Comprehensive Transcriptome Analysis of the Periodontopathogenic Bacterium *Porphyromonas gingivalis* W83

Hedda Høvik,[a] Wen-Han Yu,[b]* Ingar Olsen,[a] and Tsute Chen[b]

Department of Oral Biology, Faculty of Dentistry, University of Oslo, Oslo, Norway,[a] and Department of Molecular Genetics, The Forsyth Institute, Cambridge, Massachusetts, USA[b]

High-density tiling microarray and RNA sequencing technologies were used to analyze the transcriptome of the periodontopathogenic bacterium *Porphyromonas gingivalis*. The compiled *P. gingivalis* transcriptome profiles were based on total RNA samples isolated from three different laboratory culturing conditions, and the strand-specific transcription profiles generated covered the entire genome, including both protein coding and noncoding regions. The transcription profiles revealed various operon structures, 5′- and 3′-end untranslated regions (UTRs), differential expression patterns, and many novel, not-yet-annotated transcripts within intergenic and antisense regions. Further transcriptome analysis identified the majority of the genes as being expressed within operons and most 5′ and 3′ ends to be protruding UTRs, of which several 3′ UTRs were extended to overlap genes carried on the opposite/antisense strand. Extensive antisense RNAs were detected opposite most insertion sequence (IS) elements. Pairwise comparative analyses were also performed among transcriptome profiles of the three culture conditions, and differentially expressed genes and metabolic pathways were identified. With the growing realization that noncoding RNAs play important biological functions, the discovery of novel RNAs and the comprehensive transcriptome profiles compiled in this study may provide a foundation to further understand the gene regulation and virulence mechanisms in *P. gingivalis*. The transcriptome profiles can be viewed at and downloaded from the Microbial Transcriptome Database website, http://bioinformatics.forsyth.org/mtd.

Two important high-throughput technologies, high-density microarray and massive parallel sequencing, have recently emerged to revolutionize biological research. Based on these technologies, genomic tiling microarrays and RNA sequencing (RNAseq) have been applied for studying whole-genome or chromosome transcription for various eukaryotic and prokaryotic organisms (10, 36, 59, 65, 75, 84, 87). Comprehensive transcription studies have shown that the transcriptional landscape of all organisms is more complicated than expected, and the use of high-throughput technologies for transcript detection has caused an explosive discovery of novel RNAs. The majority of the identified novel RNAs do not encode proteins and are termed noncoding RNAs (ncRNAs), of which many are categorized as small RNAs (sRNAs). Several sRNAs have been found to regulate various biological functions in bacteria, such as expression of outer membrane proteins (24, 79), iron homeostasis (51, 81), quorum sensing (42, 77), and virulence factors (63, 76).

Periodontal diseases are a group of bacterial inflammatory diseases affecting the supporting tissues of the teeth. They include the whole range of severity from mild inflammation of the gingiva (gingivitis) to destruction of periodontal tissue and progressive alveolar bone resorption (periodontitis). Furthermore, a possible relationship between periodontal disease and increased risk of systemic conditions, such as cardiovascular diseases (64), preterm delivery (1), diabetes (55), and rheumatoid arthritis (11), has been proposed, suggesting that periodontal disease may have a considerable impact on general health.

Chronic periodontitis is caused by a complex of different bacterial species (69). Among the periodontopathogenic bacteria, *Porphyromonas gingivalis* has been studied most extensively and has been identified as one of the primary pathogens in adult periodontitis (9, 80, 86). This Gram-negative anaerobic bacterium, *P. gingivalis*, has a short rod cell shape and prefers peptides as the metabolic source for energy (40). *P. gingivalis* produces a panel of potential virulence factors involved in colonization, tissue destruction, bone resorption, and host defense perturbation (19). These virulence factors include lipopolysaccharide, polysaccharide capsule, fimbriae, hemagglutinins, and several different proteinases (16, 19, 40). In addition, several studies have detected *P. gingivalis* in aortic tissue or arterial plaque samples (17, 38, 70), suggesting a relationship between this oral pathogen and cardiovascular disease.

In 2003, the genome of *P. gingivalis* strain W83 was sequenced (56). It comprises a 2.34-Mbp circular chromosome, and according to the genomic annotation maintained by the Comprehensive Microbial Resource of the J. Craig Venter Institute (CMR-JCVI; http://cmr.jcvi.org), the genome holds 2,053 annotated genes (85.46% of all bases), of which 1,988 are protein coding genes and 65 are structural genes. Several gene expression studies in *P. gingivalis* have been carried out using a spotted microarray that was provided by The Institute for Genomic Research (now the JCVI). This *P. gingivalis*-specific microarray contains 1,907 unique 70-mer oligonucleotide probes representing 1,990 annotated open reading frames (ORFs) (90). Gene expression studies using this

microarray have reported many valuable findings (28, 44, 49, 62, 90) but are limited with regard to coverage, as the probes represent only a small portion of the genome. Transcription from noncoding sequences, including intergenic and antisense regions, has thus far not been studied comprehensively in *P. gingivalis*.

The aim of this study was to survey the genome-wide transcription activities of *P. gingivalis* at high resolution using both genomic tiling microarray and RNAseq technologies. The transcriptome profiles generated allowed genome-wide identification of transcription patterns and large-scale discovery of novel RNAs from the intergenic and antisense regions. This paper presents the transcriptome profiles of *P. gingivalis* and the results from the analysis of these profiles. The profile data are made available through the Microbial Transcriptome Database, and together with the transcriptome analysis presented in this paper, they serve as a valuable resource for further studies of the transcriptional mechanisms in this important periodontal pathogen.

MATERIALS AND METHODS

Bacterial strain and growth conditions. *P. gingivalis* strain W83 was routinely maintained at 37°C in an anaerobic chamber (80% N_2, 10% CO_2, 10% H_2) (14) on blood agar plates (BAPHK), i.e., Trypticase soy agar supplemented with defibrinated sheep blood (5%, vol/vol), hemin (1 μg/ml), and menadione (0.5 μg/ml). For the transcriptome study, cells were cultured under the same conditions in three different media: (i) BAPHK, as described above; (ii) TSB, a Trypticase soy broth supplemented with hemin (1 μg/ml) and menadione (0.5 μg/ml); and (iii) MIN, a chemically defined minimal liquid medium providing α-ketoglutarate and bovine serum albumin (BSA) as the only protein sources (54). For BAPHK cultures, colonies grown on the agar plates for 48 h were used. Liquid cultures were allowed to reach mid-log phase (A_{550}, 0.40 to 0.55) prior to harvest for RNA extraction.

Tiling microarray design. The microarray used in this study was manufactured by Roche NimbleGen, Inc. (Madison, WI), using the Maskless Array Synthesizer (MAS) technology (2, 57), which enables high flexibility for producing a wide range of custom-defined microarrays. A genomic tiling microarray probe set consisting of 385,000 unique 50-mer oligonucleotide sequences covering both strands of the *P. gingivalis* W83 genome was designed by Høvik and Chen (29) using a dynamic computer algorithm. The entire *P. gingivalis* genome was covered approximately 4 times by this probe set with an average of one probe per 12 nucleotides (nt) on both strands.

RNA extraction, labeling, and microarray hybridization. Cells grown on BAPHK were harvested by immersing the colonies on the agar plates in a solution containing a 2:1 (vol/vol) ratio of RNAprotect bacterial reagent (Qiagen, Valencia, CA) and 1× phosphate-buffered saline (PBS) and incubated in the anaerobic chamber for 5 min. The cells were then dispersed into the solution, transferred to microcentrifuge tubes, and pelleted by centrifugation at 5,000 × *g* for 7 min at 4°C. For broth cultures, mid-log-phase cells were pelleted by centrifugation at 7,000 × *g* for 7 min at 4°C. The MIN-cultured cells were immediately subjected to cell lysis while the TSB-cultured cells were dissolved in a 2:1 (vol/vol) ratio of RNAprotect bacterial reagent and 1× PBS and pelleted by centrifugation to be stored at −20°C until further processing.

RNA extraction followed by direct labeling with the Label IT Cy3 reagent (Mirus Bio, Madison, WI) and microarray hybridization were performed as described previously (89). In brief, proteinase K-based lysis of bacterial cells was performed using the MasterPure RNA purification kit (Epicentre, Madison, WI), and to remove genomic DNA, the RNA extract was treated twice with Turbo DNase (Applied Biosystems/Ambion, Austin, TX). The mirVana miRNA isolation kit (Applied Biosystems/Ambion) (BAPHK and TSB samples) or the MasterPure RNA purification kit (MIN samples) was applied to purify the RNA samples both before and after the DNase treatment. The purified total RNA was directly labeled with the Label IT Cy3 reagent followed by RNA fragmentation, mirVana purification, and microarray hybridization. Washed and dried microarray slides were immediately scanned in a GenePix 4000B scanner (Axon Instruments, Union City, CA).

Genomic DNA extraction, labeling, and hybridization. *P. gingivalis* W83 genomic DNA was extracted with the MasterPure DNA purification kit (Epicentre), and RNA was removed with the RNase A supplied in the kit. The genomic DNA was then partially fragmented with DNase I (Applied Biosystems/Ambion) and labeled with the Label IT Cy3 reagent. Conditions for labeling, microarray hybridization, and scanning were similar to those for the RNA samples. Two DNA microarray hybridization replicates were used for RNA signal adjustment as described below.

Microarray signal detection, data normalization, and analysis. NimbleScan v2.5 software was used for spot feature extraction from the scanned images. Probe intensities with repeated sequences were regressed to single-copy level prior to normalization of the RNA signal intensities using the genomic DNA signals as reference, with the Bioconductor R package tilingArray (31). Specifically, nonspecific background was estimated based on the intensity of the probes representing the intergenic regions and the genomic DNA reference signals provided experimental corrections for sequence-specific factors (88). Between-array normalization was done using the vsn algorithm (32) also provided in the tilingArray package (31). The normalization included biological replicates of independently grown cultures (two BAPHK, three TSB, and three MIN samples), and the $\log_2$ means of the normalized signal intensities from each condition were used for downstream processes.

A hybrid supervised machine learning algorithm, hidden Markov support vector machines (HM-SVM) (88), was applied to identify the boundaries of transcriptionally active regions (TARs). The HM-SVM algorithm was applied to determine the "expressed" and "nonexpressed" regions in the genome based on a set of training data derived from both ORF and intergenic regions (88).

The expression level of an annotated gene (CMR-JCVI) was determined by averaging the nucleotide intensities based on probe signals within the range of the gene. In this report, an annotated gene was considered expressed if one of these two criteria was met: (i) $\log_2$ mean signal intensity of >4.7 and coverage of >90% of gene length and (ii) $\log_2$ mean signal intensity of >5.0 and coverage of >75% of gene length. The coverage was determined by the percentage of the nucleotides within a gene that was determined to be expressed by the HM-SVM algorithm. The intensity criterion was included to increase the confidence considering a gene expressed. These criteria were also applied when identifying ORF-containing TARs.

Differential expression at ORF level was measured as the difference between the $\log_2$ mean probe signal intensities of the same ORF from two conditions. Genes that showed >2-fold mean signal intensity difference with a *P* value of <0.05 and a *Q* value of <0.1 and, in addition, displayed a >1.5-fold change in the RNAseq-based mean read count were determined as differentially expressed. The *P* and *Q* values were calculated using the SAM software (78), with default settings performing 10 or 20 permutations for including two or three sets of repeats, respectively.

Strand-specific cDNA library construction for RNAseq. An aliquot of the same RNA samples used for the microarray experiments described above was used for sequencing—one from each of the three conditions BAPHK, TSB, and MIN. The RNA was converted to strand-specific cDNA libraries based on procedures modified from those described by Lister et al. (48) and reference 32a. Specifically, the RNA was first subjected to partial removal of 16S and 23S rRNAs using the MICROBExpress bacterial mRNA enrichment kit (Applied Biosystems/Ambion). The enriched RNA was fragmented with RNA fragmentation reagents (Applied Biosystems/Ambion) in 10 μl of 1× fragmentation buffer for 25 min at 95°C. The fragmentation reaction was terminated with 1 μl stop solution and cooled on ice, followed by ethanol precipitation. Antarctic phosphatase (New England BioLabs, Cambridge, MA) was used to remove the 5′-phosphate group from the RNA by adding 5 U of the phosphatase to a

20-μl reaction mixture incubated at 37°C for 30 min. Fragmented RNA was subjected to size selection in a 15% Novex Tris-buffered EDTA (TBE)–urea polyacrylamide gel (Invitrogen, Carlsbad, CA). The gel slice corresponding to the RNA size between ~35 and 45 nucleotides was excised, and the RNA was eluted and precipitated with ethanol. The RNA was then ligated to the 3′-end oligonucleotide adapter (5′-UCGUAUGC CGUCUUCUGCUUGUidT-3′). The ligation was performed using T4 RNA ligase in 10% dimethyl sulfoxide (DMSO), incubated at 20°C for 6 h. The ligated RNA was purified again in a 15% Novex TBE-urea polyacrylamide gel as described above and then phosphorylated in a 50-μl reaction mixture containing 40 U T4 polynucleotide kinase (New England Bio-Labs) and 1 mM ATP (Epicentre) for 1 h at 37°C. After phenol-chloroform purification using Phase-Lock Gel Heavy tubes (Eppendorf, Hauppauge, NY) and ethanol precipitation, the 5′-end oligonucleotide adapter (5′-GUUCAGAGUUCUACAGUCCGACGAUC-3′) was ligated to the phosphorylated RNA under the same conditions used for the 3′-end adapter ligation. The double-ligated RNA was gel purified and converted to strand-specific cDNA by reverse transcription using a primer specific to the 3′-end adapter (5′-CAAGCAGAAGACGGCATACGA-3′). The cDNA was amplified with 15 cycles of PCR using the primer pair 5′-AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCC GA-3′ and 5′-CAAGCAGAAGACGGCATACGA-3′. The amplified product was size fractionated by electrophoresis in a 6% Novex TBE polyacrylamide gel, and the gel slice corresponding to the DNA size ~95 to 120 bp was excised and eluted in 1× gel elution buffer (Illumina). All oligonucleotides (adapters and primers), the T4 RNA ligase, the Phusion polymerase used for the PCR, and several other reagents were provided in Illumina's small RNA sample preparation kit. Finally, the resulting product was ethanol precipitated and dissolved in 10 μl resuspension buffer (Illumina) before being subjected to sequencing. The sequencing was carried out on the Illumina Genome Analyzer II platform at the Norwegian High-Throughput Sequencing Centre (Oslo, Norway). Each of the three cDNA libraries was loaded to separate lanes on Illumina flow cells and subjected to 38 (MIN) or 50 (BAPHK and TSBHK) single-end cycles of sequencing.

Sequence read alignment and normalization of read-based profiles. The Illumina sequence reads were mapped to the reference genome, *P. gingivalis* W83, using the software PerM (8) and allowing a maximum of two mismatches. Reads that did not align with the genome were trimmed from the 3′ end, one base at a time, and the mapping-trimming process was repeated until the read was aligned or the trimmed length reached 24 nucleotides (nt) in length. Reads that were not aligned after being trimmed to 24 nt were discarded. The aligned results were separated into forward and reverse-complement groups and converted to the "pileup" format by SAMtools (47). Log$_2$ read counts at each nucleotide position of the genome were plotted to generate the transcriptome profiles displayed in the figures of this paper as well as in the web-based transcriptome profile viewer (described below). For normalization, the log$_2$ read value of each nucleotide was adjusted based on the mean of values from the upper quartile range (6), excluding the sequences with repeats (e.g., rRNA and insertion sequence [IS] elements) or zero read coverage.

Transcriptome profile visualization and data availability. The RNAseq transcriptome profiles of *P. gingivalis* W83 grown under three different culture conditions are visualized both individually and together side-by-side in the web-based transcriptome profile viewer that we developed, available at the Microbial Transcriptome Database (MTD) website, http://bioinformatics.forsyth.org/mtd. In this viewer, the microarray-based transcriptome profiles are plotted alongside the RNAseq-based profiles. The microarray profiles were also available for viewing separately in the Genome Viewer of the Human Oral Microbiome Database (HOMD; http://www.homd.org). The original RNAseq-generated log$_2$-transformed and normalized data as well as the microarray-generated data can be downloaded from the MTD website.

Microarray and RNAseq data accession numbers. The microarray data were deposited in the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) database with the group accession

TABLE 1 Summary of microarray and RNAseq data based on three different culture media

Result	BAPHK	TSB	MIN
Microarray results			
Transcribed segments	764	715	772
Forward	384	352	373
Reverse complement	380	363	399
Percent transcribed genome[a]	84.00	86.19	84.51
Forward	42.78	44.99	43.54
Reverse complement	44.78	44.79	45.05
Both strands	3.56	3.59	4.08
ORF coverage[b]	85.79	88.19	85.89
Intergenic coverage[c]	11.79	11.92	12.61
RNAseq results			
Sequencing cycles	50	50	38
Total no. of reads	14,903,646	15,386,907	21,921,798
Read length range	24–50	24–50	24–38
Total hit count	41,175,929	43,207,841	68,364,977
No. of reads with hits	14,000,570	14,601,921	21,483,802
% reads aligned	93.94	94.90	98.00
No. of unique sequences	2,358,377	3,235,353	2,314,794
% of unique sequences	15.82	21.02	10.56
% genome coverage[a]	69.01	83.42	80.33
Forward	35.68	44.65	42.73
Reverse complement	35.33	44.39	42.23
Both	1.99	5.63	4.63
ORF coverage[b]	71.39	85.08	82.44
Intergenic coverage[c]	8.17	13.64	12.05

[a] Percentage of the genome sequence that is transcribed from either strand, forward, reverse complement, or both strands, respectively. Values are from RNAseq data, mapped with at least three reads.
[b] Percentage of sequences of CMR-JCVI-annotated ORFs that are expressed. Values are from RNAseq data, mapped with at least three reads.
[c] Percentage of intergenic or antisense sequence that is transcribed. Values are RNAseq data, mapped with at least three reads.

identification (ID) GPL11291. The RNAseq data, including both the original read sequences in "fastq" format and the read count in the pileup format, were deposited in the NCBI GEO database with the accession ID GSE30452.

RESULTS

Microarray-based transcriptome profiles. *P. gingivalis* W83 was grown in three different laboratory culturing media: a chemically defined minimal medium (MIN) (54), Trypticase soy broth (TSB), and sheep blood agar (BAPHK). Total RNA was extracted from each type of the cell cultures and subjected to two different technologies, genomic tiling microarray and RNA sequencing (RNAseq), for compiling strand-specific transcriptome profiles. The RNA signal intensities detected from the tiling microarrays were normalized and subjected to segmentation analysis (HM-SVM algorithm) for identifying transcriptionally active regions (TARs) (88). The identified TARs represent potential RNA units transcribed from distinct transcription start sites, and totals of 764, 715, and 772 TARs were recorded for the BAPHK, TSB, and MIN samples, respectively (Table 1). On average, the TARs cover 84.9% of the genomic sequence and are quite equally distributed on the two coding strands.

RNAseq transcriptome profiles. One sample from each of the three culture conditions was subjected to RNAseq to determine the high-resolution transcriptome profile. The RNA samples were fragmented and ligated with adapters prior to cDNA synthesis and PCR amplification. The resulting cDNA libraries were applied to
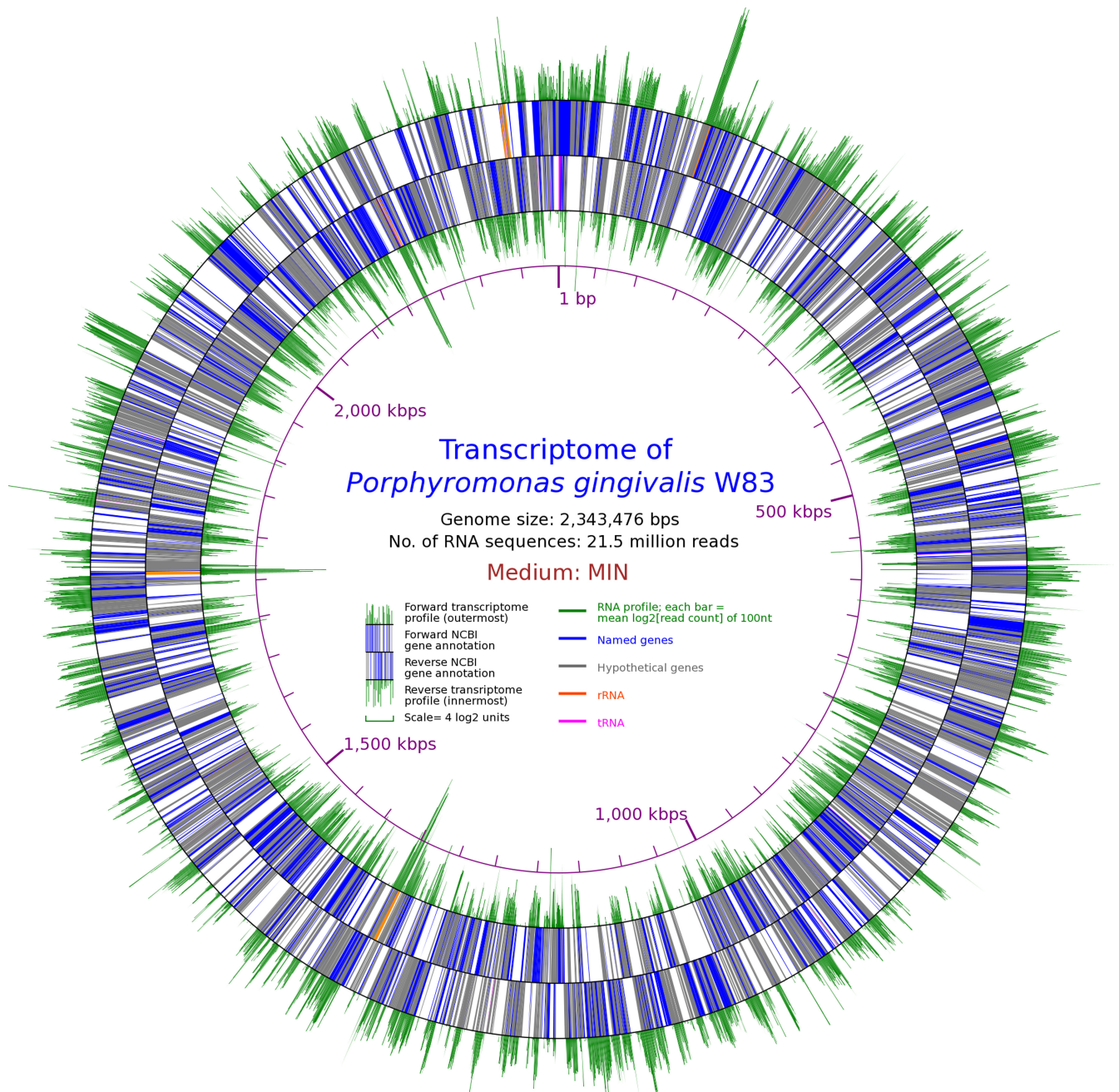
FIG 1 Circular transcriptome map showing the normalized RNAseq transcription signals derived from the MIN-cultured cells.

flow cells for cluster generation and sequencing. From the BAPHK and TSB samples, subjected to 50 cycles of sequencing, totals of 14.9 and 15.4 million sequence reads were generated, respectively, while the MIN sample, subjected to 38 cycles of sequencing, generated a higher number, 21.9 million sequence reads (Table 1). By iterative alignment and 3′-end trimming, on average 95.6% of the sequence reads were successfully mapped to the genome. Even though partial removal of rRNA significantly reduced the level of rRNA in the samples, the majority of the reads were still rRNA sequences. Consequently, the number of reads with unique sequences ranged from 2.3 to 3.2 million, comprising only ~10 to 20% of the

total reads. In addition, due to the repeated sequences in the *P. gingivalis* W83 genome, some reads were mapped to multiple genomic locations. The read counts were $\log_2$ transformed and subjected to upper quartile normalization excluding repeated and no-read sequences, as described in Materials and Methods. The percent coverage on the genome for nucleotides that were matched with more than two reads was in the range of ~69 to 83%.

Whole-genome visualization of transcriptome profiles. To view the transcription pattern at whole-genome scale, the RNAseq profiles were plotted as circular maps. Figure 1 presents the circular transcriptome map displaying the normalized RNAseq tran-
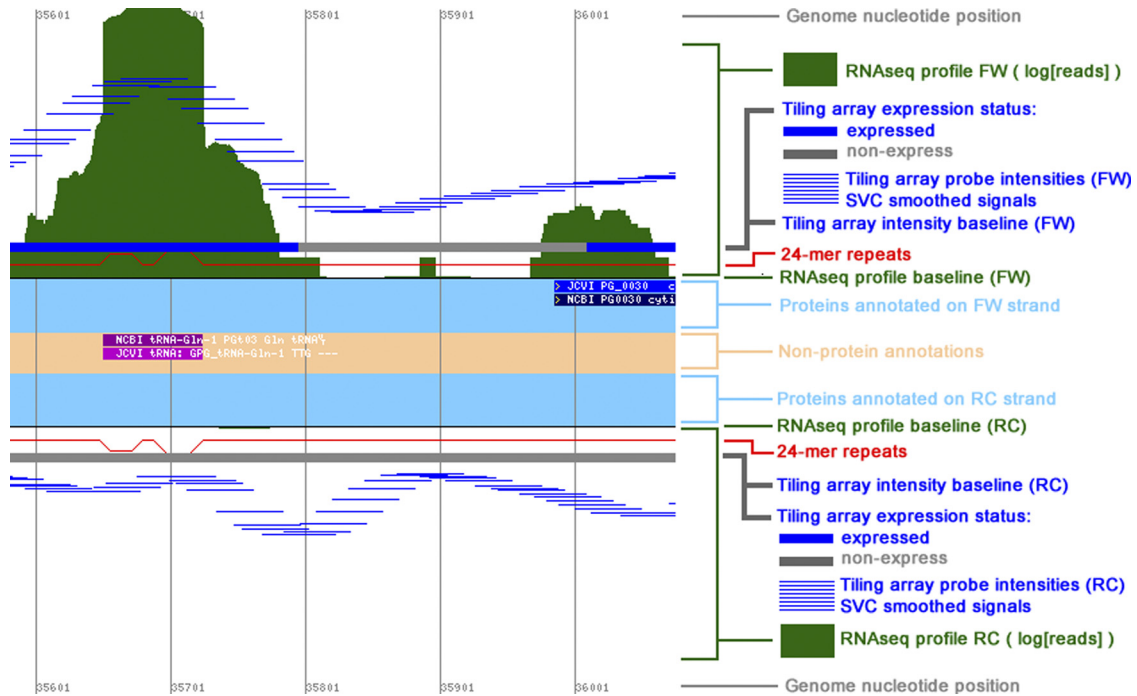
FIG 2 Screenshot of a genomic region as viewed in the Transcriptome Profile Viewer at the project website (http://bioinformatics.forsyth.org/mtd).

scription signals derived from the MIN culture. The distribution of TARs revealed in this whole-genome map generally corresponds to the layout of the NCBI-annotated ORFs on the genomic sequence. The ORFs rarely overlap each other on opposite strands and are distributed more or less evenly on the two coding strands of the genome. The global transcription patterns of the other two conditions were similar. Similarly, the microarray-based profiles were also plotted on whole-genome circular maps. At the whole-genome scale, both the microarray- and RNAseq-based transcriptome profiles follow consistent patterns. However, compared to RNAseq, the microarray-based profiles are associated with a higher level of background intensity. All whole-genome maps can be downloaded from the MTD website.

Transcription patterns of *P. gingivalis* W83. To facilitate examination and interpretation of the transcriptomic data obtained in this study, a web-based transcriptome profile viewer was developed. This transcriptome profile viewer dynamically displays the RNAseq profiles at the resolution of one nucleotide per pixel. Individual nonnormalized profiles are displayed separately, whereas the normalized profiles derived from the three conditions can be viewed together to allow easy comparisons. In addition, the HM-SVM-processed microarray-based signal intensity profiles are also plotted, probe by probe, alongside the RNAseq profiles for comparing the transcriptome profiles derived from the two technologies (Fig. 2). Since the microarray signals and the RNAseq reads that matched the repeated sequences were not excluded, a "repeat" plot based on all possible 24-mer nucleotide sequences in the genome is provided alongside the profiles (Fig. 2). In the regions where the sequence is unique, the repeat plot is at the baseline level, i.e., equal to 24. For repeating 24-mers elsewhere in the genome, the plot will be higher than 24.

Based on the microarray data, an average of 86.6% of the nucleotides within CMR-JCVI-annotated genes and 12.1% within

intergenic or antisense sequences were determined as expressed by the HM-SVM algorithm. Most regions of the genome displayed transcription from only one coding strand. However, RNA signals were detected on both strands with an average of 3.7% of the genomic sequence. In the RNAseq data, averages of 79.6% of the nucleotides within genes and 11.3% within intergenic or antisense sequences were recorded with more than two reads. Several novel transcripts were identified from the intergenic and antisense regions, but most of the RNAs detected within these regions were continuous transcription that extended upstream of the first ORF or downstream of the last ORF in a TAR, namely, 5′ and 3′ untranslated regions (UTRs). RNA signals detected on both coding strands were often caused by extended 3′ UTRs overlapping the ORF carried on the opposite strand (Fig. 3C). The different patterns of transcription observed for ORF-containing TARs are summarized in Fig. 3.

Expression of annotated protein coding genes. Of the 1,988 protein coding genes annotated by CMR-JCVI, 1,673 genes (84.15%) were considered expressed under at least one of the three culture conditions and 315 genes were not expressed under any of the conditions. The mean intensity level and sequence read count of all 1,988 protein coding genes are listed in Table S1 in the supplemental material. Of the 315 nonexpressed genes, the majority (54%) were annotated as hypothetical proteins and more than half (~58%) were not regarded as part of the core genome of *P. gingivalis* defined by Brunner et al. (5) (see Table S2 in the supplemental material). A total of 54 (~17%) nonexpressed genes were previously identified in comprehensive proteomic studies of *P. gingivalis* (strain ATCC 33277) (27, 39) (see Table S2). The nonexpressed genes were also grouped based on COG (clusters of orthologous groups) assignment (71) and subjected to clustering using the online software DAVID (The Database for Annotation, Visualization and Integrated Discovery) (30). COG annotation
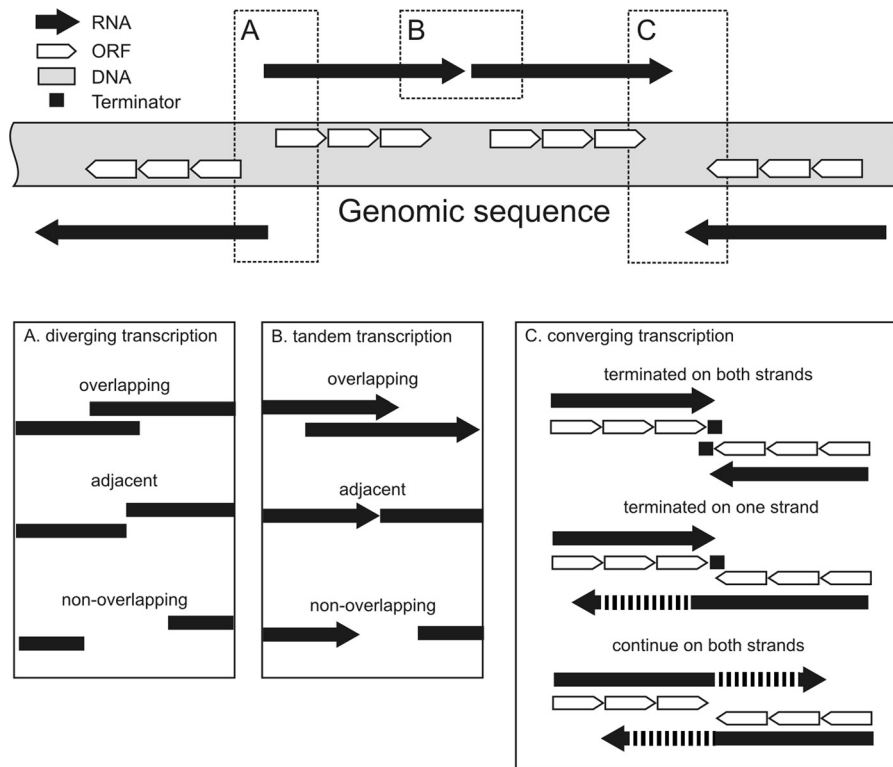
FIG 3 Different transcription patterns observed in the reported transcriptome profiles. (A) Diverging transcription at the 5′ ends. (B) Tandem transcription. (C) Converging transcription at the 3′ ends.

revealed several nonexpressed genes belonging to the functional group L (replication, recombination, and repair). Correspondingly, the most significant gene clusters grouped by DAVID include DNA binding, integration, recombination, and DNA metabolic processes. The second most significant DAVID cluster involved drug transport (e.g., genes encoding MATE family proteins).

Although the genome annotations of *P. gingivalis* maintained by CMR-JCVI and by the Los Alamos National Library (LANL Oralgen; http://www.oralgen.lanl.gov/) in general are similar, some CMR-JCVI- and LANL-specific genes have been reported (available at the LANL Oralgen website). Based on our microarray data, 65 of the 160 CMR-JCVI-specific genes and 57 of the 170 LANL-specific genes were determined as expressed by our criteria (see Table S3 in the supplemental material).

Operons and UTRs. In prokaryotes, multiple consecutive ORFs are often cotranscribed into a single transcription unit, termed an operon. Based on the HM-SVM-processed microarray data, we found that the majority of the *P. gingivalis* genes (an average of 1,468 ORFs across the samples) were expressed within TARs that contained more than one ORF (Table 2). Among these ORFs, 1,260 and 1,142 were also associated with the operons predicted by the MicrobesOnline database (60) and by the database of prokaryotic operons (DOOR) (50), respectively. All ORF-containing TARs were recorded together with their associated HM-SVM-identified transcription start and termination sites, available in the supplemental material (see Table S4).

The ends of the ORF-containing TARs can recess or protrude from the first or last ORF of the 5′ or 3′ end of the TAR, respec-

tively. The range of lengths from recessing (i.e., negative) and protruding (i.e., positive) ends detected in the transcriptome profile of cells grown in TSB is depicted in the histogram shown in Fig. 4. In general, most ends are protruding UTRs, and 3′ UTRs are longer than 5′ UTRs, with average sizes of 205 and 89 nucleotides, respectively (i.e., peak values in Fig. 4). Histograms of the 5′- and 3′-end distributions for the two other culture conditions are provided in the supplemental material (see Fig. S1).

When a recessed 5′ end is recorded for a TAR, the computer-annotated start codon of the first ORF may not be correct (Fig. 5A). When the 5′ end of a TAR is observed upstream of the first ORF, i.e., protruding, this region is named the 5′ UTR as described above (Fig. 5B). Some 5′ UTRs contain a metabolite-sensing regulatory element termed a riboswitch (83). In the LANL Oralgen database, a total of 8 riboswitches were predicted in the genome of *P. gingivalis* W83. These riboswitches include five copies of cobalamin (RF00174), two of thiamine pyrophosphate (TPP; RF00059), and one of *S*-adenosylmethionine (SAM_alpha; RF00521). RNA signals were detected in all of these riboswitches.

Novel RNAs. Many transcriptionally active but not-yet-annotated regions of the genome were detected in the transcriptome profiles. To screen for novel RNAs, protein coding and intergenic sequences of the ORFs within the TARs as well as their associated 5′ and 3′ UTRs were excluded. Table S5 in the supplemental material provides a conservative list of the novel RNAs identified in this study. The list of novel RNAs is not comprehensive, as many additional transcripts of often smaller and fragmented RNA signals were also observed in the nonannotated regions of the genome. These RNAs were not included in the list due

TABLE 2 Summary of microarray-based (HM-SVM-processed) identification of ORF-containing TARs from the three different culture conditions

| Category of TARs | No. by strand and medium | | | | | | | | |
| | Forward strand | | | Reverse complement strand | | | Both strands, combined | | |
	BAPHK	TSB	MIN	BAPHK	TSB	MIN	BAPHK	TSB	MIN
Total ORFs (NCBI/CMR-JCVI)	988	988	988	1,069	1,069	1,069	2,057	2,057	2,057
Nonexpressed ORFs	208	175	200	213	215	206	421	390	406
Expressed ORFs	780	813	788	856	854	863	1,636	1,667	1,651
Mono-ORF-containing TARs	99	89	96	82	85	98	181	174	194
TARs containing									
2 ORFs	51	50	49	38	31	35	89	81	84
3 ORFs	41	40	39	30	30	29	71	70	68
4 ORFs	23	18	18	26	27	27	49	45	45
5 ORFs	13	15	13	17	18	15	30	33	28
>5 ORFs	35	39	40	48	50	48	83	89	88
Total ORF-containing TARs	262	251	255	241	241	252	503	492	507
ORFs associated with									
Multi-ORF-containing TARs	681	724	692	774	769	765	1,455	1,493	1,457
Operons (DOOR)	539	570	549	586	589	592	1,125	1,159	1,141
Operons (MicrobesOnline)	585	619	598	658	657	663	1,243	1,276	1,261

to various reasons such as detection under one condition or in one assay only, close proximity to annotated genes, low-level signals, detection in a repeated region, or any combination of these reasons. In addition, although distinct RNA signals occasionally were observed in the 5′ or 3′ UTRs (Fig. 5F), due to their association with annotated ORFs, they were not considered novel RNAs in this report. All potential novel RNAs can be viewed in the MTD transcriptome profile viewer.

Some of the novel transcripts detected contain ORFs that were annotated only by the HOMD database and thus may encode novel proteins. However, most of the novel RNAs are expected to be noncoding. None of the novel RNAs listed were identified by the Rfam database (18).

cis-acting antisense RNAs. Two types of *cis*-acting antisense RNAs were identified. The first is bona fide antisense RNA, which is a well-isolated transcript detected on the opposite strand of one or several ORFs; the second type is the overlapping UTRs where the transcript on both ends or either end of an ORF is extended to
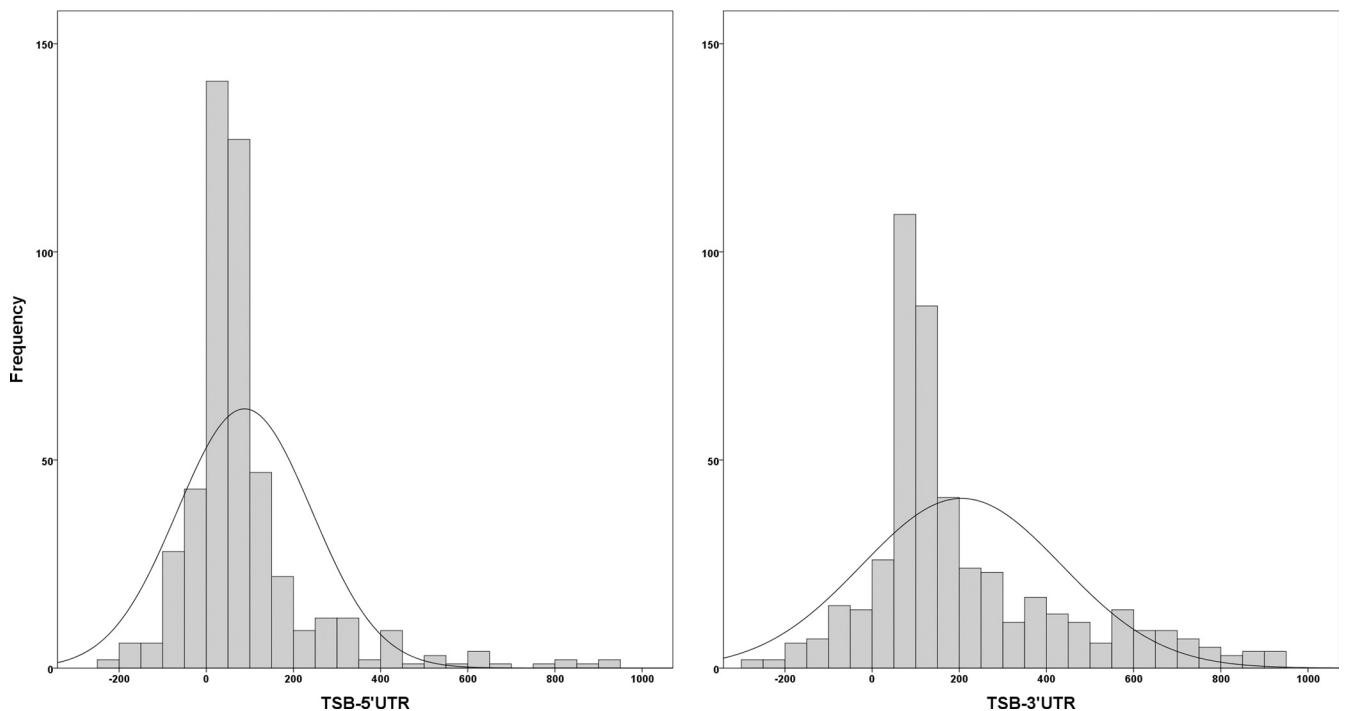

FIG 4 5′- and 3′-end distributions (TSB). 5′ and 3′ ends within the limits of recessing <300 nt and protruding <1,000 nt from their corresponding annotated ORFs were plotted as histograms. The x axis is the nucleotide size of recessing (i.e., negative) or protruding (i.e., positive) ends (/UTRs).
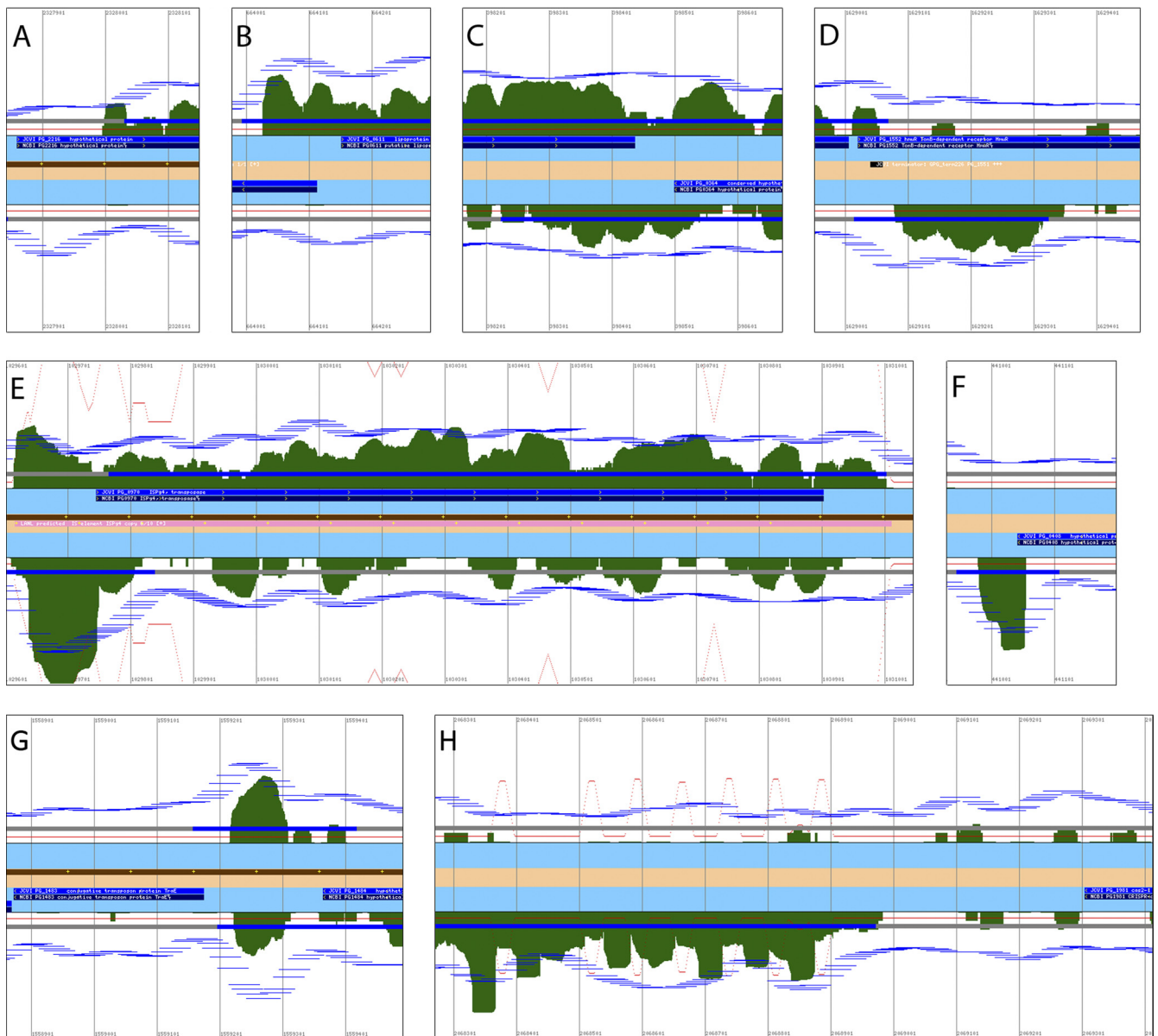
FIG 5 MTD, Transcriptome Profile Viewer screenshots. (A) Recessed 5′ end of TAR located downstream of the respective ORF (PG2216). (B) Protruding 5′ end of TAR located upstream of the respective ORF (PG0611); this 5′ UTR is overlapping the ORF on the opposite strand. (C) Example of extended 3′ UTRs antisense of annotated ORFs on both strands (PG0363-PG0364). (D) *cis*-antisense RNA opposite PG1552. (E) Antisense transcript IS*Pg4* (PG0970). (F) Distinct transcription at 3′ end of low-transcribed/nontranscribed gene (PG0408). (G) Intergenic, antisense RNA between operon PG1473 to PG1483 and PG1484 to PG1486 (encoding conjugative transposon proteins). (H) CRISPR transcription downstream of PG1981 (*cas2*).

the antisense region of the ORF carried on the opposite strand. Overlapping 3′ UTRs were observed in more than 30 genomic locations (Fig. 5C and 3C). A few instances of 5′ UTRs overlapping the ORF located on the opposite strand were also detected (Fig. 5B and 3A).

The most prominent bona fide antisense RNAs were observed on the opposite strand of insertion sequence (IS) elements. Numerous copies of transposase-encoding IS elements have been identified in the *P. gingivalis* W83 genome, designated IS*Pg1* to IS*Pg11* (56), and strong antisense RNA signals were detected for many of these IS elements. The transcriptome profiles display antisense transcription overlapping the 5′ end of IS*Pg3* and IS*Pg4*

(Fig. 5E), the 3′ end of IS*Pg2*, and virtually all of IS*Pg1*. Another bona fide antisense transcript was located opposite the *hmuR* gene (Fig. 5D). RNA signals were also identified from both strands of the intergenic region between two loci (PG1473 to PG1483 and PG1484 to PG1486) encoding conjugative transposon proteins of which the transposon genes displayed no or low mRNA abundance (Fig. 5G).

CRISPRs. Non-coding RNAs were also identified from the genomic sequences annotated as CRISPRs, which stands for "clustered regularly interspaced short palindromic repeats." These CRISPRs were predicted based on the unique alternating repeated and variable spacer sequences known for other CRISPRs (22).
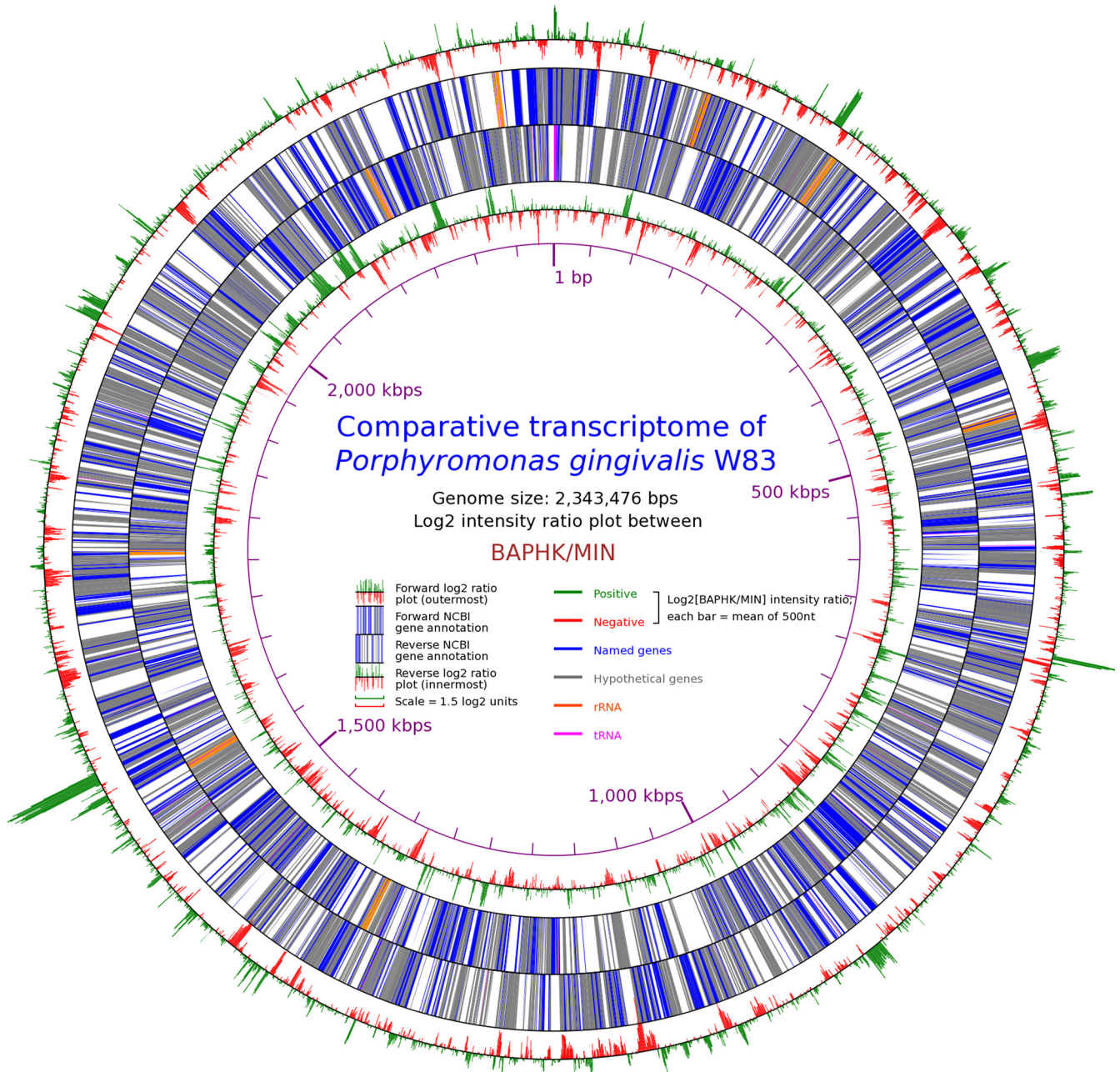
FIG 6 Circular genomic map of the differential expression between BAPHK and MIN based on the microarray signal intensity ratios.

CRISPRs are associated with resistance to phage infections (83) and both strong microarray signals and high-number RNAseq reads were detected from the CRISPR regions under all three culture conditions (an example is shown in Fig. 5H).

Comparative transcriptomics. To compare the transcriptome profiles between different growth conditions at whole-genome scale, the pairwise $\log_2$ ratios were plotted as circular genomic maps. Figure 6 displays the transcriptome comparative map between BAPHK and MIN based on the microarray $\log_2$ signal intensity ratios. Mean $\log_2$ intensity ratios were calculated from normalized signal intensities every 500 bp of the genomic sequences and plotted on the circular coordinate representing the genomic position. Positive $\log_2$ ratios were plotted in green and negative ratios were plotted in red to indicate the up- and downregulation when comparing BAPHK to MIN, respectively. Corresponding circular maps of the other two pairwise comparisons (i.e., BAPHK to TSB and TSB to MIN) are available at the MTD website. The whole-genome ratio maps revealed clusters of differentially expressed RNAs in *P. gingivalis*. Similar comparison plots based on the RNAseq profile data are also provided at the MTD website. By and large, microarray- and RNAseq-based data resulted in similar $\log_2$ ratio patterns in terms of the distribution of differentially expressed regions.

When the differential expression was compared at the single-

TABLE 3 Percent differentially expressed (DE) genome in *P. gingivalis* W83

DE[a]	% genome					
	Forward			Reverse complement		
	TSB/MIN	BAPHK/MIN	BAPHK/TSB	TSB/MIN	BAPHK/MIN	BAPHK/TSB
Range, log(ratio)						
7.5–8.0	0	0.02	0	0	0	0
7.0–7.5	0	0	0.02	0	0	0
6.5–7.0	0	0.02	0.04	0	0	0
6.0–6.5	0	0.06	0.06	0	0	0.06
5.5–6.0	0.04	0.02	0.17	0	0	0.28
5.0–5.5	0.02	0.09	0.04	0	0.04	0.06
4.5–5.0	0	0.02	0.06	0.02	0.13	0.09
4.0–4.5	0.02	0.49	0.28	0.04	0.23	0.15
3.5–4.0	0	0.41	0.45	0.04	0.36	0.32
3.0–3.5	0.06	0.28	0.43	0.28	0.49	0.34
2.5–3.0	0.13	0.28	0.77	0.55	0.62	0.64
2.0–2.5	0.26	0.55	1.02	0.77	1.41	0.77
1.5–2.0	1.49	1.92	1.62	1.88	2.50	1.79
1.0–1.5	3.95	3.78	3.71	3.76	3.73	3.18
0.5–1.0	9.05	6.00	6.81	7.85	6.57	6.53
0.0–0.5	24.60	15.19	13.74	25.05	15.02	15.43
0.0–−0.5	23.92	26.90	26.07	24.73	25.28	26.22
−0.5–−1.0	7.68	9.75	10.84	7.72	8.83	10.09
−1.0–−1.5	3.50	4.67	4.80	3.16	4.63	6.06
−1.5–−2.0	1.22	2.15	2.01	0.66	2.07	2.18
−2.0–−2.5	0.64	0.45	0.58	0.45	0.70	0.60
−2.5–−3.0	0.66	0.15	0.32	0.19	0.21	0.15
−3.0–−3.5	0.23	0.13	0.09	0.23	0	0.06
−3.5–−4.0	0.06	0.02	0	0	0	0
−4.0–−4.5	0	0	0	0	0	0
−4.5–−5.0	0	0	0.02	0	0	0
−5.0–−5.5	0	0	0.02	0	0	0
−5.5–−6.0	0	0	0.06	0	0	0
−6.0–−6.5	0	0	0	0	0	0
−6.5–−7.0	0	0	0	0	0	0
−7.0–−7.5	0	0	0	0	0	0
−7.5–−8.0	0	0	0	0	0	0
%						
>0.5 or <−0.5	29.02	31.26	34.22	27.61	32.54	33.35
>0.5	15.02	13.93	15.49	15.19	16.09	14.21
<−0.5	14.00	17.32	18.73	12.42	16.45	19.14

[a] DE, differentially expressed.

nucleotide level by the $\log_2$ ratio of RNAseq read coverage, approximately 30% of the genomic sequence displayed greater than 1.4-fold signal differences (i.e., $\log_2$ ratio of >0.5 or <−0.5, Table 3) on either coding strand, depending on the conditions compared. Similar results were recorded when the comparison was done based on the microarray data (data not shown).

Differentially expressed genes. Pairwise comparisons of the expression signals from CMR-JCVI-annotated protein coding genes were performed between all three culture conditions. Comparing two conditions at a time, the ORFs with >2-fold mean microarray signal intensity ratios, P values of <0.05, Q values of <0.1, and >1.5-fold change in the mean RNAseq read count were reported as differentially expressed. The total numbers of differentially expressed genes based on these criteria were 223, 218, and 267 for the pairwise comparisons of BAPHK and TSB, BAPHK and MIN, and MIN and TSB, respectively. Complete lists of the differentially expressed genes are provided in the supplemental material (Table S6). The differentially expressed genes were

grouped based on COG assignment (71) and subjected to clustering using the online software DAVID (30).

Combining the results from all three pairwise comparisons, COG annotation revealed that several genes from the functional group C (energy production and conversion) were upregulated in BAPHK and TSB compared to MIN. Genes upregulated in BAPHK and TSB also clustered to the "thiamine metabolism" KEGG pathway. These genes are organized in an operon (PG2107 to PG2111) with an LANL-predicted TPP (thiamine pyrophosphate) riboswitch located in the 5′ UTR. This TPP riboswitch was identified in our transcriptome profiles displaying a high level of transcription under all three culture conditions.

Two of the CRISPRs annotated in *P. gingivalis* have adjacent CRISPR-associated (*cas*) genes that are organized in the loci PG1981 to PG1989 and PG2013 to PG2020. These loci, containing both core *cas* genes and genes encoding RAMPs (repeat-associated mysterious proteins), were upregulated in BAPHK and TSB compared to MIN. In addition, the *rag* locus (PG0185-*ragA* and

PG0186-*ragB*) and the locus containing PG0063 to PG0066, encoding membrane efflux proteins, also displayed higher expression levels in BAPHK and TSB.

Comparing BAPHK and MIN to TSB, the extracytoplasmatic function (ECF) sigma factor-encoding genes PG0162, PG0214, and PG1660 with the putative anti-sigma factor PG1659 (13) and the consecutive genes PG0214 to PG0218 and PG0287 to PG0292 were upregulated in BAPHK and MIN. At the greater range, consecutive genes PG0679 to PG0685, annotated as encoding membrane efflux proteins and ABC transporter proteins (putative), were expressed at higher levels in TSB than in BAPHK and MIN, while the locus containing PG1661 to PG1667, also annotated as encoding membrane efflux and ABC transporter proteins, was upregulated in BAPHK compared to TSB. In addition, several genes within the K-antigen capsule-encoding locus PG0106 to PG0120 were upregulated in TSB compared to both MIN and BAPHK, whereas the downstream gene (PG0121) encoding DNA-binding protein HU was highly expressed under all culture conditions, but the RNA levels were higher in MIN and BAPHK than in TSB.

In BAPHK, IS*Pg1* genes and degenerated/truncated versions of IS*Pg1* displayed higher expression than in TSB and MIN. The consecutive genes PG0611 to PG0614 also displayed significantly higher RNA abundance in BAPHK than in TSB and MIN.

Additional differential expression patterns observed between only two conditions are presented as follows. Comparing BAPHK to TSB, COG annotation revealed several genes belonging to the functional group M (cell envelope biogenesis, outer membrane) upregulated in BAPHK. Two putative hemagglutinin genes (PG0411 and PG1326) as well as *hagA* (PG1837) were also upregulated in BAPHK compared to TSB.

Comparing BAPHK to MIN, the hemagglutinin-encoding genes, PG1837 (*hagA*), PG1844 (*hagD*), PG2024 (*hagE*), and PG0411 and PG1326 (both putative), as well as the arginine-specific cysteine proteinase gene PG0506 (*rgpB*), were upregulated in BAPHK and clustered together by the DAVID online software. The locus containing PG1503 to PG1509 displayed a significantly higher level of RNA signal in BAPHK, whereas the loci containing PG1252 to PG1254, PG1310 to PG1312, and PG1343 to PG1348 and two genes within the *bat* locus, *batA* and *batC*, were expressed at higher levels in MIN. Some genes known to be involved in oxidative stress protection (44, 52) were also observed to be differentially expressed and included genes encoding ferritin (PG1286), superoxide dismutase (PG1545), and thiol peroxidase (PG1729). These genes were upregulated in BAPHK compared to MIN whereas alkylhydroperoxide reductase (subunit PG0618) was significantly downregulated in BAPHK compared to MIN. However, these four genes displayed a higher degree of variation between the biological repeats in MIN (see Table S6B in the supplemental material).

Comparing TSB to MIN, COG annotation revealed several genes belonging to the functional groups J (translation, ribosomal structure, and biogenesis) and C (energy production and conversion) upregulated in TSB. Similar results were obtained from DAVID clustering of the genes with increased expression in TSB from which two significant KEGG pathways, "ribosome" and "oxidative phosphorylation," were identified. The oxidative phosphorylation-associated genes included PG1803 to PG1807 coding for v-type ATPase subunits. In addition, three hemagglutinin-encoding genes, *hagB*, *hagC*, and *hagE*, were up-

regulated in TSB compared to MIN. DAVID clustering of the genes that were upregulated in MIN resulted in one cluster of seven genes encoding tetratricopeptide repeat (TPR) domain proteins and a second cluster of genes associated with the cell external encapsulating structure. In addition, six glycosyl transferase-encoding genes and the *bat* locus genes PG1581 to PG1585 were upregulated in MIN compared to TSB.

DISCUSSION

This report presents the first comprehensive transcriptome analysis of *P. gingivalis*. The transcriptome profiles reported in this paper were derived from RNA signals detected by two distinct high-throughput technologies, the hybridization-based genomic tiling microarray and the sequencing-based RNAseq method. The generated profiles are strand specific, and for the RNAseq data, the resolution is at single-nucleotide level. Since the transcriptome of an organism is dynamic and condition dependent, we studied the transcriptome of *P. gingivalis* under three distinct laboratory culturing conditions so that more condition-specific RNA expression could be detected. Based on the microarray profiles, there is still 10.53% of the genomic sequence with no transcriptional activity under any of the three conditions. Whether or not these untranscribed regions contain any TARs for other specific conditions remains to be investigated.

Microarray versus RNAseq profiles. The transcriptome profiles derived from the RNAseq method provide single-nucleotide resolution and are in general superior for exact transcript boundary detection. Areas without transcriptional activity can also be determined distinctively in the RNAseq profiles simply as no-read coverage. However, the RNAseq profiles do not come without pitfalls. First, the genomic regions with single-read coverage in our profiles were on average 8.5% on the forward and 8.7% on the reverse strand, and whether or not these sequence reads represent actual RNA products in the cell remains to be determined. Second, in our profiles we experienced fluctuating sequence read coverage, and at times the read profiles were fragmented by gaps without read coverage even within regions with high read counts. Other RNAseq-based transcriptome studies have also reported fluctuating read coverage (15, 23, 59), which may be caused by stochastic artifacts introduced during sample preparation or be due to possible degradation of RNA (85). During the cDNA library preparation, sequencing biases may be introduced due to different efficiencies in fragmentation, reverse transcription, and amplification, all of which can be affected by sequence composition as well as the level of transcripts. If the TARs were to be determined solely based on the gaps between the RNAseq signals, numerous small pieces of TARs would have been recorded, leading to overestimation of transcription units and false interpretation of the transcription patterns. For this reason, and because the microarray transcription profiles were compiled based on hybridization signals from two to three biological repeats, the reported TARs were primarily determined based on the HM-SVM-processed tiling microarray data, while the sequencing data were used to validate and support the results generated.

Transcription activity in *P. gingivalis* W83. The transcriptome profiles presented in this report were based on RNA isolated from cells grown in three basic but distinctly different laboratory culturing media. Overall, we detected 84.15% of the CMR-JCVI-annotated genes encoding proteins expressed across the three culture conditions. The majority of the expressed genes were tran-

scribed within TARs containing more than one ORF. Cotranscribed genes are often defined as operons, and many of the ORFs within multigene TARs were associated with operons annotated by the MicrobesOnline (60) and the DOOR (50) databases. However, we recorded a higher number of ORFs within multigene TARs compared to the number of MicrobesOnline/DOOR-annotated operon-ORFs. Our HM-SVM-based TAR analysis identified the regions of continuous transcriptional activity, and in most cases, the TARs reflect real transcriptional units. The higher number of ORFs within multigene TARs can be due to transcript boundaries that were not detected by the HM-SVM algorithm, as closely localized neighboring transcription units of different transcriptional levels may not have been separated by the algorithm. The original microarray intensity data, the normalization procedures, and the HM-SVM algorithm can all affect the detection of real transcript boundaries. Transcript boundaries may have been missed, resulting in overestimation of the number of ORFs associated with multigene TARs—potential operon structures. Genes with similar functions are often organized in operon structures, and identifying the operon organization of a genome under specific conditions will help the understanding of both gene regulation and function (58). Between the different growth conditions, we recorded that some TARs contained various numbers of ORFs, suggesting different gene regulation mechanisms. Recent transcriptome studies have reported frequent condition-dependent changes of expression pattern as well as modifications of operon structures leading to alternative transcripts in both bacteria and archaea (23, 37).

Several of the recorded TARs were previously verified as operons in *P. gingivalis*, such as the locus containing the genes PG1333 to PG1335 in strain W83 (82) and the *htrABCD* locus (PG0645 to PG0648) (68) and the *rag* locus (PG0185-PG0186) (26) in strain W50. Johnson et al. (35) recently reported the transcriptional start site of the *bcp-recA-vimA-vimE-vimF* operon in *P. gingivalis* W83 to be located at the genomic position nt 942713, downstream from the start of the current CMR-JCVI/LANL-annotated *bcp* gene. Close but not identical, our sequencing profiles displayed distinct transcription start sites at nt 942710 (TSB and MIN) and nt 942704 (BAPHK), based on a cutoff of more than 2 read counts. Both our microarray- and sequencing-based transcriptome profiles also indicate the *bcp-recA-vimA-vimE-vimF* operon genes to be part of the same transcriptional unit.

5′ and 3′ UTRs. The transcriptome profiles revealed several 5′ and 3′ UTRs of various lengths and transcription patterns. 5′ UTRs can have regulatory functions and mechanisms such as riboswitches with metabolite-sensing regulatory 5′-UTR RNA structures, *trans*-acting small RNAs that bind to the 5′ UTR affecting the translation and/or stability of the mRNA, and temperature-dependent changes of 5′-UTR secondary structure masking the Shine-Dalgarno sequence (21, 72). All of the eight computationally predicted riboswitches in the *P. gingivalis* W83 genome showed significant RNA signals. Two of these riboswitches were categorized as TPP riboswitches, which respond to thiamine pyrophosphate (TPP) and regulate the genes responsible for importing and synthesizing thiamine and its phosphorylated derivatives (66). In the MIN transcriptome profiles, a significantly lower level of RNA signals was detected from the genes within one of the operons with a TPP riboswitch located in the 5′ UTR. The predominant mechanisms of TPP-dependent regulation are translation inhibition and premature transcription termination

(66), suggesting that the latter mechanism may be an important factor in the downregulation of thiamine metabolic genes observed in the MIN cultures.

Although less frequent than the 3′ UTRs, some 5′ UTRs do overlap each other. When the 5′ UTRs of two neighboring and oppositely transcribed TARs overlap or are very close to each other (Fig. 3A), the promoter sequences controlling the transcription initiation are bound to be located within the antisense regions of the opposite TAR. How the antisense sequence of an ORF also serves as the promoter of another transcription event is an interesting phenomenon for future studies.

In contrast to most 5′-UTR RNAs, which displayed distinct and intact transcriptional signals, many of the identified 3′ UTRs were fragmented with gradually fading transcriptional signals. The fragmentation may be caused by degradation targeted at the 3′ end or may be due to technical biases. Nevertheless, a possible function of the 3′ UTRs is the stabilization of the mRNA, as longer 3′ UTRs may form secondary structures, making them less susceptible to 3′ exoribonuclease degradation (21, 61). The 3′ UTRs may also have a role in regulation by potentially giving rise to small RNAs (21). It is also possible that some of the 3′ UTRs are simply the products of continued transcription due to the lack of a functional transcriptional terminator and may not have any biological function.

***cis*-antisense RNAs.** Many novel RNAs within the intergenic and antisense regions were detected in the transcription profiles. Most of these transcripts are expected to be noncoding RNAs and may have regulatory functions. Regulatory RNAs are known to be important in rapid response to changing environmental conditions and for controlling bacterial virulence (21, 74). Some of these known regulatory RNAs bind to and modulate protein activity, but the majority function by base pairing with specific target mRNAs. The base-pairing RNAs can be transcribed from a genomic location different from or at the opposite strand of their target mRNA, i.e., in *trans* or in *cis*, respectively (21). *trans*-acting antisense RNAs are difficult to identify based on the profile data only. *cis*-acting antisense RNAs, directly transcribed from the opposite strand, have complete complementarity with their target RNAs and may have regulatory effects on them. We detected several *cis*-acting antisense RNAs, including a high number of antisense RNAs opposite transposase-encoding genes. The *P. gingivalis* genome contains many copies of different categories of IS elements, which encode transposase necessary for transposition activities (56). Transposase-associated antisense RNAs have been reported to inhibit mRNA translation of the transposases encoded by IS*10* and IS*30* (4). Antisense RNAs opposite transposase-encoding genes have also been detected in other genome-scale transcription studies (3, 34, 41), which suggests that an important function of antisense RNAs may be to inhibit transposition (74). The presence of antisense RNAs potentially repressing transposition activities of the IS elements may explain why, despite the high number of IS elements in *P. gingivalis*, the genome was known to be quite stable (12). From mapping the distribution of IS elements in several *P. gingivalis* laboratory strains and clinical isolates, Califano et al. (7) suggested a preserved strain relationship reflecting the fact that neither recent transposition nor homologous recombination between IS copies had occurred extensively in the isolates studied. However, Califano et al. (7) emphasized that their results do not imply that genomic variation caused by IS elements never occurred or that the transposition between IS elements is not im-

portant in adapting to new environments. Evidence of genomic rearrangement has been detected where a gene is interrupted with an IS sequence (45). The multiple copies of the IS elements in the genome also serve as evidence of replicative transposition events in the past. Under what conditions the IS elements are activated and how they respond to the presence of antisense RNA will be an interesting topic for future research.

Antisense transcription was also detected opposite other genes, e.g., *hmuR*. This gene belongs to the *hmuYRSTUV* operon (PG1551 to PG1556), encoding a hemin uptake system (46). Under all three conditions, *hmuY* (PG1551) was expressed but the consecutive genes (PG1552 to PG1556) within the operon were recorded as nonexpressed. Significantly higher expression levels of *hmuY* than of *hmuR* have also been reported in other studies, and this differential expression may be explained by the presence of a transcription terminator, suggested based on the sequence composition downstream of the *hmuY* ORF (46, 91). However, in the transcription profiles we detected a well-defined antisense transcript localized in the 5′ end of the *hmuR* ORF. This antisense transcript is likely to affect the differential expression recorded within the *hmu* operon, although the exact mechanism of this antisense transcript remains to be investigated.

Differential expression revealed by comparative transcriptomics. This report describes the first genome-wide comparison of the RNA expression in *P. gingivalis* from three different culture conditions. The microarray approach permits the comparison of two profiles at the probe-by-probe level, and with the RNAseq method, $\log_2$ read count ratios can be calculated between two profiles for each nucleotide position in the genome. On the larger scale, measuring the differential expression between corresponding TARs is difficult because TARs are dynamic entities of the transcriptome repertoire and different growth conditions generate different number of TARs of various nucleotide lengths. For this reason, the comparison of the differential expression in this study was done at ORF level. One may also think that comparison can be done for the intergenic regions, but the fact that many intergenic regions consist of 5′- and 3′-UTR sequences which are cotranscribed with their associated ORFs makes direct comparison between the signals within intergenic regions less meaningful. This is not to say that the comparison at the nucleotide, probe, TAR, or intergenic level cannot be done and will not generate useful results. Contrarily, we believe that differentially expressed RNAs that may have important functions can be easily identified, especially if a smaller genomic locus of interest is focused.

For the CMR-JCVI-annotated protein coding genes determined to be differentially expressed, several mRNAs encoding ribosome subunit and translation-related proteins were detected at higher levels in TSB than in MIN. Comprehensive proteomic studies using *P. gingivalis* strain ATCC 33277 have reported increased levels of ribosome subunit and translation-related proteins both for cells grown in a multispecies community (39) and for *P. gingivalis* cells internalized in human gingival epithelial cells (27), indicating that these environments provide sufficient energy for increased translational activity. In our study, the upregulation of genes involved in protein synthesis may be explained by better energy supply and higher growth rate in TSB medium than in the more nutrient-limited MIN medium. In addition, several genes in the COG functional group C (energy production and conversion) were also upregulated in TSB compared to MIN, another indication of a more nutrient-rich environment. Some of these genes

clustered to the KEGG pathway (oxidative phosphorylation). The clustered genes belong to an operon holding seven genes (PG1801 to PG1807) annotated as encoding v-type ATPase subunits. Meuric et al. (53) recently reported that the functional mechanisms of the ATPase in *P. gingivalis* are unknown but that the genes in this operon are the only genes in *P. gingivalis* that are homologous to known ATPase genes. If the protein products of these genes organize into an ATPase, they may allow ATP synthesis (53).

The *P. gingivalis* W83 genome contains six extracytoplasmatic function (ECF) sigma factors, and under the conditions studied, we found an increased mRNA abundance of the ECF sigma factor-encoding genes PG0214, PG0162, and PG1660 with putative cognate anti-sigma factor gene PG1659 (13) in both BAPHK and MIN compared to TSB. ECF sigma factors are known to regulate gene expression in response to stress conditions, and Dou et al. (13) recently reported that mutants defective in PG1660 protein were more sensitive to oxidative stress (i.e., H_2O_2 exposure). They also found that ECF sigma factors PG0162 and PG1660 may regulate the activity of known virulence factors. The higher abundance of mRNA coding for ECF sigma factors may indicate a stress response due to nutrient-limited medium (i.e., MIN). With regard to the colony-forming cells grown on blood agar plates (i.e., BAPHK), cells on the surface of the colonies may both have reduced access to nutrients and be more exposed to environmental changes, which may induce a stress response reaction.

CRISPR/Cas systems are important in bacterial defense against phage and plasmid invasion. The CRISPR spacer sequences function as a memory of past invasions and provide the ability to inhibit a second invasion by the same phage or plasmid. The mechanism of action involves cleavage of the CRISPR transcript to short guide RNAs (i.e., spacer sequences) that are complementary to and direct the destruction of the target invading nucleic acid (20). CRISPR-associated *cas* genes are involved in several important functions, including processing of guide RNAs, degradation of invading nucleic acids, and acquisition of new spacer sequences (20). The condition-dependent transcription profiles revealed that several *cas* genes were upregulated in BAPHK and TSB compared to MIN. The biological reason for these differential expression patterns is unknown and remains to be investigated.

Hemagglutinins are adhesins that when expressed on the bacterial surface are among the components that facilitate bacterial coaggregation and host adhesion (33, 73). The two genes *hagC* and *hagB* display a high degree of sequence homology and are coding for almost identical proteins (43). The expression of these genes was upregulated in TSB compared to MIN. The sequence homology explains the similar differential expression patterns. Both the growth phase and the level of hemin have been reported to affect the transcription of these two genes (43). The hemin levels in the TSB and MIN media were similar, but the lower mRNA abundance in MIN may reflect the slower growth caused by the nutrient-limited condition. Higher expression of *hagA*, *hagD*, *hagE*, *rgpB*, and the two putative hemagglutinin genes was observed in BAPHK than in MIN. The two putative hemagglutinin genes and *hagA* were also upregulated in BAPHK compared to TSB. Most of these hemagglutinin genes share different homology regions, and the two genes *hagA* and *hagD* were reported to be expressed under excess-hemin conditions and at higher levels at late exponential and stationary phases (25). The differential expression patterns observed for these hemagglutinin-encoding

genes may be explained by the various growth stages represented in the *P. gingivalis* cells grown on blood agar plates compared to liquid cultures and by the different medium contents. The defibrinated sheep blood contained in the BAPHK medium may have affected the expression of hemagglutinin-encoding genes, as adhesin domains encoded by the hemagglutinin genes are involved in agglutination to erythrocytes and further in the binding to hemoglobin for heme acquisition (67).

Concluding remarks. We have compiled the *P. gingivalis* transcriptome profiles for three different laboratory culture conditions using two different technologies—genomic tiling microarrays and RNA sequencing. The transcriptome profiles generated are strand specific and allowed genome-wide identification of transcription patterns and large-scale discovery of novel transcripts from the intergenic and antisense regions. The transcriptomic data compiled in this study hold a tremendous amount of information that remains to be explored and analyzed in different ways. Ultimately, the transcriptome profiles can provide a valuable resource for future research to further investigate unknown transcriptional events and gene regulation mechanisms in *P. gingivalis.*

ACKNOWLEDGMENTS

REFERENCES

1. **Agueda A, Echeverria A, Manau C.** 2008. Association between periodontitis in pregnancy and preterm or low birth weight: review of the literature. Med. Oral Patol. Oral Cir. Bucal **13**:E609–E615.
2. **Albert TJ, et al.** 2003. Light-directed 5′→3′ synthesis of complex oligonucleotide microarrays. Nucleic Acids Res. **31**:e35.
3. **Beaume M, et al.** 2010. Cartography of methicillin-resistant S. aureus transcripts: detection, orientation and temporal expression during growth phase and stress conditions. PLoS One **5**:e10725.
4. **Brantl S.** 2007. Regulatory mechanisms employed by cis-encoded antisense RNAs. Curr. Opin. Microbiol. **10**:102–109.
5. **Brunner J, et al.** 2010. The core genome of the anaerobic oral pathogenic bacterium Porphyromonas gingivalis. BMC Microbiol. **10**:252.
6. **Bullard JH, Purdom E, Hansen KD, Dudoit S.** 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics **11**:94.
7. **Califano JV, Arimoto T, Kitten T.** 2003. The genetic relatedness of Porphyromonas gingivalis clinical and laboratory strains assessed by analysis of insertion sequence (IS) element distribution. J. Periodontal Res. **38**:411–416.
8. **Chen Y, Souaiaia T, Chen T.** 2009. PerM: efficient mapping of short sequencing reads with periodic full sensitive spaced seeds. Bioinformatics **25**:2514–2521.
9. **Colombo AP, et al.** 2002. Subgingival microbiota of Brazilian subjects with untreated chronic periodontitis. J. Periodontol. **73**:360–369.
10. **David L, et al.** 2006. A high-resolution map of transcription in the yeast genome. Proc. Natl. Acad. Sci. U. S. A. **103**:5320–5325.
11. **de Pablo P, Chapple IL, Buckley CD, Dietrich T.** 2009. Periodontitis in systemic rheumatic diseases. Nat. Rev. Rheumatol. **5**:218–224.
12. **Dong H, et al.** 1999. Genomic loci of the Porphyromonas gingivalis insertion element IS1126. Infect. Immun. **67**:3416–3423.
13. **Dou Y, Osbourne D, McKenzie R, Fletcher HM.** 2010. Involvement of extracytoplasmic function sigma factors in virulence regulation in Porphyromonas gingivalis W83. FEMS Microbiol. Lett. **312**:24–32.
14. **Duncan, MJ, Nakao S, Skobe Z, Xie H.** 1993. Interactions of Porphyromonas gingivalis with epithelial cells. Infect. Immun. **61**:2260–2265.
15. **Filiatrault MJ, et al.** 2010. Transcriptome analysis of Pseudomonas syringae identifies new genes, noncoding RNAs, and antisense activity. J. Bacteriol. **192**:2359–2372.
16. **Fitzpatrick RE, Wijeyewickrema LC, Pike RN.** 2009. The gingipains: scissors and glue of the periodontal pathogen, Porphyromonas gingivalis. Future Microbiol. **4**:471–487.
17. **Gaetti-Jardim E Jr, Marcelino SL, Feitosa AC, Romito GA, Avila-Campos MJ.** 2009. Quantitative detection of periodontopathic bacteria in atherosclerotic plaques from coronary arteries. J. Med. Microbiol. **58**:1568–1575.
18. **Gardner PP, et al.** 2009. Rfam: updates to the RNA families database. Nucleic Acids Res. **37**:D136–D140.
19. **Gibson FC, Genco CA.** 2006. The genus Porphyromonas, p 428–454 *In* Dworkin M, Falkow S, Schleifer K-H, Rosenberg E, Stackebrandt E (ed), The prokaryotes, vol 7. Springer, New York, NY.
20. **Gottesman S.** 2011. Microbiology: dicing defence in bacteria. Nature **471**:588–589.
21. **Gripenland J, et al.** 2010. RNAs: regulators of bacterial virulence. Nat. Rev. Microbiol. **8**:857–866.
22. **Grissa I, Vergnaud G, Pourcel C.** 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics **8**:172.
23. **Guell M, et al.** 2009. Transcriptome complexity in a genome-reduced bacterium. Science **326**:1268–1271.
24. **Guillier M, Gottesman S.** 2006. Remodelling of the Escherichia coli outer membrane by two small regulatory RNAs. Mol. Microbiol. **59**:231–247.
25. **Han N, Lepine G, Whitlock J, Wojciechowski L, Progulske-Fox A.** 1998. The Porphyromonas gingivalis prtP/kgp homologue exists as two open reading frames in strain 381. Oral Dis. **4**:170–179.
26. **Hanley SA, Aduse-Opoku J, Curtis MA.** 1999. A 55-kilodalton immunodominant antigen of Porphyromonas gingivalis W50 has arisen via horizontal gene transfer. Infect. Immun. **67**:1157–1171.
27. **Hendrickson EL, Xia Q, Wang T, Lamont RJ, Hackett M.** 2009. Pathway analysis for intracellular Porphyromonas gingivalis using a strain ATCC 33277 specific database. BMC Microbiol. **9**:185.
28. **Hosogi Y, Duncan MJ.** 2005. Gene expression in Porphyromonas gingivalis after contact with human epithelial cells. Infect. Immun. **73**:2327–2335.
29. **Høvik H, Chen T.** 2010. Dynamic probe selection for studying microbial transcriptome with high-density genomic tiling microarrays. BMC Bioinformatics **11**:82.
30. **Huang DW, Sherman BT, Lempicki RA.** 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. **4**:44–57.
31. **Huber W, Toedling J, Steinmetz LM.** 2006. Transcript mapping with high-density oligonucleotide tiling arrays. Bioinformatics **22**:1963–1970.
32. **Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M.** 2002. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics **18**(Suppl 1):S96–S104.
32a. **Illumina.** 2009. Small RNA sample preparation guide—1004239 Rev B. Illumina, San Diego, CA.
33. **Ito R, Ishihara K, Shoji M, Nakayama K, Okuda K.** 2010. Hemagglutinin/adhesin domains of Porphyromonas gingivalis play key roles in coaggregation with Treponema denticola. FEMS Immunol. Med. Microbiol. **60**:251–260.
34. **Jager D, et al.** 2009. Deep sequencing analysis of the Methanosarcina mazei Go1 transcriptome in response to nitrogen availability. Proc. Natl. Acad. Sci. U. S. A. **106**:21878–21882.
35. **Johnson NA, McKenzie RM, Fletcher HM.** 2011. The bcp gene in the bcp-recA-vimA-vimE-vimF operon is important in oxidative stress resistance in Porphyromonas gingivalis W83. Mol. Oral Microbiol. **26**:62–77.
36. **Kapranov P, et al.** 2002. Large-scale transcriptional activity in chromosomes 21 and 22. Science **296**:916–919.
37. **Koide T, et al.** 2009. Prevalence of transcription promoters within archaeal operons and coding sequences. Mol. Syst. Biol. **5**:285.
38. **Kozarov EV, Dorn BR, Shelburne CE, Dunn WA Jr, Progulske-Fox A.** 2005. Human atherosclerotic plaque contains viable invasive Actinobacillus actinomycetemcomitans and Porphyromonas gingivalis. Arterioscler. Thromb. Vasc. Biol. **25**:e17–e18.

39. **Kuboniwa M, et al.** 2009. Proteomics of Porphyromonas gingivalis within a model oral microbial community. BMC Microbiol. **9**:98.

40. **Lamont RJ, Jenkinson HF.** 1998. Life below the gum line: pathogenic mechanisms of Porphyromonas gingivalis. Microbiol. Mol. Biol. Rev. **62**:1244–1263.

41. **Landt SG, et al.** 2008. Small non-coding RNAs in Caulobacter crescentus. Mol. Microbiol. **68**:600–614.

42. **Lenz DH, Miller MB, Zhu J, Kulkarni RV, Bassler BL.** 2005. CsrA and three redundant small RNAs regulate quorum sensing in Vibrio cholerae. Mol. Microbiol. **58**:1186–1202.

43. **Lepine G, Progulske-Fox A.** 1996. Duplication and differential expression of hemagglutinin genes in Porphyromonas gingivalis. Oral Microbiol. Immunol. **11**:65–78.

44. **Lewis JP, Iyer D, Anaya-Bergman C.** 2009. Adaptation of Porphyromonas gingivalis to microaerophilic conditions involves increased consumption of formate and reduced utilization of lactate. Microbiology **155**:3758–3774.

45. **Lewis JP, Macrina FL.** 1998. IS195, an insertion sequence-like element associated with protease genes in Porphyromonas gingivalis. Infect. Immun. **66**:3035–3042.

46. **Lewis JP, Plata K, Yu F, Rosato A, Anaya C.** 2006. Transcriptional organization, regulation and role of the Porphyromonas gingivalis W83 hmu haemin-uptake locus. Microbiology **152**:3367–3382.

47. **Li H, et al.** 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**:2078–2079.

48. **Lister R, et al.** 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell **133**:523–536.

49. **Lo AW, et al.** 2009. Comparative transcriptomic analysis of Porphyromonas gingivalis biofilm and planktonic cells. BMC Microbiol. **9**:18.

50. **Mao F, Dam P, Chou J, Olman V, Xu Y.** 2009. DOOR: a database for prokaryotic operons. Nucleic Acids Res. **37**:D459–D463.

51. **Masse E, Vanderpool CK, Gottesman S.** 2005. Effect of RyhB small RNA on global iron use in Escherichia coli. J. Bacteriol. **187**:6962–6971.

52. **Meuric V, Gracieux P, Tamanai-Shacoori Z, Perez-Chaparro J, Bonnaure-Mallet M.** 2008. Expression patterns of genes induced by oxidative stress in Porphyromonas gingivalis. Oral Microbiol. Immunol. **23**:308–314.

53. **Meuric V, Rouillon A, Chandad F, Bonnaure-Mallet M.** 2010. Putative respiratory chain of Porphyromonas gingivalis. Future Microbiol. **5**:717–734.

54. **Milner P, Batten JE, Curtis MA.** 1996. Development of a simple chemically defined medium for Porphyromonas gingivalis: requirement for alpha-ketoglutarate. FEMS Microbiol. Lett. **140**:125–130.

55. **Nagasawa T, et al.** 2010. Relationship between periodontitis and diabetes—importance of a clinical study to prove the vicious cycle. Intern. Med. **49**:881–885.

56. **Nelson KE, et al.** 2003. Complete genome sequence of the oral pathogenic bacterium Porphyromonas gingivalis strain W83. J. Bacteriol. **185**:5591–5601.

57. **Nuwaysir EF, et al.** 2002. Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. Genome Res. **12**:1749–1755.

58. **Okuda S, et al.** 2007. Characterization of relationships between transcriptional units and operon structures in Bacillus subtilis and Escherichia coli. BMC Genomics **8**:48.

59. **Passalacqua KD, et al.** 2009. Structure and complexity of a bacterial transcriptome. J. Bacteriol. **191**:3203–3211.

60. **Price MN, Huang KH, Alm EJ, Arkin AP.** 2005. A novel method for accurate operon predictions in all sequenced prokaryotes. Nucleic Acids Res. **33**:880–892.

61. **Rasmussen S, Nielsen HB, Jarmer H.** 2009. The transcriptionally active regions in the genome of Bacillus subtilis. Mol. Microbiol. **73**:1043–1057.

62. **Rodrigues PH, Progulske-Fox A.** 2005. Gene expression profile analysis of Porphyromonas gingivalis during invasion of human coronary artery endothelial cells. Infect. Immun. **73**:6169–6173.

63. **Romby P, Vandenesch F, Wagner EG.** 2006. The role of RNAs in the regulation of virulence-gene expression. Curr. Opin. Microbiol. **9**:229–236.

64. **Saini R, Saini S, Sharma S.** 2010. Periodontal disease linked to cardiovascular disease. J. Cardiovasc. Dis. Res. **1**:161–162.

65. **Selinger DW, et al.** 2000. RNA expression analysis using a 30 base pair resolution Escherichia coli genome array. Nat. Biotechnol. **18**:1262–1268.

66. **Serganov A, Polonskaia A, Phan AT, Breaker RR, Patel DJ.** 2006. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. Nature **441**:1167–1171.

67. **Shi Y, et al.** 1999. Genetic analyses of proteolysis, hemoglobin binding, and hemagglutination of Porphyromonas gingivalis. Construction of mutants with a combination of rgpA, rgpB, kgp, and hagA. J. Biol. Chem. **274**:17955–17960.

68. **Slakeski N, et al.** 2000. A Porphyromonas gingivalis genetic locus encoding a heme transport system. Oral Microbiol. Immunol. **15**:388–392.

69. **Socransky SS, Haffajee AD.** 2005. Periodontal microbial ecology. Periodontol. 2000 **38**:135–187.

70. **Stelzel M, et al.** 2002. Detection of Porphyromonas gingivalis DNA in aortic tissue by PCR. J. Periodontol. **73**:868–870.

71. **Tatusov RL, Galperin MY, Natale DA, Koonin EV.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. **28**:33–36.

72. **ten Broeke-Smits NJ, et al.** 2010. Operon structure of Staphylococcus aureus. Nucleic Acids Res. **38**:3263–3274.

73. **Tezuka A, Hamajima S, Hatta H, Abiko Y.** 2006. Inhibition of Porphyromonas gingivalis hemagglutinating activity by IgY against a truncated HagA. J. Oral Sci. **48**:227–232.

74. **Thomason MK, Storz G.** 2010. Bacterial antisense RNAs: how many are there, and what are they doing? Annu. Rev. Genet. **44**:167–188.

75. **Toledo-Arana, A, et al.** 2009. The Listeria transcriptional landscape from saprophytism to virulence. Nature **459**:950–956.

76. **Toledo-Arana A, Repoila F, Cossart P.** 2007. Small noncoding RNAs controlling pathogenesis. Curr. Opin. Microbiol. **10**:182–188.

77. **Tu KC, Bassler BL.** 2007. Multiple small RNAs act additively to integrate sensory information and control quorum sensing in Vibrio harveyi. Genes Dev. **21**:221–233.

78. **Tusher VG, Tibshirani R, Chu G.** 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. U. S. A. **98**:5116–5121.

79. **Valentin-Hansen P, Johansen J, Rasmussen AA.** 2007. Small RNAs controlling outer membrane porins. Curr. Opin. Microbiol. **10**:152–155.

80. **van Winkelhoff AJ, Loos BG, van der Reijden WA, van der Velden U.** 2002. Porphyromonas gingivalis, Bacteroides forsythus and other putative periodontal pathogens in subjects with and without periodontal destruction. J. Clin. Periodontol. **29**:1023–1028.

81. **Vecerek B, Moll I, Blasi U.** 2007. Control of Fur synthesis by the noncoding RNA RyhB and iron-responsive decoding. EMBO J. **26**:965–975.

82. **Walters S, Rodrigues P, Belanger M, Whitlock J, Progulske-Fox A.** 2009. Analysis of a band 7/MEC-2 family gene of Porphyromonas gingivalis. J. Dent. Res. **88**:34–38.

83. **Waters LS, Storz G.** 2009. Regulatory RNAs in bacteria. Cell **136**:615–628.

84. **Wilhelm BT, et al.** 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature **453**:1239–1243.

85. **Wurtzel O, et al.** 2010. A single-base resolution map of an archaeal transcriptome. Genome Res. **20**:133–141.

86. **Yang HW, Huang YF, Chou MY.** 2004. Occurrence of Porphyromonas gingivalis and Tannerella forsythensis in periodontally diseased and healthy subjects. J. Periodontol. **75**:1077–1083.

87. **Yoder-Himes DR, et al.** 2009. Mapping the Burkholderia cenocepacia niche response via high-throughput sequencing. Proc. Natl. Acad. Sci. U. S. A. **106**:3976–3981.

88. **Yu WH, Høvik H, Chen T.** 2010. A hidden Markov support vector machine framework incorporating profile geometry learning for identifying microbial RNA in tiling array data. Bioinformatics **26**:1423–1430.

89. **Yu WH, Høvik H, Olsen I, Chen T.** 2011. Strand-specific transcriptome profiling with directly labeled RNA on genomic tiling microarrays. BMC Mol. Biol. **12**:3.

90. **Yuan L, Rodrigues PH, Belanger M, Dunn WA Jr, Progulske-Fox A.** 2008. Porphyromonas gingivalis htrA is involved in cellular invasion and in vivo survival. Microbiology **154**:1161–1169.

91. **Yukitake H, et al.** 2011. Effects of non-iron metalloporphyrins on growth and gene expression of Porphyromonas gingivalis. Microbiol. Immunol. **55**:141–153.