DNA Research, 2018, 25(2), 207–215
doi: 10.1093/dnares/dsx050
Advance Access Publication Date: 2 December 2017
Full Paper

OXFORD

Full Paper

Small RNA repertoires and their intraspecies variation in *Aggregatibacter actinomycetemcomitans*

Yuichi Oogai[1], Yasuhiro Gotoh[2], Yoshitoshi Ogura[2], Miki Kawada-Matsuo[1], Tetsuya Hayashi[2], and Hitoshi Komatsuzawa[1,*]

[1]Department of Oral Microbiology, Kagoshima University Graduate School of Medical and Dental Sciences, 8-35-1, Sakuragaoka, Kagoshima 890-8544, Japan, and [2]Department of Bacteriology, Faculty of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

*To whom correspondence should be addressed. Tel. +81-99-275-6150. Fax. +81-99-275-6158.
Email: hkomatsu@dent.kagoshima-u.ac.jp

Edited by Dr. Katsumi Isono

Received 6 April 2017; Editorial decision 5 November 2017; Accepted 15 November 2017

Abstract

Aggregatibacter actinomycetemcomitans is a major periodontal pathogen that has several virulence factors such as leukotoxin and cytolethal distending toxin. Although the genes responsible for virulence have been identified, little is known about their regulatory mechanisms. Small RNA (sRNA) has been recognized as an important factor for gene regulation. To identify new regulatory mechanisms via sRNA in *A. actinomycetemcomitans* HK1651, we performed a systematic search for sRNAs by RNA-seq and identified 90 intergenic region sRNAs and 30 antisense sRNAs. Of the 85 analysable sRNAs, we successfully detected and quantified 70 sRNAs by developing an RT-PCR system, and we identified 17 sRNAs that were differentially expressed during different growth phases. In addition, we found notable intraspecies variation in the sRNA repertoire of *A. actinomycetemcomitans*, thus suggesting that frequent acquisition or deletion of sRNAs occurred during the evolution of this species. The predicted target genes of the intergenic region sRNAs indicated the possibility of sRNA interaction with several virulence genes including leukotoxin and cytolethal distending toxin. Our results should serve as an important genomic and genetic basis for future studies to fully understand the regulatory network in *A. actinomycetemcomitans* and provide new insights into the intraspecies variation of the bacterial sRNA repertoire in bacteria.

Key words: small RNA, RNA-seq, *Aggregatibacter actinomycetemcomitans*

1. Introduction

Aggregatibacter actinomycetemcomitans is a Gram-negative facultative anaerobe. It is a major periodontal pathogen that is often associated with aggressive periodontitis.[1] Two exotoxins, leukotoxin and cytolethal distending toxin (Cdt), have been reported as the major virulence factors of *A. actinomycetemcomitans*. Leukotoxin is a membrane-active toxin that specifically kills leukocytes and contributes to the evasion of the host immune response.[2] Cdt inhibits the proliferation of host cells, including lymphocytes and gingival fibroblasts, by arresting the cell cycle, and it can also induce apoptosis.[3]

© The Author 2017. Published by Oxford University Press on behalf of Kazusa DNA Research Institute.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

In the past decade, a group of non-coding RNAs known as small RNAs (sRNAs) have been identified as novel regulatory factors of gene expression in various bacteria, including several pathogens.[4] The regulatory functions of sRNAs can be categorized into two types: *trans*-acting and *cis*-acting. The *trans*-acting regulation is mediated by sRNAs that are encoded in the intergenic region (IGR) and have been termed IGR sRNAs. Most IGR sRNAs contain a stem–loop structure. Via short, imperfect base paring between the loop and the target RNAs, IGR sRNAs regulate the expression of multiple genes located at several genome positions. Both negative and positive regulations by IGR sRNA have been demonstrated.[4] In negative regulation, IGR sRNA base pairs with a ribosome-binding site (RBS) in the target mRNA and inhibits translation by preventing the ribosome from recognizing an mRNA.[5] IGR sRNAs can also affect the stability of mRNA by binding to target mRNAs and recruiting endoribonucleases to accelerate the degradation of the target mRNAs.[6] Some IGR sRNAs can positively regulate gene expression via translational activation or RNA protection. For example, the *Staphylococcus aureus* sRNA RNAIII is a global regulator of virulence genes that enhances the translation of alpha-toxin (Hla). The binding of RNAIII to the 5′ untranslated region (UTR) of *hla* mRNA affects the secondary structure of the mRNA, thus enhancing ribosomal access and translation.[7] The *cis*-acting sRNA regulation is mediated by 'antisense sRNAs (asRNAs)', which are encoded on the opposite strand of the same loci as the target RNA. Several regulatory mechanisms have been reported for asRNAs including target RNA degradation and translational inhibition.[8] For example, the degradation of the *isrA* mRNA of *Synechocystis* PCC 6803 is enhanced by the binding of an asRNA (IsiR).[9] In the SymE–SymR toxin–antitoxin system of *Escherichia coli*, the SymR asRNA binds the RBS of *symE* mRNA and inhibits its translation.[10]

Systematic searches for bacterial sRNAs have recently been performed by using various comprehensive methods such as computational prediction, a tiling array analysis, and RNA sequencing (RNA-seq) with high-throughput next generation sequencers.[11,12] In *E. coli*, more than 400 sRNAs have been identified, and some of them have been characterized.[13,14] In *Pseudomonas aeruginosa*, more than 500 sRNAs have been identified by RNA-seq, nearly 90% of which have no orthologues in other bacterial species.[15]

Efforts to identify sRNAs in *A. actinomycetemcomitans* have also been undertaken. Jorth et al. have used northern blotting to identify 9 sRNAs from 35 sRNA candidates predicted by bioinformatics in strain VT1169.[16] The same group has performed a more systematic screening of sRNA by using RNA-seq in strain 624 and has identified 202 sRNAs.[17] In strain HK1651, four sRNAs that are expressed under iron-limiting conditions have been identified.[18]

In the present study, we systematically screened sRNAs by RNA-seq to determine the repertoire of sRNAs in *A. actinomycetemcomitans* strain HK1651, which is a clinical isolate from an aggressive periodontitis patient and a JP2 clonal isolate.[19] We subsequently validated the expression of the identified sRNAs by RT-PCR, analysed the distribution of sRNAs in four other *A. actinomycetemcomitans* strains and determined the expression pattern of the sRNAs during growth.

2. Materials and methods

2.1. Bacterial strains and growth conditions

We analysed five *A. actinomycetemcomitans* strains: HK1651 (serotype b), ATCC29523 (serotype a), Y4 (serotype b), NCTC9710 (serotype c), and IDH781 (serotype d) in the present study. Of these, the genomes of HK1651 and IDH781 have been sequenced (CP008984.1 and CP016553.1). All strains were grown in Todd Hewitt broth (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) supplemented with 1% (wt/vol) yeast extract (Nacalai Tesque, Kyoto, Japan) (THY broth) at 37°C under 5% CO_2 without shaking.

2.2. RNA isolation

A small aliquot (10^8 CFU) of an overnight culture of strain HK1651 was inoculated into 30 ml of THY broth and grown to early-exponential phase [optical density at 660 nm (OD_{660}) of 0.1] or late-exponential phase (OD_{660} of 0.2) (Supplementary Fig. S1). The total RNA was extracted using the acidic phenol method[20] with several modifications. Briefly, the bacterial culture was pelleted by centrifugation (3,000 × *g* for 10 min) and re-suspended in ice-cold resuspension buffer (50 mM sodium acetate, 10 mM magnesium acetate). Total RNA was extracted twice with acidic phenol (saturated with citrate buffer, pH 4.3) extraction, followed by a chloroform extraction and ethanol precipitation. The sample was treated with TURBO™ DNase (Thermo Fisher Scientific, Inc., Waltham, MA, USA) (20 units/ml) to remove the DNA. One microgram of RNA was electrophoresed on a TBE-Urea (9M)-acrylamide (8%) gel and visualized by ethidium bromide staining to confirm the RNA quality.

2.3. RNA-seq analysis

RNA-seq was performed on two independently prepared samples from early-exponential phase cells. Sequencing libraries were prepared from each RNA preparation without fragmentation by using the NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's protocol, with a minor modification. This kit is specialized for sequencing eukaryotic sRNAs that contain a 5′ monophosphate, such as miRNA and siRNA. Therefore, to analyse sRNAs containing a 5′ monophosphate terminus or a triphosphate terminus, the total RNA was treated with RNA 5′ poly phosphatase (Epicentre, Madison, WI, USA) according to the manufacturer's instructions before a sequencing library was prepared (Fig. 1). Paired-end sequencing (2 × 300 bp) was performed on the MiSeq platform (Illumina, San Diego, CA, USA) according to the manufacturer's protocol. After removal of the low-quality reads and trimming of the adaptor sequences using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html), the sequence reads were strand-specifically mapped to the HK1651 genome sequence by using the Burrows–Wheeler Aligner,[21] and the data were visualized with Integrative Genomics Viewer.[22] The sequence reads obtained have been deposited in the GenBank/EMBL/DDBJ database (DRA submission number, DRA005568; DRR accession number, DRR088289).

2.4. Northern blot analysis

Twenty micrograms of total RNA extracted from HK1651 harvested during the early-exponential phases was electrophoresed on a TBE-Urea (9M)-acrylamide (8%) gel. The separated RNA was transferred onto a positively charged nylon membrane (Roche Diagnostics, Basel, Switzerland) in 20× SSC. The membrane was baked for 2 h at 80°C. After prehybridization using DIG Easy Hyb (Roche Diagnostics), the membrane was hybridized with DIG Easy Hyb containing a 25 pmol of single-stranded DNA probe (Supplementary Table S1) labelled with digoxigenin using DIG Oligonucleotide Tailing Kit, 2nd generation (Roche Diagnostics) for 18 h at 45°C. After hybridization, the membrane was washed and blocked using

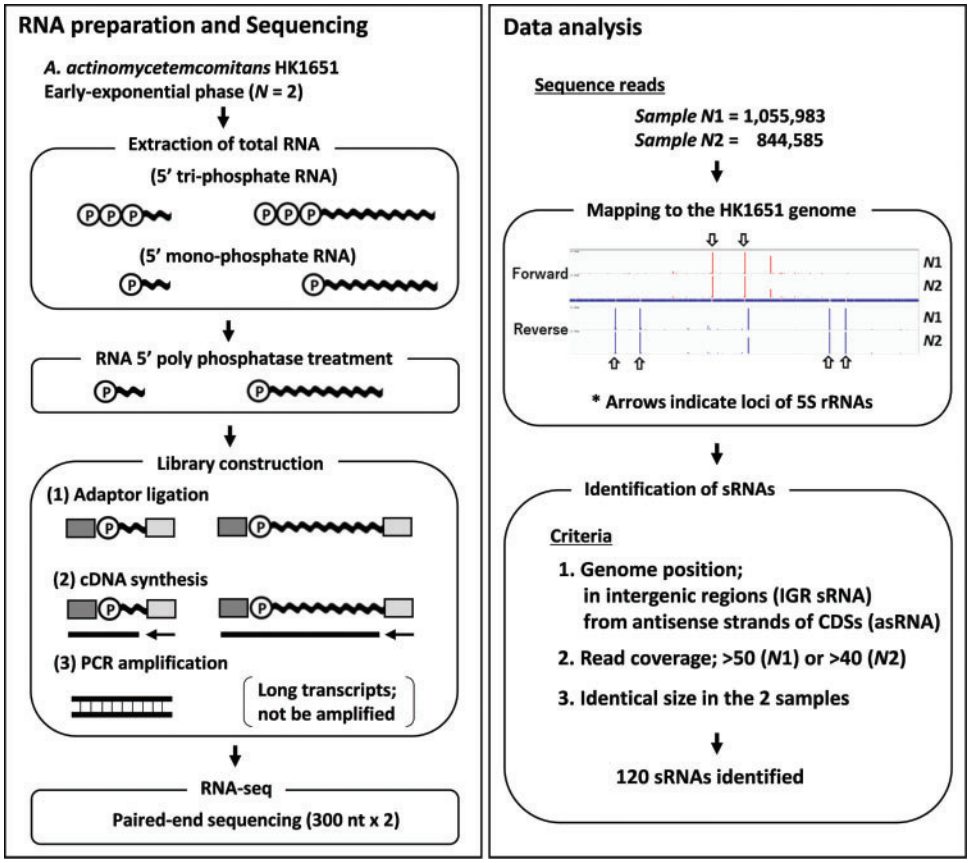

Figure 1. A flowchart illustrating sRNA identification in *A. actinomycetemcomitans* HK1651.

DIG Wash and Block Buffer Set (Roche Diagnostics), and then the hybridization signals were detected using DIG Luminescent Detection Kit (Roche Diagnostics) and ChemiDoc™ XRS+ (Bio-Rad Laboratories, Inc., Hercules, CA, USA).

2.5. RT-PCR and quantitative RT-PCR

Total RNA extracted from five *A. actinomycetemcomitans* strains (HK1651, ATCC29523, Y4, NCTC9710, and IDH781) harvested during the early- and late-exponential phases was polyadenylated using *E. coli* poly (A) polymerase (New England Biolabs) at 37°C for 30 min. This was followed by phenol/chloroform extraction and ethanol precipitation. One microgram of the RNA was then converted to cDNA with a Transcriptor First Strand cDNA Synthesis Kit (Roche Diagnostics) with an oligo d(T) primer containing a sequence tag at the 5′-end (5′-AAGCAGTGGTATCAACGCAGAGTACTT TTTTTTTTTTTTTTTTTVN-3′). The cDNA was amplified using FastStart Essential DNA Green Master reaction mix (Roche Diagnostics) with 5′-primers specific to each sRNA and a 3′-primer (5′-CAGTGGTATCAACGCAGAGT-3′) that targeted the previously mentioned sequence tag. The sRNA specific primers are listed in Supplementary Table S2. The specific primers were designed within 50 bp from the 5′ terminus of each sRNA to amplify the nearly full-length sRNAs. In cases in which the sRNAs had high homology to other loci, the primers were designed to contain two nucleotide mismatches at the 3′ termini to prevent non-specific amplification. The thermal cycling conditions used were a 10-min incubation at 95°C followed by 25–45 cycles of 10 s at 95°C and 30 s at 58–62°C. The number of amplification cycles and the annealing/elongation temperatures were altered according to the amplification efficiency of each sRNA. The RT-PCR products were electrophoresed on a 2% agarose gel and visualized with ethidium bromide staining. Each sRNA was analysed two or three times using independently prepared RNA samples.

Quantitative RT-PCR (qPCR) was performed using a LightCycler Nano (Roche Diagnostics) and the same experimental procedures described for the RT-PCR assay (polyadenylation, cDNA conversion, and PCR primers). The transcription level was normalized to 5S rRNA (5′-primer; 5′-TGAAACCATACCGAACTCAG-3′) in a 1,000-fold diluted cDNA sample. Each sRNA was analysed by qPCR three times using independently prepared RNA samples. The EZR version 1.32 software (http://www.jichi.ac.jp/saitama-sct/SaitamaHP.files/statmedEN.html) was used for statistical analyses.

2.6. Genomic DNA PCR

Aggregatibacter actinomycetemcomitans strains were cultivated in 5 ml of THY broth at 37°C under 5% CO_2 without shaking for 24 h. The genomic DNA was extracted using a DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany), and 100 ng was used as a template for PCR. PCR was performed using BIOTAQ™ DNA Polymerase (Bioline Reagents Ltd, London, UK) with the 5′ primers listed in Supplementary Table S2 and the 3′ primers listed in Supplementary Table S3. The 3′ primers were designed ~100 bp downstream of the 5′ primer for each target genomic region. The thermal cycling conditions were a 2-min incubation at 94°C followed by 30 cycles of 20 s at 94°C, 30 s at 60°C, and 30 s at 72°C. The

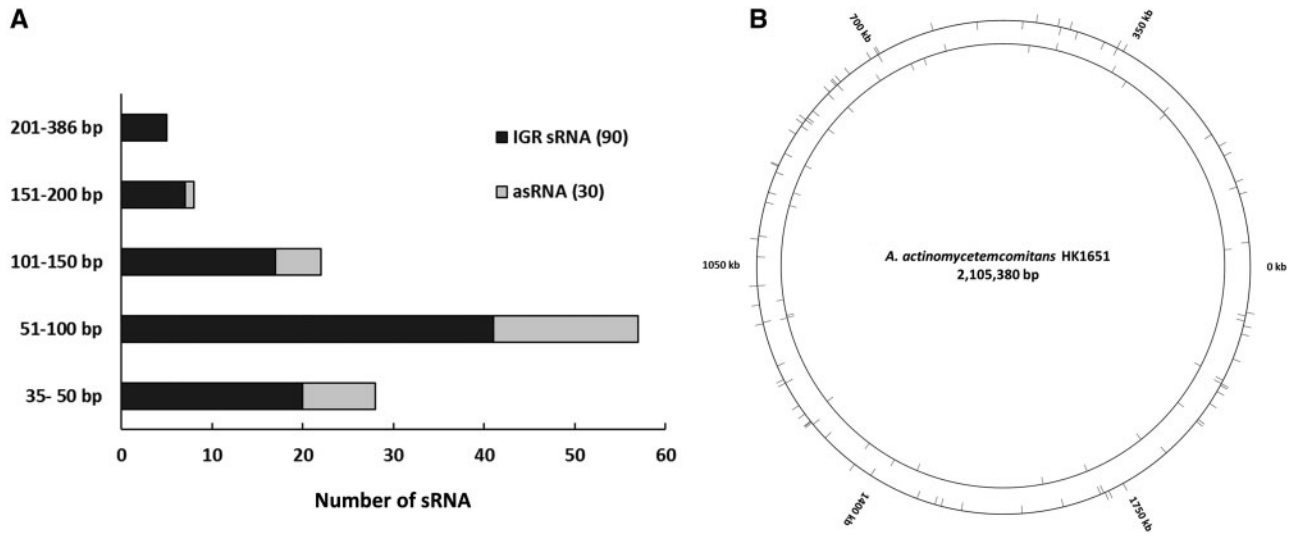

Figure 2. sRNAs identified in *A. actinomycetemcomitans* HK1651. (A) Length distribution of 120 sRNAs identified in the present study. (B) The circular genome mapping of the sRNAs identified. Outer and inner circles indicate the loci of the IGR sRNAs and the asRNAs, respectively. The outward and inward lines indicate the sRNAs located on the forward and reverse strands, respectively.

PCR products were electrophoresed and visualized as described above.

2.7. Multilocus sequence analysis

Six housekeeping genes (*pgi*, *recA*, *adk*, *frdB*, *atpG*, and *mdh*) were chosen according to a previous study.[23] Internal fragments of ~500 bp were amplified by PCR with the primers listed in Supplementary Table S4. PCR was performed using TaKaRa Ex Taq® (TAKARA BIO INC., Kusatsu, Shiga, Japan). The reactions contained genomic DNA (100 ng) that was purified from the three *A. actinomycetemcomitans* strains (Y4, ATCC29523, and NCTC9710) as described above. The thermal cycling conditions were 2 min of incubation at 94°C followed by 35 cycles of 20 s at 94°C, 30 s at 60°C, and 1 min at 72°C. The PCR products were purified using a Microcon YM-100 (Merck Millipore Corporation, Darmstadt, Germany) and sequenced using a BigDye® Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific Inc.) and an ABI PRISM 3130xl Genetic Analyser (Thermo Fisher Scientific Inc.). The primers used for PCR amplification were used as sequencing primers. The sequences of the six genes (*pgi*; 514 bp, *recA*; 480 bp, *adk*; 432 bp, *frdB*; 538 bp, *atpG*; 516 bp, and *mdh*; 457 bp) were concatenated for each of the three strains, and the corresponding sequences of HK1651 (CP008984.1), IDH781 (CP016553.1), and *Aggregatibacter aphrophilus* NJ8700 (CP001607.1) were obtained from the GenBank/EMBL/DDBJ database. The concatenated sequences were aligned for a phylogenetic analysis by the maximum likelihood method using MEGA ver. 7.0.21.[24]

2.8. Target prediction of sRNAs

The target mRNAs of each IGR sRNA were predicted by IntaRNA,[25] using the sequences encompassing 75 bp upstream and downstream of the start codon of each protein coding sequence (CDS), which were extracted from the *A. actinomycetemcomitans* HK1651 RefSeq sequence (NZ_CP007502.1). The parameters were set to a minimum of seven base pairs in the seed region. We defined the threshold for target prediction at a maximal energy of −14.2 kcal/mol. The maximal energy was the sum of the hybridization free energy of the interacting subsequences and the free energies required to unfold the interaction sites in both RNA molecules.

3. Results and discussion

3.1. Identification of sRNAs in *A. actinomycetemcomitans* HK1651

The sRNA identification process is shown in Fig. 1. The primary transcripts contained a triphosphate at the 5′ terminus, whereas the processed transcripts had a 5′ monophosphate. To identify both types of sRNA molecules,[26] we converted the 5′ triphosphate RNA to a 5′ monophosphate RNA and then constructed a library for RNA-seq. Because most known sRNAs are 30–500 nucleotides long, we did not perform RNA fragmentation; thus, large transcripts were ignored. We obtained 1,055,983 (Experiment 1) and 844,585 (Experiment 2) pairs of sequence reads from the RNA-seq analyses of 2 independently prepared RNA samples from early exponentially growing HK1651 cells. Although many reads were derived from the *rrn* loci (mainly 5S rRNAs), 471,017 (Experiment 1) and 473,797 (experiments 2) read pairs were mapped to non-*rrn* loci. Using these reads, we searched for sRNAs on the basis of the following criteria: (i) transcribed from IGRs (defined as IGR sRNA) or opposite strands of CDSs (asRNA); (ii) with >50 (Experiment 1) and >40 (Experiment 2) mapping coverage to normalize for the different numbers of sequence reads between 2 samples (1,055,983 and 844,585 reads); and (iii) of identical size in the two samples. In addition, we excluded data with gradually decreased coverage towards the ends of transcripts, to remove false positives derived from degraded RNA. Through these criteria, we identified a total of 120 sRNAs that were 35–386 bp long (Fig. 2A) and broadly distributed throughout the entire chromosome (Fig. 2B). Of the 120 sRNAs, 90 were IGR sRNAs, including two riboswitches and one tmRNA, and 30 were asRNAs (Supplementary Tables S5 and S6). To confirm the expression of the sRNAs, we performed northern blot analysis. We selected 10 sRNAs based on the high expression by RNA-seq results, and found the bands of 10 sRNAs (Fig. 3). Most of their lengths were almost consistent with those of the RNA-seq, but the lengths of

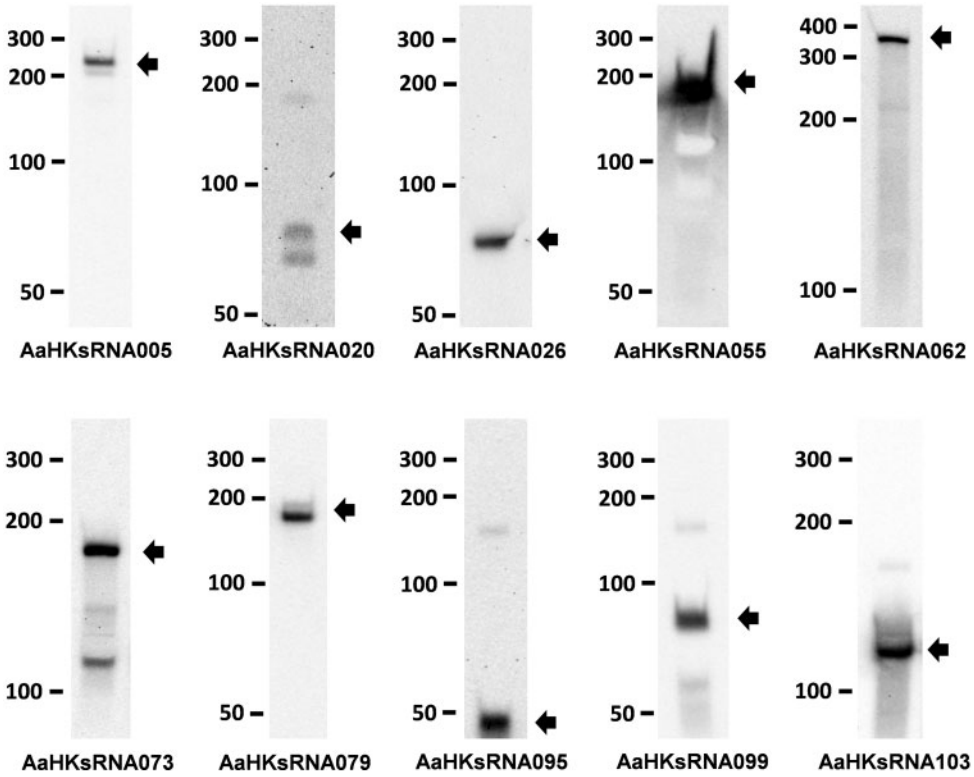

Figure 3. Northern blot analysis of sRNAs in *A. actinomycetemcomitans* HK1651. Ten sRNAs were validated by Northern blot analysis. The RNA sizes (bp) indicate on the left of the blot. Arrows indicate the predicted sRNAs.

AaHKsRNA099 and 103 were slightly different between northern analysis and RNA seq. This difference may be due to the definition of sRNA with our criteria described above. We found that the three sets of IGR sRNAs (AaHKsRNA031 and 032, AaHKsRNA046 and 047, AaHKsRNA063 and 064) were located in both strands and were partially overlapping (Supplementary Table S5). In these cases, it was possible for one of the transcripts to be a protein-encoding mRNA and the other one to be an asRNA, such as in a toxin–antitoxin system.[10] However, we did not find a functional CDS in these sRNAs. Many IGR sRNAs have a Rho-independent terminator, which is an important structure for interaction with the RNA chaperone Hfq.[27] Therefore, we assessed the presence of Rho-independent terminators in sRNAs by using ARNold (http://rna.igmors.u-psud.fr/toolbox/arnold/) and found 34 Rho-independent terminators in 90 IGR sRNAs (Supplementary Table S5).

We compared our results against the sRNAs identified in previous studies[16–18] and found that 60 of the 90 IGR sRNAs were newly identified. Of the 75 IGR sRNA detected in strain 624 by Jorth et al.[17] under various conditions including an *in vivo* model, 49 were not found in our IGR sRNA set. An analysis of these 49 sRNAs revealed that the sequences corresponding to 19 sRNAs were absent in the HK1651 genome, 8 sRNAs were in CDSs, and 7 were located in repetitive sequences. The 13 IGR sRNAs were detected in our RNA-seq analysis, but their mapping coverage was lower than the study thresholds (below 50 in Experiment 1 and/or below 40 in Experiment 2). No reads were obtained for the remaining two IGR sRNAs in our RNA-seq analysis. Of the four Fur regulated IGR sRNAs (JA01, JA02, JA03, and JA04) identified in strain HK1651,[18] only JA04 was identified under our experimental conditions, and no reads were detected for the other three.

Of the 30 asRNAs identified in the present study, as many as 29 were newly identified. Although in the previously mentioned study of strain 624, the authors have reported 127 asRNAs,[17] only one of these (asRNA corresponding to D7S_0307.1) was identified in the present study (AaHKsRNA084). In addition, the lengths of D7S_0307.1 (1,067 bp) and AaHKsRNA084 (56 bp) differed significantly. Of the remaining 126 asRNAs identified in the previous study, 1 asRNA (D7S_2363.1) showed sequence homology to 2 tRNA genes, and sequences corresponding to 22 asRNAs were absent in the HK1651 genome. Although sequences corresponding to the remaining 103 asRNAs were conserved in the HK1651 genome, the mapping coverage of 29 asRNAs was below the threshold, and no reads were detected for 73 asRNAs in our RNA-seq analysis. A portion of one sRNA reported as an asRNA in strain 624 (D7S_2289.1; 295 bp) corresponded to IGR sRNA020 (77 bp) identified in the present study. Thus, the asRNA set identified in HK651 was significantly different from that of strain 624 (17). Two large asRNAs (D7S_1370.1; 996 bp and D7S_0307.1; 1,624 bp) identified in strain 624 were undetectable in our experimental analysis. A total of 32 asRNAs shorter than 30 bp were identified in strain 624 that were not identified in strain HK1651. Although systematic searches for asRNAs have been performed in several bacterial species,[28–30] asRNAs shorter than 30 bp have rarely been reported, including in our study.

3.2. Development of RT-PCR to detect sRNA and confirmation of the expression of the sRNAs identified in strain HK1651

We attempted to develop an RT-PCR-based, simple detection system for sRNAs to confirm the expression of sRNAs identified by

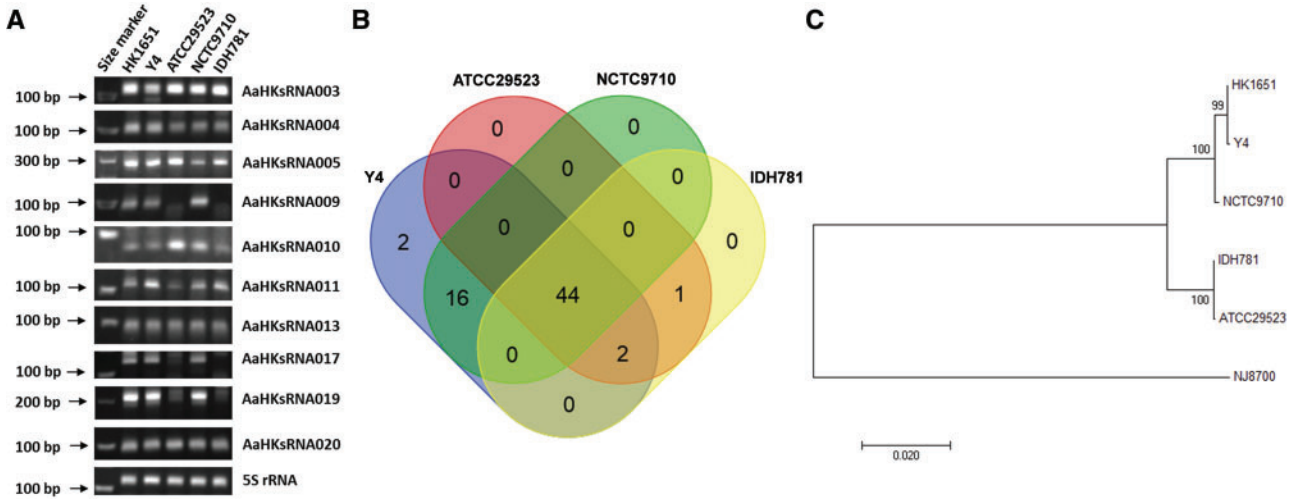

Figure 4. Distribution of the 70 sRNAs identified in strain HK1651 in other *A. actinomycetemcomitans* strains and their phylogenetic relationship with strain HK1651. (A) The RT-PCR results of 10 of the 70 representative sRNAs in *A. actinomycetemcomitans* strains. The 5S rRNA was used as a loading control. (B) A Venn diagram shows the distribution/conservation pattern of the 70 sRNAs in five *A. actinomycetemcomitans* strains. Five sRNAs were specific to strain HK1651. (C) A maximum likelihood tree for the five *A. actinomycetemcomitans* strains, which was constructed on the basis of the sequences of the six housekeeping genes *pgi, recA, adk, frdB, atpG,* and *mdh. Aggregatibacter aphrophilus* NJ8700 was used as an outgroup. The bootstrap values calculated from 1,000 replicates are indicated at the corresponding nodes. The scale bar corresponds to 0.02 estimated nucleotide substitutions per site.

RNA-seq. However, it is impossible to distinguish an asRNA from its cognate mRNA by conventional RT-PCR. It is also difficult to design forward and reverse primers for many IGR sRNAs because of their short lengths. Therefore, we first polyadenylated the RNA molecules, converted them to cDNA by using a tagged oligo d(T) primer, and then detected sRNAs by PCR using a sequence tag-specific primer and a primer specific to each sRNA.[31] Because we designed specific primers within 50 bp from the 5′ terminus of each sRNA to amplify nearly full-length sRNAs, we were unable to design optimal primers for three sRNAs. To analyse these three sRNAs, the primers were designed slightly further downstream from the 5′ terminus (AaHKsRNA020, 40–58 bp from the 5′ terminus; AaHKsRNA037, 26–53 bp; and AaHKsRNA040, 34–60 bp). In addition, we were not able to design specific primers for 35 sRNAs, because identical or nearly identical sequences to any primer candidate sequences were found in other genomic loci. Therefore, we finally constructed an RT-PCR system to detect 85 of the 120 sRNAs identified by RNA-seq. Using this system, we confirmed the expression of 70 sRNAs (53 IGR sRNAs and 17 asRNAs) (Fig. 4A; Supplementary Tables S5 and S6). Our inability to detect the expression of other 15 sRNAs might have been due to the lower sensitivity of this RT-PCR system and/or the incomplete optimization of the PCR primers and/or conditions.

3.3. Comparison of sRNA repertoires among *A. actinomycetemcomitans* strains

A comparison of the sRNA sets identified in strains 624 and HK1651 suggested that a significant difference in sRNA repertoire between *A. actinomycetemcomitans* strains might exist. Therefore, we used the RT-PCR system described above to analyse the expression of 70 sRNAs in four *A. actinomycetemcomitans* strains (Y4, ATCC29523, NCTC9710, and IDH781). Figure 4A and B and Supplementary Table S7 show that 44 sRNAs (37 IGR sRNAs and seven asRNAs) were detected in all five *A. actinomycetemcomitans* strains, whereas 5 sRNAs were specifically detected in HK1651. The remaining 21 sRNAs were variably detected in the four strains, but the number of sRNAs detected in Y4 and NCTC9710 (64 and 60, respectively) was higher than in ATCC29523 and IDH781 (47 in both).

Because as much as 37% of the sRNAs analysed (26 of 70) was variably detected in the five strains, we investigated the existence of genomic loci encoding the 26 sRNAs in the strains by PCR. Of the six sRNAs not detected by RT-PCR in strain Y4, three were not detected by PCR of genomic DNA. Similarly, in the other strains, 5 of 10 (NCTC9710), 15 of 23 (ATCC29523), and 16 of 23 (IDH781) sRNAs were not detected by PCR of genomic DNA (Supplementary Table S7). Because the complete genome sequence was available for IDH781, we further examined the IDH781 sequences corresponding to the 16 primers that did not yield a product in the genomic DNA PCR assays for this strain. This analysis revealed that whereas three primers did not work because of mismatches in the primer sequence (Supplementary Table S2), the genomic loci encoding 13 sRNAs were absent in IDH781 (Supplementary Table S7). This result indicated that, although our inability to detect some sRNAs was due to a primer sequence mismatch, a significant number of sRNAs were not detected by RT-PCR because their genes were actually absent in the strain. Although seven sRNAs were found in the chromosome of IDH781, their expression was not detected by RT-PCR without a primer mismatch. We checked the upstream regions of the seven sRNAs and found that four sRNAs contained one or two single-nucleotide polymorphisms (SNPs) within the −50 position of each sRNA, when compared with the HK1651 sequence (AaHKsRNA019; −47C/T, AaHKsRNA028; −10C/T and −39G/A, AaHKsRNA054; −45C/T, and AaHKsRNA076; −3C/G), and AaHKsRNA097 showed no homology at the promoter region, when compared with the HK1651 sequence, whereas the other two sRNAs had no SNPs. These SNPs or the deletion found in the 5′ UTR region may be associated with the intraspecies variation of the expression of the sRNAs.

These data, together with those from the comparison of sRNAs between strains 624 and HK1651, indicated that the sRNA repertoire is notably different among *A. actinomycetemcomitans* strains. A phylogenetic tree based on five housekeeping genes showed that HK1651 was closely related to Y4 and NCTC9710 but not to ATCC29523 and

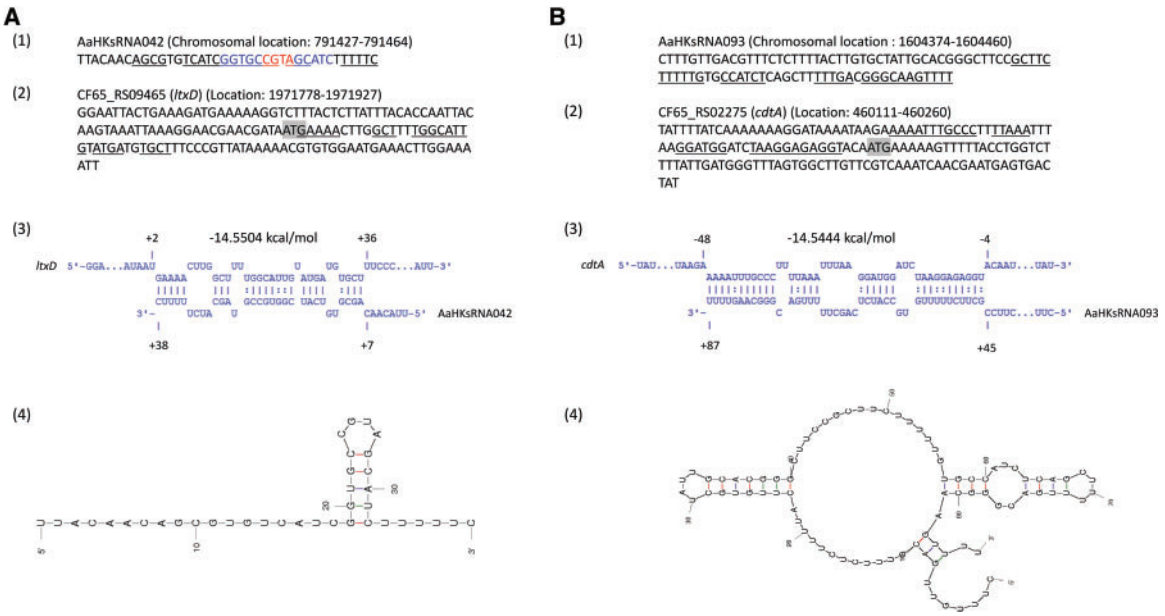

Figure 5. Putative sRNA interaction with mRNA encoding *ltxD* (A) and *cdtA* (B). Nucleotide sequences of sRNA (1) and the region of the target gene upstream and downstream of the start codon (2) were shown. Underlines indicate the region of sRNA–mRNA binding predicted by IntaRNA. Boxed nucleotides indicate start codon in target gene. Blue and red nucleotides indicate rho independent terminator in sRNA. (3) Base pairing of sRNA and target mRNA was predicted by IntaRNA. Binding energy of sRNA–mRNA is shown. (4) Secondary structure prediction of sRNA by mFold. The lowest free energy secondary structure is shown.

IDH781 (Fig. 4C). The distribution/conservation patterns of the sRNAs of the five strains were consistent with their phylogenetic relationships; strains more closely related to HK1651 (Y4 and NCTC9710) shared more sRNAs with HK1651 (Fig. 4B and C), thus suggesting the lineage-dependent acquisition or deletion of these sRNAs during the evolution of these *A. actinomycetemcomitans* strains. A similar result has been reported for other species such as *Campylobacter jejuni*, which shows a differential distribution pattern of sRNAs among strains.[32] Although intraspecies variation in the sRNA repertoire has not been well elucidated in other bacteria, similar strain-to-strain variation may be observed in many other bacterial species.

Aggregatibacter actinomycetemcomitans have been classified into seven serotypes, and certain characteristics of their virulence properties differ among the serotypes. Specifically, serotype b strains have stronger virulence than other serotypes.[1,33,34] Recently, the phylogenetic distance of 31 strains of *A. actinomycetemcomitans* including HK1651 has been analysed using 397 core genes, and their phylogeny has been found to be divisible into five clades, one of which is constructed in the serotype b strains.[35] In a sRNAs distribution/conservation pattern analysis, HK1651 and Y4, which are serotype b, exhibited the closest pattern, when compared with the other three strains that showed other serotypes (NCTC9710; serotype c, ATCC29523; serotype a and IDH781; serotype d) (Fig. 4B). Additionally, the phylogenetic distances of HK1651 and Y4, constructed using the five housekeeping genes, were the closest (Fig. 4C). Therefore, the presence of sRNA and its regulatory mechanism may be a determinant of virulence among the different serotypes.

3.4. Expression of sRNAs during growth

The expression levels of the RT-PCR products of the 70 sRNAs detected by RT-PCR in HK1651 were further analysed in the early- and late-exponential phases. As shown in Supplementary Table S8, we identified 17 sRNAs exhibiting greater than 2-fold changes during late-exponential phase compared with early-exponential phase (nine were upregulated and eight were downregulated). A previous study has demonstrated that the expression of sRNA Aa35 in *A. actinomycetemcomitans* strain VT1169, which corresponds to AaHKsRNA003 in strain HK1651, is drastically decreased during the stationary phase compared with the exponential phase.[16] Consistent with this, the expression of AaHKsRNA003 (Aa35) exhibited a 3.5-fold decrease during late-exponential phase. To analyse the *cis*-acting regulation by asRNA, the transcriptional levels of genes located on opposite strands of four asRNAs (downregulated sRNAs; AaHKsRNA011, 066 and 070, upregulated sRNA; AaHKsRNA106) that showed altered expression during growth were analysed by qPCR. Three of the four genes (HK1651_09435, HK1651_05805, and HK1651_05685) showed slightly decreased transcription in the late-exponential phase (data not shown). Generally, many of asRNA negatively affects the gene expression.[8] However, some sRNAs were reported as positive regulator such as GadY. GadY asRNA stabilized *gadX* transcripts through interacting with 3′ UTR of *gadX* mRNA.[36] Although two asRNAs (AaHKsRNA066 and 070) may regulate the expression of the target gene positively, AaHKsRNA066 and 070 locate in a CDS, not in 3′ UTR, of the opposite strand transcripts, suggesting AaHKsRNA066 and 070 may have a different regulation mechanism with GadY asRNA.

3.5. Target prediction of IGR RNAs

Finally, the target mRNAs of the 83 IGR sRNAs identified in the present study (excluding 7 IGR sRNAs; tmRNA, 6S RNA, RNase P RNA, 2 riboswitches, and 2 signal recognition particle RNAs) were predicted by the IntaRNA programme.[25] Lists of the predicted targets are provided in Supplementary Table S9. Although the predicted targets are involved in a wide range of biological functions, the most notable finding was that several IGR sRNAs were predicted to bind to mRNAs encoding known virulence factors. For example, AaHKsRNAs042 was predicted to interact with the mRNA of the leukotoxin export protein (LtxD) (CF65_RS09465) (Fig. 5A), AaHKsRNA093 was predicted to interact

with CdtA (CF65_RS02275) (Fig. 5B), and AaHKsRNA051 was predicted to interact with the mRNA of fimbrial protein (Flp1) (CF65_RS02930)[37] (Supplementary Fig. S2A). In addition, eight sRNAs [AaHKsRNA004, 015, 025, 036, 046, 065, 068, and 105 (Supplementary Fig. S2B)] were predicted to interact with mRNA of the virulence-associated protein D, which encodes a homologue of *Haemophilus influenza* VapD that is a part of a toxin–antitoxin system.[38] Of these sRNAs, AaHKsRNA065 was expressed only in serotype b strains (HK1651 and Y4) (Supplementary Table S7). Therefore, the interaction of this sRNA and *vapD* mRNA may be one of the determinants of strong virulence in the serotype b.

In IDH781, the 34 detectable IGR sRNAs (excluding 7 IGR sRNAs; tmRNA, 6S RNA, RNase P RNA, 2 riboswitches, and 2 signal recognition particle RNAs) by our RT-PCR were compared with the corresponding sRNAs in HK1651. We found that the 18 sRNAs contained 1–7 SNPs (Supplementary Table S10). We predicted the targets of the 18 sRNAs to analyse the switching of the target mRNAs due to the SNPs by using the HK1651 RefSeq sequence because the IDH781 RefSeq sequence was not supported in IntaRNA. Some of the prediction patterns were drastically altered by a few SNPs (Supplementary Table S10), thus suggesting that SNPs in the sRNA may affect the regulatory systems among the strains.

These data, together with other data obtained in this study, should serve as an important genomic and genetic basis for future studies to fully understand the sRNA-involved regulatory network in *A. actinomycetemcomitans* and provide new insights into the intra-species variation of the bacterial sRNA repertoire in bacteria. Functional analyses of each sRNA in different conditions and sRNA repertoire analyses of a wider range of strains, including those in closely related species, will enhance our understanding of the regulatory network in *A. actinomycetemcomitans*, especially that regulates the virulence of this important periodontal pathogen.

Funding

This study was supported in part by Grants-in-Aid for Scientific Research (C) (Grant No: 25462868) from the Ministry of Education, Culture, Sports, Sciences, and Technology of Japan.

Accession numbers

DRA submission number, DRA005568; DRR accession number, DRR088289.

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at *DNARES* online.

References

1. Fine, D.H., Markowitz, K. and Furgang, D. 2007, *Aggregatibacter actinomycetemcomitans* and its relationship to initiation of localized aggressive periodontitis: longitudinal cohort study of initially healthy adolescents, *J. Clin. Microbiol.*, **45**, 3859–69.
2. Kachlany, S.C. 2010, *Aggregatibacter actinomycetemcomitans* leukotoxin: from threat to therapy, *J. Dent. Res.*, **89**, 561–70.
3. Tan, K.S., Song, K.P. and Ong, G. 2002, Cytolethal distending toxin of *Actinobacillus actinomycetemcomitans*. Occurrence and association with periodontal disease, *J. Periodont. Res.*, **37**, 268–72.
4. Waters, L.S. and Storz, G. 2009, Regulatory RNAs in bacteria, *Cell*, **136**, 615–28.
5. Bouvier, M., Sharma, C.M., Mika, F., Nierhaus, K.H. and Vogel, J. 2008, Small RNA binding to 5′ mRNA coding region inhibits translational initiation, *Mol. Cell.*, **32**, 827–37.
6. Massé, E., Escorcia, F.E. and Gottesman, S. 2003, Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*, *Genes Dev.*, **17**, 2374–83.
7. Morfeldt, E., Taylor, D., von Gabain, A. and Arvidson, S. 1995, Activation of alpha-toxin translation in *Staphylococcus aureus* by the *trans*-encoded antisense RNA, RNAIII, *EMBO J.*, **14**, 4569–77.
8. Georg, J. and Hess, W.R. 2011, *cis*-antisense RNA, another level of gene regulation in bacteria, *Microbiol. Mol. Biol. Rev.*, **75**, 286–300.
9. Dühring, U., Axmann, I.M., Hess, W.R. and Wilde, A. 2006, An internal antisense RNA regulates expression of the photosynthesis gene *isiA*, *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 7054–8.
10. Kawano, M., Aravind, L. and Storz, G. 2007, An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin, *Mol. Microbiol.*, **64**, 738–54.
11. Sridhar, J. and Gunasekaran, P. 2013, Computational small RNA prediction in bacteria, *Bioinform. Biol. Insights*, **7**, 83–95.
12. Sharma, C.M. and Vogel, J. 2009, Experimental approaches for the discovery and characterization of regulatory small RNA, *Curr. Opin. Microbiol.*, **12**, 536–46.
13. Hershberg, R., Altuvia, S. and Margalit, H. 2003, A survey of small RNA-encoding genes in *Escherichia coli*, *Nucleic Acids Res.*, **31**, 1813–20.
14. Shinhara, A., Matsui, M., Hiraoka, K., et al. 2011, Deep sequencing reveals as-yet-undiscovered small RNAs in *Escherichia coli*, *BMC Genomics*, **12**, 428.
15. Gómez-Lozano, M., Marvig, R.L., Molin, S. and Long, K.S. 2012, Genome-wide identification of novel small RNAs in *Pseudomonas aeruginosa*, *Environ. Microbiol.*, **14**, 2006–16.
16. Jorth, P. and Whiteley, M. 2010, Characterization of a novel riboswitch-regulated lysine transporter in *Aggregatibacter actinomycetemcomitans*, *J. Bacteriol.*, **192**, 6240–50.
17. Jorth, P., Trivedi, U., Rumbaugh, K. and Whiteley, M. 2013, Probing bacterial metabolism during infection using high-resolution transcriptomics, *J. Bacteriol.*, **195**, 4991–8.
18. Amarasinghe, J.J., Connell, T.D., Scannapieco, F.A. and Haase, E.M. 2012, Novel iron-regulated and Fur-regulated small regulatory RNAs in *Aggregatibacter actinomycetemcomitans*, *Mol. Oral Microbiol.*, **27**, 327–49.
19. Haubek, D., Poulsen, K. and Kilian, M. 2007, Microevolution and patterns of dissemination of the JP2 clone of *Aggregatibacter* (*Actinobacillus*) *actinomycetemcomitans*, *Infect. Immun.*, **75**, 3080–8.
20. Zubay, G. 1962, The isolation and fractionation of soluble ribonucleic acids, *J. Mol. Biol.*, **4**, 347–56.
21. Li, H. and Durbin, R. 2010, Fast and accurate long-read alignment with Burrows–Wheeler transform, *Bioinformatics*, **26**, 589–95.
22. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., et al. 2011, Integrative genomics viewer, *Nat. Biotechnol.*, **29**, 24–6.
23. Eriksen, K.T., Haubek, D. and Poulsen, K. 2005, Intragenomic recombination in the highly leukotoxic JP2 clone of *Actinobacillus actinomycetemcomitans*, *Microbiology (Reading, Engl.)*, **151**, 3371–9.
24. Tamura, K., Dudley, J., Nei, M. and Kumar, S. 2007, MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0, *Mol. Biol. Evol.*, **24**, 1596–9.
25. Busch, A., Richter, A.S. and Backofen, R. 2008, IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions, *Bioinformatics*, **24**, 2849–56.
26. Sharma, C.M., Hoffmann, S., Darfeuille, F., et al. 2010, The primary transcriptome of the major human pathogen *Helicobacter pylori*, *Nature*, **464**, 250–5.

27. Otaka, H., Ishikawa, H., Morita, T. and Aiba, H. 2011, PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action, *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 13059–64.
28. Yachie, N., Numata, K., Saito, R., Kanai, A. and Tomita, M. 2006, Prediction of non-coding and antisense RNA genes in *Escherichia coli* with Gapped Markov Model, *Gene*, **372**, 171–81.
29. Georg, J., Voß, B., Scholz, I., Mitschke, J., Wilde, A. and Hess, W.R. 2009, Evidence for a major role of antisense RNAs in cyanobacterial gene regulation, *Mol. Syst. Biol.*, **5**, 305.
30. Beaume, M., Hernandez, D., Farinelli, L., 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.
31. Ro, S. and Yan, W. 2010, Detection and quantitative analysis of small RNAs by PCR, *Methods Mol. Biol.*, **629**, 295–305.
32. Dugar, G., Herbig, A., Förstner, K.U., et al. 2013, High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates, *PLoS Genet.*, **9**, e1003495.
33. Yang, H.W., Huang, Y.F., Chan, Y. and Chou, M.Y. 2005, Relationship of *Actinobacillus actinomycetemcomitans* serotypes to periodontal condition: prevalence and proportions in subgingival plaque, *Eur. J. Oral Sci.*, **113**, 28–33.
34. Kawamoto, D., Ando, E.S., Longo, P.L., Nunes, A.C., Wikström, M. and Mayer, M.P. 2009, Genetic diversity and toxic activity of *Aggregatibacter actinomycetemcomitans* isolates, *Oral Microbiol. Immunol.*, **24**, 493–501.
35. Kittichotirat, W., Bumgarner, R.E. and Chen, C. 2016, Evolutionary divergence of *Aggregatibacter actinomycetemcomitans*, *J. Dent. Res.*, **95**, 94–101.
36. Opdyke, J.A., Kang, J.G. and Storz, G. 2004, GadY, a small-RNA regulator of acid response genes in *Escherichia coli*, *J. Bacteriol.*, **186**, 6698–705.
37. Clock, S.A., Planet, P.J., Perez, B.A. and Figurski, D.H. 2008, Outer membrane components of the Tad (tight adherence) secretion of *Aggregatibacter actinomycetemcomitans*, *J. Bacteriol.*, **190**, 980–90.
38. Ren, D., Walker, A.N. and Daines, D.A. 2012, Toxin–antitoxin loci *vapBC-1* and *vapXD* contribute to survival and virulence in nontypeable *Haemophilus influenzae*, *BMC Microbiol.*, **12**, 263.