied, we performed linear discriminant analysis of effect size (LEfSe), as proposed by Segata et al. (26), with default settings.

**Differential expression analysis.** To assess differential expression in genes within a specific species, we normalized the transcript counts by the relative frequency of the species in the metagenome database. In the case of Gene Ontology (GO) term analysis, we did not normalize by relative abundance since we were treating the whole community as a single organ-

---

**TABLE 1 Sample collection scheme and clinical characteristics of progressing and stable sites**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Site</th>
<th>Visit duration (mo)</th>
<th>Pocket depth (mm)</th>
<th>Clinical attachment loss (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>361</td>
<td>0</td>
<td>3.0</td>
<td>2.0</td>
</tr>
<tr>
<td>1</td>
<td>361</td>
<td>2</td>
<td>5.0</td>
<td>4.0</td>
</tr>
<tr>
<td>2</td>
<td>353</td>
<td>0</td>
<td>2.5</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>353</td>
<td>2</td>
<td>3.0</td>
<td>2.0</td>
</tr>
<tr>
<td>2</td>
<td>473</td>
<td>0</td>
<td>3.0</td>
<td>2.0</td>
</tr>
<tr>
<td>2</td>
<td>473</td>
<td>2</td>
<td>5.0</td>
<td>4.0</td>
</tr>
<tr>
<td>3</td>
<td>241</td>
<td>0</td>
<td>3.0</td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>241</td>
<td>2</td>
<td>3.0</td>
<td>2.0</td>
</tr>
<tr>
<td>3</td>
<td>273</td>
<td>0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>3</td>
<td>273</td>
<td>2</td>
<td>8.5</td>
<td>8.0</td>
</tr>
<tr>
<td>3</td>
<td>441</td>
<td>0</td>
<td>4.0</td>
<td>3.0</td>
</tr>
<tr>
<td>3</td>
<td>441</td>
<td>2</td>
<td>3.0</td>
<td>2.0</td>
</tr>
<tr>
<td>4</td>
<td>151</td>
<td>0</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>4</td>
<td>151</td>
<td>2</td>
<td>5.0</td>
<td>5.5</td>
</tr>
<tr>
<td>5</td>
<td>373</td>
<td>0</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>5</td>
<td>373</td>
<td>2</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>6</td>
<td>143</td>
<td>0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>6</td>
<td>143</td>
<td>4</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>6</td>
<td>151</td>
<td>0</td>
<td>1.5</td>
<td>2.5</td>
</tr>
<tr>
<td>6</td>
<td>151</td>
<td>2</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>7</td>
<td>253</td>
<td>0</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>7</td>
<td>253</td>
<td>4</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>7</td>
<td>141</td>
<td>0</td>
<td>2.0</td>
<td>3.0</td>
</tr>
<tr>
<td>7</td>
<td>141</td>
<td>2</td>
<td>3.0</td>
<td>3.5</td>
</tr>
<tr>
<td>8</td>
<td>253</td>
<td>0</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>8</td>
<td>253</td>
<td>2</td>
<td>5.5</td>
<td>5.5</td>
</tr>
<tr>
<td>9</td>
<td>353</td>
<td>0</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>9</td>
<td>353</td>
<td>2</td>
<td>2.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

*The first two digits indicate the tooth number according to the FDI World Dental Federation two-digit notation; third digit indicates the site position: 1, mesiobuccal; 3, distobuccal.*
ism. To identify differentially expressed (DE) genes from RNA libraries, we applied nonparametric tests to the normalized counts using the NOISeqBio function of the R package NOISeq default conditions \((k = 0.5, kc = 0.05, nm = 0, 0 \text{ and } “tmm” \text{ normalization, using the threshold value for significance suggested by the authors of } q = 0.95, \text{ which is equivalent to a false discovery rate [FDR] of } < 0.05)\) (27, 28).

GO assignments of genes and sRNAs. To evaluate functional activities differentially represented in health or disease, we mapped the differentially expressed genes to known biological ontologies based on the GO project (http://www.geneontology.org/). GO terms to which the different open reading frames (ORFs) belong were obtained from the PATRIC database (http://patricbrc.org/portal/portal/patric/Home). GO terms not present in the PATRIC database and whose annotation was obtained from the HOMD database or from the J. Craig Venter Institute were acquired using the program blast2GO under the default settings (29).

For analysis of GO terms associated with DE sRNAs, we used the mapping file from the GO project (http://geneontology.org/external2go/rfam2go), which maps Rfam families of version 11.0 to GO terms. We used the REVIGO web page (30) to summarize and remove redundant GO terms from the results. Only GO terms with an FDR of <0.05 were used. REVIGO plots were obtained for biological processes categories.

Rfam families of small RNA enrichment analysis. To evaluate differentially represented Rfam families of small RNAs in health or disease, we mapped the DE IGRs to known families on the Rfam database. A total of 986 Rfam families were identified in the database. A total of 896 Rfam families had at least one hit in the sRNA database. Here, only families that represented >0.5% of the total number of hits in the database are shown. The Rfam families are listed in decreasing order of abundance.

Computational target identification of DE sRNAs and GO terms enrichment analysis of the putative targets. Putative target sequences for DE sRNAs were identified using Rsearch (36). The energy cutoff was \(-45 \Delta G (\text{kcal/mol}),\) which is based on the receiver-operating characteristic (ROC) curve presented by the authors to maximize true positives and minimize the rate of true negatives (36). Target identification was performed against the ORFs of the different genomes, using only the sections of IGRs that aligned with the query sequences (see Table S1 in the supplemental material). The aligned sequences were clustered using “usearch” (37). Enrichment analysis on these sets was performed using the R package “GOSeq,” which accounts for biases due to overdetection of long and highly expressed transcripts (38). Gene sets with \(\leq 10\) genes were excluded from analysis.

RESULTS

Abundance distribution of small RNAs during periodontitis progression. We determined the changes in expression of sRNAs during periodontitis progression. In a previous study we analyzed the patterns of expression of mRNA transcripts during progression of this polymicrobial disease and found signatures that define the initial stages of progression in an already periodontally diseased patient (23).

A total of 140,713 transcripts from IGRs were identified. Overall, the small RNA-Seq data set displayed the broad dynamic range characteristic of NGS data sets with read counts spanning 5 orders of magnitude (from 1 to \(>1,000,000\) mapped reads for the least and most highly expressed sRNAs, respectively). Nonetheless, the 3,000 most highly expressed sRNAs represented between 90% up to \(>99%\) of the total number of hits in all samples. Most of IGR hits had an unknown function. In fact, 84% of the IGRs in the database have no significant match in the Rfam database. In the small RNA-Seq data set, we detected several classes of sRNA. Of a total 140,713 hits, 22,982 had significant homology to the Rfam families of sRNAs. A total of 986 Rfam families were represented in the sequence data, although a large fraction (639 families) had fewer than 10 hits. The most abundant family was tRNAs, which represented 36% of all hits assigned to an Rfam
family (Fig. 1; see also Table S2 in the supplemental material). We also found representatives of T-box leader, bacterial RNases P, and several riboswitches (Fig. 1).

Rank abundance distribution of the phylogenetic origin of the sRNAs showed a uniform distribution, with none of the species expressing large numbers of sRNAs in comparison with the rest of the community (see Fig. S1 in the supplemental material). The species with the most hits to the IGR database, Streptococcus mutans, had less than 4% of the total of all hits. The red complex, which appears later in biofilm development, comprises three species that are considered to be the major periodontal pathogens: Porphyromonas gingivalis, Treponema denticola, and Tannerella forsythia (39, 40). Interestingly, two members of the red complex, P. gingivalis and T. forsythia, expressed a large number of transcripts from IGRs (see Fig. S1 in the supplemental material).

In addition, we analyzed the phylogenetic assignment of the sRNA metatranscriptomes. Large number of members of the genera Campylobacter, Anaerococcus, Leptotrichia, Treponema, and Selenomonas expressed significantly larger numbers of sRNAs in progression (Fig. 2), whereas members of the genus Pseudomonas expressed significantly larger numbers of sRNAs at baseline.

Characterization of DE sRNAs and associated Rfam families during periodontitis progression. We identified DE sRNAs by comparing the profiles of expression at baseline and at the endpoints of samples with periodontitis progression, wherein the endpoint was the moment when the tooth broke down (developed disease). We first normalized the transcript numbers based on the frequency of the different species in the biofilm. A total 12,097 RNA reads from IGRs were differentially expressed. The largest fraction of DE RNAs from IGRs were upregulated (n/H11005). We then identified a list of Rfam families that were significantly enriched during progression of the disease. We used the R package GOseq for enrichment analysis of these families. To predict enriched families, an FDR of 0.05 was applied. We found that 20 Rfam sRNA families were overrepresented in progression and 23 were overrepresented at baseline (Table 2). Bacterial RNases P, both archaeal and bacterial, were overrepresented in progressing sites. In addition, several transfer mRNAs (alphaproteobacteria, betaproteobacteria, and cyanobacteria) were also overrepresented.
in the progression sites. At baseline, catalytic introns (self-splicing ribozymes), CRISPR RNA direct repeat elements, and RNA antitoxin were overrepresented families of sRNAs.

Three different riboswitches were overrepresented in the conditions studied. One, overrepresented in progression, was a flavin mononucleotide (FMN) riboswitch, an element found frequently in the 5′-untranslated regions of mRNAs that encode FMN. The other two riboswitches were overrepresented at baseline and were associated with regulation of thiamine pyrophosphate synthesis (TPP riboswitch) and purine biosynthesis or uptake (purine riboswitch).

In addition, we assigned GO terms to the DE sRNAs using the mapping tool "rfam2go" which is based on their corresponding Rfam families, thus identifying global activities that could be controlled by these sRNAs. We performed the analysis using the mapping file generated by GO Consortium, which matches Rfam accession numbers with GO terms (see Materials and Methods). As shown in Fig. 3, there were certain activities that appeared to be associated with both upregulated and downregulated sRNAs. Among these, we found amino acid metabolism, ethanolamine catabolism, signal recognition particle (SRP)-dependent cotranslational protein targeting to membrane, group I and II intron splicing, carbohydrate metabolism, control of the plasmid copy number, and response to stress. Quorum sensing and response to iron ion activities were only associated with upregulated sRNAs (Fig. 3A).

Integration of sRNA expression profiles with mRNA-Seq profiles. Integrating microbiological functions from different “omics” is still one of the challenges in this kind of bioinformatic analysis. Using the package mixOmics, we identified a large number of correlated and anticorrelated expression profiles between DE sRNAs and expressed mRNAs. There were several clusters of profiles of DE sRNAs and expressed mRNAs that followed similar patterns, as shown on the hit map (Fig. 4). Positive correlations (dark red in Fig. 4) would indicate that these sRNAs somehow are related to an increase of these transcripts by an increase in transcription or stabilization of the mRNAs, whereas negative correlations (dark blue in Fig. 4) would indicate the opposite, either a repression of expression of these genes or an increase in degrading these transcripts. The patterns of these associations were extremely complex (see Fig. S2 in the supplemental material). Thus, to simplify the visualization of the results, we identified genes that correlated and anticorrelated with profiles of expression of DE sRNAs. We then assigned GO numbers to the genes that correlated, positively or negatively, with profiles of expression of sRNAs based on their abundance. Using REVIGO, we summarized the functional activities that followed similar patterns of expression as the DE sRNAs in the database (Fig. 5). Functional activities of genes that correlated positively with profiles of expression of DE sRNAs comprise pathogenesis, proteolysis, ferrous iron transport, cobalamin biosynthesis, chemotaxis, chloride transport, oligopeptide transport, and potassium ion transport (Fig. 5A). There was a redundancy in the target activities of these DE sRNAs. Negative correlation was also observed for genes involved in proteolysis, ferrous iron transport, cobalamin biosynthesis, chemotaxis, chloride transport, and potassium ion transport but not for activities directly involved in oligopeptide transport and pathogenesis (Fig. 5B).

Community-wide analysis of targets of the DE sRNAs. To better characterize the role that the DE sRNAs could have in regulating metabolic activities during periodontitis progression, we generated a set of target predictions using RIsearch with a cutoff value of −45 ∆G (kcal/mol), which maximizes true positives according to the ROC curve presented by the authors of this software (36). A total of 148,987 genes were identified as putative targets for the 12,097 DE sRNAs in the database. Most of the targets were identified as hypothetical proteins, followed by transporters and different transcriptional regulators, including the TetR, LysR, AraC, IcIR, MarR, and GntR families (see Table S3 in the supplemental material). In addition, we observed a large number of targets identified as mobile element proteins.

However, the main motivation of the analysis was to identify global functional activities that could be regulated by sRNAs by looking at genes that could be targeted by these DE sRNAs. Thus, GO enrichment analysis corresponding to target mRNA genes showed enriched annotations of activities that have been previously associated with pathogenesis in periodontitis. As in the previous section, we summarized these functional activities using REVIGO and found that genes involved in pathogenesis, oligopeptide transport, proteolysis, cobalamin biosynthesis, cell adhesion, and response to antibiotic were all enriched target functions of the DE sRNAs during periodontitis progression (Fig. 6). Another interesting activity that was over-represented overrepresented metabolism. GO terms associated with cellular copper ion homeostasis, copper ion transport, and response to copper ion represented a significant fraction of targets’ functions (Fig. 6A). Accordingly, we found a large fraction target genes identified as

<table>
<thead>
<tr>
<th>TABLE 2 Annotations enriched in sRNAs during periodontitis progression*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overrepresentation</strong></td>
</tr>
<tr>
<td>5.8S rRNA</td>
</tr>
<tr>
<td>5S rRNA</td>
</tr>
<tr>
<td>6S/SrS RNA</td>
</tr>
<tr>
<td>Archaeal RNase P</td>
</tr>
<tr>
<td>Archaeal signal recognition particle RNA</td>
</tr>
<tr>
<td>Bacterial large signal recognition particle RNA</td>
</tr>
<tr>
<td>Bacterial RNase P class A</td>
</tr>
<tr>
<td>Bacterial RNase P class B</td>
</tr>
<tr>
<td>Bacterial small signal recognition particle RNA</td>
</tr>
<tr>
<td>Bacteroidales-1 RNA</td>
</tr>
<tr>
<td>Betaproteobacteria transfer mRNA</td>
</tr>
<tr>
<td>Cyanobacteria transfer mRNA</td>
</tr>
<tr>
<td>FMN riboswitch (RFN element)</td>
</tr>
<tr>
<td>L17 ribosomal protein downstream element</td>
</tr>
<tr>
<td>MALAT1-associated small</td>
</tr>
<tr>
<td>cytoplasmic RNA/MEN beta RNA</td>
</tr>
<tr>
<td>Proteobacterial sRNA sX4</td>
</tr>
<tr>
<td>Pseudoknot of the domain G (G12)</td>
</tr>
<tr>
<td>of 235 rRNA</td>
</tr>
<tr>
<td>Transfer mRNA</td>
</tr>
<tr>
<td>tRNA</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Enrichment analysis was performed on differentially expressed sRNAs using GOseq with an FDR of <0.05.
FIG 3 Summary of GO terms associated with Rfam families identified in the DE sRNAs during periodontitis progression. (A) GO terms of Rfam families associated with upregulated sRNAs. (B) GO terms of Rfam families associated with downregulated sRNAs. The value bar indicates the number of GO terms assigned to the different Rfam families in each of the displayed functional activities.
copper-translocating P-type ATPases and multicopper oxidases (see Table S3 in the supplemental material).

Role of sRNAs in regulation of the metabolism of major periodontal pathogens during periodontitis progression. Porphyromonas gingivalis, Tannerella forsythia, and Treponema denticola (the “red complex”) are considered major periodontopathogens and are highly associated with severe chronic periodontitis (39). Moreover, P. gingivalis can modulate, even at small concentrations, the behavior of the whole community turning it dysbiotic (41, 42). To identify metabolic activities controlled by sRNAs during progression of the disease, we searched for potential targets of the DE sRNAs of these members of the community.

As mentioned above, a large number of sRNAs were expressed by the members of the red complex (see Fig. S1 in the supplemental material). Of these 2,171 were DE sRNAs that were used to look for potential targets using Rfam. We identified a total of 1,377 putative mRNA targets for these DE sRNAs. The pattern observed was very similar to the observations on the whole microbial community. Most putative targets corresponded to hypothetical proteins followed by mobile element proteins. Targets related to iron metabolism such as TnB-dependent receptor proteins were frequently identified in the red complex but not as frequently in the community as a whole (see Table S4 in the supplemental material). Another differential feature of the red complex target profile is the high frequency of transposases and conjugative transposon proteins identified as potential targets for the DE sRNAs.

The list of potential targets was later used to identify GO terms associated with these putative targets. In accordance with the observations exposed above, ferrous ion transport was highly represented among the red complex’s sRNA targets (see Fig. S3 in the supplemental material). In addition, the response to antibiotics, including beta-lactam antibiotic catabolism, was also highly represented in the target data set.

**DISCUSSION**

A key element to survival of microorganisms in any environment is their capacity to rapidly adapt to new conditions. This is especially true in the oral cavity, where conditions change daily in short periods of time due to the ingestion of nutrients and to the fact that the oral cavity is an open entrance in contact with the external environment, hence the importance of regulatory elements such as sRNAs, which are rapidly synthesized and capable of modulating the metabolism faster than regulation involving protein synthesis (43, 44). It is now well established that sRNAs are capable of performing a wide variety of physiological roles, including adapting to new environmental conditions through quorum-sensing systems or the development of biofilms (45). Recent reports have shown that sRNAs also play important roles in microbial virulence and infection (46, 47). However, to this date, the biological functions of most bacterial sRNAs are largely unknown. We conducted a metatranscriptomic analysis of the sRNA fraction to understand the role that they play in the regulation of metabolic activities leading to dysbiosis in the oral microbiome and to assess whether they are involved in the regulation of activities that we had previously associated with progression of the disease.

Our understanding of the sRNA metatranscriptome in complex microbial communities is even more limited than our understanding of the role of sRNA for specific bacteria. Community-wide studies have been mainly circumscribed to environmental settings (21, 48). Thus, the first community-wide metatranscriptomic analysis of sRNAs by Shi et al. (21) was performed in the ocean’s water column, where these researchers found a large number of new groups of putative sRNAs previously unknown. Recently, Gosalbes et al. reported the presence of sRNAs in the metatranscriptome of human fecal microbiota, although these researchers did not identify the possible functions of the sRNAs (49).

Through integrative analysis of parallel sequencing of both small RNAs and gene-coding transcripts from the same samples, we were able to infer the importance that these elements play in regulation of gene expression in progression of periodontitis, a polymicrobial disease where a dysbiotic community is the causative agent of the disease (42, 50).

In the database we identified a large number of Rfam families of sRNAs. Large diversity in the sRNA population has been observed in other complex microbial communities (21, 48), as well as in isolated organisms (51, 52). In accordance with other sRNA metatranscriptomic analysis (48), tRNAs were the most frequently observed sRNAs in the sequence libraries. Several riboswitches were not only abundant in the sRNA database but also differentially expressed during progression (e.g., FMN and TPP riboswitches). Riboswitches control the production of proteins in response to the concentration of a specific effector that binds to the mRNA. The FMN-riboswitch has been implicated in the stimulation of virulence in Listeria monocytogenes (53), while the TPP-riboswitch belongs to the most widespread class of these regulatory elements (54). Also abundant in the libraries, and differentially represented, were T-box leaders and CRISPR RNA direct repeat elements. CRISPR-Cas systems have been implicated in regulating virulence and stress responses in different pathogenic bacteria (55). Similarly, toxin-antitoxin systems in Enterococcus faecalis have been linked to the regulation of virulence and stress response (56).

Direct assignment of GO terms to DE sRNAs showed common themes whose activities are probably tightly regulated by the coordinate action of these sRNAs. Among these activities, we identified ethanolamine catabolism, amino acid metabolism, the
FIG 5 Summary of GO terms associated with genes whose gene expression profiles correlated with DE sRNAs in periodontitis progression. GO terms from genes whose patterns of expression correlated or anticorrelated with patterns of expression of DE sRNAs were used to construct networks using REVIGO. (A) GO terms of genes whose profiles of expression correlated positively with profiles of expression of sRNAs were used to construct networks using REVIGO. (B) GO terms of genes whose profiles of expression correlated negatively with profiles of expression of sRNAs were used to construct networks using REVIGO. The intensity of the colors represents the significance of the GO representation.
regulation of carbohydrate metabolism, SRP-dependent cotranslational protein targeting to the membrane, gene silencing, and regulation of plasmid copy number as being represented in both sets (up- and downregulated) of DE sRNAs. A wide variety of studies have revealed that sRNAs are implicated in carbon metabolism, transport, and amino acid metabolism (45), protein transport by SRP-dependent cotranslational protein targeting (57), and control of plasmid copy number (58, 59). Some of these processes have been linked to bacterial pathogenesis. For example, ethanolamine, an abundant phospholipid in membranes of both bacteria and human cells, can be used by a wide variety of bacteria (60, 61). In the intestine, it has been suggested that use of ethanolamine could contribute to the pathogenesis of bacterial species either by providing a useful carbon and/or nitrogen source that promotes successful colonization or by disrupting the innate immune system of the intestine (60). Interestingly, we observed in our previous metatranscriptomic analysis of mRNA from the same samples used here that the synthesis of cobalamin (vitamin B<sub>12</sub>) is associated with disease progression (23). The ethanolamine deamination systems of <i>Escherichia coli</i> and <i>Klebsiella aerogenes</i> are cobalamin dependent (62). Moreover, adenosyl-cobalamine is a cofactor required in the first enzymatic step of ethanolamine catabolism (62), and in <i>Enterococcus faecalis</i> the ethanolamine utilization (<i>eut</i>) locus, containing at least 19 genes distributed over four polycistronic messenger RNAs, appears to be regulated by a single adenosyl-cobalamine-responsive riboswitch that controls gene expression by sequestration of a response regulator (63, 64).

We identified several clusters of sRNAs whose expression was highly correlated with expression profiles of mRNAs. The functions of the correlated genes were strikingly similar to signature functions important during periodontitis progression identified in a previous study on the mRNA metatranscriptome (23), reinforcing the idea that the identified DE sRNAs play a key role in regulating these functions. Among these, we found proteolysis, ferrous iron transport, protein secretion and import, potassium ion transport, and cobalamin biosynthesis (23).

Finally, we identified putative targets for the DE sRNAs and characterized their putative global biological functions. When dealing with a large number of genomes, as in the present case, in <i>silico</i> prediction of sRNA targets is computationally intensive; thus, we decided to use Rlsearch under highly stringent energy interaction conditions. This software identifies putative duplexes of RNAs based on an implementation of a simplified Turner energy model, significantly reducing run-time, while at the same time maintaining accuracy (36). Among the enriched GO terms for these targets was pathogenesis, which confirms the hypothesis that sRNAs play a key role in the regulation of bacterial activity during periodontitis progression. The regulatory role of sRNA on virulence has been previously reported for individual species of bacteria (65). In addition, we also observed an enrichment in tar-

FIG 6 Summary of GO terms associated with target genes to DE sRNAs in periodontitis progression identified using Rlsearch. GO terms from target genes whose patterns of expression correlated or anticorrelated with patterns of expression of DE sRNAs were used to construct networks using REVIGO. (A) GO terms of target genes of DE sRNAs overrepresented in progression. (B) GO terms of target genes of DE sRNAs overrepresented at baseline. The circle size is proportional to the frequency of the GO term, while the color indicates the log<sub>10</sub> P value (red higher, blue lower).
gets associated with oligopeptide transport, proteolysis, cobalamin biosynthesis, beta-lactam catabolism, and cell adhesion, whose activities had been previously associated with progression and severe periodontitis (22, 23).

We also observed a large representation of different families of transcription factors in the putative target data set. It is now known that it is quite common for sRNAs to target transcriptional regulators. A classic example is spoS, which is regulated directly by several Hfq-binding sRNAs (e.g., DsrA, KprA, and ArcZ) and has been shown to regulate virulence in *Vibrio cholerae* (66). Other examples include *E. coli* CsgD and Lrp, as well as LuxR and AphA in *Vibrio* spp. (44).

A large set of targets was related to copper metabolism. Among the most abundant targets, we found copper-translocating P-type ATPases and multicopper oxidases. Interestingly, both type of proteins have been associated with a mechanism of defense of bacteria against the host immune response. Copper-translocating P-type ATPases appear to be critical for bacterial virulence by overcoming high phagosomal metal levels and are required for the assembly of periplasmic and secreted metalloproteins that are essential for survival in extreme oxidant environments (67, 68). As for multicopper oxidases, in *Mycobacterium tuberculosis* are required for virulence by reducing the internal copper amount (69). They also play a role in the virulence and survival of *Salmonella enterica* (70). As a whole, these results seem to indicate that sRNA play a key role in regulating mechanisms of defense against the host immune system.

Regarding the three major pathogens that form the red complex, we observed that transposases and genes from conjugative transposons were among the most targeted mRNA in the database. These results are in agreement with what we observed in the mRNA transcript database, where we found high levels of expression of both transposases and conjugative transposons genes in the red complex.

High levels of expression of transposases have been observed in oral microorganisms such as *Fusobacterium nucleatum* and *Treponema denticola* in parallel with the expression of other virulence factors (71, 72). sRNAs have been identified as regulatory elements in conjugative transposons of *Bacteroides* spp. (73). Furthermore, Philips et al. found that some antisense regulatory sRNAs of *P. gingivalis* are located within putative conjugative transposons (74).

In summary, these results seem to indicate that sRNAs may be involved in regulating the transition of the oral microbiome from a commensal state to a symbiotic one, controlling essential activities of the community involved both in virulence and in defense against the host immune response. However, these are in silico predictions, and the individual role of the identified sRNAs must be confirmed experimentally in the laboratory.

**ACKNOWLEDGMENTS**

Research reported in this publication was supported by the National Institute of Dental and Craniofacial Research of the National Institutes of Health under awards DE021553 and DE021127.

**REFERENCES**


Duran-Pinedo et al.


