DB
2418
2008
HG

Identification and Functional Analysis of Non-coding RNAs in *Bacillus subtilis*

A Dissertation Submitted to
the Graduate School of Life and Environmental Sciences,
the University of Tsukuba
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Science

Shinichi SAITO

寄贈
斉藤伸一氏

09909204

TABLE OF CONTENTS

ABBREVIATIONS

bp	base pair(s)
Bsr	Bacillus small RNA
DIG	digoxigenin
dNTP	deoxynucleoside triphosphate
IGR	intergenic region
kb	kilobases
LB	Luria-Bertani (medium)
ncRNA	non-coding RNA
nt	nucleotide(s)
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polynucleotide chain reaction
SD	Shine-Dalgarno
sRNA	small RNA

ABSTRACT

Small, non-coding RNAs (ncRNAs) perform diverse functions in a variety of organisms, but few ncRNAs have been identified in *Bacillus subtilis*. To search the *B. subtilis* genome for genes encoding ncRNAs, I focused on 123 intergenic regions (IGRs) over 500 bp in length and analyzed expression from these regions. Seven IGRs termed *bsrC*, *bsrD*, *bsrE*, *bsrF*, *bsrG*, *bsrH* and *bsrI* expressed RNAs smaller than 380 nt. All small RNAs except BsrD RNA were expressed in transformed *Escherichia coli* cells harboring a plasmid with PCR-amplified IGRs of *B. subtilis*, indicating that their own promoters independently express small RNAs. Under the non-stressed condition, depletion of the genes for the small RNAs did not affect growth. Although their functions are unknown, gene expression profiles at several time points showed that most of the genes except for *bsrD* were expressed during the vegetative phase (4 - 6 hours), but undetectable during the stationary phase (8 hours). Mapping the 5′ ends of the 6 small RNAs revealed that the genes for BsrE, BsrF, BsrG, BsrH, and BsrI RNAs are preceded by a recognition site for RNA polymerase sigma factor σ^A. These small RNAs might lack an SD sequence and exert their actions as ncRNAs.

INTRODUCTION

Bacterial cells contain many RNAs of 70-500 nucleotides that are not translated (Chen *et al.*, 2002). Such RNAs in eukaryotic cells range in size from 18 to over 10,000 nucleotides (Costa, 2007). Because of their size, these RNAs have also been referred to as sRNAs. Most sRNAs in bacterial cells do not encode proteins but they do function as RNA molecules. The numbers of ncRNA genes and their functions vary from species to species. Many of them, such as the OxyS, MicF, Spot 42, RyhB and DsrA RNAs in *Escherichia coli*, act by basepairing to activate or repress translation or to destabilize mRNA (Gottesman, 2004; Guillier *et al.*, 2006). A few are integral parts of RNA-protein complexes, such as signal recognition particle (SRP) and RNase P (Frank and Pace, 1998; Hainzl *et al.*, 2005). The 4.5S RNA in *Escherichia coli* and scRNA (small cytoplasmic RNA) in *Bacillus subtilis* are components of bacterial SRP (Fang *et al.*, 2001; Sharkady and Nolan, 2001). Some ncRNAs function in a unique manner, for example, the CsrB and CsrC RNAs that contain multiple repeats of the CsrA protein-binding sequence found in several mRNAs (Romeo, 1998) regulate diverse cellular functions. Often, they are identified as regulatory molecules that are involved in fine-tuning the cellular response to stress by integrating environmental signals into global regulation (Massé *et al.*, 2003a,b; Gottesman, 2005; Rasmussen *et al.*, 2005; Mattick and Makunin, 2006).

The genome sequences of many organisms have been completely

determined and can be systematically analyzed. Thus, investigators have been encouraged to develop a search for novel ncRNA genes. Wassarman *et al.* (1999) showed that at least 10 ncRNAs are encoded within the *E. coli*. They then applied sequence conservation coupled with promoter and rho-independent terminator prediction and identified 17 new ncRNAs in the same species (Wassarman *et al.*, 2001). Computational analysis of ncRNAs in *E. coli* by comparative genomics predicted 275 candidate structural RNA loci and experimentally identified at least 11 loci that apparently express small ncRNAs (Rivas *et al.*, 2001). Computational and experimental "RNomics" applied to model organisms from eubacteria to mice have identified numerous functionally important ncRNAs (Marker *et al.*, 2002; Vogel *et al.*, 2003; Zhang *et al.*, 2004; Washietl *et al.*, 2005a,b). For instance, 86 candidates have been identified in the Archaeon *Archaeoglobus fulgidus* (Tang *et al.*, 2002), 66 in *Drosophila melanogaster* (Yuan *et al.*, 2003) and 201 novel non-messenger RNA candidates in mice (Hüttenhofer *et al.*, 2001). The control of translation and mRNA degradation are important parts of the regulation of gene expression. Small RNA molecules are clearly ubiquitous and effective modulators of gene expression in many eukaryotic cells. The small RNAs that control gene expression can be either micro RNAs (miRNAs; endogenous or exogenous) or short interfering RNAs (siRNAs) and they affect mRNA degradation and translation. Therefore,

more ncRNAs need to be identified to elucidate the genetic network in model organisms.

The following ncRNAs in addition to tRNA and rRNA have been identified in *B. subtilis*: SRP RNA, tmRNA, RNase P RNA, BS190 RNA, BS203 RNA, SR1 RNA, RatA RNA, SurA RNA, SurC RNA, and the RNA transcribed from *polC-ylxS* intergenic region (Ando *et al.*, 2002; Barrick *et al*, 2005; Nakamura *et al.*, 1992; Fang *et al.*, 2001; Ito *et al.*, 2002; Silvaggi *et al.*, 2006; Suzuma *et al.*, 2002; Trotochaud and Wassarman, 2005). Suzuma *et al.* (2002) previously identified BS190 RNA that is transcribed as a 201-nt-long precursor and processed into a 190-nt-long mature product in *B. subtilis*. The gene for the BS190 RNA is located in an intergenic region (IGR) that is > 600 bp long, including the extension for a promoter and a transcriptional terminator. Therefore, I considered that IGRs of over 500 bp are candidate regions containing a novel ncRNA gene. The entire genome of *B. subtilis* 168 (sequenced in 1997) comprises an estimated 4,103 genes among which, those encoding proteins have been thoroughly annotated and functionally analyzed. Moreover, Kobayashi *et al.* (2003) estimated the minimal gene set required to sustain living cells. In prokaryotes, open reading frames (ORFs) belonging to the same operon are transcribed together into a single mRNA molecule. De Hoon *et al.* (2005) predicted the operon structure of the *B. subtilis* genome using operon length,

intergene distance and gene expression information. The predicted number of operons based on their method was 651 and is available from the DBTBS database (http://dbtbs.hgc.jp). Thus, 651 intergenic regions are located between operons. This information will help to reliably identify novel ncRNA genes in the intergenic regions of the *B. subtilis* genome.

The major aim of the present study was to detect and examine the functions of novel sRNAs of *B. subtilis* that are expressed from the intergenic regions. I selected 123 intergenic regions of 500 bp or more as candidate sRNA genes and identified seven unique RNA transcripts. Sequence analyses of the transcripts revealed that six out of the seven transcripts are ncRNAs without any ORFs harboring Shine-Dalgarno (SD) sequences. Deletion mutant analyses suggested that their functions are not essential for normal cell growth and sporulation.

MATERIALS AND METHODS

Bacterial strains, plasmids and their construction

Table 1 lists the strains used in this study. Mutants were constructed by replacing IGRs with the chloramphenicol resistance gene. Disruption cassettes were constructed by PCR fusion using a strategy similar to that described by Kuwayama *et al.* (2002). Briefly, DNA fragments of the upstream and downstream genes were amplified by PCR from *B. subtilis* chromosomal DNA using the primers listed in Table 2. A DNA fragment containing the chloramphenicol resistance gene was amplified from plasmid pDH88 (Henner, 1990). Three fragments were mixed and amplified using the outside primers to form a single fragment consisting of the upstream and downstream genes of the region encoding the sRNA. Chloramphenicol-resistant isolates were generated from double-crossover events and confirmed by Northern and Southern hybridization (data not shown). To examine whether or not the IGRs that produce detectable transcripts in *B. subtilis* have their own transcription units, DNA fragments corresponding to the IGRs were amplified by PCR and cloned into the T-plasmid, pGEM-T (Promega), prepared by cutting pGEM-5Zf(+) with *Eco*RV and adding a 3′ terminal thymidine to both strands. This system can be used to directly clone PCR-amplified fragments. The recombinant plasmid was used to transform *E. coli* for selection in L-broth-agar plates containing 50 μg ml^{-1}

ampicillin. The nucleotide sequences of the cloned region were confirmed by DNA sequencing.

Growth conditions and isolation of total RNA

Overnight cultures of *B. subtilis* strain 168 were diluted 1:100 with 100 ml of LB medium in 200 ml Erlenmeyer flasks, incubated at 37°C on a reciprocal shaker and then harvested at the times indicated and immediately frozen at –80°C until use. Total RNA was isolated from frozen cells essentially as described by Igo and Losick (1986). The *E. coli* transformants were grown in liquid LB medium containing 50 μg ml^{-1} ampicillin and total RNA was isolated as described by Chuang *et al.* (1993).

Northern blot analysis

I defined an IGR as the region between the last nucleotide of the translation stop codon of an upstream gene and the first nucleotide of the promoter sequence of a downstream gene in accordance with the annotations in the *Bacillus subtilis* genome database (BSORF; http://bacillus.genome.jp/). I serially numbered 123 IGRs of > 500 bp as sraA1 through sraA123 clockwise

from the replication origin of the *B. subtilis* genome (Table 3). Each candidate IGR was amplified by PCR and then labeled using digoxigenin (DIG) DNA labeling mix (Roche Diagnostics). The sequences of the PCR primers and the predicted sizes of the amplification fragments for each IGR are listed in Table 4. After electrophoresis on 6% polyacrylamide gels containing 6 M urea, total RNAs (10 μg in each lane) were electrotransferred onto Gene Screen Plus membranes (NEN Research Products). Specific transcripts on the membranes were detected using DIG detection kits (Roche Diagnostics) according to the manufacturer's instructions.

Primer extension analysis

Primer extension analysis proceeded essentially as described (Silvaggi *et al.*, 2006). I end-labeled 0.2 pmol of each oligonucleotide with [γ-^{32}P]-ATP (GE Healthcare) using T4 polynucleotide kinase (Takara) [6C#2 (5′-GAGAGGTAG-GGCTTTCGTTG-3′), k61C#2 (5′-CATTAATGCTAATGCTTGCA-3′), 62#1 (5′-GTATTGAGGCGATGCACCACACGC-3′), k72_1 (5′-TTCGTAAACAGTC-ATTTTCCCA-3′), k80#2 (5′-GCTTGAAATGTTGACACGTG-3′), k101C#1 (5′-GTGTGACAGCTGCATCGCTTT-3′)]. Thereafter, 20 μl of extension mix [50 mM Tris–HCl, pH 8.3, 75 mM KCl, 5 mM $MgCl_2$, 10 mM dithiothreitol,

0.125 mM dNTPs, 20 U ribonuclease inhibitor (Takara), 200 U reverse transcriptase M-MLV (Takara)] was added to the annealing mix and reverse transcription proceeded at 42°C for 60 min followed by 70°C for 10 min and ethanol precipitation. Total RNA (15 μg) isolated from 4-hour cultures of *B. subtilis* strain 168 grown in LB medium was used for reverse transcription. Sequencing ladders were generated using the same radiolabeled oligonucleotide that was used for primer extension and DNA was sequenced using 7-DEAZA Sequencing Kits (Takara) according to the manufacturer's instructions. The sizes of extended fragments were determined by 6% polyacrylamide/7 M urea gel electrophoresis followed by autoradiography.

Accession numbers

The nucleotide sequences of the BsrC, BsrE, BsrF, BsrG, BsrH, and BsrI RNAs have been deposited in GenBank (http://www.ncbi.nlm.nih.gov/Genbank/) under accession numbers BK006371, BK006376, BK006372, BK006373, BK006374, and BK006375, respectively.

RESULTS AND DISCUSSION

Experimental identification of sRNAs

I determined whether or not sRNAs were transcribed from the IGRs using Northern blot analyses. Each intergenic region was amplified by PCR using primers that can produce approximately 400 nt-DNA fragments. The PCR products were then DIG-labeled and used as hybridization probes. I prepared RNA from cells that had been growing in LB medium for various periods. The results showed that among 123 IGRs, 7 new RNA transcripts ranging in size of 70 to 380 nt were generated from IGRs sraA6, sraA7, sraA61, sraA62, sraA72, sraA80, and sraA101, and these IGRs are referred to as *bsrC*, *bsrD*, *bsrE*, *bsrF*, *bsrG*, *bsrH*, and *bsrI*, respectively (Fig. 1A, Table 5). Figure 1B shows the features of the RNAs with respect to expression time. All of them were detectable at 4 hours after inoculation (Fig. 1B, lane 2). The transcription levels of *bsrD*, *bsrG*, *bsrH*, and *bsrI* seemed to persist throughout the incubation and were detectable during the stationary phase (8 hours).

I also examined whether or not RNA transcripts were expressed from the 123 IGRs in cells grown in sporulation medium (2 × SG medium, Leighton and Doi, 1971) for 11, 13, and 15 hours. However, I did not detect any RNA transcripts at those time points (data not shown). Licht *et al.* (2005) analyzed potential promoters and terminators in the intergenic regions of the *B. subtilis*

genome *in silico* and identified SR1 RNA comprising of 205 nts that are expressed from the *slp-speA* intergenic region in the absence of glucose. When *B. subtilis* is grown in TY medium, SR1 RNA expression is maximal during the stationary phase; that is, sugar in the yeast extract is drained. Using a DNA probe corresponding to the *slp-speA* region (sraA40), I could detect a 210 nt long transcript with a transcription start site that was identical to that of SR1 RNA.

Identification of transcription unit for each small RNA gene

Although I detected seven novel sRNAs, the possibility could not be dismissed that the transcripts were produced by the cleavage of longer transcripts of upstream ORFs. To confirm that the potential genes encoding sRNAs constitute complete transcription units, *E. coli* strains harboring a plasmid with the IGRs of *B. subtilis* were constructed as described in Materials and Methods. Northern blotting using DIG-labeled probes for each region showed that *E. coli* did not contain genes that are homologous to the RNA transcripts detected in *B. subtilis* (Fig. 2, lanes 2 in each panel). Moreover, RNA bands were detected in *E. coli* transformants harboring recombinant plasmids with the PCR-amplified IGRs, sraA6 (*bsrC*), sraA61 (*bsrE*), sraA62 (*bsrF*), sraA72 (*bsrG*), sraA80 (*bsrH*), and sraA101 (*bsrI*) of *B. subtilis* (Fig. 2 A, C, D E, F and G). The RNA bands

detected in almost of all transformant and the sizes of RNA bands were the same as those in *B. subtilis*, except in that with the IGR sraA72 (Fig. 2 A, C, D, F, G and H). In our laboratory, it was previously reported that *bsrA* encodes BS190 RNA that is transcribed as a 201-nt-long precursor and processed into a 190-nt-long mature product in *B. subtilis* (Suzuma *et al.*, 2002). Figure 2H shows only a larger band corresponding to the RNA size marker at around 200-nt in the *E. coli* transformant harboring the plasmid with *B. subtilis bsrA* gene. This suggests that the precursor of BS190 RNA is not processed in *E. coli*. Unlike in *B. subtilis*, the RNA transcript from sraA7 was undetectable in *E. coli* (Fig. 2B, lane 3). DNA fragments corresponding to the IGRs between *guaA-yebB* were also amplified and ligated to pGEM-T, this region was not transcribed in *B. subtilis*. Northern blot analysis showed that they are not expressed in *E. coli* (data not shown), suggesting that the minor transcript expressed from the promoter in the plasmid vector, if present, could not be detected. These results demonstrated that the IGRs, sraA6, sraA61, sraA62, sraA72, sraA80, and sraA101 expressed sRNA via their own promoters rather than the sRNAs being simply the read-through products of preceding genes. Several smaller bands were also evident in addition to the major band in *E. coli* transformants harboring the recombinant pGEM-T plasmid with the *bsrG* gene (Fig. 2E, lane 3). Non-specific processing or degradation could have occurred in *E. coli*. RNA transcripts

expressed from the *bsrD* gene were undetectable in *E. coli* (Fig. 2B, lane 3). Although the 70-nt-long transcript was evident in Northern blots using the DIG-labeled probe specific for the *bsrD* gene in *B. subtilis*, no potential promoter sequence in sraA7 IGR was recognized by RNA polymerase containing σ^A. Moreover, I was unable to identify recognition sequences of a sigma factor unique to *B. subtilis* in the DBTBS database (http://dbtbs.hgc.jp, Sierro *et al.*, 2008) of transcription factor- and sigma factor-binding sites of *B. subtilis*.

Determination of the 5′ ends of the small RNAs

The 5′ ends of each sRNA were determined by primer extension analysis. The longest extension products of 129, 217, 81, 75, 92, and 100 nt were obtained from BsrC, BsrE, BsrF, BsrG, BsrH, and BsrI RNA, respectively (Fig. 3 A to F, filled arrowheads). These results suggest that the transcriptional start sites of BsrC, BsrE, BsrF, BsrG, BsrH, and BsrI RNAs are located at 196, 136, 469, 172, 311, and 299 bp downstream from the 3' end of the respective upstream gene. Consistent with the notion that the possible promoter is inactive, I could not detect an extension product from the BsrD RNA (data not shown). A 215-nt-long RatA RNA (Silvaggi *et al.*, 2006) is expressed from the IGR between *yqbM-txpA* (sraA80, in our study), and BsrH RNA is expressed from the same region (Fig.

3E). However, our primer extension analysis revealed that BsrH RNA and RatA RNA are encoded on different strands (Fig. 3E), and they do not overlap.

Effects of sRNA depletion on cell growth

To study the essential cellular functions of each RNA, I constructed null mutant strains as described in Materials and Methods (Table 1). The absence of intracellular sRNAs did not affect cell growth in liquid LB medium (data not shown). Under stressed conditions such as high temperature (48°C), high salt (4% NaCl) and 4% ethanol, the growth ratios of the null mutants NA6, NA7 and NA8 were the same as that of the parent strain, suggesting that the functions of BsrF, BsrG, and BsrH RNAs are not essential for growth under these conditions (data not shown).

CONCLUSION

I identified seven new RNA transcripts and showed that all of them except BsrD RNA are expressed via their own promoters in the IGR.

The secondary structure of each sRNA was predicted by using the RNAfold program (Hofacker, 2003), their 3′ ends were deduced from the 5′ ends and sizes calculated from electrophoretic migration (Fig. 1A). The predicted secondary structures of BsrC, BsrE, BsrF, BsrG, BsrH, and BsrI RNAs contain a distinctive ρ-independent terminator structure near their 3′ ends with minimal free energy values (ΔG) of -68.23, -132.35, -48.81, -106.74, -114.58, and -71.41 kcal/mol, respectively (Fig. 4). Transcription of the sRNA genes is likely to stop at the ρ-independent terminators.

A search of the Breaker Laboratory Intergenic Sequence Server (BLISS; http://bliss.biology.yale.edu, Corbino *et al.*, 2005) revealed that sequences to *bsrC*, *bsrE* and *bsrG* are conserved in other microbial genomes whereas *bsrF*, *bsrH* and *bsrI* are unique to *B. subtilis*. The similar sequence of *bsrC* is found in various species such as *Bacillus halodurans*, *Bacillus anthracis*, *Bacillus cereus*, *Xylella fastidiosa*, and *Methanosarcina acetivorans*. The sequence similar to *bsrE* is conserved only in *Bacillus anthracis* and in *Bacillus cereus*, suggesting that *bsrE* is restricted in Bacillus species. The sequence similar to *bsrG* is conserved in some Gram-positive bacteria and in Gram-negative cyanobacterium Nostoc PCC 7120. Interestingly, the nucleotide sequence of BsrE RNA shows

66.3% and 69.5% identity to the complementary sequences of BsrG RNA and BsrH RNA, respectively. The sequence of BsrG RNA shows 56.5% identity with that of BsrH RNA in a 283-nt overlap. Their complementary sequences indicate that BsrE, BsrG and BsrH RNA anneal to each other. Two highly similar sRNAs, GlmY and GlmZ in *E. coli* (Urban and Vogel, 2008), function in *glmS* gene expression. GlmY RNA antagonizes GlmZ RNA in this mechanism by hybridizing to GlmZ RNA. The functions of BsrE, BsrG and BsrH RNAs remain unknown and I do not have any evidence that they anneal to each other within the cell.

To test whether Bsr RNAs are untranslated RNAs, I searched for ORFs in which a putative initiation codon (ATG, TTG, or GTG) was preceded at an appropriate distance by a SD sequence. I found a weak SD sequence (5′-AGGA-3′) which shows a minimal, but identical, degree of complementarity to the 3′ end of 16S rRNA in the sequence of BsrC, BsrE, and BsrI RNA. Band and Henner (1984) showed that a message with the weak SD sequence shows no detectable translation initiation in *B. subtilis*. In addition, the distance from the weak SD sequence to the ATG of BsrC, BsrE, and BsrI RNA were 21, 19, and 25 nucleotides, respectively. Vellanoweth and Rabinowitz (1992) showed that messages with spacings greater than 10 nucleotides were translated less efficiently compared to the spacing optimum of 9 nucleotides. On the other hand,

there were no sequences that resembled a SD sequence in the sequence of BsrF, BsrG, and BsrH RNA. Judging from the SD strength and the distance between SD sequence and potential initiation codon, it is unlikely that Bsr RNAs identified here encode protein products.

In summary, I identified and characterized six genes that encode novel sRNAs. Since they have neither an apparent SD sequence (5′-AGGAGG-3′) nor an ORF, they appeared to exert action as ncRNAs. Although none of them had any effect on growth in LB liquid medium, understanding the mode of action of these gene products should provide new insights into the control of gene expression in *B. subtilis*.

TABLES

Table 1

Bacterial strains and plasmids used in this study

Strain or plasmid	Relevant genotype	Source or reference
***B. subtilis* strains**		
168	*trpC2*	Laboratory stock
NA3	168 [*bsrC*: :*cat*]	This study
NA4	168 [*bsrD*: :*cat*]	This study
NA5	168 [*bsrE*: :*cat*]	This study
NA6	168 [*bsrF*: :*cat*]	This study
NA7	168 [*bsrG*: :*cat*]	This study
NA8	168 [*bsrH*: :*cat*]	This study
NA9	168 [*bsrI*: :*cat*]	This study
Plasmids		
pGEM-T	f1 *ori*, ColE1 *ori*, *lacZ*, MCS, Ap^R, T7- and SP6-Promoter	Promega (Madison, Wis., USA)
pBSR1	pGEM-T carrying *bsrA*	This study
pBSR3	pGEM-T carrying *bsrC*	This study
pBSR4	pGEM-T carrying *bsrD*	This study
pBSR5	pGEM-T carrying *bsrE*	This study
pBSR6	pGEM-T carrying *bsrF*	This study
pBSR7	pGEM-T carrying *bsrG*	This study
pBSR8	pGEM-T carrying *bsrH*	This study
pBSR9	pGEM-T carrying *bsrI*	This study

Table 2

PCR primers used in gene disruption experiment

Primer*	Sequence (5′→3′)
F6U	CTACAGTCTGTTAAAGCGTG
F6D	ACCTCTGACACATGACCGTCCGGCGT
R6U	GAGCGTACGCGAATATCGCCCCTG
R6D	GCCGGGATCCAATTCCGTTGATCAAACCCG
F7U	GCGCAAGCTTGCGTTCAGTTGCTAATTACA
R7U	ATTTGAGCGTACGCAGCAAACTCA
F7D	CCATGACCCAGACGGTCGGCGGT
R7D	GCCGGGATCCCAGCGGAAATTGCTGCAAAG
F61U	GCGCAAGCTTCCTCTCTAATTGTTGT
R61U	AGGATTTGAGCGTACGCGTCCTGTAATTGGGCG
F61D	AAACCTCTGACACATGCAGAGAGACTTGGAAAGC
R61D	GCCGGGATCCCAGAAGGTGCAAGTC
F62U	GCGCAAGCTTGGTCACCATCA
R62U	GGATTTGAGCGTACGCGGGTCGGCGTATCCAC
F62D	GCAGCTCCCGGAGACGGGATCCGGCGGGGCGG
R62D	GCCGGGATCCCACCCGAGCGTCACA
F72U	GCGCAAGCTTGATCATCTAAGGATGGC
R72U	GAGCGTACGCGAAAAGGAGGATTTAGGA
F72D	ATGCAGCTCCCGGATTGGAACCCAGT
R72D	GCCGGGATCCCACAGGTGATAGCAA
F80U	GCGCAAGCTTCAAGCAATCGAGCCC
R80U	AGGATTTGAGCGTACGCGTGGCAGCACTCTAGC
F80D	AAACCTCTGACACATGCAGTCGTAAATATCTTC
R80D	GCCGGGATCCGGTACGGCCCCTG
F101U	GCGCAAGCTTGAGCCGCGTGAAG
R101U	AGGATTTGAGCGTACGCATGTGCAAGGCACGA
F101D	AAACCTCTGACACATGCAGAACAAATGAGAATGG
R101D	GCCGGGATCCCTGGCATTACGGT

*, F; Forward primer, R; Reverse primer, U; upstream region of sRNA gene, D; downstream region of sRNA gene. A number separating two alphabets indicates the serial number of sRNA gene.

Table 3

Candidates for sRNAs subjected to experimental confirmation

Candidate*	5′ end†	3′ end†	Length†	5′ Gene	Strand	3′ Gene	Strand
sraA1	176308	177082	774	*rrnG-5S*	>	*ybaR*	>
sraA2	204874	205394	518	*adaB*	>	*ndhF*	>
sraA3	210930	211844	912	*ybcI*	>	*ybcL*	>
sraA4	375964	376536	573	*yckH*	>	*srfAA*	>
sraA5	429666	430202	537	*phrC*	>	*yclM*	<
sraA6	473779	474283	505	*ydaG*	>	*ydaH*	>
sraA7	549383	549952	570	*yddM*	>	*yddN*	<
sraA8	552872	553423	552	*yddQ*	>	*yddR*	>
sraA9	555693	556475	783	*yddS*	>	*yddT*	>
sraA10	557163	558121	959	*yddT*	>	*ydeA*	>
sraA11	559179	559864	686	*cspC*	>	*ydeB*	<
sraA12	560327	560893	567	*ydeB*	<	*ydzE*	<
sraA13	578648	579254	607	*ydeS*	>	*ydeT*	>
sraA14	586786	587457	672	*ydfG*	<	*ydfH*	>
sraA15	599058	599942	885	*ydzH*	<	*ydfR*	<
sraA16	634095	634825	731	*ydhU*	<	*trnE-Arg*	>
sraA17	651585	653147	1563	*groEL*	>	*ydiM*	>
sraA18	653529	654694	1166	*ydiM*	>	*ydiN*	>
sraA19	658089	659339	1251	*ydiQ*	>	*ydiR*	>
sraA20	662745	663317	573	*ydjA*	>	*ydjB*	>
sraA21	675575	676158	584	*ydjJ*	<	*ydjK*	>
sraA22	821839	822410	572	*yfmG*	>	*yfmF*	<
sraA23	874309	875931	1623	*yfjO*	>	*yfjN*	>
sraA24	874309	1018451	531	*phoA*	<	*lytE*	>
sraA25	1070719	1071517	799	*prsA*	<	*yhaK*	<
sraA26	1158554	1159394	841	*asnO*	>	*yisP*	>
sraA27	1218578	1219320	743	*trpS*	<	*oppA*	>
sraA28	1232568	1233084	517	*yjbG*	>	*yjbH*	<
sraA29	1262410	1263172	763	*trnSL-Val2*	>	*yjcM*	<
sraA30	1264849	1265357	509	*yjcN*	>	*yjcO*	>

*, Candidate names are arbitrary codes, not final gene names. Candidate loci are relative to the *B. subtilis* genome database, BSORF (http://bacillus.genome.jp/).

†, The length is equal to the absolute value of the '3′ end' minus the '5′ end'. The intergenic region is defined by the both sides of 'adjacent genes'. For example, *sraA5* is encoded in the intergenic region delineated by the *phrC* gene on the left-hand side, which is transcribed from the minus strand, and the *yclM* gene on the right-hand side, which is transcribed from the plus strand.

Table 3-continued

Candidate*	5′ end†	3′ end†	Length†	5′ Gene	Strand	3′ Gene	Strand
sraA31	1269552	1270101	550	*yjdB*	<	*manR*	>
sraA32	1278365	1278983	619	*yjdJ*	<	*ctaO*	<
sraA33	1356885	1357401	517	*ykcC*	>	*htrA*	<
sraA34	1394974	1395478	505	*ykzD*	>	*ykoK*	>
sraA35	1441255	1441812	558	*ykvM*	>	*ykvN*	<
sraA36	1445803	1446716	914	*ykvQ*	>	*ykvR*	>
sraA37	1472438	1473003	566	*patA*	<	*cheV*	>
sraA38	1482876	1483520	645	*ykuI*	>	*ykuJ*	>
sraA39	1526155	1526689	535	*ykrA*	>	*ykrB*	<
sraA40	1533107	1533683	577	*slp*	<	*speA*	<
sraA41	1779589	1780596	1008	*mutL*	>	*ymcC*	<
sraA42	1783999	1784501	503	*pksC*	>	*pksD*	>
sraA43	1787342	1787979	638	*pksE*	>	*pksF*	>
sraA44	1861955	1862689	735	*aprX*	<	*ymaC*	>
sraA45	1872821	1873443	623	*cwlC*	<	*spoVK*	>
sraA46	1882933	1883491	559	*ynaD*	>	*ynaE*	>
sraA47	1884895	1885496	602	*ynaG*	>	*ynaI*	>
sraA48	1885782	1886605	824	*ynaI*	>	*xynP*	>
sraA49	1899769	1900371	603	*yncF*	>	*ynzH*	<
sraA50	1900633	1901473	841	*ynzH*	<	*thyA*	>
sraA51	1903519	1904249	731	*yncM*	<	*cotC*	<
sraA52	1916353	1916893	541	*yndN*	>	*lexA*	<
sraA53	2027836	2928683	848	*yoaG*	<	*yoaH*	<
sraA54	2930289	2031015	727	*yoaH*	<	*yoaI*	>
sraA55	2035976	2936853	878	*yoaM*	>	*yoaN*	<
sraA56	2040283	2040829	547	*yoaQ*	>	*yozF*	<
sraA57	2053184	2053852	669	*pps*	<	*xynA*	<
sraA58	2054495	2055531	1037	*xynA*	<	*yobD*	>
sraA59	2058894	2059491	598	*yobF*	<	*yozJ*	<
sraA60	2059948	2061404	1457	*yozJ*	<	*rapK*	>
sraA61	2068285	2069498	1214	*yobI*	<	*yobJ*	<
sraA62	2077882	2078468	587	*yobO*	>	*csaA*	<
sraA63	2091560	2092153	594	*yocG*	>	*yocH*	<
sraA64	2095041	2095604	564	*yocI*	<	*yocJ*	<
sraA65	2117162	2117766	605	*yojk*	<	*yojJ*	<

Table 3-continued

Candidate*	5′ end†	3′ end†	Length†	5′ Gene	Strand	3′ Gene	Strand
sraA66	2159228	2159807	580	*yosR*	<	*yosQ*	<
sraA67	2211738	2212323	586	*yopE*	<	*yopD*	<
sraA68	2220629	2221580	952	*yonR*	<	*yonP*	>
sraA69	2224857	2226537	1681	*yonN*	>	*yonK*	>
sraA70	2240378	2241005	628	*yomU*	>	*yomT*	>
sraA71	2246584	2247129	546	*yomL*	<	*yomK*	>
sraA72	2272605	2273229	625	*yolA*	<	*yokL*	<
sraA73	2277320	2277842	523	*yokH*	<	*yokG*	<
sraA74	2281727	2282376	650	*yokD*	>	*yokC*	<
sraA75	2433580	2434244	665	*ypuB*	<	*ypzD*	>
sraA76	2465404	2465949	548	*yqzH*	>	*yqjV*	>
sraA77	2515136	2515666	531	*yqiH*	<	*yqiG*	>
sraA78	2516786	2517300	515	*yqiG*	>	*spoOA*	<
sraA79	2646379	2647185	807	*yqeG*	<	*yqeF*	<
sraA80	2677690	2678411	722	*yqdB*	>	*yqbM*	<
sraA81	2701413	2701956	544	*yrkS*	<	*yrkR*	<
sraA82	2705885	2706395	511	*yrkO*	>	*yrkN*	>
sraA83	2708689	2709270	582	*yrkK*	<	*yrkJ*	<
sraA84	2713410	2714201	792	*yrkD*	<	*yrkC*	<
sraA85	2733413	2734221	809	*yrdB*	<	*yrdA*	<
sraA86	2739463	2740625	1163	*yrpD*	>	*yrpE*	<
sraA87	2749619	2750150	532	*yraK*	>	*yraJ*	<
sraA88	2755072	2755578	507	*yraB*	<	*adhA*	>
sraA89	2765110	2765825	716	*levR*	<	*aapA*	>
sraA90	2777690	2778192	503	*yrhH*	<	*yrzI*	<
sraA91	2813395	2814007	613	*yrvM*	<	*aspS*	<
sraA92	2896072	2896886	815	*ilvB*	<	*ysnD*	>
sraA93	2897847	2898466	620	*ysnE*	>	*ysnF*	>
sraA94	3074246	3074752	507	*ytgP*	<	*ytfP*	>
sraA95	3128109	3128610	502	*metK*	<	*pckA*	>
sraA96	3177713	3178375	663	*rrnB-16S*	<	*yuaJ*	>
sraA97	3212399	3212922	524	*tgl*	>	*yugU*	<
sraA98	3250283	3250799	517	*mrpD*	>	*mrpF*	>
sraA99	3334049	3334808	760	*pucM*	>	*pucE*	<
sraA100	3353994	3354651	658	*yurT*	<	*yurU*	<

Table 3-continued

Candidate*	5′ end†	3′ end†	Length†	5′ Gene	Strand	3′ Gene	Strand
sraA101	3359840	3360363	524	*yurY*	<	*yurZ*	<
sraA102	3420128	3420833	706	*yvsH*	<	*yvsG*	>
sraA103	3534534	3535071	538	*pbpE*	<	*sacB*	>
sraA104	3630633	3631209	577	*yvyd*	<	*fliT*	<
sraA105	3665568	3666268	701	*gtaB*	>	*ggaB*	<
sraA106	3670436	3672623	2188	*ggaA*	<	*tagH*	<
sraA107	3699685	3700467	783	*ywsC*	<	*rbsR*	>
sraA108	3732329	3732908	580	*ywqC*	<	*ywqB*	>
sraA109	3768163	3769158	996	*ureA*	<	*csbD*	<
sraA110	3855225	3855815	591	*thrZ*	<	*mmr*	<
sraA111	3906111	3906639	529	*ywcI*	<	*vpr*	>
sraA112	3917058	3917667	610	*qoxA*	<	*ywzA*	>
sraA113	3921760	3922715	956	*ywcC*	<	*ywcB*	>
sraA114	3950456	3951068	613	*menA*	<	*dltA*	>
sraA115	3977990	3978544	555	*cydA*	<	*yxkJ*	>
sraA116	3995100	3995620	521	*pepT*	>	*yxjJ*	>
sraA117	4003926	4004540	615	*yxjB*	<	*yxjA*	>
sraA118	4039665	4040271	607	*yxiA*	<	*hutP*	>
sraA119	4095701	4096208	508	*yxbC*	<	*yxbB*	>
sraA120	4129253	4129785	533	*fbp*	>	*yydD*	<
sraA121	4165797	4166317	521	*yybS*	>	*cotF*	<
sraA122	4170561	4172321	1761	*yybO*	<	*yybN*	>
sraA123	4186383	4186888	506	*yyaP*	>	*tetB*	>

Table 4

PCR primers used for amplification of the intergenic regions

Primer*	Sequence (5′→3′)	Primer*	Sequence (5′→3′)	Size (bp)†
sraA1F	CGATTTGCTTTCGTCAGTAT	sraA1R	TACGGGAGAAAATGAAAAGA	651
sraA2F	CGTAACTTAATAATCGATGC	sraA2R	TTCCAACAAAAAAGCCGACA	319
sraA3F	CTGCCGATTCTTTTGATTTA	sraA3R	AAACACATCGTAATATGTCG	319
sraA4F	GGTACACATAGTCATGTAAA	sraA4R	CTCTTTCTTATCCATATCAG	319
sraA5F	GCCTTCATGGAGATTACGTT	sraA5R	CTCTTGTCCACTATTATCCC	386
sraA6F	TTTCTGTTTTGAGAGAGAAG	sraA6R	GGAGGTTTCAGTACGATGT	371
sraA7F	GAACGACCAAACAATGGT	sraA7R	TGAACGAATAGGACATGGT	408
sraA8F	GTTCAGAAAAATACTTCTGG	sraA8R	GTATAATATCAGCAGTGTTC	431
sraA9F	TAGTAAGGAGTTGACATGG	sraA9R	AAGCAATAACAGCGAGG	651
sraA10F	GTAAAATCCCTCGTCTGAGG	sraA10R	ATACTCTATCTTCCGTACCT	838
sraA11F	ATAAAAAACCCTACTCAACG	sraA11R	TGTTTAGAACATTAACGACC	561
sraA12F	CAAGCAAAAACAATAGGGA	sraA12R	TTCCCCTTTACTTCTTTATG	394
sraA13F	AAGATTTTTGTTGCTGAGCG	sraA13R	CCAAATGATTATACATTCCC	494
sraA14F	CATTTTACCAATGTGCGGT	sraA14R	CGAACCCATATTTCAACGTG	572
sraA15F	TCATTGCATCAACATAATGG	sraA15R	TGCCGTTTTGTTTAAGCAAC	750
sraA16F	GCGCTCCAATAATATAATGA	sraA16R	CGAATCTTTTTCAATTCGAC	614
sraA17F	AAAAGAGAAGGTCTTTCATC	sraA17R	AAGTGAAAGATATATCAGGC	1442
sraA18F	GCAGGCTATATTGTTTTCC	sraA18R	TCTAGAAGCCTTAGATCAC	1066
sraA19F	CGATCCCATTAGATGGAAGT	sraA19R	GGAATTGTTTGAGAGACCAA	1093
sraA20F	AAAACAGTATTTTGAGCTGG	sraA20R	GCAAAGAAGCACAAGATTTA	427
sraA21F	CCTGAGTCTGAAACAATTGA	sraA21R	TTCCTCACATTTCGTAAGCA	467
sraA22F	CAAAATTATCAGGGATTGAG	sraA22R	AACTCGTCATTCTTCTTATG	369
sraA23F	ATAAACAATTCCGATTTCCG	sraA23R	TTAAAAACTGACCATGCTTC	1514
sraA24F	CGGCTGTTGATCATGTAAA	sraA24R	TATAAGGATAGATTCCCATC	410
sraA25F	TAACATATTGTGGCTTTCTG	sraA25R	TTGGTAATGTAATATGGGAC	690
sraA26F	AGGATAGATGAGCAGGGA	sraA26R	AGTAACAAGACGCACAGCAT	722
sraA27F	CCCTGTAAAAAAGGGACG	sraA27R	ATATCAGACAAATGAACCTG	639
sraA28F	AAGGTTTTCCGGTGATTC	sraA28R	ACGTTCCATTTTTAGCACTA	417
sraA29F	AGTTCAATAGTAACAAACGG	sraA29R	TGGGAAAGCAATCCCCT	653
sraA30F	CAGAGAAATTTGTGCTTCG	sraA30R	GCTTTTTGTTCTGCGATA	398

*, F; Forward primer, R; Reverse primer, Letters followed by 'F' or 'R' indicates the target region.

†, The sizes of the PCR products used for preparation of the DIG-labeled DNA probes.

Table 4-continued

Primer*	Sequence (5′→3′)	Primer*	Sequence (5′→3′)	Size (bp)†
sraA31F	GTGAACGATCAATTGATTTG	sraA31R	GTTTCGTTTTTTACCGAAGC	356
sraA32F	ACCACCGTATGGGTTAT	sraA32R	TCTAATAGAAATGTCATGGC	517
sraA33F	ATCTAAAGCAAAAGATGCTG	sraA33R	GACCTAATTGTTCTATGAAAC	414
sraA34F	CTTTTTTGAGTTGACATAGC	sraA34R	CTTTTTTGAGTTGACATAGC	403
sraA35F	ATCCGAACATTCAGCGT	sraA35R	AACAGATGAAGAACAGTCAG	457
sraA36F	CTGTCAATCACGATAAAAAG	sraA36R	CAATGAATCTGTTCTTAACG	810
sraA37F	TATTTCGATTGTATGCGCCT	sraA37R	TGGCAGCTCTTCAAAATCGA	422
sraA38F	TGACATTGATACTGAGAATCA	sraA38R	CAAGTTCTATTGAGCATCAT	536
sraA39F	AAAGGTGCAACGAAACGA	sraA39R	TCACGTTAAATCCCCCA	435
sraA40F	AGAGACGGCTTAATCTGC	sraA40R	CAACAGGACATTGACATTG	477
sraA41F	GTACTCCTTCGAGTGTC	sraA41R	TGCTATACTATTTTCGTCAG	902
sraA42F	AAGATTCAATCGTTCGATTG	sraA42R	TCCATGCATGTTGTTAAAGC	374
sraA43F	CCCGAAGCATAAAATGGCTT	sraA43R	TAAATCCTCTAGCACCATCG	507
sraA44F	AGGTTTGGTAAGTTTGTTTG	sraA44R	CATCAGTGAAAAAGTTCCC	601
sraA45F	CATTCAACTGAAAAATGAGG	sraA45R	ACAAACGCCTATCATGTGT	517
sraA46F	TATGACAGTGATTAAATGGG	sraA46R	AATTGTAATTCCTTCCTGTG	425
sraA47F	CTTCAGAATTGCTTGGC	sraA47R	GGAGATCAATATCTTCTGG	487
sraA48F	CATAGTGTTTAACACAGCC	sraA48R	CTCGAATTTACGGTCAG	697
sraA49F	TTTCAGCCAAGGTACATATA	sraA49R	TATTTATCGACGAACTCAGA	500
sraA50F	GGTATGGTCAAATGAAATAAG	sraA50R	CGAATGCAATATTATCGCAA	739
sraA51F	GAATGATGTAAGGTCGAAAC	sraA51R	GTATATTTCTGAAGCTTCCC	600
sraA52F	GAGACTTTATGTTTGGGG	sraA52R	GGG GGT TTG TTG TTC AA	441
sraA53F	CTTTTATCGAGATGTCTTCC	sraA53R	CTGTGGTTGTTTCTAAAAAC	729
sraA54F	TAATCCCCAGCAATTTTGTT	sraA54R	GTTCAATCATGGTCCTTC	630
sraA55F	GGTGCCAATCCCTCTAGT	sraA55R	AGCTGAATTTTGGCCTGT	778
sraA56F	TGTTAAGTTGAAATGCGCAA	sraA56R	TTGTGTCAAACCAAAAGAAG	443
sraA57F	TTAACTAGAAAGCGCAC	sraA57R	CGTTAGTAATGGTTAAAGGT	505
sraA58F	CTAACGCCAAAGCTTTTGAT	sraA58R	AATTTTCACGATTTCTGGCG	880
sraA59F	AAATTGCTCCGTTCTC	sraA59R	TCGATAACGATCGCAGA	494
sraA60F	TCCATTAGTCACTTCCAAAG	sraA60R	GGATGTAAAGTTTTCTGC	1332
sraA61F	CTGGAGCAAAGAACTAA	sraA61R	GCTTTCCAAGTCTCTTGA	1113
sraA62F	GATCTCTTTTCTAGGTTC	sraA62R	TGCACCACACGCCACAT	494
sraA63F	CGTTTGAGAACCTCGTCGCA	sraA63R	TGTTCTGCTTAGAGCTGT	496
sraA65F	CGCGATATCTTACAGGC	sraA65R	GGGCCTTAAGAATCCTA	486

Table 4-continued

Primer*	Sequence (5′→3′)	Primer*	Sequence (5′→3′)	Size (bp)†
sraA66F	GGCTCCACTGTTGCTTT	sraA66R	CGAAACGGGAGGCATCC	472
sraA67F	GGTGATCTTGTGGAAGA	sraA67R	CGGACAGTTTTTGTTTGC	456
sraA68F	CGGGAGAAACTAAAGGT	sraA68R	CGCTAGTGTGTTAATAGC	806
sraA69F	ATGGAGGAAATTGTCGAC	sraA69R	GCCCGATGATCCGAAGC	1546
sraA70F	CACCTCTGCAAAAGATGATG	sraA70R	AGATCAACTGTACCGGCTGC	543
sraA71F	GCAATTACAGCGAGGTA	sraA71R	GGTTTCTGAGAGTACGT	427
sraA72F	GAACCTGCATTAAGCAAG	sraA72R	CACCAAGCAATAGTATTGC	491
sraA73F	GGGATGTTTCGGCATCT	sraA73R	GGAAAAGTCAGCACCATAG	427
sraA74F	CATTGAGCTTGGAAGGAAAT	sraA74R	AGAATGAAGAGGCGAGCCGC	485
sraA75F	ATCTTCAGTCGTCCAAC	sraA75R	GCTAGAGTTGATTAGCG	548
sraA76F	CACCGCCTGAAACAAATAAA	sraA76R	CCTTCTCCAATAGATCGTTC	443
sraA77F	CGAGCAAGTTCCATGATTTT	sraA77R	CGACTGAATATCTTGAGTGC	379
sraA78F	AGAGCTCCCCACCAACT	sraA78R	GCGCTGATAAATAGGAG	380
sraA79F	TCGAACTTGCGCCTAAA	sraA79R	GGAGGCGCTTCACATGTAT	696
sraA80F	GGTACCAACTATAAGCTT	sraA80R	CCGTACCAAATGGAAAC	543
sraA81F	AAAAACCGATGGTTTGTACC	sraA81R	CTCTATCTGTTTAATGTCAG	385
sraA82F	CGGACAGATTTGATATATGG	sraA82R	CTGCGAGGCCTTTTCACAAA	387
sraA83F	GAGAGAAGAATCTGCCT	sraA83R	GACACTTGGCCGCGTTC	432
sraA84F	GTTGCCAGTTCATAACTG	sraA84R	CCAGCATGAAACATGGG	639
sraA85F	TTGTAGTTTTCGGCGAG	sraA85R	CCAAACCACTTCAAATGAC	591
sraA86F	GCTTAACATGCCCCTTT	sraA86R	GAGACCATTCAGTATCC	1024
sraA87F	CTGAGAGGTTACACGAC	sraA87R	TGGTTTTTGCTGCGATG	382
sraA88F	CAAGTGGTTAGCTGCTG	sraA88R	GGTGTCTTCAACATAACG	374
sraA89F	AGTCCGCGTTGAGATTT	sraA89R	TACATGTCTTGTGGCTC	556
sraA90F	GTGAAGGAAGGTACCA	sraA90R	TACAATGGAGAAGGACC	367
sraA91F	AAAGCCCTTCCGCCCAC	sraA91R	GGATTGCAAGATGATCTG	663
sraA93F	CACCTTTATGATACACGC	sraA93R	GCCGTTATCAATAGGAG	480
sraA94F	GCAATCCTGTGAACATTCAT	sraA94R	TAAATTCAGAAAACGCGCTG	357
sraA95F	CTCGCTCCCTCTTATAC	sraA95R	GTGTGTGCATTTCCATTG	354
sraA96F	TCGTGCAGGCCCTGTTT	sraA96R	ACGCGGAAGGATGAGAA	560
sraA97F	CCAAAATATTGGTTCGCAGC	sraA97R	ACGCTGCGGCTTTATCGAAT	381
sraA98F	CGCATGGCATCTCAAATTTT	sraA98R	CCTCGATATCCATTGCATGA	410
sraA99F	GATCCTAGGACAGTACT	sraA99R	CGTTGTGGGAAAACACA	658
sraA100F	CTGGCTACAAGCAAAAC	sraA100R	TAAAATGGCCGGCGTGC	557

Table 4-continued

Primer*	Sequence (5′→3′)	Primer*	Sequence (5′→3′)	Size (bp)†
sraA101F	CGTGCCTTGCACATTCC	sraA101R	CGTCTCATCTGCCGCTA	384
sraA102F	GAAACACACACCCTTGA	sraA102R	CAACAGGACTTTGGACG	548
sraA103F	ATATACCTGCCGTTCAC	sraA103R	ATGTAAGGGGTGACGCC	338
sraA104F	TATTGGCGGTTCGGAGG	sraA104R	TTACCATCGAAGAGGGG	387
sraA105F	AACGGCCTGCCGCAAAAATT	sraA105R	CACTTCAAATGTTGGGAGAA	585
sraA106F	AGAAGCCTTGCCTCTTCTAA	sraA106R	AGCTCTTTATCCTGCATCGC	2045
sraA107F	ATGAATTCTCAATCGGCAT	sraA107R	CAATCCGAAGTTAGCAATAA	672
sraA108F	GAAACCGCATTTCTCCA	sraA108R	TATATCGCAACACATCGGCC	477
sraA109F	GGCGGGACATCACCTCCCTT	sraA109R	AGCTGATTTGGTCAAGGTAA	849
sraA110F	TTCCACCCAAGTTCCCGACA	sraA110R	CACTGATAAGGACACGAGAT	447
sraA111F	CAAGAAAAGACGGCCGATTT	sraA111R	ATCCTTGTGTCATTTGGACC	390
sraA112F	CTCTACTTTGTTATGCTATC	sraA112R	CCATCCGATAATAATAGCG	506
sraA113F	CGGTTATGCTATTCGGT	sraA113R	CTGAAAGTCCGATTTATGC	826
sraA114F	CTCCTTTTGCTTCCTTATC	sraA114R	CGCGGTGTGAAAACCAT	504
sraA115F	GCTCACTACTTCACAAATAG	sraA115R	TCAACTTTTCTCCTCTTACC	425
sraA116F	CCTCCGTAAAACAAGTATCA	sraA116R	TTGCTCTTGTTGATGATTGG	396
sraA117F	AAACCTAGGCCTATCTC	sraA117R	CACAAATTCTGGAAGTGC	507
sraA118F	CATATTCATCCAGATAGGTG	sraA118R	ATTGTAATAGCGTGACGCAG	412
sraA119F	GTTGTGCGTGATAGCCTATT	sraA119R	CAAACGTTAATCGTTTTGGA	375
sraA120F	CTGTTTGCCTTGGGGAG	sraA120R	CCCATAATGTCAAGGGTG	406
sraA121F	TTTTGTTCATCGCTGCAACA	sraA121R	CACCTTATATAGGATACGGG	366
sraA 122F	CATTTACCTTCTCCTCG	sraA122R	TAAATGAAGACCGTAATGGG	1596
sraA123F	CATAAATAATGGGTTCCCC	sraA123R	CATGGCATCAATAGCATG	388

Table 5

Summary of newly identified sRNAs

Gene*	Acc. No.†	Adjacent genes‡	Strand	Size (nt)§
bsrC	BK006371	*ydaG* (>) *ydaH* (>)	<	280
bsrD		*yddM* (>) *yddN* (<)	N.D.	70
bsrE	BK006376	*yobI* (<) *yobJ* (<)	<	380
bsrF	BK006372	*yobO* (>) *csaA* (<)	>	110
bsrG	BK006373	*yolA* (<) *yokL* (<)	>	300
bsrH	BK006374	*yqdB* (>) *yqbM* (<)	>	290
bsrI	BK006375	*yurY* (<) *yurZ* (<)	<	280

*, The column labeled "Gene" indicates gene names assigned to candidates that were experimentally confirmed by Northern analysis and submitted to DDBJ/EMBL/GenBank database.

†, Accession number of sRNA under DDBJ/EMBL/GenBank database.

‡, The names of adjacent genes and transcriptional direction of adjacent genes. The gene transcribed from each plus or minus strand in *B. subtilis* genome is indicated by > or < respectively.

§, The sizes of sRNA estimated from Northern analysis using oligonucleotide probes.

FIGURES

Figure 1. (A) Identification of novel sRNAs by Northern blot analysis. Total RNA was isolated from *B. subtilis* 168 cells grown in LB medium for 4 hours at 37°C. Total RNA (20 μg) was loaded in each lane and separated on 6% polyacrylamide gels containing 6 M urea. Arrows indicate predominant bands. Positions of RNA standard markers (nt, nucleotides) are indicated on left of each picture. (B) Expression profiles of sRNAs. Harvest time points are indicated above each lane. Samples of total RNAs equivalent to 10 μg of RNA, were loaded in each lane. The same amounts of total RNAs applied to Northern blots were resolved in 0.8% agarose gels and 16S and 23S rRNA bands are shown.

Figure 1

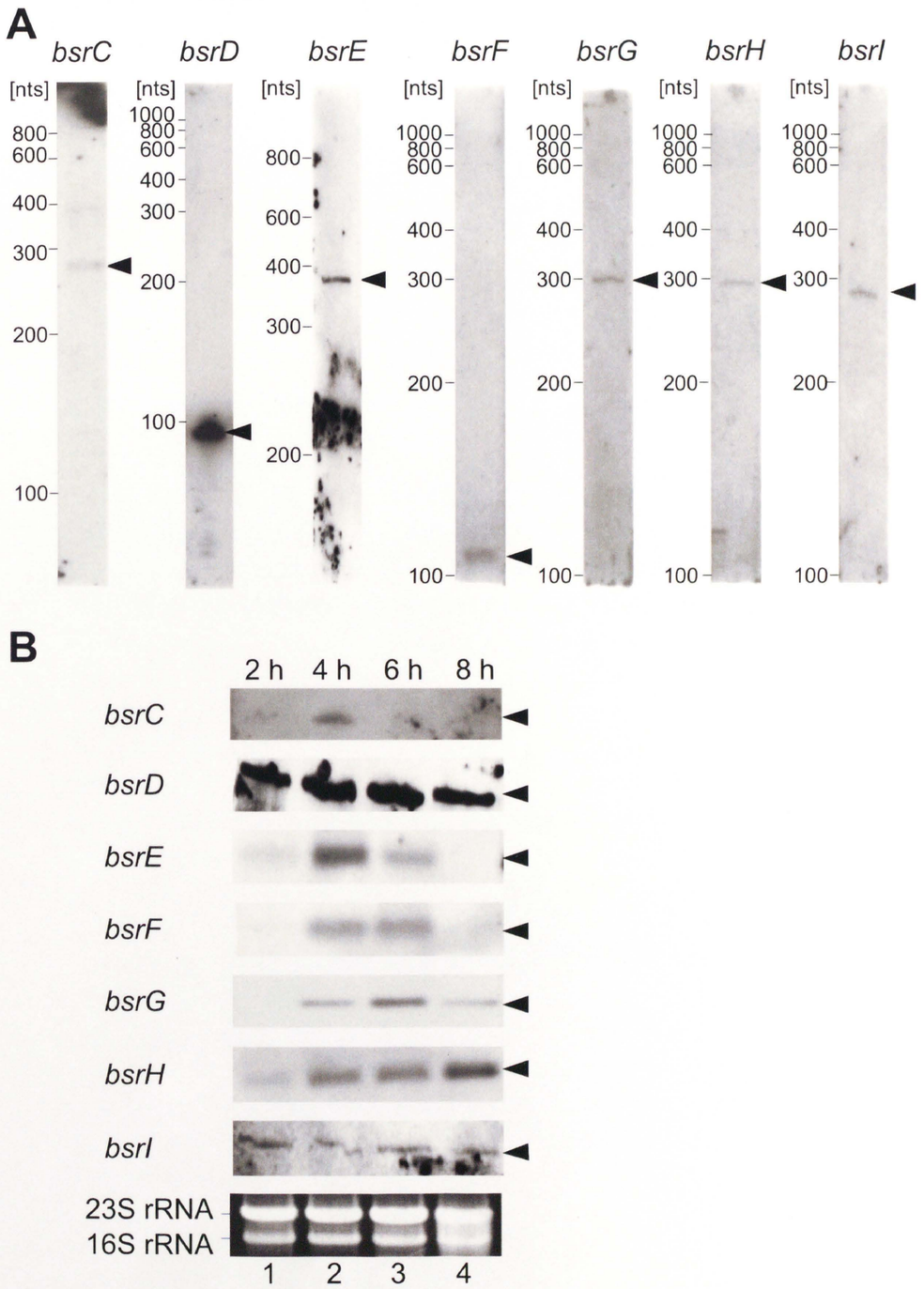


Figure 2. Expression of Bsr RNAs in *E. coli* cells.

Expression of each sRNA was analyzed in *B. subtilis* strain 168 (lanes 1) and *E. coli* strain JM109 harboring pGEM-T without (lanes 2) or with (lanes 3) sRNA gene, respectively (panel A, *bsrC*; panel B, *bsrD*; panel C, *bsrE*; panel D, *bsrF*; panel E, *bsrG*; panel F, *bsrH*; panel G, *bsrI*). Arrows indicate predominant bands. Positive control is *B. subtilis bsrA* gene encoding BS190 RNA (panel H, lane 3) transcribed as 201-nt precursor and processed into 190-nt mature product (Suzuma *et al.*, 2002).

Figure 2

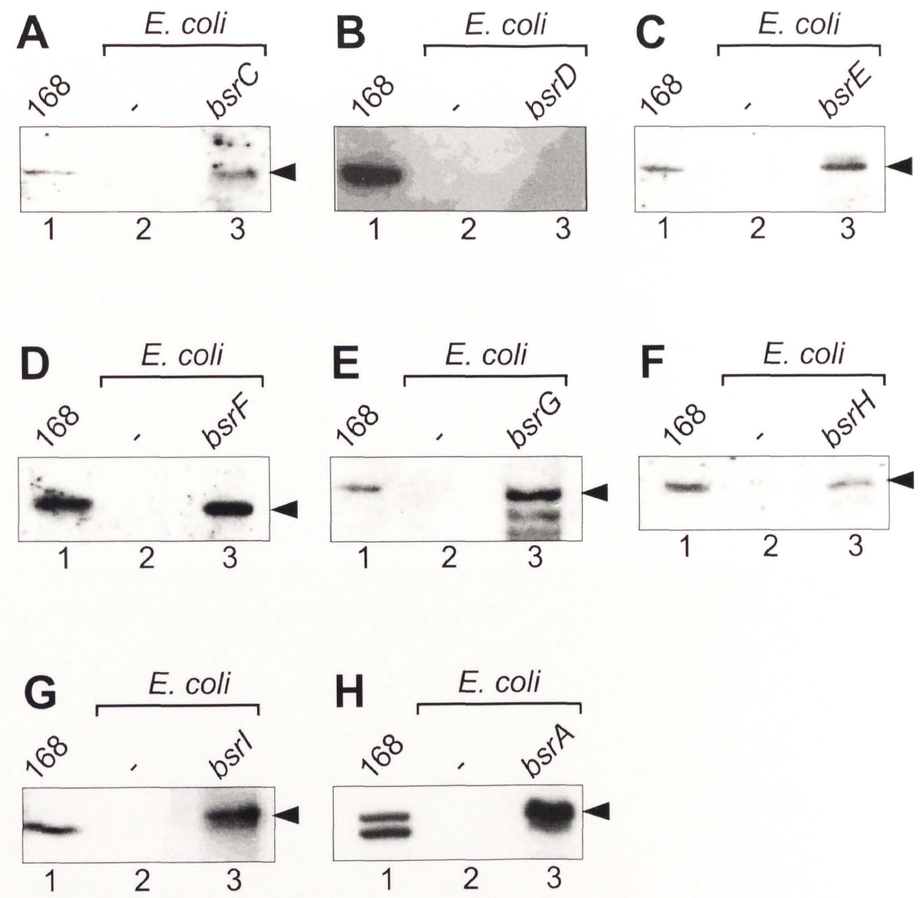
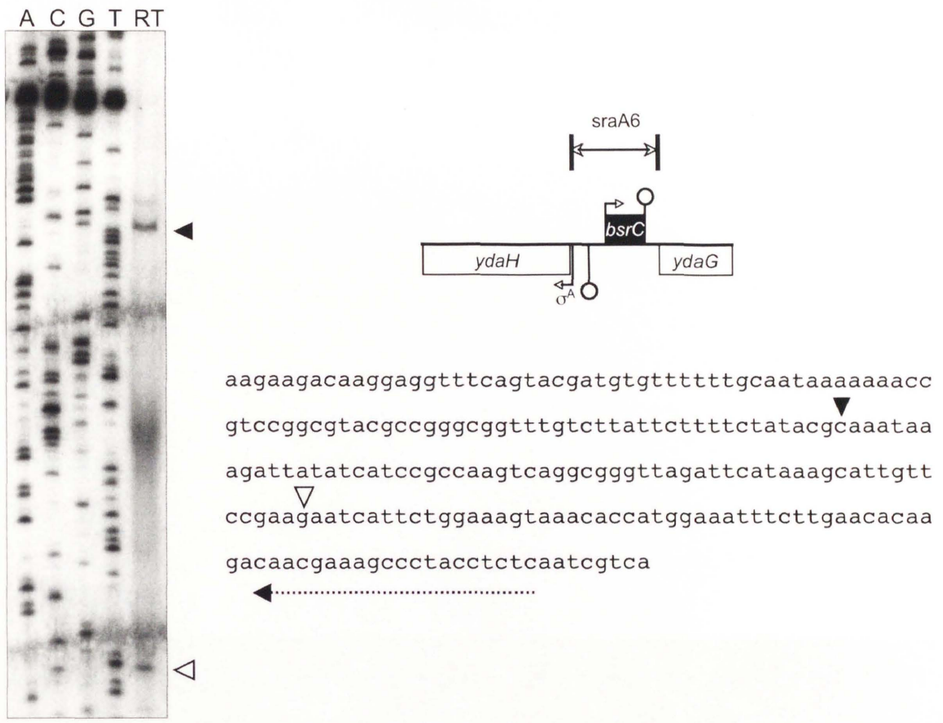


Figure 3. Determination of 5′ ends of Bsr RNAs.

Transcriptional start sites of BsrC (A), BsrE (B), BsrF (C), BsrG (D), BsrH (E), and BsrI (F) RNAs were determined by primer extension analysis. Reverse transcription products (lanes RT) and sequencing ladders (lanes A, C, G, and T) separated by 6% polyacrylamide gel electrophoresis, are shown in left side of each panel. Filled arrowheads, major bands; open arrowheads, additional smaller bands. Upper right of each panel shows diagrammatic map with adjacent genes. Top shows boundaries of intergenic regions. Bent arrows indicate the directions of transcription. Lollipops indicate predicted stem-loop terminators. Diagrams are drawn to scale. Lower right of each panel shows nucleotide sequence upstream of *bsr* gene. Putative - 35 and - 10 promoter sequences are boxed. Filled and open arrowheads above nucleotide sequences indicate position of 5′ end of each cDNA fragment determined by primer extension analysis. Discontinuous arrows show location of primers used in primer extension and sequencing reactions.

Figure 3

A



B

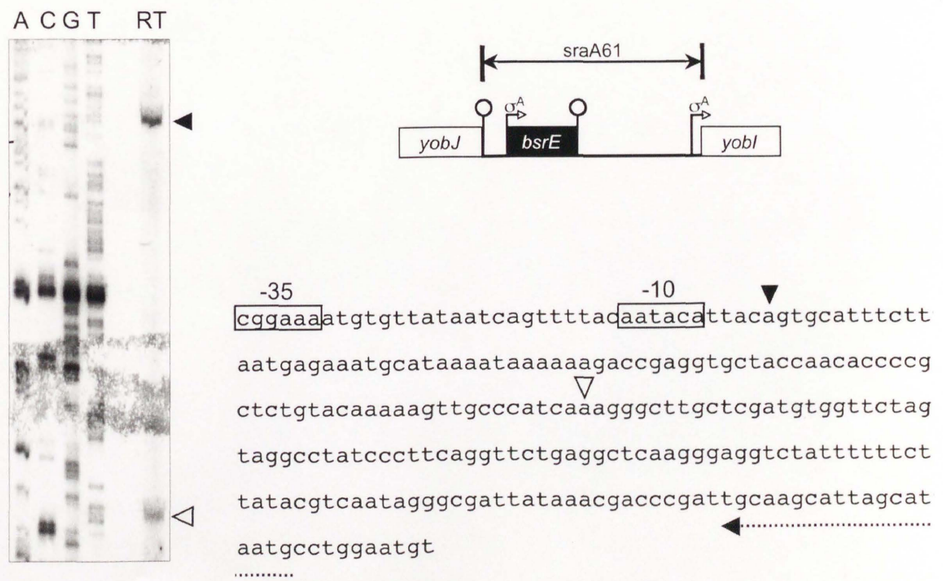


Figure 3-continued

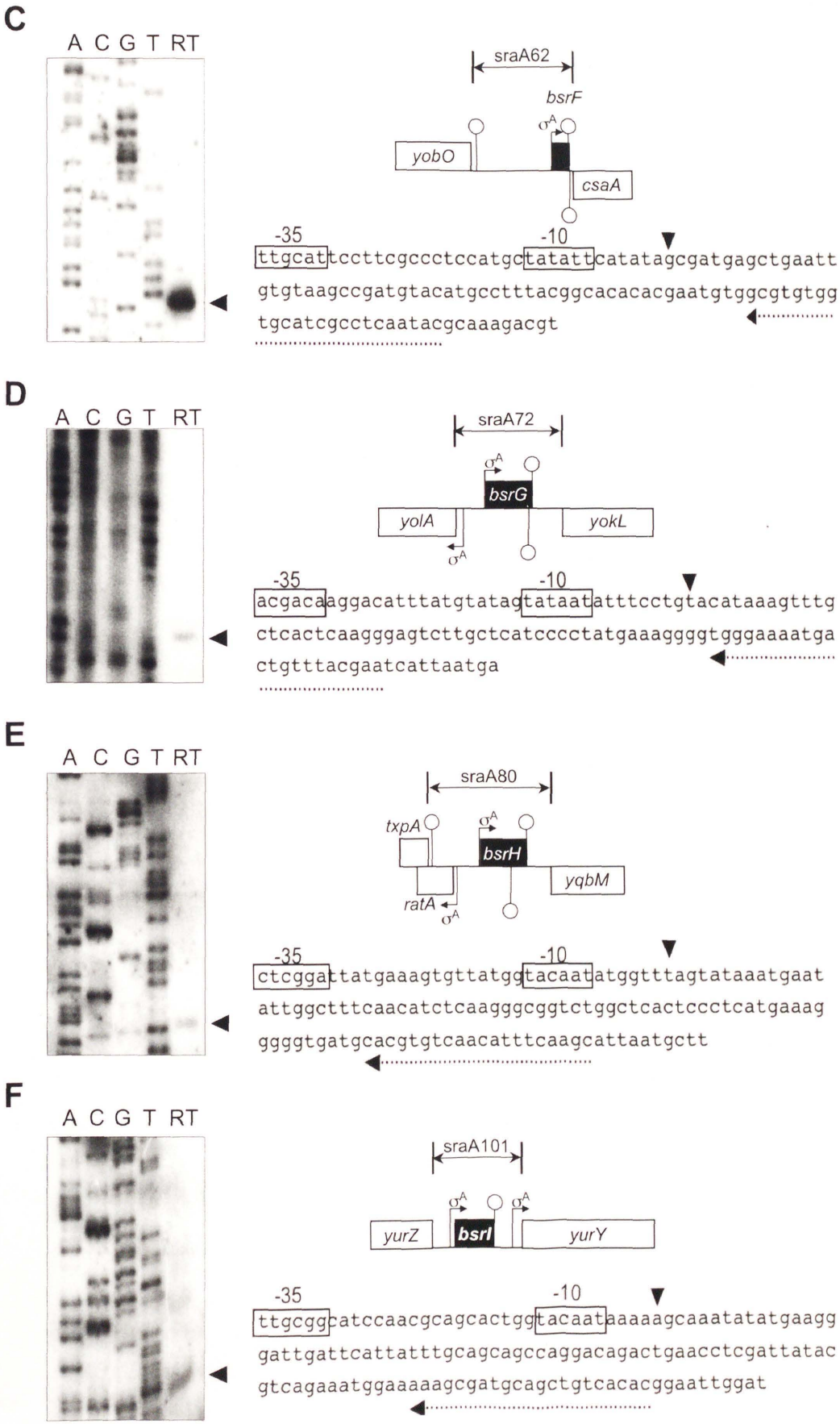


Figure 4. Predicted secondary structures of BsrC (A), BsrE (B), BsrF (C), BsrG (D), BsrH (E), and BsrI (F) RNAs. Putative sRNA sequence predicted by the RNAfold program (Hofacker, 2003) is from BSORF database (http://bacillus.genome.jp/).

Figure 4

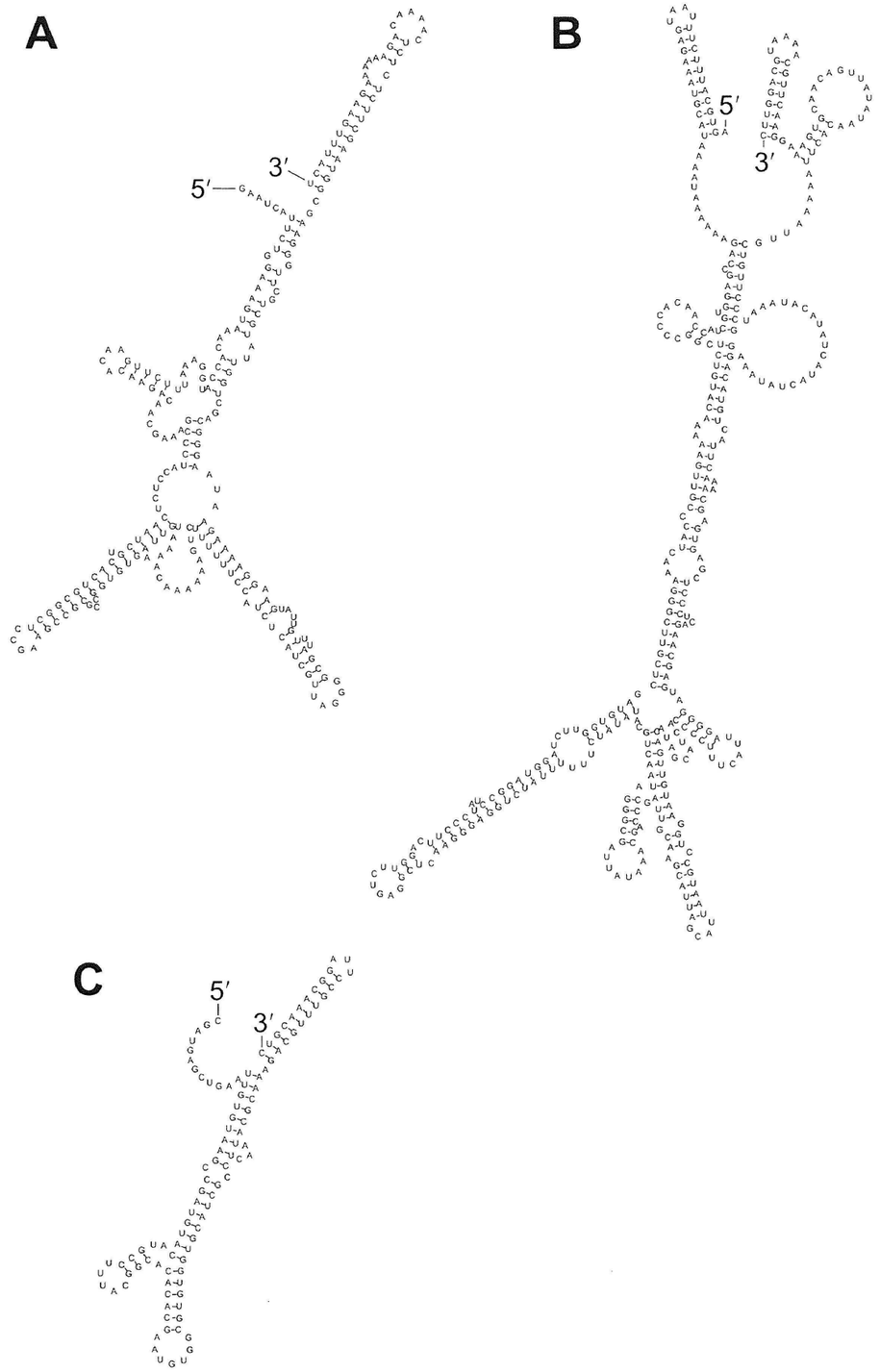
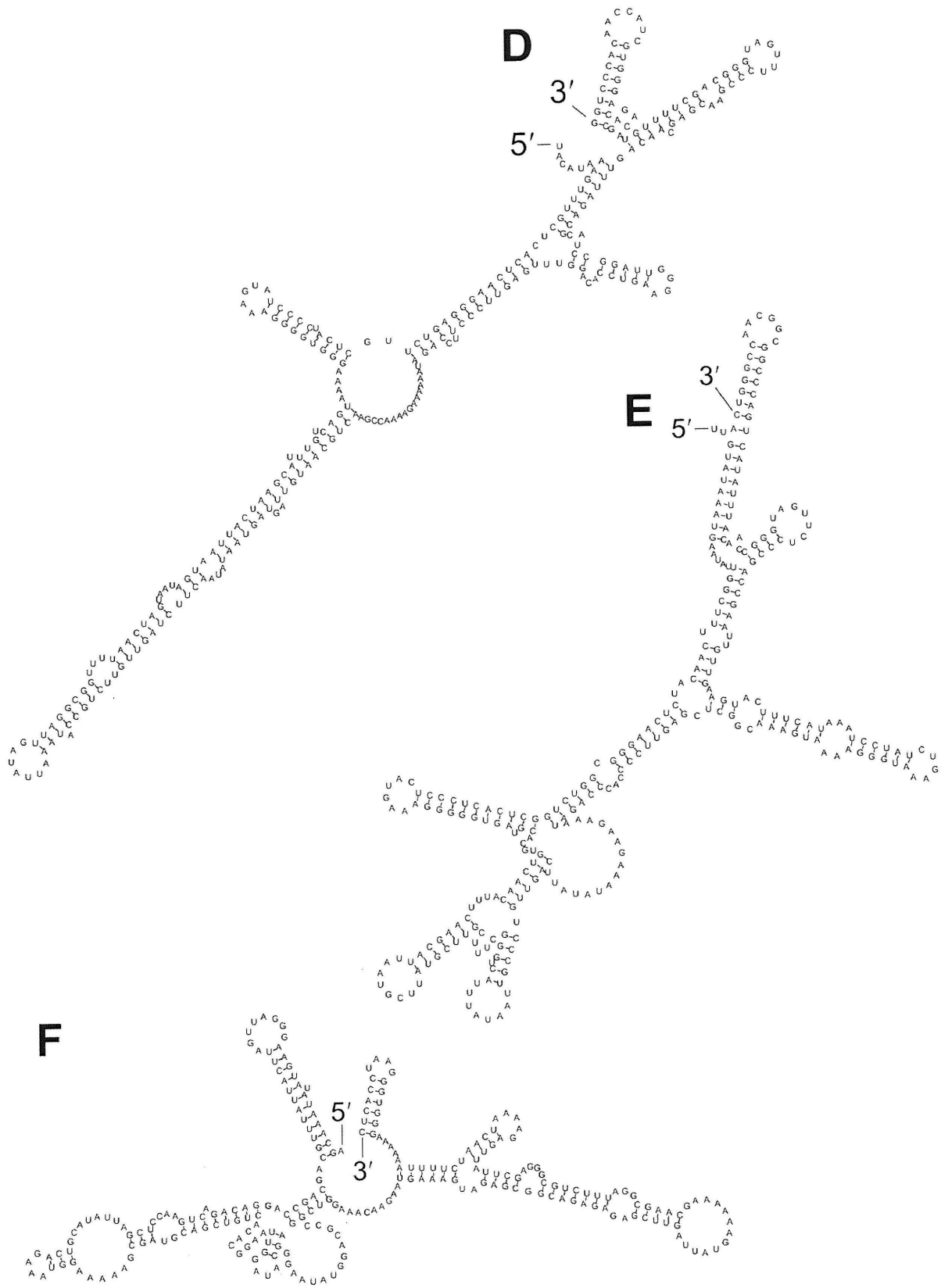


Figure 4-continued



REFERENCES

Ando, Y., Asari, S., Suzuma, S., Yamane, K., Nakamura, K., 2002. Expression of a small RNA, BS203 RNA, from the *yocI-yocJ* intergenic region of *Bacillus subtilis* genome. FEMS Microbiol. Lett. 207, 29–33.

Band, L., Henner, D.J., 1984. *Bacillus subtilis* requires a "stringent" Shine-Dalgarno region for gene expression. DNA. 3, 17–21.

Barrick, J.E., Sudarsan, N., Weinberg, Z., Ruzzo, W.L., Breaker, R.R., 2005. 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. RNA. 11, 774–784.

Chen, S., Lesnik, E.A., Hall, T.A., Sampath, R., Griffey, R.H., Ecker, D.J., Blyn, L.B., 2002. A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. Biosystems 65, 157–177.

Chuang, S.E., Daniels, D.L., Blattner, F.R., 1993. Global regulation of gene expression in *Escherichia coli*. J. Bacteriol. 175, 2026-2036.

Corbino, K.A., Barrick, J.E., Lim, J., Welz, R., Tucker, B.J., Puskarz, I., Mandal, M., Rudnick, N.D., Breaker, R.R., 2005. Evidence for a second class of

S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. Genome Biol. 6, R70.

Costa, F.F., 2007. Non-coding RNAs: lost in translation? Gene 386, 1–10.

De Hoon, M.J., Makita, Y., Nakai, K., Miyano, S., 2005. Prediction of transcriptional terminators in *Bacillus subtilis* and related species. PLoS Comput. Biol. 1, e25.

Fang, X.W., Yang, X.J., Littrell, K., Niranjanakumari, S., Thiyagarajan, P., Fierke, C.A., Sosnick, T.R., Pan, T., 2001. The *Bacillus subtilis* RNase P holoenzyme contains two RNase P RNA and two RNase P protein subunits. RNA. 7, 233–241.

Frank, D.N., Pace, N.R., 1998. Ribonuclease P: unity and diversity in a tRNA processing ribozyme. Annu. Rev. Biochem. 67, 153–180.

Gottesman, S., 2004. The small RNA regulators of *Escherichia coli*: roles and mechanisms. Annu. Rev. Microbiol. 58, 303–328.

Gottesman, S., 2005. Micros for microbes: non-coding regulatory RNAs in bacteria. Trends Genet. 21, 399–404.

Guillier, G., Gottesman, S., Storz, G., 2006. Modulating the outer membrane with small RNAs. Genes Dev. 20, 2338–2348.

Hainzl, T., Huang, S., Sauer-Eriksson, A.E., 2005. Structural insights into SRP RNA: an induced fit mechanism for SRP assembly. RNA 11, 1043–1050.

Henner, D.J., 1990. Inducible expression of regulatory genes in *Bacillus subtilis*. Methods Enzymol. 185, 223–228.

Hüttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J.P., Brosius, J., 2001. RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. EMBO J. 20, 2943–2953.

Hofacker, I.L., 2003. Vienna RNA secondary structure server. Nucleic Acids Res. 31, 3429–3431.

Igo, M.M., Losick, R., 1986. Regulation of a promoter that is utilized by minor forms of RNA polymerase holoenzyme in *Bacillus subtilis*. J. Mol. Biol. 191, 615–624.

Ito, K., Tadaki, T., Lee, S., Takada, K., Muto, A., Himeno, H., 2002. Trans-translation mediated by *Bacillus subtilis* tmRNA. FEBS Lett. 516, 245–252.

Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G., Andersen, K.K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P., Boland, F., Brignell, S.C., Bron, S., Bunai, K., Chapuis, J., Christiansen, L.C., Danchin, A., Débarbouille, M., Dervyn, E., Deuerling, E., Devine, K., Devine, S.K., Dreesen, O., Errington, J., Fillinger, S., Foster, S.J., Fujita, Y., Galizzi, A., Gardan, R., Eschevins, C., Fukushima, T., Haga, K., Harwood, C.R., Hecker, M., Hosoya, D., Hullo, M.F., Kakeshita, H., Karamata, D., Kasahara, Y., Kawamura, F., Koga, K., Koski, P., Kuwana, R., Imamura, D., Ishimaru, M., Ishikawa, S., Ishio, I., Le Coq, D., Masson, A., Mauël, C., Meima, R., Mellado, R.P., Moir, A., Moriya, S., Nagakawa, E., Nanamiya, H., Nakai, S., Nygaard, P., Ogura, M., Ohanan, T., O'Reilly, M., O'Rourke, M., Pragai, Z., Pooley, H.M., Rapoport, G., Rawlins, J.P., Rivas, L.A., Rivolta, C., Sadaie, A., Sadaie, Y., Sarvas, M., Sato, T., Saxild, H.H.,

Scanlan, E., Schumann, W., Seegers, J.F., Sekiguchi, J., Sekowska, A., Séror, S.J., Simon, M., Stragier, P., Studer, R., Takamatsu, H., Tanaka, T., Takeuchi, M., Thomaides, H.B., Vagner, V., van Dijl, J.M., Watabe, K., Wipat, A., Yamamoto, H., Yamamoto, M., Yamamoto, Y., Yamane, K., Yata, K., Yoshida, K., Yoshikawa, H., Zuber, U., Ogasawara, N., 2003. Essential Bacillus subtilis genes. Proc. Natl. Acad. Sci. U.S.A. 100, 4678–4683.

Kuwayama, H., Obara, S., Morio, T., Katoh, M., Urushihara, H., Tanaka, Y., 2002. PCR-mediated generation of a gene disruption construct without the use of DNA ligase and plasmid vectors. Nucleic Acids Res. 30, E2.

Leighton, T.J., Doi, R.H., 1971. The stability of messenger ribonucleic acid during sporulation in *Bacillus subtilis*. J. Biol. Chem. 246, 3189–3195.

Licht, A., Preis, S., Brantl, S., 2005. Implication of CcpN in the regulation of a novel untranslated RNA SR1 in *Bacillus subtilis*. Mol. Microbiol. 58, 189 – 206.

Marker, C., Zemann, A., Terhörst, T., Kiefmann, M., Kastenmayer, J.P., Green, P., Bachellerie, J.P., Brosius, J., Hüttenhofer, A., 2002. Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant

Arabidopsis thaliana. Curr. Biol. 12, 2002–2013.

Massé, E., Escorcia, F.E., Gottesman, S., 2003a. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. Genes Dev. 17, 2374–2383.

Massé, E., Majdalani, N., Gottesman, S., 2003b. Regulatory roles for small RNAs in bacteria. Curr. Opin. Microbiol. 6, 120–124.

Mattick, J.S., Makunin, I.V., 2006. Non-coding RNA. Hum. Mol. Genet. 15 Spec No 1:R17–29.

Nakamura, K., Imai, Y., Nakamura, A., Yamane, K., 1992. Small cytoplasmic RNA of *Bacillus subtilis*: functional relationship with human signal recognition particle 7S RNA and *Escherichia coli* 4.5S RNA. J. Bacteriol. 174, 2185–2192.

Rasmussen, A.A., Eriksen, M., Gilany, K., Udesen, C., Franch, T., Petersen, C., Valentin-Hansen, P., 2005. Regulation of *ompA* mRNA stability: the role of a small regulatory RNA in growth phase-dependent control. Mol. Microbiol. 58, 1421–1429.

Rivas, E., Klein, R.J., Jones, T.A., Eddy, S.R., 2001. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. Curr. Biol. 11, 1369–1373.

Romeo, T., 1998. Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. Mol. Microbiol. 29, 1321–1330.

Sharkady, S.M., Nolan, J.M., 2001. Bacterial ribonuclease P holoenzyme crosslinking analysis reveals protein interaction sites on the RNA subunit. Nucleic Acids Res. 29, 3848–3856.

Sierro, N., Makita, Y., De Hoon, M., Nakai, K., 2008. DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. Nucleic Acids Res. 36, D93–96.

Silvaggi, J.M., Perkins, J.B., Losick, R. 2006. Genes for small, noncoding RNAs under sporulation control in *Bacillus subtilis*. J. Bacteriol. 188, 532–541.

Suzuma, S., Asari, S., Bunai, K,, Yoshino, K., Ando, Y., Kakeshita, H., Fujita, M.,

Nakamura, K., Yamane, K., 2002. Identification and characterization of novel small RNAs in the *aspS-yrvM* intergenic region of the *Bacillus subtilis* genome. Microbiology 148, 2591–2598.

Tang, T.H., Bachellerie, J.P., Rozhdestvensky, T., Bortolin, M.L., Huber, H., Drungowski, M., Elge, T., Brosius, J., Hüttenhofer, A., 2002. Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. Proc. Natl. Acad. Sci. U.S.A. 99, 7536–7541.

Trotochaud, A.E., Wassarman, K.M., 2005. A highly conserved 6S RNA structure is required for regulation of transcription. Nat. Struct. Mol. Biol. 12, 313–319.

Urban, J.H., Vogel, J., 2008. Two seemingly homologous noncoding RNAs act hierarchically to activate glmS mRNA translation. PLoS Biol. 6, e64

Vellanoweth, R.L., Rabinowitz, F.C., 1992. The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli in vivo*. Mol. Microbiol. 6, 1105–1114.

Vogel, J., Bartels, V., Tang, T.H., Churakov, G., Slagter-Jäger, J.G., Hüttenhofer,

A., Wagner, E.G., 2003. RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. Nucleic Acids Res. 31, 6435–6443.

Washietl, S., Hofacker, I.L., Lukasser, M., Hüttenhofer, A., Stadler, P.F., 2005a. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. Nat. Biotechnol. 23, 1383–1390.

Washietl, S., Hofacker, I.L., Stadler, P.F., 2005b. Fast and reliable prediction of noncoding RNAs. Proc. Natl. Acad. Sci. U.S.A. 102, 2454–2459.

Wassarman, K.M., Zhang, A., Storz, G., 1999. Small RNAs in *Escherichia coli*. Trends Microbiol. 7, 37–45.

Wassarman, K.M., Repoila, F., Rosenow, C., Storz, G., Gottesman, S., 2001. Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev. 15, 1637–1651.

Yuan, G., Klämbt, C., Bachellerie, J.P., Brosius, J., Hüttenhofer, A., 2003.

RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. Nucleic Acids Res. 31, 2495–2507.

Zhang, Y., Zhang, Z., Ling, L., Shi, B., Chen, R., 2004. Conservation analysis of small RNA genes in *Escherichia coli*. Bioinformatics 20, 599–603.

ACKNOWLEDGEMENTS

This study was started in 2001 and ended in 2008 in the Gene Research Center, Graduate School of Life and Environmental Sciences, University of Tsukuba in the University of Tsukuba.

I owe my deepest gratitude to Associate Professor Kouji Nakamura who suggested the present study to me, provided me with the facilities, and guided me during these years. I admire his unflagging ability to read my manuscripts during all his busy working hours.

I wish to express my greatest gratitude to Dr. Hideko Urushihara, Dr. Osamu Numata and Dr. Tomoki Chiba for their valuable suggestions and encouragement through my doctoral program.

My respectful thanks to Dr. Hiroshi Kakeshita, Dr. Tatsuhiko Someya and all colleagues of Nakamura's laboratory for their keen interest in my studies and valuable discussions about studies on gene regulatory mechanisms.

Finally, I would like to appreciate my family and friends for their encouragement and great support for my research.