Deep sequencing of *Porphyromonas gingivalis* and comparative transcriptome analysis of a LuxS mutant

**Takanori Hirano**, **David A. C. Beck**, **Donald R. Demuth**, **Murray Hackett** and **Richard J. Lamont**

1 Center for Oral Health and Systemic Disease, School of Dentistry, University of Louisville, Louisville, KY, USA
2 Department of Chemical Engineering, University of Washington, Seattle, WA, USA
3 eScience Institute, University of Washington, Seattle, WA, USA

**INTRODUCTION**

*Porphyromonas gingivalis*, a Gram-negative anaerobe, is a major pathogen in periodontal disease, one of the most common bacterial infections of humans in developed countries (Sabbah et al., 2010). Moreover, emerging evidence indicates an association between periodontal infections and serious systemic conditions such as coronary artery disease and preterm delivery of low birth weight infants (Baharak et al., 2007; Humphrey et al., 2008; Katz et al., 2009). The pathogenicity of *P. gingivalis* is multifactorial and is associated with adhesins such as fimbriae that mediate attachment to surfaces and invasion of epithelial cells; proteases such as the gingipains that degrade the structural components of the periodontium and immune effector molecules; and toxic metabolites that can induce apoptotic and necrotic cell death (Lamont and Jenkinson, 1998).

A major ecological niche of *P. gingivalis* is the subgingival crevice or periodontal pocket (Lamont and Jenkinson, 1998), and this region experiences fluctuation in oxygen levels, temperature, pH, and nutrient availability. *P. gingivalis* can sense and respond at the transcriptional level to these environmental cues (Diaz et al., 2006; Okano et al., 2006; Duran-Pinedo et al., 2007; Meuric et al., 2008; Johnson et al., 2011), although the mechanistic basis of gene regulation has not been fully defined.

*Porphyromonas gingivalis* can accumulate into complex multispecies consortia with other organisms (Kuboniwa and Lamont, 2010) resulting in the differential expression of a large percentage of its genome and proteome (Kuboniwa et al., 2009; Redanz et al., 2011). *P. gingivalis* and its community partner *S. gordonii* participate in inter-species communication based on the soluble mediator AI-2 (Chung et al., 2001; McNab et al., 2003), produced by the action of the AI-2 synthase, LuxS (Schauer et al., 2001). LuxS deficient mutants of *P. gingivalis* and *S. gordonii* co-adhere normally but do not accrete into microcolonies, indicating that AI-2 is required for *P. gingivalis-S. gordonii* community development (McNab et al., 2003). In *P. gingivalis*, LuxS regulates the expression of a number of genes involved in the uptake of hemin and inorganic iron, including *hmuR*, *fetB*, *feoB1*, and *fn*, which are negatively regulated, and *tir* and *kpg* which are positively regulated (James et al., 2006). Expression of the luxS gene is controlled by a phospho-dependent signaling pathway that converges on the transcriptional regulator CdhR (Maeda et al., 2008). Moreover, AI-2 also participates in a positive feedback loop that suppresses CdhR expression (Chawla et al., 2010). A microarray analysis of *P. gingivalis* has shown that LuxS is involved in the control of resistance to stress (Yuan et al., 2005); however, little else is known regarding the broader LuxS regulon.

High-throughput sequencing of cDNAs (deep sequencing) is a powerful technique for quantitative transcriptional profiling. As deep sequencing can detect and quantify RNA present in low levels, the technique is more sensitive than microarrays. A number of bacteria have been analyzed at the transcriptional level by deep sequencing approaches (Oliver et al., 2009; Perkins et al., 2010). As deep sequencing can detect and quantify RNA present in low levels, the technique is more sensitive than microarrays. A number of bacteria have been analyzed at the transcriptional level by deep sequencing approaches (Oliver et al., 2009; Perkins et al., 2010).
Hirano et al. Deep sequencing of P. gingivalis 2009; Yoder-Himes et al., 2009; Beck et al., 2011; Isabella and Clark, 2011) and the methodology has provided novel insights into pathogenic regulons, as well as identifying small non-coding RNAs. To analyze the P. gingivalis transcriptome and elucidate the P. gingivalis LuxS/AI-2 regulatory circuit in more detail, we employed transcriptome Sequencing by Oligonucleotide Ligation and Detection (SOLiD) deep sequencing. We applied RNA-Seq in a semi-targeted fashion as an alternative to a quantitative microarray experiment. We quantitated differences in gene expression in a luxS mutant, identified novel regions under control of this system and showed that loss of LuxS affects both AI-2 dependent signaling and non-signal dependent systems in P. gingivalis. In addition, we acquired more qualitative information about the transcriptome that would not have been possible using a targeted array approach. We have identified many putative novel small non-coding RNAs (sRNAs), and we have validated the expression of previously hypothetical genes.

RESULTS AND DISCUSSION
SEQUENCING THE P. GINGIVALIS TRANSCRIPTOME

In order to characterize the P. gingivalis transcriptome by SOLiD based RNA-Seq, RNA was prepared from three biological replicate cultures of ATCC 33277 and its ΔluxS mutant. We previously showed that AI-2 synthesis in P. gingivalis peaks in early exponential phase (James et al., 2006), therefore cultures were grown to early exponential phase (OD$_{600}$ 0.5). Reads from the RNA-Seq data were mapped to the annotated P. gingivalis ATCC 33277 genome (NCBI GI:188593544), and gene expression was quantified as reads per kilobase of coding sequence per million reads sequenced (RPKM). Expression data for each replicate were plotted against each other (Figure 1). The replicates were in good agreement with each other, with correlations ranging from 0.8 to 0.95. The number of 50 bp reads in the samples ranged from 44 to 76 million. On average, BWA software (see Materials and Methods) was able to assign 43.3% of the color space reads to the P. gingivalis chromosome. The majority of mapped reads were attributed to ribosomal RNA with the mean across conditions of 80%.

The average per base coverage is shown in Table S1 in Supplementary Material. The coverage of protein-coding regions for each sample was strongly correlated to rRNA removal efficacy and the size of the read pool. In each sample, the transcriptome penetration was such that all non-hypothetical ORFs were detected. When the replicates for each condition were pooled, every predicted ORF (Naito et al., 2008) was detected. The entire comparative gene expression profiles for WT and luxS mutant are shown in Figure S1 in Supplementary Material. Transcript abundance for WT samples sorted by RPKM ratio is shown in Table S2 in Supplementary Material. Genes for three hypothetical proteins, PGN_0215, PGN_2047, and PGN_0483 had exceptionally high reads compared to other genes. Interestingly, these genes are located in front of 16S ribosomal RNA, PGN_r0001, PGN_r0010, and PGN_r0004, respectively, and may be co-transcribed, suggesting that these gene products might be important for ribosome function. Another 16S ribosomal RNA gene, PGN_r0009 has no such hypothetical gene, indicating a specific role for these hypothetical proteins. Predictably, genes for outer membrane proteins, ribosome proteins, and cysteine proteinases had high reads. Genes encoding the major (fimA) and minor (mfa1) fimbriae also had high numbers of reads mapped.

**RESULTS AND DISCUSSION**

**SEQUENCING THE P. GINGIVALIS TRANSCRIPTOME**

In order to characterize the P. gingivalis transcriptome by SOLiD based RNA-Seq, RNA was prepared from three biological replicate cultures of ATCC 33277 and its ΔluxS mutant. We previously showed that AI-2 synthesis in P. gingivalis peaks in early exponential phase (James et al., 2006), therefore cultures were grown to early exponential phase (OD$_{600}$ 0.5). Reads from the RNA-Seq data were mapped to the annotated P. gingivalis ATCC 33277 genome (NCBI GI:188593544), and gene expression was quantified as reads per kilobase of coding sequence per million reads sequenced (RPKM). Expression data for each replicate were plotted against each other (Figure 1). The replicates were in good agreement with each other, with correlations ranging from 0.8 to 0.95. The number of 50 bp reads in the samples ranged from 44 to 76 million. On average, BWA software (see Materials and Methods) was able to assign 43.3% of the color space reads to the P. gingivalis chromosome. The majority of mapped reads were attributed to ribosomal RNA with the mean across conditions of 80%.

The average per base coverage is shown in Table S1 in Supplementary Material. The coverage of protein-coding regions for each sample was strongly correlated to rRNA removal efficacy and the size of the read pool. In each sample, the transcriptome penetration was such that all non-hypothetical ORFs were detected. When the replicates for each condition were pooled, every predicted ORF (Naito et al., 2008) was detected. The entire comparative gene expression profiles for WT and luxS mutant are shown in Figure S1 in Supplementary Material. Transcript abundance for WT samples sorted by RPKM ratio is shown in Table S2 in Supplementary Material. Genes for three hypothetical proteins, PGN_0215, PGN_2047, and PGN_0483 had exceptionally high reads compared to other genes. Interestingly, these genes are located in front of 16S ribosomal RNA, PGN_r0001, PGN_r0010, and PGN_r0004, respectively, and may be co-transcribed, suggesting that these gene products might be important for ribosome function. Another 16S ribosomal RNA gene, PGN_r0009 has no such hypothetical gene, indicating a specific role for these hypothetical proteins. Predictably, genes for outer membrane proteins, ribosome proteins, and cysteine proteinases had high reads. Genes encoding the major (fimA) and minor (mfa1) fimbriae also had high numbers of reads mapped.

**RESULTS AND DISCUSSION**

**SEQUENCING THE P. GINGIVALIS TRANSCRIPTOME**

In order to characterize the P. gingivalis transcriptome by SOLiD based RNA-Seq, RNA was prepared from three biological replicate cultures of ATCC 33277 and its ΔluxS mutant. We previously showed that AI-2 synthesis in P. gingivalis peaks in early exponential phase (James et al., 2006), therefore cultures were grown to early exponential phase (OD$_{600}$ 0.5). Reads from the RNA-Seq data were mapped to the annotated P. gingivalis ATCC 33277 genome (NCBI GI:188593544), and gene expression was quantified as reads per kilobase of coding sequence per million reads sequenced (RPKM). Expression data for each replicate were plotted against each other (Figure 1). The replicates were in good agreement with each other, with correlations ranging from 0.8 to 0.95. The number of 50 bp reads in the samples ranged from 44 to 76 million. On average, BWA software (see Materials and Methods) was able to assign 43.3% of the color space reads to the P. gingivalis chromosome. The majority of mapped reads were attributed to ribosomal RNA with the mean across conditions of 80%.

The average per base coverage is shown in Table S1 in Supplementary Material. The coverage of protein-coding regions for each sample was strongly correlated to rRNA removal efficacy and the size of the read pool. In each sample, the transcriptome penetration was such that all non-hypothetical ORFs were detected. When the replicates for each condition were pooled, every predicted ORF (Naito et al., 2008) was detected. The entire comparative gene expression profiles for WT and luxS mutant are shown in Figure S1 in Supplementary Material. Transcript abundance for WT samples sorted by RPKM ratio is shown in Table S2 in Supplementary Material. Genes for three hypothetical proteins, PGN_0215, PGN_2047, and PGN_0483 had exceptionally high reads compared to other genes. Interestingly, these genes are located in front of 16S ribosomal RNA, PGN_r0001, PGN_r0010, and PGN_r0004, respectively, and may be co-transcribed, suggesting that these gene products might be important for ribosome function. Another 16S ribosomal RNA gene, PGN_r0009 has no such hypothetical gene, indicating a specific role for these hypothetical proteins. Predictably, genes for outer membrane proteins, ribosome proteins, and cysteine proteinases had high reads. Genes encoding the major (fimA) and minor (mfa1) fimbriae also had high numbers of reads mapped.

**RESULTS AND DISCUSSION**

**SEQUENCING THE P. GINGIVALIS TRANSCRIPTOME**

In order to characterize the P. gingivalis transcriptome by SOLiD based RNA-Seq, RNA was prepared from three biological replicate cultures of ATCC 33277 and its ΔluxS mutant. We previously showed that AI-2 synthesis in P. gingivalis peaks in early exponential phase (James et al., 2006), therefore cultures were grown to early exponential phase (OD$_{600}$ 0.5). Reads from the RNA-Seq data were mapped to the annotated P. gingivalis ATCC 33277 genome (NCBI GI:188593544), and gene expression was quantified as reads per kilobase of coding sequence per million reads sequenced (RPKM). Expression data for each replicate were plotted against each other (Figure 1). The replicates were in good agreement with each other, with correlations ranging from 0.8 to 0.95. The number of 50 bp reads in the samples ranged from 44 to 76 million. On average, BWA software (see Materials and Methods) was able to assign 43.3% of the color space reads to the P. gingivalis chromosome. The majority of mapped reads were attributed to ribosomal RNA with the mean across conditions of 80%.

The average per base coverage is shown in Table S1 in Supplementary Material. The coverage of protein-coding regions for each sample was strongly correlated to rRNA removal efficacy and the size of the read pool. In each sample, the transcriptome penetration was such that all non-hypothetical ORFs were detected. When the replicates for each condition were pooled, every predicted ORF (Naito et al., 2008) was detected. The entire comparative gene expression profiles for WT and luxS mutant are shown in Figure S1 in Supplementary Material. Transcript abundance for WT samples sorted by RPKM ratio is shown in Table S2 in Supplementary Material. Genes for three hypothetical proteins, PGN_0215, PGN_2047, and PGN_0483 had exceptionally high reads compared to other genes. Interestingly, these genes are located in front of 16S ribosomal RNA, PGN_r0001, PGN_r0010, and PGN_r0004, respectively, and may be co-transcribed, suggesting that these gene products might be important for ribosome function. Another 16S ribosomal RNA gene, PGN_r0009 has no such hypothetical gene, indicating a specific role for these hypothetical proteins. Predictably, genes for outer membrane proteins, ribosome proteins, and cysteine proteinases had high reads. Genes encoding the major (fimA) and minor (mfa1) fimbriae also had high numbers of reads mapped.
HYPOTHETICAL PROTEINS

We identified 854 hypothetical genes, which were expressed with at least average read counts of 10 in either the WT or \(\Delta luxS\) mutant samples (Table S3 in Supplementary Material). In both strains, three genes, PGN_2047, PGN_0215, and PGN_0438, which are fully conserved at the amino acid level, had large numbers of reads mapped. The corresponding predicted proteins are annotated as cytoplasmic proteins with unknown function. Moreover, the nucleotide sequences \(5^{\prime}\) of the 43rd nt are identical to the sequence spanning the region \(5^{\prime}\) upstream to the first 32 nucleotides of PGN_r0009 16S rRNA. Hypotheticals PGN_0217, PGN_0440, PGN_2049, and PGN_1398, also fully conserved at the amino acid level, were highly expressed in both strains. These results suggest that several of the relatively smaller proteins might have physiologically important functions as yet to be determined.

IDENTIFICATION OF SMALL (s)RNAs

Small RNAs (sRNA) are widespread in bacteria and the significance of sRNAs in gene regulation is increasingly recognized (Storz et al., 2011). Several potential sRNAs in \(P. gingivalis\) strain W83 have been identified by transcriptome analysis but the \(5^{\prime}\) and \(3^{\prime}\) ends have not been determined experimentally (Hovik et al., 2012). Analysis of the SOLiD data (see Materials and Methods) revealed over 2000 intergenic transcripts that will include sRNAs (Table S4 in Supplementary Material). To identify sRNAs and define their \(5^{\prime}\) and \(3^{\prime}\), we employed a modified rapid amplification of cDNA ends (RACE) technique. Firstly we validated the modified RACE technique in \(P. gingivalis\) cDNA ends (RACE) technique. Secondly we focused on IGRs that showed reading counts higher than 1000 in the WT, did not appear as UTR of orthologs in the \(P. gingivalis\) strain W83\(^1\), and were not adjacent to rRNA genes. Thirteen IGRs were examined by RACE and 11 sRNAs were identified with lengths between 50 and 783 bases in 10 IGRs (Table 1), none of which are listed in rnam database. Two stable fragments were identified that started from the same position in IGR0247 and ended at 268 and 218 bases, suggesting sRNA processing or read-through. The fragments in IGR0960 and IGR1662 located as part of the 3′-UTR but could be processed to the stable sRNA fragment since these two miRNA have unusually long 3′-UTR regions reading to 300 and 700 bases beyond the translational stop, respectively. The forward strand in IGR0960 also encodes trNA-Agr-4. tRNA-Lys-1 in IGR1524 has been predicted to be 73 and 76 bases long in the rnam database web site and the Oralgen database web site, respectively, however a 221 base fragment including the 76 bases segment was detected by RACE.

COBALAMIN RIBOWITCH

Riboswitches are regulatory elements that modulate expression of the downstream gene or operon (Breaker, 2011). We identified a potential cobalamin-type riboswitch sequence cbl-1 (BLAST e-value of 9.71e-21) that was located between PGN_0703 and PGN_0704 and is predicted to be transcribed in the same direction as PGN_0704. PGN_0704 encodes a BtuB homolog, an outer membrane protein that is required for the uptake of vitamin B12 (Storz et al., 2011).

---

Table 1 | List of sRNAs identified by RACE.

| ID in SOLiD | Start \(^2\) | Stop | Length (bases) | Coding \(^3\) | IGR \(^2\) | S′ Gene | Coding | 3′ Gene | Coding | Notes |
|------------|------------|------|---------------|-------------|-----------|---------|-------|---------|-------|-------|
| ID158      | 146754     | 146582 | 173           | R           | IGR0121   | F       | PGN_0141 | F       | PGN_0142 | F       | RF01693 (Bacteroidales-1) |
| ID266      | 254941     | 255103 | 163           | F           | IGR0204   | F       | PGN_0235 | F       | PGN_0236 | F       |          |
| ID317      | 307824     | 308091 | 268           | F           | IGR0247   | F       | PGN_0286 | R       | mfa1     | F       |          |
| ID317      | 307824     | 308042 | 218           | F           | IGR0247   | F       | PGN_0286 | R       | mfa1     | F       |          |
| ID319      | 311532     | 311483 | 50            | R           | IGR0249   | F       | PGN_0286 | F       | PGN_0289 | F       |          |
| ID989      | 821427     | 820645 | 783           | R           | IGR0648   | F       | PGN_0748 | F       | PGN_0749 | F       |          |
| ID321      | 1229785    | 1229553 | 233         | R           | IGR0960   | F       | PGN_1103 | F       | PGN_1104 | R       | tRNA-Arg-4 on F-strand |
| ID1520     | 141017     | 1409844 | 274         | R           | IGR1099   | R       | PGN_1262 | R       | PGN_1263 | R       |          |
| ID2004     | 1865668    | 1865314 | 355         | R           | IGR1456   | R       | PGN_1669 | R       | PGN_1670 | F       |          |
| ID2089     | 1953887    | 1954115 | 229         | F           | IGR1520   | F       | PGN_1739 | F       | PGN_1740 | R       |          |
| ID2261     | 2123160    | 2123379 | 220         | N/D         | IGR1662   | F       | hagB     | F       | PGN_1905 | R       |          |
| ID2381     | 2238240    | 2238533 | 294         | F           | IGR1746   | F       | PGN_1999 | R       | PGN_2000 | R       |          |
| ID2096     | 1957637    | 1957857 | 221         | F           | IGR1524   | F       | PGN_1744 | R       | PGN_1745 | F       | tRNA-Lys-1 on F-strand |

---

1. From Table S4 in Supplementary Material.
2. Designation from http://www.oralgen.lanl.gov
3. F, forward; R, reverse; N/D, not determined Figure.

Frontiers in Cellular and Infection Microbiology www.frontiersin.org June 2012 | Volume 2 | Article 79 | 3
membrane cobalamin receptor protein for which the riboswitch mechanism has been well characterized in other organisms (Nahvi et al., 2004). Since riboswitches are long 5’ non-coding regulatory regions of the downstream coding region, we examined co-transcription of cbl-1 with PGN_0704, by reverse transcription PCR (RT-PCR) using specific primer sets binding to the riboswitch region and the downstream protein-coding region. RT-PCR showed cbl-1 was co-transcribed with downstream coding regions (Figure 2B). Hereafter PGN_0704 is designated as btuB, and the transcriptional start site for btuB was determined by 5’-RACE. The potential 5’-end position from three independent cbl-1-btuB 5ʼ-RACE clones was found at 430 nt upstream from ATG translational start site of btuB (Figure 2A).

The structure of riboswitches can be modulated upon binding to a chemical, usually one of metabolites of the pathway encoded by the downstream genes. The cobalamin transporter BtuB is part of the vitamin B12 synthetic pathway, and in E. coli and Salmonella one of the metabolites of this pathway, Coenzyme B12, binds to the riboswitch region of btuB mRNA (Ravnum and Andersson, 1997; Nahvi et al., 2004). Subsequent structural changes lead to susceptibility to cleavage causing transcriptional arrest, and/or to inhibition of binding to the ribosome. Hence, we examined the stability of cbl-1-btuB at the transcriptional and the translational levels in P. gingivalis. To assess the effects of Coenzyme B12, cells were grown in the presence or absence of 100 μM Coenzyme B12. qRT-PCR using the same primer sets as for RT-PCR was carried out to show the relative amount of cbl-1 riboswitch transcripts compared to those of the btuB coding region (Figures 2C-E). qRT-PCR for the cbl-1 region showed the highest levels were at early log phase with a gradual decrease to almost 60% during culture to OD600 1 (Figure 2C). This signature as well as copy numbers at each sampling point were not significantly changed by the addition of Coenzyme B12. btuB transcripts levels were the highest at early log phase and then similarly decreased in the absence of Coenzyme B12 (Figure 2D). However, culture in the presence of Coenzyme B12 resulted in a significant decrease in the level of btuB mRNA (Figure 2D). The ratios of cbl-1 to btuB were similar in the absence of Coenzyme B12, but significantly higher in the presence of Coenzyme B12 through the growth phase (Figure 2E). These results indicate that under Coenzyme B12 limitation, transcription of the cbl-1-btuB region is most active at early log phase (OD600 0.3) followed by a gradual transcriptional stalling over the growth curve. In contrast, in the presence of Coenzyme B12, although promoter activity of cbl-1-btuB regulator is unchanged, transcription of the btuB coding region stalls more frequently upon Coenzyme B12 binding to cbl-1 riboswitch region.

To verify btuB expression at the translational level, BtuB protein was C-terminally tagged with FLAG. Western blotting of whole cell lysates with anti-FLAG monoclonal antibody demonstrated reduced BtuB protein levels in the presence of Coenzyme B12 (Figure 3A). Levels of FLAG-tagged BtuB decreased to almost undetectable at OD 1.0. Western blotting with antibodies to the Mfa1 fimbrial protein of P. gingivalis was performed as a loading control (data not shown). FLAG/Mfa1 ratios (Figure 3B) confirmed the inhibition of BtuB expression in the presence of Coenzyme B12, and the decrease in BtuB as the cells progress through the growth curve. These results corroborate the concept that cbl-1 upstream of btuB acts as a conventional cobalamin-type riboswitch in P. gingivalis.

**EFFECT OF luxS MUTATION ON THE P. GINGIVALIS TRANSCRIPTOME**

LuxS AI-2 synthase is responsible for the production of the AI-2 family of signaling molecules, and the enzyme is a component of the activated methyl cycle (Schauder et al., 2001); however, the LuxS regulon in P. gingivalis has not been fully defined. Comparative transcriptome analysis of WT and ΔluxS strains identified 57 ORFs as differentially expressed with a p-value cutoff of 0.05 (Figure S2 in Supplementary Material). The differentially regulated ORFs are depicted by red dots in Figure 1 from which it is clear that they are well reproduced by the replicates for each condition with inter-replicate correlations better than all ORFs, ranging from 0.86 to 0.98. Of these 57, 17 were annotated as hypothetical proteins, and 15 of these were flagged as possessing signal peptides by JGI’s annotation pipeline (Markowitz et al., 2008), with 10 as transmembrane domain containing proteins.

Nearly half of the differentially abundant ORFs identified were found in clusters of two or more neighbor or near-neighbor ORFs in the genome that were similarly differentially abundant. The largest such cluster spanned six ORFs from PGN_0553 to PGN_0558, encompassing the hmu hemin uptake operon, which was up-regulated in the ΔluxS mutant strain. This is consistent with earlier work showing that transcription of hmuB (PGN_0557) is elevated in the same ΔluxS mutant strain as compared to WT (James et al., 2006). Also up-regulated in the ΔluxS mutant were the potential operons spanning PGN_1791 to PGN_1790, and PGN_1472 to PGN_1468. The former region encodes a flavodoxin FlmA and a hypothetical protein, respectively. The latter region comprises genes involved in diverse functions such as tRNA-guanine transglycosylase, probable YigP/YigQ family permease, ssrA RNA binding protein, along with dpp and lipoyl synthase.

In the P. gingivalis chromosome, luxS (PGN_1474) is encoded on the negative strand, preceded and co-transcribed with pfs which encodes an enzyme that synthesizes the substrate for LuxS. mRNA for pfs was differentially abundant, over threefold higher in the mutant. This result suggests the existence of a regulatory feedback mechanism that senses the degree of production of AI-2 and acts to maintain optimal levels.

Down-regulated genes contained insertion sequences: ISPg4 (one), and hypothetical proteins (10). In addition, PGN_0481 and PGN_0482, encoding a hypothetical protein and an immunoreactive protein respectively, and PGN_1403 to PGN_1401, involved in arginine and proline metabolic pathways, were down-regulated.

No systematic bias in COG (Tatusov et al., 2000), pfam (Finn et al., 2010), or KEGG (Kanehisa and Goto, 2000) classifications was observed for the differentially abundant ORFs. We mapped all of the detected ORFs onto the KEGG pathway map. The depictions of these pathways can be found in the Supplementary Material (Figures S3A–C in Supplementary Material) with genes shown in red and green (up and down in the ΔluxS mutant relative to WT, respectively) and the statistical significance as enzyme background (dark to light, increasing). The streptomycin biosynthesis pathway (KEGG pathway pgn00521, Figure S3A in Supplementary Material) had several enzyme transcripts that were more abundant.
Deep sequencing of P. gingivalis

Hirano et al.

A

| 0.5 kb- | 1.0 kb- | genomic cDNA No RT | genomic cDNA No RT | genomic cDNA No RT | genomic cDNA No RT |
|--------|--------|---------------------|---------------------|---------------------|---------------------|
| cbl-1  | btuB   | cbl-1-btuB          |

AAAGAATATT TATCTAAGAG ATATCGGTTAT GATTAACAAA CATAATACGA
K E Y L S R W C D Stop
TGATCTTAT TTGGCAAGAT GCAGATTGTA CTGTGCCCTA CAGTCTATA

AAAGCTATAC TTTTGCCCTT GAAAAAGCTG AAGTACCCGC GGTGCAATTC CGGAGCTGAG CCCTCAGCTG

FIGURE 2 | Transcriptional organization and regulation of the cbl-1-btuB (PGN_0704) region in the P. gingivalis chromosome. (A) Based on 5'-RACE, the transcriptional start site for cbl-1-btuB mRNA was mapped to a G (in bold italic and labeled +1) 430 bp upstream of the ATG start codon of btuB. Primers in the cbl-1 region used for RTPCR in (B) and qRTPCR in (C–E) are underlined. (B) RTPCR of genomic DNA or cDNA with primers amplifying cbl-1, btuB or cbl-1, and btuB. The presence of an amplicon with cDNA demonstrates co-transcription of cbl-1 and btuB. The No RT lanes are controls without reverse transcriptase. (C) cbl-1 levels were not affected by 100 µM Coenzyme B12. (D) btuB transcripts were significantly reduced by culture with 100 µM Coenzyme B12. (E) Ratio of cbl-1 to btuB mRNA copy number. Results in (C–E) are means with standard deviation of three biological replicates (n = 9).
in the ΔluxS mutant, although not all were statistically significant. One product of this pathway is L-rhamnose (pgn00520), a saccharide that has been shown to be important for virulence in some oral bacteria (Henderson et al., 2009). *P. gingivalis* has been reported as streptomycin-resistant (Loos et al., 1990). However, since *P. gingivalis* does not have key enzymes downstream from dTDP-L-rhamnose in the pathway, the cell does not produce streptomycin. Deletion of dTDP-L-rhamnose biosynthesis genes in the rml operon (PGN_0546 to PGN_0549) is lethal in *P. gingivalis* (Shibata et al., 1999). dTDP-L-rhamnose is a component of lipopolysaccharide (LPS) in *P. gingivalis*, and thus another role of LuxS may be to regulate LPS biogenesis. In the lipoic pathway (pgn00785, Figure S3B in Supplementary Material), the transcript for LipA (Lipoic acid synthase, PGN_1468) which is important for biosynthesis of cofactors, prosthetic groups, and carriers including lipoate (Kondo et al., 2010), was observed to be more abundant in the ΔluxS mutant. However, most of the KEGG pathway reconstructions had limited numbers of changed transcripts and exhibited abundance change trends, e.g., pgn00550, the peptidoglycan synthesis pathway (Figure S3C in Supplementary Material). The lack of clear trends may be a consequence, at least in part, of the dual function of LuxS in signaling and metabolism (see below).

**VERIFICATION OF DIFFERENTIALLY EXPRESSED GENES AND COMPLEMENTATION BY DPD**

To corroborate the sequencing results, expression of a subset of genes was determined by qRT-PCR analysis (Figure 4). RNA samples were prepared from WT and ΔluxS cultures grown in the same conditions as those for the deep sequencing and converted to cDNA for qRT-PCR. Although the magnitude of fold differences was different between the two techniques, the up- or down-regulated profiles for the genes tested were concordant. In order to distinguish between differential regulation due to a defect in AI-2 signaling or disruption of the methyl cycle, exogenous 4,5-dihydroxy-2,3-pentanedione (DPD, AI-2 precursor) was introduced into the mutant culture, and complementation of the phenotype examined (Figure 5). WT and luxS mutant strains at OD600 = 0.3 were cultured anaerobically in the presence or absence of 4 μM DPD and total RNAs were purified for qRT-PCR. Expression levels of PGN_0482 and PGN_1403 in the luxS mutant were not restored by exogenous DPD, suggesting that regulation of these genes is independent of AI-2 signaling (Figures 5A,B). In contrast, expression levels of PGN_1791 and hmuR in the luxS mutant were significantly (*p* = 0.0001 and *p* = 0.0012, respectively) reduced by culture with DPD for 60 min (Figures 5C,D). These results indicate that LuxS has both an AI-2 signaling function, and a metabolic function through the activated methyl cycle, with independent regulons. Similar AI-2 dependent and independent functions for LuxS have also been shown in *Streptococcus suis* (Cao et al., 2011) and the potential dual role of LuxS may partially explain why no common theme has emerged for the larger LuxS regulon across bacterial species, and indeed different bacteria may utilize LuxS for different functions (Fong et al., 2003; Zhu et al., 2003; Vendeville et al., 2005; Doherty et al., 2006; Lebeer et al., 2007).
et al., 2007; Siller et al., 2008; Learman et al., 2009). In *P. gingivalis*, a mutation in the *luxS* gene affects hemin and iron acquisition, protease activity, and also impacts stress response pathways (Burgess et al., 2002; Yuan et al., 2005; James et al., 2006). Moreover, hemin levels in *P. gingivalis* can affect virulence (Marsh et al., 1994; Genco, 1995; Genco et al., 1995) and resistance to oxidative stress (Diaz and Rogers, 2004), and the HmuR protein is required for maximal community formation by *P. gingivalis* with *F. nucleatum* and *S. gordonii* (Kuboniwa et al., 2009). Hence AI-2 signaling has the potential to impact a number of important pathogenic properties of *P. gingivalis*.

The non-signal effects of LuxS mutation may be related to disruption of the activated methyl cycle. As previously described (James et al., 2006), the ΔluxS mutant was generated by insertion of permF::ermF fragment into the luxS coding region and it is unlikely that the gene disruption would have a positive polar effect (i.e., up-regulation) on neighboring gene transcription, especially on the region upstream of the insertion site. Given that the ΔluxS strain lacks the ability to catalyze the reaction to convert the LuxS substrate, SRH, to homocysteine and DPD, we predict that SRH would accumulate in the cell. Since Pfs is required for the conversion of S-adenosyl homocysteine, SAH, to SRH, this also

**FIGURE 5** | Exogenous DPD complementation of the ΔluxS phenotype. Levels of differentially expressed genes PGN_0482 (A), PGN_1403 (B), PGN_1791 (C), and hmuR (D), in the ΔluxS mutant strain were examined by qRT-PCR following incubation with 4 μM of synthetic DPD for the indicated times. Results are representative of three independent assays. *p*-values were calculated by unpaired *t*-test. Error bars represent standard deviations (*n* = 3).
**Deep sequencing of *P. gingivalis***

Hirano et al.

CONCLUSION

The deep sequencing data provided insights into global regulation in *P. gingivalis*. A small percentage of the total transcriptome was regulated by LuxS, nonetheless the sequencing and analyses were sufficiently sensitive for detection of these differentially abundant transcripts. As the total data set of comparisons is orders of magnitude greater, this demonstrates the utility of RNA-Seq for the study of systems, like this one, where very subtle changes in transcript abundance on a global scale can have biologically important effects on pathogenic potential or specific regulatory or metabolic pathways. Furthermore, the results validated the involvement of LuxS in regulation of the *hmu* operon described previously, giving additional weight to the newly recognized components of the LuxS regulon. Several novel sRNAs were identified in *P. gingivalis* which are likely to be important components of gene regulatory pathways.

MATERIALS AND METHODS

**BACTERIAL STRAINS AND RNA PREPARATION FOR SEQUENCING**

Wild-type (WT) *P. gingivalis* ATCC 33277 and isogenic ΔluxS (James et al., 2006) were used in this study. Bacteria were cultured in trypticase soy broth (TSB) supplemented with yeast extract (1 mg ml⁻¹), hemin (5 μg ml⁻¹), and menadione (1 μg ml⁻¹), anaerobically, at 37°C. Overnight cultures were diluted in TSB to OD₆₀₀ 0.25 and grown to OD₆₀₀ 0.5. Three biological replicates were prepared for each strain. RNA was isolated by Trizol according to the manufacturer’s instructions (Invitrogen). RNA was treated with Turbo-DNA free (Ambion) at 37°C for 1 h, and then purified by an RNeasy mini kit (Qiagen). 16S and 23S rRNAs were removed by an Microb Express Bacterial mRNA Enrichment Kit according to the manufacturer’s instructions (Ambion). Double-stranded cDNAs were synthesized by a SuperScript Double-stranded cDNA synthesis kit according to the manufacturer’s instructions (Invitrogen).

**SOLiD SEQUENCING AND ANALYSIS**

After amplification, beads containing clonally amplified template DNA were P2-enriched and extended with a bead linker by terminal transferase. The quantity of the beads was determined using a NanoDrop ND1000 spectrophotometer. Approximately 96 M beads of each enriched library were deposited on slides and sequenced using “sequencing by ligation” chemistry and 50 bp protocol on a SOLiD v3 plus sequencer (Applied Biosystems, CA, USA) at the Interdisciplinary Centre of Biotechnology, University of Florida. Results were obtained as good and best beads as color space FASTQ files. Each library was sequenced using 1/4th region of the slide. Figure S4 in Supplementary Material shows an example of the raw reads for all replicates, prior to normalization, for the luxS transcript (PGN_1474).

**RNA-SEQ ALIGNMENT AND COVERAGE**

The 50 base unpaired SOLiD short reads were aligned in color space to the *P. gingivalis* genome scaffold using BWA version 0.5.8 using default parameters (Li and Durbin, 2009) This process began by combining the separate color space read and quality files from the sequencer into a single color space FASTQ file with the “solid2fastq.pl” script distributed with BWA. Next, the *P. gingivalis* scaffold was indexed by BWA, and the alignments were generated. The SAMtools software was used to generate a “pileup” file from the BWA output (Li et al., 2009). Finally, the pileup data were loaded in a MySQL database and the reads mapped per ORF. Statistics were generated by a standard SQL “SUM over GROUP BY” operation.

Differential abundance analyses were computed using R version 2.11.1 (R Development Core Team, 2010). In order to normalize the samples prior to differential analysis, we applied “quantile normalization,” to assure the reads mapped per ORF distribution had the same empirical distribution across all replicates and experiments (Bolstad et al., 2003). The quantile normalization was performed with the “normalizeBetweenArrays” function of LIMMA (Smyth et al., 2003). In addition, we calculated RPKMs from the normalized reads mapped per ORF as follows: the quantile normalized number of reads mapped was divided by the size...
of each gene in kb and the count of the millions of reads mapped in the quantile normalized sample (Mortazavi et al., 2008).

To evaluate the significance of differential abundance between experiments we used a t-test with the Welch approximation to account for unequal variances. The p-values derived from the t-statistics were corrected for multiple hypothesis testing using the R package "Idfrtool" (Strimmer, 2008). This package calculates Storey’s q-values (Storey, 2003) and Efron's local false discovery rate (IIdr; Efron et al., 2001) from a set of p-values.

The sequence data have been deposited in the GEO repository2, accession numbers GSM844728–GSM844733.

### QUANTITATIVE REAL-TIME PCR

Total RNA was isolated from three independent cultures at OD\textsubscript{600} 0.5. RNA was converted to cDNA with an iScript CDNA synthesis kit (Bio-Rad). qRT-PCR was performed by StepOne plus by the ∆∆Ct method using 16S rRNA as an internal control. Three experimental replicates were analyzed for each biological sample. Primers used are listed in Table S5 in Supplementary Material. Data analysis was performed using both DataAssist v3.0 and StepOne software v2.2.2 from Applied Biosystems.

### ΔluxS COMPLEMENTATION BY SYNTHETIC DPD

Overnight cultures of WT and ΔluxS strains were diluted in TSB and the cells were grown to OD\textsubscript{600} = 0.3 to avoid accumulation of endogenous AI-2. The cells were washed once in PBS and resuspended in TSB supplemented with 4 µM synthetic DPD (OMM Scientific). The cells were further cultured anaerobically and samples collected after 5, 30, and 60 min. Total RNAs were purified for qRT-PCR.

### tRNA IDENTIFICATION

Reads mapped per base “pileup” created with SAMtools from the BWA alignments was used to create a list of all possible intergenic non-coding RNAs as follows. A loop over the base pair positions in the genome was used to scan for continuous stretches of sequence with the running average number of reads mapped greater than a minimum determined by the size of each dataset. In addition, the stretches had to be within 500 bp upstream of a coding region and 100 bp downstream. All loci within 1 kb of each potential ncRNA were flagged for further manual inspection. The potential ncRNAs were assigned unique identifiers based on their position in the genome. For each ncRNA stretch, the RPKM was computed as abundance, and log ratios between conditions were calculated in the same manner as for the ORFs.

tRNAs showing RPKM reading greater than 1000 were chosen for further analysis. Modified RACE for simultaneous identification of 5′- and 3′-end was performed (Britton et al., 2007). Total RNAs were purified from the cells grown in the same condition as for SOLID analysis. Five µg of total RNAs were treated with 20 U of Tobacco Acid Phosphatase (TAP; Epicenture Biotechnologies) in 1xTAP buffer at 37 C for 1 h. Two µg of TAP-treated RNA were self-ligated by 40 U of RNA ligase I (New England Biolabs) at 17 C overnight. Reverse transcription was performed by using either random primer or IGR-specific forward or reverse primer with the High Capacity cDNA reverse transcription kit (Life Technologies). Nested PCR was carried out by using two sets of PCR primers (Table S5 in Supplementary Material) with GoTaq Green master mix (Promega). Fragments specifically amplified from TAP-treated RNA were TA-cloned into the TOPO-pCR2.1 vector for sequencing (Life technologies).

### 5′ RACE

Total RNA was isolated using Trizol from WT culture grown to OD\textsubscript{600} 0.5 followed by RNaseasy mini kit including DNase I treatment to remove genomic DNA. The transcriptional start site was determined by 5′-RACE using the FirstChoice RLM-RACE kit (Ambion). Primers used for nested PCR are listed in Table S5 in Supplementary Material. The fragments were ligated into pCR2.1 TOPO vector (Invitrogen) and the ligated DNA was transformed into E. coli Top10 (Invitrogen). The inserts were confirmed by colony PCR using M13 primers and sequenced by the University of Louisville sequencing facility.

### TRANSCRIPTIONAL AND TRANSLATIONAL ANALYSIS OF btuB

C-terminally in-frame FLAG-tagged btuB (PGN_0704) was constructed in P. gingivalis. For efficient exposure of the flag tag sequence from the BtuB protein structure, a linker sequence (GAGAGA) was introduced in between BtuB and three tandem repeats of flag sequences (FLAGx3). FLAGx3-tag and ermF alleles were amplified using genomic DNA from a laboratory stock strain as a template to make GAGAGA::FLAGx3::ermF. 700 bases of upstream and downstream of the 3′ end of btuB coding region were amplified using P. gingivalis genomic DNA. The resulting three fragments were fused by PCR and electroprotected into electrocompeptent cells. Transformants were selected for ermF insertion on blood agar plates containing 5 µg/ml erythromycin and 150 µg/ml gentamycin. The insertion was confirmed by both PCR and sequencing of the recombined region.

To examine the regulation of BtuB, the P. gingivalis strain harboring BtuB::FLAGx3 was inoculated at OD\textsubscript{600} 0.2 and incubated anaerobically in the presence or absence of 100 µM of Coenzyme B12 (Sigma-Aldrich). Total RNA samples and whole cell lysates were prepared at OD\textsubscript{600} 0.3, 0.5, 0.7, and 1.0. qRT-PCR was performed to examine the transcriptional level of both the 5′ leader region and the coding region of btuB. For quantifying translational levels of BtuB, lysates from equal cell numbers were Western blotted and probed with anti-FLAG monoclonal antibody (Sigma-Aldrich) or anti-Mfa antibody (Park et al., 2005) followed by probing with HRP conjugated secondary antibody (Cell signaling). Signal was illuminated by the Pierce ECL kit (Thermo) and the bands were quantified by Image lab software on a ChemiDoc™XRS + system (Bio-Rad).

### ACKNOWLEDGMENTS

We thank Erik L. Hendrickson for help with the pathway analysis, and Fred Taub for database programming. This work was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system, supported in part by the University of Washington eScience Institute. The support of the NIDCR through DE14605 (Donald
SUPPLEMENTAL MATERIAL

The Supplemental Material for this article can be found online at http://www.frontiersin.org/Cellular_and_Infection_Microbiology/10.3389/fcimb.2012.00079/abstract

Table S1 | RNA-Seq coverage statistics by RPKM.

Table S2 | Transcript abundance in WT samples sorted by RPKM.

Table S3 | Hypothetical proteins expressed in either WT or ΔluxS strain.

Table S4 | Intergenic transcripts sorted by WT reads (large to small).

Table S5 | Primers used in this study.

REFERENCES

Bahekar, A. A., Singh, S., Saha, S., Molnar, J., and Arora, R. (2007). The prevalence and incidence of coronary heart disease is significantly increased in periodontitis: a meta-analysis. Am. Heart J. 154, 830–837.

Beck, D. A., Hendrickson, E. L., Vorobec, A., Wang, T., Lin, S., Kalyuzhnaya, M. G., Lidstrom, M. E., Hackert, M., and Chistoserdova, L. (2011). An integrated proteomics/transcriptomics approach points to oxygen as the main electron sink for methanol metabolism in Methylobacterium mobilis. J. Bacteriol. 193, 4758–4765.

Boestad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19, 185–193.

Breakey, R. R. (2011). Prospects for riboswitch discovery and analysis. Mol. Cell 43, 867–879.

Britton, R. A., Wen, T., Schaefer, L., Pellegri, O., Ucker, W. C., Matzy, N., Tobin, C., Daou, R., Szyk, J., and Condon, C. (2007). Maturation of the 5’ end of 16S rRNA by the essential ribonuclease YtfK/RnrNaseJ1. Mol. Microbiol. 63, 127–138.

Burgess, N. A., Kirke, D. E., Williams, P., Winzer, K., Hardie, K. R., Meyers, N. L., Aduse-Opoku, J., Curtis, M. A., and Camara, M. (2002). LuxS-dependent quorum sensing in Porphyromonas gingivalis modulates protease and haemagglutinin activities but is not essential for virulence. Microbiology 148, 763–772.

Can, M., Peng, F., Wang, C., Zheng, F., Li, M., Liao, H., Mao, Y., Pan, X., Wang, J., Hu, D., Hu, F., and Tang, J. (2011). Functional definition of LuxS, an autinducer-2 (AI-2) synthase and its role in full virulence of Streptococcus suis serotype 2. J. Microbiol. 49, 1000–1011.

Chawla, A., Hirano, T., Bainbridge, B. W., Demuth, D. R., Xie, H., and Lamont, R. J. (2010). Community signalling between Streptococcus gordonii and Porphyromonas gingivalis is controlled by the transcriptional regulator CdhR. Mol. Microbiol. 78, 1510–1522.

Chung, W. O., Park, Y., Lamont, R. J., McNab, R., Barbieri, B., and Demuth, D. R. (2011). Signaling system in Porphyromonas gingivalis based on a LuxS protein. J. Bacteriol. 183, 3903–3909.

Diaz, P. L., and Rogers, A. H. (2004). The effect of oxygen on the growth and physiology of Porphyromonas gingivalis. Oral Microbiol. Immunol. 19, 88–94.

Diaz, P. J., Slakseri, N., Reynolds, E. C., Morona, R., Rogers, A. H., and Kolenbrander, P. E. (2006). Role of oxyS in the oral anaerobe Porphyromonas gingivalis. J. Bacteriol. 188, 2454–2462.

Doherty, N., Holdman, T. M., Quzi, S. N., Williams, P., and Winzer, K. (2006). Functional analysis of luxS in Staphylococcus aureus reveals a role in metabolism but not quorum sensing. J. Bacteriol. 188, 2885–2897.

Duncan-Pinedo, A. E., Nishikawa, K., and Duncan, M. J. (2007). The RprY response regulator of Porphyromonas gingivalis. Mol. Microbiol. 64, 1061–1074.

Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001). Empirical bayes analysis of a microarray experiment. J. Am. Stat. Assoc. 96, 1151–1160.

Finn, R. D., Mistry, I., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R., and Bateman, A. (2010). The Pfam protein families database. Nucleic Acids Res. 38, D211–D222.

Fong, K. P., Gao, L., and Demuth, D. R. (2003). luxS and arcB control aerobic growth of Actinobacillus actinomycetemcomitans under iron limitation. Infect. Immun. 71, 298–308.

Genco, C. A. (1995). Regulation of hemin and iron transport in Porphyromonas gingivalis. Adv. Dent. Res. 9, 41–47.

Genco, C. A., Simpson, W., Fong, R. Y., Egal, M., and Odusanya, B. M. (1995). Characterization of a Tn4351-generated hemin uptake mutant of Porphyromonas gingivalis: evidence for the coordinate regulation of virulence factors by hemin. Infect. Immun. 63, 2459–2466.

Henderson, B., Curtis, M., Seymour, R., and Donos, N. (2009). Periodontal Medicine and Systems Biology. Oxford: John Wiley & Sons-Blackwell.

Humphrey, L. L., Fu, R., Buckley, D. I., Freeman, M., and Helfand, M. (2008). Periodontal disease and coronary heart disease incidence: a systematic review and meta-analysis. J. Gen. Int. Med. 23, 2079–2089.

Isabella, V. M., and Clark, V. L. (2011). Deep sequencing-based analysis of the anaerobic stimulation in Neisseria gomorrieae. BMC Genomics 12, 51. doi:10.1186/1471-2164-12-51.

Johnson, N. A., McKenzie, R. M., and Fletcher, H. M. (2011). The bcp gene in the bcp-recA-vimS-vimE-vimF operon is important in oxidative stress resistance in Porphyromonas gingivalis W83. Mol. Oral Microbiol. 26, 62–72.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30.

Kat, J., Chigeni, N., Shiverick, K. T., and Lamont, R. J. (2009). Localization of P. gingivalis in preterm delivery placenta. J. Dent. Res. 88, 575–578.

Kondo, Y., Ohara, N., Sato, K., Yoshimura, M., Yukitake-Hatai, N., Fujwara, T., and Nakayama, K. (2010). Tetrazoliotetide repeat protein-associated proteins contribute to the virulence of Porphyromonas gingivalis. Infect. Immun. 78, 2846–2856.
