6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter

JEFFREY E. BARRICK,[1] NARASIMHAN SUDARSAN,[2] ZASHA WEINBERG,[3] WALTER L. RUZZO,[3,4] and RONALD R. BREAKER[2]

[1]Department of Molecular Biophysics and Biochemistry and [2]Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA
[3]Department of Computer Science and Engineering and [4]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA

ABSTRACT

6S RNA is an abundant noncoding RNA in *Escherichia coli* that binds to σ^{70} RNA polymerase holoenzyme to globally regulate gene expression in response to the shift from exponential growth to stationary phase. We have computationally identified >100 new 6S RNA homologs in diverse eubacterial lineages. Two abundant *Bacillus subtilis* RNAs of unknown function (BsrA and BsrB) and cyanobacterial 6Sa RNAs are now recognized as 6S homologs. Structural probing of *E. coli* 6S RNA and a *B. subtilis* homolog supports a common secondary structure derived from comparative sequence analysis. The conserved features of 6S RNA suggest that it binds RNA polymerase by mimicking the structure of DNA template in an open promoter complex. Interestingly, the two *B. subtilis* 6S RNAs are discoordinately expressed during growth, and many proteobacterial 6S RNAs could be cotranscribed with downstream homologs of the *E. coli ygfA* gene encoding a putative methenyltetrahydrofolate synthetase. The prevalence and robust expression of 6S RNAs emphasize their critical role in bacterial adaptation.

Keywords: covariance model; natural aptamer; methenyltetrahydrofolate synthetase; noncoding RNA; stationary phase

INTRODUCTION

Escherichia coli survives nutrient limitation by compacting its nucleoid, altering the promoter specificity of RNA polymerase (RNAP), and sequestering ribosomes in inactive 100S dimers to globally reduce and adapt gene expression (Ishihama 1999). Levels of the perhaps ~1000 growth-related genes expressed in an exponentially dividing bacterium generally decrease, while ~100 genes are specifically activated for stationary phase maintenance. Much of this adaptation is accomplished by increasing the population and activity of RNAP holoenzymes that direct transcription initiation with the stationary phase promoter-specific sigma factor (σ^S) relative to the housekeeping sigma factor (σ^{70}). Changes in cytoplasmic solute composition, an increase in σ^S levels, and expression of Rsd, an anti-σ^{70} factor, all contribute to this overall shift in promoter specificity (Ishihama 2000).

Reprint requests to: Ronald R. Breaker, Department of Molecular, Cellular, and Developmental Biology, Yale University, P.O. Box 208103, New Haven, CT 06520, USA; e-mail: ronald.breaker@yale.edu; fax: (203) 432-6604.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.7286705.

E. coli 6S RNA participates in the transcriptional response to starvation by binding to σ^{70}-containing RNAP holoenzyme (Wassarman and Storz 2000). Its expression increases 11-fold during stationary phase to a maximum of ~10,000 copies per cell when >75% of σ^{70} holoenzymes are associated with 6S RNA. The molecular details of this recognition are unknown, but the extended hairpin structure proposed for 6S RNA resembles DNA template in an open promoter complex with RNAP (Wassarman 2002). 6S RNA is necessary for the repression of σ^{70}-dependent promoters that contain extended −10 sequences under nutrient limitation and concomitant activation of certain σ^S-dependent promoters (Trotochaud and Wassarman 2004). Despite this widespread regulatory role, 6S RNA knockouts exhibit only subtle growth defects. Deletion of 6S RNA causes reduced viability compared to wild-type cells after >20 d of continuous culture, and cells lacking 6S are at a competitive disadvantage when cocultured with wild-type cells after several days of growth (Wassarman 2002).

Although *E. coli* 6S RNA was the first noncoding RNA to be sequenced >30 yr ago (Brownlee 1971), additional 6S RNA homologs have only been reported in *Pseudomonas aeruginosa* (Vogel et al. 1987) and *Haemophilus influenzae* (Brosius 1996). All currently known 6S RNA sequences

identified by bioinformatics in the Rfam database are likewise restricted to species of γ-proteobacteria (Griffiths-Jones et al. 2003). We have computationally identified numerous additional homologs of 6S RNA in >100 bacterial species representing diverse eubacterial lineages. A comparative analysis of 6S RNAs has allowed us to elaborate on how this RNA could mimic an open promoter to bind RNAP holoenzyme, to examine the evolution of two functionally divergent copies of 6S in some Gram-positive bacteria, and to predict that 6S RNA is cotranscribed with the reading frame for a protein regulating folate levels in many proteobacteria.

RESULTS AND DISCUSSION

Identification of 6S RNA homologs

In the course of investigating new RNA motifs in *Bacillus subtilis* by genomic comparisons of intergenic regions (Barrick et al. 2004), we rediscovered two noncoding RNAs, BsrA and BsrB. BsrA RNA had been isolated as a highly expressed transcript in total *B. subtilis* RNA separated on polyacrylamide gels (Suzuma et al. 2002). The 201-nt BsrA RNA, encoded by the *aspS-yrvM* intergenic region, is slowly processed into a 190-nt RNA by the removal of 11 nt from its 5′ end. At the same time, an abundant 203-nt transcript from the *yocI-yocJ* intergenic region of *B. subtilis* was recovered and named BsrB (Ando et al. 2002). Preliminary biochemical investigations did not reveal the functions of BsrA or BsrB. We manually aligned BLAST hits between the *aspS-yrvM* intergenic region and sequences upstream of *yrvM* homologs in other Gram-positive species. The common secondary structure model for this RNA family was essentially the same as that predicted previously for BsrA by thermodynamic calculations (Suzuma et al. 2002).

We used covariance models (Eddy and Durbin 1994) trained on this alignment to search the complete and unfinished microbial genomes available in GenBank for more divergent BsrA homologs. Surprisingly, BsrB was present in the expanded collection of hits from Gram-positive organisms even though the secondary structure predicted for BsrB RNA is completely different (Ando et al. 2002). These searches also uncovered convincing similarity to cyanobacterial genomes that overlapped annotations of an RNA named 6Sa. Reminiscent of BsrA and BsrB, 6Sa RNA was identified as an abundant noncoding transcript of unknown function with a size of 185 nt from *Synechococcus* sp. PCC6301 (Watanabe et al. 1997). Note, however, that all genomic annotations of 6Sa RNA in cyanobacterial genomes are on the incorrect strand (e.g., *Nostoc* sp. PCC 7120, GenBank accession NC_003272.1). The suggestive name of 6Sa RNA, presence of this RNA family in distant bacterial lineages, and common ~200 nt length of these RNAs encouraged us to look for similarity between these noncoding RNAs and *E. coli* 6S RNA.

Indeed, 6S RNA was originally isolated >35 yr ago as a small stable RNA of 184 nt that formed a distinct band after polyacrylamide gel electrophoresis (PAGE) of *E. coli* total RNA (Hindley 1967), and its function was entirely cryptic until recently. Despite an average pairwise similarity between 6S RNA and the *B. subtilis* BsrA and BsrB RNAs of only 46%, conservation of key secondary structure and nucleotide sequence elements provide strong evidence that these noncoding RNAs are structural homologs (Fig. 1). An improved covariance model, trained with known 6S sequences from the Rfam database (Griffiths-Jones et al. 2003) and with our original sequences from Gram-positive bacteria and cyanobacteria, identified 6S RNA sequences in almost every major group of Eubacteria.

6S RNA sequences from α-proteobacteria species were conspicuously absent from these expanded results. We initially investigated the presence of 6S RNA in this clade by isolating total RNA from *Caulobacter cresentus* cultures grown to mid-log and stationary phase and looking for ~200-nt bands on polyacrylamide gels stained with ethidium bromide. These experiments revealed a likely 6S candidate. An abundant ~180-nt RNA did not match the predicted sizes of annotated noncoding RNAs, and its expression increased during stationary phase (data not shown).

With this indication that our bioinformatics searches were not identifying all 6S homologs, we adopted a targeted strategy. Alignments of the other 6S RNA matches indicated that a conserved bulge in the terminal loop of 6S RNA was missing in certain lineages (Figs. 1, 2). We used an alignment of only these sequences from β-proteobacteria, δ-proteobacteria, and spirochetes to create covariance models with the diverse terminal loops explicitly modeled as variable insertions and conducted unfiltered searches against selected α-proteobacterial genomes. Among the matches in *C. cresentus* and *Agrobacterium tumafaciens* were sequences that clearly matched the consensus features of 6S RNA but had terminal loops truncated to a single stem–loop. Incorporating these sequences and a similar loop-truncated 6S RNA from *Aquifex* into the multiple sequence alignment and repeating the search readily identified other high-scoring 6S homologs from α-proteobacteria with this variation.

Figure 1 is an alignment of 17 representative 6S RNA homolog sequences. Our final curated alignment contains 121 sequences, and covariance models built from this alignment find hundreds of additional 6S RNA sequences in microbial genomes and environmental sequences (Venter et al. 2004). The curated seed alignment and an automated alignment of all matches are available upon request. Additionally, the seed alignment has been submitted to the Rfam database (Griffiths-Jones et al. 2003) to update the model for the 6S RNA family (accession RF00013).

6S RNA nomenclature

Independent descriptions of unrecognized 6S RNA homologs in different bacteria and the presence of multiple

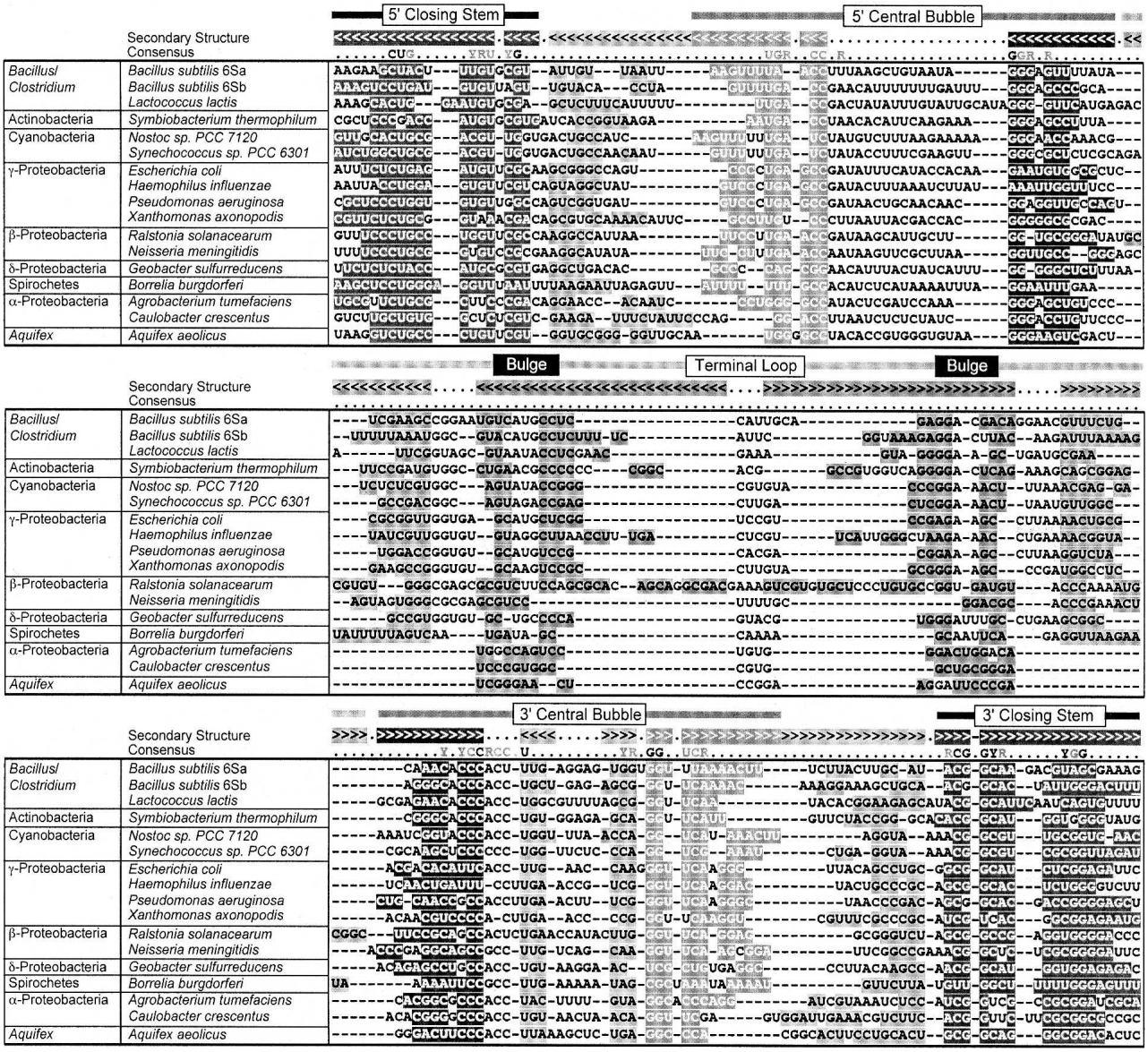

FIGURE 1. 6S RNA sequence alignment. Representative 6S RNAs from diverse bacterial groups were aligned to common primary sequence and secondary structure features. The three structural domains of 6S RNA (closing stem, central bubble, and terminal loop) and the conserved "bulge" present in the terminal loop sequences of certain bacterial lineages are labeled. Letters in the consensus line identify nucleotides that are conserved in >80% (gray) and >95% (black) of all 6S RNA sequences. Purine (R = A or G) and pyrimidine (Y = C or U) designations are used when a single nucleotide is not >80% conserved. Putative base pairing in individual sequences is highlighted with shaded backgrounds corresponding to paired angle brackets in the consensus secondary structure line.

copies of 6S RNA within a single genome complicate 6S RNA nomenclature. We suggest using the *E. coli* 6S RNA and *ssrS* gene designations for all organisms. The *Synechococcus ssaA* gene and 6Sa RNA names can be directly replaced with *E. coli* equivalents in this naming scheme. We discriminate between multiple 6S sequences within one genome by appending a single letter to each name in order of 6S RNA gene distance from the genomic origin. Accordingly, *ssrSA* and *ssrSB* are updated synonyms for the *B. subtilis bsrB* and *bsrA* genes, and they encode 6Sa RNA and 6Sb RNA, respectively. These nomenclature recommendations are used throughout the remainder of this report.

Conserved 6S RNA features

The overall structure of 6S RNA (Fig. 2A) can be divided into three conserved domains separated by variable stems:

1. The *closing stem* is the outer boundary of the 6S RNA hairpin. It consists of a ≥15-nt-long stem with conserved base pairs and bulges. The identities of both nucleotides in eight of these base pairs are constrained, including two G-U wobble base pairs. The 1-nt bulge interrupting the inner stretch of conserved bases on the 5′ side is highly conserved, and a bulged nucleotide on the 3′ side between the conserved base pairs is com-

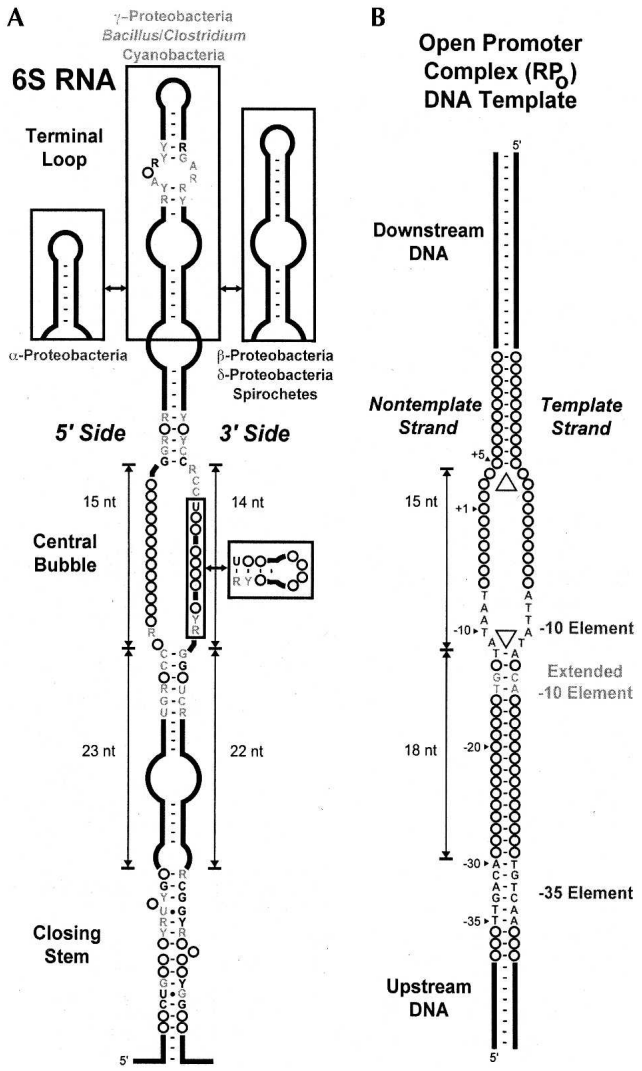

FIGURE 2. 6S RNA consensus secondary structure compared with open promoter DNA. (*A*) Consensus secondary structure model for 6S RNA. Nucleotide symbols and shading are the same as in Figure 1 (consensus line). Certain nucleotides whose identity is not conserved but are present in >60% of sequences are represented as empty circles. Solid lines represent variable regions of the structure. Three parallel *insets* show lineage-specific terminal loop structures and sequence conservation, and the boxed nucleotides on the 3′ side of the central bubble can alternately form the pictured base-paired stem in many sequences. Annotated nucleotide distances are the median lengths between conserved segments. (*B*) Schematic of DNA template in the open promoter complex with RNA polymerase (RP_o) as described elsewhere (Murakami et al. 2002).

monly present. The outer margin of the closing stem pairing is highly variable in individual sequences.

2. The *central bubble* is a large internal loop. The 5′ side of the bubble appears to be completely single-stranded and does not contain conserved sequences. However, its length is relatively constant, ranging from 12–21 nt with a median length of 15 nt. The poor secondary structure potential of this stretch of nucleotides is underscored by its abnormally low guanosine content of only 10.9% compared with an overall frequency of 25.1% in the entire alignment. In contrast, the 3′ side of the central bubble includes four conserved bases on its inner side, and the remainder can fold into a short stem–loop in many sequences. Two conserved G-C base pairs surround the central bubble on both sides. There is additional conservation of the identities of nucleotides forming three base pairs further along the outer stem, and a common purine-purine mismatch as the fourth base pair of the inner stem in certain groups of bacteria.
3. The *terminal loop* is a lineage-specific extension of the innermost base pairing elements of the 6S hairpin. In γ-proteobacteria, cyanobacteria, and the *Bacillus*/*Clostridium* group, there is a variable penultimate base-paired stem and bulge separating the central bubble from a final conserved stem. This final stem contains a characteristic purine-rich asymmetric bulge with 3 nt on the 5′ side and 2 nt on the 3′ side surrounded by preferred base pairs. Spirochetes, β-proteobacteria, and δ-proteobacteria preserve this general arrangement but do not have the cosnerved bulge in their final stems. In α-proteobacteria, the terminal loop is truncated to a single, short stem without any apparent sequence conservation.

The most recently published secondary structure model for *E. coli* 6S (Wassarman and Storz 2000) must be modified slightly to match the conserved structure. Specifically, the previously proposed pairing of CAA to UUG on the inner side of the central bubble must be disrupted, and the optional stem–loop on the 3′ side of the central bubble also can be formed by the *E. coli* variant.

Two features of 6S RNA conservation are unusual for a bacterial noncoding RNA. First, much of 6S primary sequence conservation occurs in canonical base-paired stems. Assuming that the central bubble's facultative stem–loop does not form and excluding the lineage-specific conservation in the terminal loop, 88% of the highly-conserved positions (≥80% conservation of a specific nucleotide) are in putative base pairs. For comparison, only 47% and 59% of conserved nucleotides within the aptamer domains of the metabolite-binding glycine (Mandal et al. 2004) and coenzyme B_{12} (Nahvi et al. 2004) riboswitches are paired, respectively. Second, large unconstrained loops like the central bubble 5′ strand are rare in structured bacterial RNAs. Typically, a putative single-stranded region of a functional RNA has conserved nucleotides that actually mediate the formation of a pseudoknot or tertiary packing. For example, all four of the single-stranded regions in the consensus minimal structure of bacterial RNAse P RNAs of comparable size to the 5′ central bubble strand of 6S RNA contain universally conserved nucleotides, and two combine in a pseudoknot (Frank and Pace 1998).

6S RNA resembles open promoter DNA

The conserved features of 6S RNA homologs support the hypothesis that 6S RNA mimics the structure of DNA template in an open promoter complex (RP_o) with RNAP (Wassarman 2002), and suggest further possibilities for more detailed binding models (Fig. 2B). RNAP holoenzyme melts double-stranded DNA template around the −10 element from position −11 to +4 relative to the transcription initiation site at +1 so that the template strand can weave through the polymerase active site (Murakami et al. 2002). The 15 nt of single-stranded nucleic acid in RP_o correspond remarkably well to the dimensions of the central bubble in 6S RNA homologs. The unstructured 5′ side has 15 nt, and unwinding the optional 3′ strand's stem–loop would free a total of 14 nt (median lengths).

Presenting a premelted promoter bubble may give 6S RNA a general affinity for core RNAP (lacking a σ-factor). Similar DNA templates constructed with arbitrary nonpromoter sequences and ~10-bp single-stranded bubbles are capable of directing RNA synthesis with core RNAP (Helmann and deHaseth 1999). Much of the affinity of RNAP for DNA templates is mediated by electrostatic interactions with the phosphate backbone that will be preserved by an RNA template. For example, single abasic substitutions at positions −11 to −7 on the nontemplate strand of fork junction DNA do not reduce its affinity for RNAP, although they do extenuate the subsequent formation of heparin-resistant complexes (Fenton and Gralla 2003). It is thought that RNAP recognizes the geometry of single-stranded/double-stranded DNA junctions in bubble templates. The observed conservation of two strong G-C base pairs flanking each side of the 6S central bubble might enforce its boundaries to favor these interactions. If the 6S central bubble binds core RNAP like an open promoter, then its surrounding base-paired stems will naturally follow the paths of upstream and downstream DNA template over basic surfaces in the polymerase structure (Murakami et al. 2002).

It is possible that the closing stem of 6S RNA acts as a replacement for the upstream DNA template so that its sequence conservation can interact with σ^{70}. The spacing between the central bubble and consensus elements in 6S RNA is broadly reminiscent of a typical DNA promoter (Fig. 2B). The conserved UGR/UCR base pairs located directly outside the central bubble might engage σ^{70} like an extended −10 element. However, there is no corresponding sequence similarity between the 6S closing stem and the usual −35 TTGACA consensus box. It seems more likely that σ^{70} forms a novel distal contact with 6S RNA's closing stem conservation. Unlike the other possible interactions we have described, this contact could directly contribute to the observed specificity of 6S RNA for σ^{70} holoenzyme and its inability to bind σ^S holoenzyme.

Orienting 6S RNA within RNAP with the closing stem in the direction of the DNA promoter is also appealing because it distinguishes the conserved 3′ side of the central bubble as the DNA template strand mimic. This architecture positions its conserved RCCU sequence near the site where transcription initiates on a DNA promoter. In this context, the optional stem–loop might masquerade as the short DNA/RNA hybrid helix normally present within the transcription bubble during elongation. Its placement also resembles that of stem–loops formed within the nascent RNA during the process of intrinsic transcription termination (Yarnell and Roberts 1999). Finally, this choice of template strand relegates the flexible 5′ strand of the bubble to a role as nontemplate strand and suggests that its length (and not its sequence) is conserved because it does not traverse the active site of RNAP.

It is not clear how the lineage-specific 6S RNA terminal loop could contribute to holoenzyme recognition. The purine-rich asymmetric bulge of the γ-proteobacterial loop type resembles a tertiary interaction motif far more than any other conservation in 6S RNA and might interact with a downstream site on RNAP or fold back on 6S RNA. On the other hand, terminal loops from other bacterial groups appear to lack any sequence conservation and would therefore seem incapable of participating in specific interactions. We also note that there is no obvious evolutionary correlation between the type of terminal loop and the domain structure of RNAP in different bacterial lineages (Iyer et al. 2004).

Specific binding of bacterial DNA-dependent RNAPs to templates composed of ribonucleotides is not unprecedented. In fact, some RNAs are able to act as true promoters to direct the synthesis of complementary RNA transcripts. Certain RNA sequences selected from random copolymer mixtures are capable of autocatalytic replication by *E. coli* RNAP holoenzyme through unknown intermediates (Wettich and Biebricher 2001). Also, a stem–loop derived from the peach latent mosaic viroid can initiate efficient in vitro transcription by *E. coli* RNAP from one strand of its hairpin in a reaction thought to recapitulate the natural replication of this single-stranded RNA in plants (Pelchat et al. 2002).

Recently, mouse B2 RNAs have been shown to repress general transcription in heat-shocked cells by binding directly to RNAP II (Allen et al. 2004; Espinoza et al. 2004). Although B2 RNA is similar in size and function to 6S RNA, these RNAs do not appear to be evolutionarily related. B2 RNAs are encoded by SINE elements that are thought to be derived from Ser-tRNA (Deininger and Daniels 1986), and they do not have the consensus features of known 6S RNA homologs. Both RNAs halt transcription before initiation, but B2 RNA binds to a remote docking site on Pol II and stalls polymerase while it is engaged to DNA template at the active site (Espinoza et al. 2004) whereas 6S RNA probably directly competes with DNA template for RNAP binding. It will be interesting to compare the molecular mechanisms of

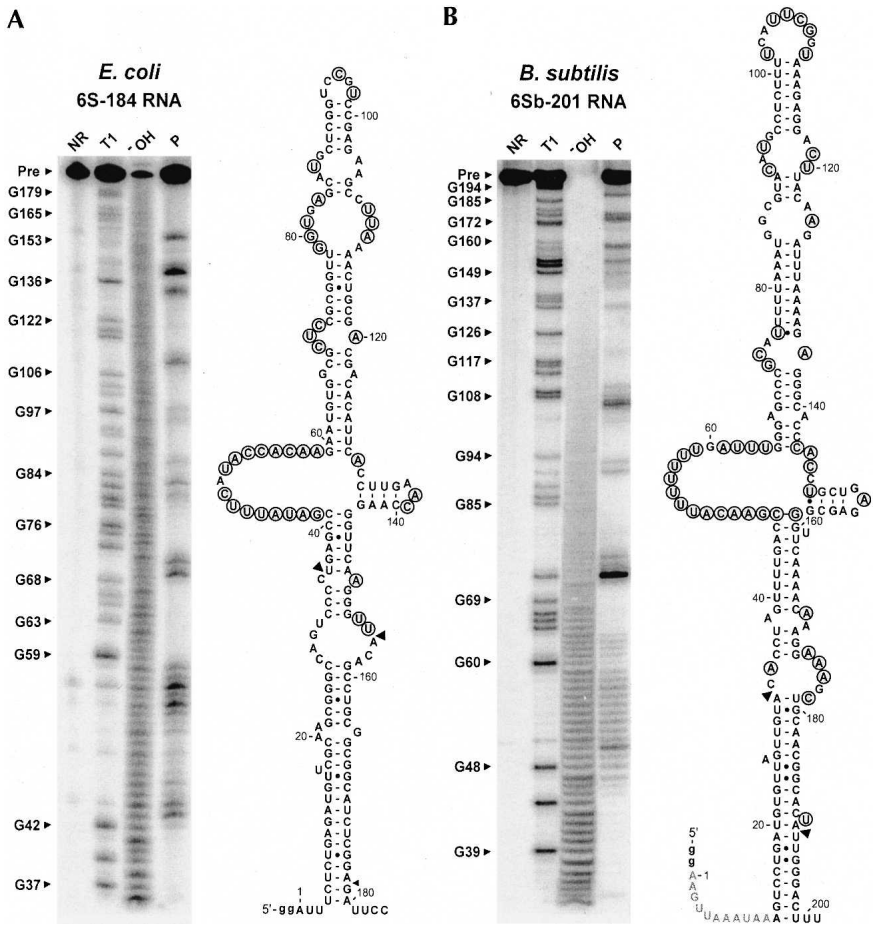

FIGURE 3. In-line probing of 6S RNA structures. (*A*) Sequence, secondary structure, and in-line probing data for *E. coli* 6S-184 RNA. Levels of spontaneous RNA cleavage at backbone linkages within the construct depicted to the *right* were measured by separating 5′-radiolabeled degradation products on a polyacrylamide gel. In-line probing gel lanes are as follows: NR, no incubation; T1, partial digestion with RNase T1 (cleaves 3′ of G nucleotides); ⁻OH, partial alkaline digestion; P, spontaneous cleavage during a 40-h in-line probing reaction incubated at 25°C. Pre identifies the full-length RNA. Bands corresponding to certain T1 cleavage products are identified as position markers. In the secondary structure model, shaded circles identify nucleotides whose 3′ linkage undergoes a high level of spontaneous cleavage relative to most other linkages. Filled triangles mark the extent of the region where cleavages were mapped. Lowercase letters identify unnatural guanosine nucleotides added for efficient in vitro transcription with T7 RNAP. (*B*) Sequence, secondary structure, and in-line probing data for *B. subtilis* 6Sb-201 RNA. Details as in *A*. Slow in vivo processing of 201 nt 6Sb RNA cleaves off 11 nucleotides (gray), resulting in the 190-nt form of 6Sb RNA.

these convergent solutions to widely inhibit transcription under stress conditions while allowing specific subsets of promoters to escape repression.

Structural probing of 6S RNAs

We subjected the 184-nt *E. coli* (6S-184) and *B. subtilis* (6Sb-201) 6S RNAs to in-line probing to verify that they folded into the structures predicted by comparative sequence analysis (Fig. 3). In this assay, spontaneous transesterification of 5′ radiolabeled RNA produces an RNA cleavage pattern that reflects the relative sampling of backbone conformations with the correct geometry for in-line attack of each ribose 2′-OH on the adjacent bridging phosphate (Soukup and Breaker 1999). Flexible regions of the RNA such as bulges and loops allow nucleotides to sample the in-line conformation and yield RNA degradation products (identified as bands upon autoradiography after PAGE), while base-paired regions are rigidly held in a structure that precludes in-line attack and are consequently resistant to degradation.

Both RNAs produce in-line probing patterns that agree with their predicted secondary structures. Cleavage at positions 56–58 in the *E. coli* construct (Fig. 3A) supports the omission predicted by comparative sequence analysis of three base pairs on the inner side of the central bubble in the previous structural model (Wassarman and Storz 2000). Reduced spontaneous cleavage of nucleotide linkages on the 3′ side of the bubble indicates that the optional stem–loop forms in both 6S RNAs, although it is difficult to precisely map pairing in this region of the in-line probing gels. In contrast, the RNA backbone in the extended internal loop on the 5′ side of the central bubble is consistently susceptible to spontaneous cleavage in both constructs. As predicted for mimicking an open promoter, this region is single-stranded in isolated 6S RNA and not involved in any higher-order structure. Overall, the agreement of the in-line probing patterns clearly indicates that the *E. coli* 6S and *B. subtilis* 6Sb RNAs adopt the same consensus structure derived from comparative sequence analysis.

Phylogenetic distribution of 6S RNA homologs

We constructed a distance-based phylogenetic tree from the curated multiple sequence alignment, restricting the analysis to majority ungapped positions and excluding highly-variable regions like the terminal loop (Fig. 4). This tree supports an ancient origin and uninterrupted evolution for 6S RNA within the Eubacteria. The 6S RNA phylogenetic tree generally reproduces the standard bacterial taxonomy based on 16S ribosomal RNA (Cole et al. 2003), and there are no obvious cases of horizontal gene transfer. The clustering of 6S RNA terminal loop synapomorphies into

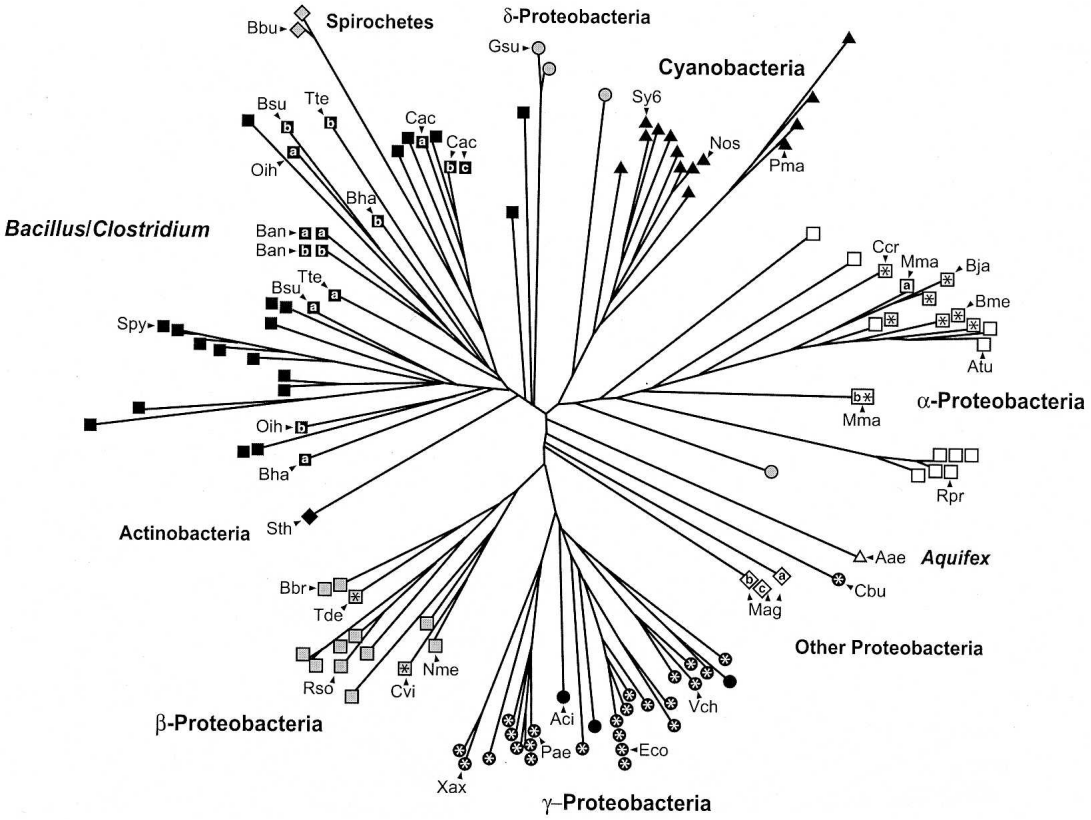

FIGURE 4. Phylogenetic tree of 6S RNA homologs. An unrooted phylogenetic tree was constructed from the final seed alignment of 121 sequences using distance methods. Symbols represent the taxonomic classification of the genomes containing each 6S RNA sequence. They are *Bacillus/Clostridium* (filled squares), actinobacteria (filled diamond), spirochetes (shaded diamonds), cyanobacteria (filled triangles), α-proteobacteria (open squares), β-protebacteria (shaded squares), γ-proteobacteria (filled circles), δ-proteobacteria (shaded circles), other proteobacteria (open diamonds), and *Aquifex* (open triangle). Groups with the same shading share terminal loop types as shown in Figure 2. Lowercase letters identify multiple 6S RNA sequences within one genome, and 6S RNA genes that are upstream of *E. coli ygfA* homologs are starred. Certain bacterial species are labeled with abbreviations as follows: Aae, *Aquifex aeolicus*; Aci, *Acinetobacter* sp. ADP1; Atu, *Agrobacterium tumefaciens*; Ban, *Bacillus anthracis*; Bbr, *Bordetella bronchiseptica*; Bbu, *Borrelia burgdorferi*; Bha, *Bacillus halodurans*; Bja, *Bradyrhizobium japonicum*; Bme, *Brucella melitensis*; Bsu, *Bacillus subtilis*; Cac, *Clostridium acetobutylicum*; Cbu, *Coxiella burnetii*; Ccr, *Caulobacter crescentus*; Cvi, *Chromobacterium violaceum*; Eco, *Escherichia coli*; Gsu, *Geobacter sulfurreducens*; Mag, *Magnetococcus* sp. MC-1; Mma, *Magnetospirillum magnetotacticum*; Nme, *Neisseria meningitidis*; Nos, *Nostoc* sp. PCC 7120; Oih, *Oceanobacillus iheyensis*; Pae, *Pseudomonas aeruginosa*; Pma, *Prochlorococcus marinus*; Sth, *Symbiobacterium thermophilum*; Tde, *Thiobacillus denitrificans*; Tte, *Thermoanaerobacter tengcongensis*; Rpr, *Rickettsia prowazekii*; Rso, *Ralstonia solanacearum*; Spy, *Streptococcus pyogenes*; Sy6, *Synechococcus* sp. PCC 6301; Vch, *Vibrio cholerae*; and Xax, *Xanthomonas axonopodis*. Other species names are omitted for clarity.

branches corresponding to evolutionarily related bacteria—even though this portion of the sequence was not included in tree calculations—further supports the large-scale features of this tree topology. It also suggests that the expanded terminal loop with the conserved bulge present in *E. coli* 6S RNA is the ancestral state, since the alternative hypothesis that an identical structure evolved separately in the cyanobacteria, γ-proteobacteria, and low-GC Gram-positive bacteria (*Bacillus/Clostridium* group) is unlikely. Presumably, the terminal loop has atrophied or become modified in other lineages. As has been the case with identifying microbial RNase P RNAs (Li and Altman 2004), further targeted experimental and bioinformatic efforts are likely to detect 6S sequence variants in other genomes, but we currently lack proof of such RNAs.

Generally, there is one copy of 6S RNA per microbial genome. However, two divergent 6S RNAs are present in several low-GC Gram-positive bacteria, including *B. subtilis*, *Bacillus halodurans*, *Clostridium acetobutylicum*, *Oceanobacillus iheyensis*, and *Thermoanaerobacter tengcongensis*. The probable phylogenetic relationships of these multiple copies in the context of all 6S RNAs from this clade indicate that at least one gene duplication must have occurred within this lineage (Fig. 4, e.g., Bsu 6Sa and 6Sb). The presumptive functional diversification in these select instances seems to have been accompanied by more widespread loss of the second 6S RNA copy in most branches. In contrast, the almost identical copies of 6S RNA in *Bacillus anthracis*, and *C. acetobutylicum* (which contains a total of three 6S homologs) are probably the result of very recent gene duplications. The only bacteria where we have identified multiple 6S RNA copies outside of the *Bacillus/Clostridium* group are

Magnetococcus sp. MC-1 and *Magnetospirillum magnetotacticum* sp. MS-1. Each of these unfinished genomes encodes at least two divergent 6S sequences, and one *Magnetococcus* homolog is duplicated.

Growth phase–dependent expression of *B. subtilis* 6S RNAs

We wondered how encoding two copies of 6S RNA could benefit some bacteria enough to be preserved during evolution. It had been previously reported that *B. subtilis* 6Sa RNA levels dramatically decrease during saturating growth after fresh innoculation (Ando et al. 2002). Since this pattern is opposite the normal increase in *E. coli* 6S RNA during stationary phase, and the timing of *B. subtilis* 6Sb expression was unknown, we probed Northern blots of total RNA isolated after different intervals of growth for both 6S homologs (Fig. 5).

The total levels of 6Sb RNA increase ~18-fold between early log and stationary phase growth (Fig. 5B). Precursor 6Sb-201 RNA transiently accumulates relative to processed 190-nt RNA (6Sb-190) as overall 6Sa levels increase, peaking at 60% of the total 6Sb RNA in mid-log phase under these growth conditions (Suzuma et al. 2002). In lag phase cells recovering from stationary phase in the culture used for innoculation (1-h time point) >90% of the 6Sb RNA has been cleaved to 190 nt. We observe a peak in 6Sa expression during mid-log phase where two to three times as much RNA is present as in early log phase under our growth conditions (Fig. 5C). After this point, the previously reported decrease in 6Sa RNA levels occurs, and stationary phase 6Sa RNA levels are reduced to at most one-eighth of the mid-log peak. Regardless, 6Sb RNA is the major 6S RNA species in *B. subtilis*. At the peak of 6Sa expression (5 h) there is still roughly twice as much 6Sb RNA as determined by ethidium bromide staining of gels (data not shown).

We conclude that *B. subtilis* 6Sb RNA is the ortholog of *E. coli* 6S RNA. *B. subtilis* 6Sa has functionally diverged at least with respect to the timing of its expression during growth, perhaps to more finely tune the transcriptional response to the approach of nutrient limitation.

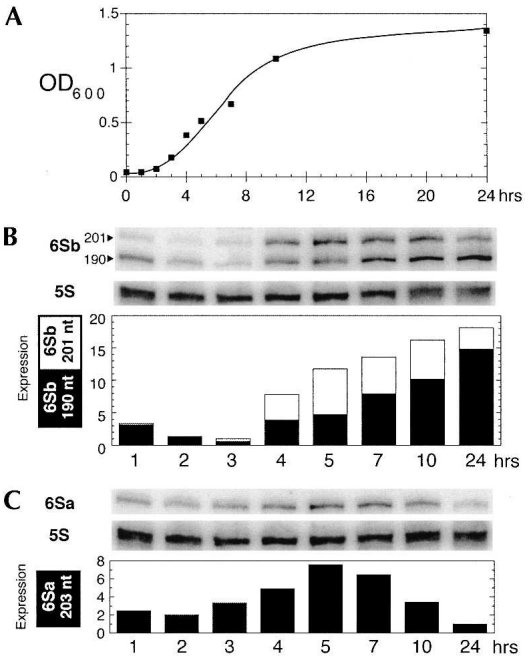

FIGURE 5. Expression of *B. subtilis* 6S RNAs during growth. (*A*) Growth curve for cultures from which total RNA was extracted to measure 6S RNA abundance. (*B*) Expression of 6Sb RNA. Northern blots were hybridized with radiolabeled probes specific for this 6S RNA and 5S RNA. Band intensities were quantitated, corrected for 5S RNA loading controls, and normalized to the 3-h time point. Levels of the 201-nt precursor and 190-nt processed 6Sb RNA bands are displayed as unfilled and filled bars, respectively. (*C*) Expression of 6Sa RNA. Details as in *B* except RNA levels were normalized to the 24-h time point.

Conservation of a 6S RNA-*ygfA* operon

There is experimental evidence that *E. coli* 6S RNA is rapidly cleaved by an unknown mechanism from the 5′ end of a transcript that includes the coding region for the *ygfA* gene (Hsu et al. 1985). We noticed that despite the marked divergence between the α- and γ-proteobacterial 6S RNA sequences, there is a widespread occurrence of *ygfA* homologs directly downstream of *ssrS* in both groups as well as some β-proteobacteria (Fig. 4). In other bacterial genomes, there is no apparent conservation of the genes found adjacent to 6S RNA homologs. The conserved juxtaposition of 6S RNA and *ygfA* in *E. coli* and other proteobacteria implies that it has functional relevance, most likely as a way of linking *ygfA* and 6S RNA expression. It is interesting in this respect that microarray experiments indicate that *ygfA* expression increases five- to eightfold in *E. coli* cells growing as biofilms (Ren et al. 2004), which represent another condition where poor nutrients availability may limit growth.

YgfA proteins share sequence similarity with eukaryotic methenyltetrahydrofolate synthetases (MTHFSs). These enzymes have been implicated in folate degradation and convert 5-formyltetrahydrofolate, which is believed to be a stable storage form of reduced folate, into 5,10-methenyltetrahydrofolate. In human cell culture, increased MTHFS activity correlates with a decrease in cellular folate pools that cannot be overcome by increased folate concentrations in the growth medium (Anguera et al. 2003). Since folate derivatives shuttle one-carbon units from degradative pathways into the synthesis of key metabolic intermediates such as purines, thymidylate, SAM, and formylmethionine-tRNA, depletion of folate may be a way for cells to globally restrict metabolic flux. If MTHFS serves a similar function in prokaryotic folate regulation, then many proteobacteria may adapt to nutrient limitation in stationary phase by the

concerted expression of YgfA and 6S RNA to slow one-carbon metabolism and RNA transcription.

Comparison of 6S RNA homologs

Previously studied noncoding RNAs of unknown function that we have recognized as 6S RNA homologs appear to differ in some respects from *E. coli* 6S RNA. Preliminary experiments have investigated the expression, processing, and dispensability of *B. subtilis* 6Sa RNA (Suzuma et al. 2002), *B. subtilis* 6Sb RNA (Ando et al. 2002), and *Synechococcus* sp. PCC6301 6S RNA (Watanabe et al. 1997). Notably, *Synechococcus* 6S RNA is abundant during exponential growth and later decreases in stationary phase. This timing is at odds with the normal regulation of *E. coli* 6S RNA and our observation that although *B. subtilis* expresses 6Sa RNA with an unusual timing, expression of the more abundant 6Sb RNA increases in stationary phase. A more detailed analysis will be necessary in all cases to determine whether regulation of other 6S RNAs is accomplished by producing transcripts from two promoter sites as in *E. coli* (Kim and Lee 2004). No growth defects have been detected in deletion mutants of *B. subtilis* 6Sa and *Synechococcus* 6S RNA under conditions where *E. coli* 6S RNA knockouts grow normally. In contrast, growth of *B. subtilis* 6Sb RNA deletion mutants is compromised during exponential phase, and mutant cultures are unable to reach densities as high as wild-type cultures during stationary phase. This is the first known instance where a defective 6S RNA results in a phenotype that is potentially useful for genetic studies.

The diversity of observed 6S RNA processing suggests that it is not important for functional maturation. Rather, exonuclease trimming of neighboring unstructured RNA regions up to the stable 6S closing stem could be incidental, as this mechanism is common for other stable bacterial RNAs (Li et al. 1998). RNase E and/or RNase G cleave *E. coli* 6S RNA at its 5′ end to produce a mixture of mature 6S RNA sequences with 5′ ends at positions −1, +1, and +2 (Kim and Lee 2004), and cleavage by an uncharacterized mechanism liberates its 3′ end from the 5′ untranslated region of an mRNA encoding the YgfA protein (Hsu et al. 1985). *B. subtilis* 6Sb RNA accumulates as a 201-nt transcript that is slowly cleaved to 190 nt by the loss of 11 nt from the 5′ end by an unknown RNase. Only the 185-nt version of *Synechococcus* 6S RNA has been observed. Similarly, no intermediates have been observed for *B. subtilis* 6Sa, although it atypically retains an extra 3′ stem–loop after its closing stem. This hairpin is probably the remnant of an intrinsic transcription terminator, and putative terminators are common directly downstream of 6S RNA sequences in certain groups of bacteria (data not shown).

Conclusion

The absence of a strong deletion phenotype and lack of comparative sequence information have historically been obstacles to understanding the function of *E. coli* 6S RNA. We have shown that 6S RNA is an ancient and conserved regulator of RNAP function. Only a small number of noncoding RNAs with critical cellular roles, including rRNAs, 5S RNA, tRNA, RNAse P RNA, SRP RNA, tmRNA, and some riboswitches, are as widely distributed as 6S RNA across different bacterial groups. An analysis of 6S RNA sequences suggests more detailed models for how it might mimic an open DNA promoter. It also raises new questions about the purposes of multiple 6S homologs in some genomes and the significance of conserved cotranscription with a downstream methylenetetrahydrofolate synthetase gene. The recognition of a 6S RNA homolog in *B. subtilis* whose deletion causes a dramatic growth defect and knowledge of conserved regions within the structure of 6S RNA should greatly enable future genetic and molecular studies of its interactions with the transcriptional machinery.

MATERIALS AND METHODS

Covariance model searches

Manual multiple alignments of 6S RNA sequences were used to construct covariance models (Eddy and Durbin 1994) using the Infernal v.0.55 software package. Filtering techniques were applied to accelerate searches of covariance models against sequence databases (Weinberg and Ruzzo 2004a,b). Score thresholds that reliably predicted new 6S RNA homologs were determined by scoring the input sequences and examining marginally scoring matches for false positives that overlapped conserved genes. A complete alignment of all 6S RNA homologs was generated by using Infernal to automatically align reliable matches to a covariance model trained on a seed alignment of 121 sequences. The final seed and complete alignments are available in Stockholm format from the authors. Sequences in the seed alignment were weighted before calculating the reported consensus sequence, length, and composition statistics to reduce biases from similar sequences with Infernal's internal implementation of the GSC algorithm (Gerstein et al. 1994).

Phylogenetic tree

We created a covariance model from the final seed alignment with 133 manually annotated consensus columns, encompassing conserved stems and the central bubble, and used it to automatically realign this set of 121 sequences. The same consensus columns were extracted from the new alignment and input into the dnadist and fitch programs from the Phylip (Phylogeny Inference Package) v.3.62 software package (University of Washington, Seattle) using the Jukes-Cantor distance method and all other parameters set to their default values to create an unrooted phylogenetic tree of 6S RNA homologs. The complete tree data file is available upon request.

In-line probing

DNA templates for the in vitro transcription of 6S RNAs were amplified by whole-cell PCR from *E. coli* strain MG1655 and *B.*

subtilis strain 168 (BGSC no. 1A1; Bacillus Genetic Stock Center, Columbus, OH). Details of the in-line probing analysis have been reported elsewhere (Soukup and Breaker 1999).

Northern blotting

B. subtilis 168 was grown at 37°C in Difco nutrient broth (Becton, Dickinson and Company) starting from an overnight culture diluted to an initial OD_{600} of 0.02. At each time point, 3 OD_{600} of cells were collected and stored at −80°C. Cell pellets were resuspended in 100 µL of 4 mg/mL lysozyme in TE buffer (10 mM Tris-HCl at pH 7.5 and 25°C; 1 mM EDTA) and incubated for 10 min at 25°C before isolating total RNA with 1 mL of TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Total RNA samples (2 µg) were heated at 90°C for 2 min in 1× gel loading buffer (45 mM Tris-borate, 4 M urea, 10% sucrose [w/v], 5 mM EDTA, 0.05% SDS, 0.025% xylene cyanol FF, 0.025% bromophenol blue), separated on a denaturing 10% (8 M urea) polyacrylamide gel, and transferred overnight to a nylon Hybond-N+ membrane (Amersham Biosciences). Blots were simultaneously probed at 37°C with 5′ [^{32}P]-labeled oligonucleotides specific for *B. subtilis* 5S RNA (5′-AACGGGTGTGACCTCTTCGCTATCGCCA) and 6Sa RNA (5′-CGCTACGTCTTGCCGTATGCAAGTAAGAAA), or 5S RNA and 6Sb RNA (5′-TTCCTTTGTTTTGAACCCGCT CTCAGCAGG) in Rapid-hyb buffer (Amersham Biosciences) and analyzed with a PhosphorImager (Molecular Dynamics).

ACKNOWLEDGMENTS

We thank C. Jacobs-Wagner for the gift of *C. crescentus* cultures, M. Cabeen and C. Guerrier-Takada for technical advice, and E. Regulski and E. Butler for helpful discussions. This work was funded by grants from the National Institutes of Health (GM 068819 to R.R.B; R01 HG02602 and HG-00035 to W.L.R) and National Science Foundation (EIA-0323510 to R.R.B). J.E.B. is a Howard Hughes Medical Institute predoctoral fellow.

Received December 23, 2004; accepted February 1, 2005.

REFERENCES

Allen, T.A., Von Kaenel, S., Goodrich, J.A., and Kugel, J.F. 2004. The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat. Struct. Mol. Biol.* **11:** 816–821.

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

Anguera, M.C., Suh, J.R., Ghandour, H., Nasrallah, I.M., Selhub, J., and Stover, P.J. 2003. Methenyltetrahydrofolate synthetase regulates folate turnover and accumulation. *J. Biol. Chem.* **278:** 29856–29862.

Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., et al. 2004. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl. Acad. Sci.* **101:** 6421–6426.

Brosius, J. 1996. More *Haemophilus* and *Mycoplasma* genes. *Science* **271:** 1302–1302.

Brownlee, G.G. 1971. Sequence of 6S RNA of *E. coli*. *Nat. New Biol.* **229:** 147–149.

Cole, J.R., Chai, B., Marsh, T.L., Farris, R.J., Wang, Q., Kulam, S.A., Chandra, S., McGarrell, D.M., Schmidt, T.M., Garrity, G.M., et al. 2003. The Ribosomal Database Project (RDP-II): Previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res.* **31:** 442–443.

Deininger, P.L. and Daniels, G.R. 1986. The recent evolution of mammalian repetitive DNA elements. *Trends Genet.* **2:** 76–80.

Eddy, S.R. and Durbin, R. 1994. RNA sequence analysis using covariance models. *Nucleic Acids Res.* **22:** 2079–2088.

Espinoza, C.A., Allen, T.A., Hieb, A.R., Kugel, J.F., and Goodrich, J.A. 2004. B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat. Struct. Mol. Biol.* **11:** 822–829.

Fenton, M.S. and Gralla, J.D. 2003. Effect of DNA bases and backbone on σ70 holoenzyme binding and isomerization using fork junction probes. *Nucleic Acids Res.* **31:** 2745–2750.

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

Gerstein, M., Sonnhammer, E.L.L., and Chothia, C. 1994. Volume changes in protein evolution. *J. Mol. Biol.* **236:** 1067–1078.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R. 2003. Rfam: An RNA family database. *Nucleic Acids Res.* **31:** 439–441.

Helmann, J.D. and deHaseth, P.L. 1999. Protein-nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners. *Biochemistry* **38:** 5959–5967.

Hindley, J. 1967. Fractionation of ^{32}P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. *J. Mol. Biol.* **30:** 125–136.

Hsu, L.M., Zagorski, J., Wang, Z., and Fournier, M.J. 1985. *Escherichia coli* 6S RNA gene is part of a dual-function transcription unit. *J. Bacteriol.* **161:** 1162–1170.

Ishihama, A. 1999. Modulation of the nucleoid, the transcription apparatus, and the translation machinery in bacteria for stationary phase survival. *Genes Cells* **4:** 135–143.

———. 2000. Functional modulation of *Escherichia coli* RNA polymerase. *Annu. Rev. Microbiol.* **54:** 499–518.

Iyer, L.M., Koonin, E.V., and Aravind, L. 2004. Evolution of bacterial RNA polymerase: Implications for large-scale bacterial phylogeny, domain accretion, and horizontal gene transfer. *Gene* **335:** 73–88.

Kim, K.S. and Lee, Y. 2004. Regulation of 6S RNA biogenesis by switching utilization of both sigma factors and endoribonucleases. *Nucleic Acids Res.* **32:** 6057–6068.

Li, Y. and Altman, S. 2004. In search of RNase P RNA from microbial genomes. *RNA* **10:** 1533–1540.

Li, Z., Pandit, S., and Deutscher, M.P. 1998. 3′ Exoribonucleolytic trimming is a common feature of the maturation of small, stable RNAs in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **95:** 2856–2861.

Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L., and Breaker, R.R. 2004. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306:** 275–279.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. 2002. Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. *Science* **296:** 1285–1290.

Nahvi, A., Barrick, J.E., and Breaker, R.R. 2004. Coenzyme B_{12} riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res.* **32:** 143–150.

Pelchat, M., Grenier, C., and Perreault, J.P. 2002. Characterization of a viroid-derived RNA promoter for the DNA-dependent RNA polymerase from *Escherichia coli*. *Biochemistry* **41:** 6561–6571.

Ren, D., Bedzyk, L.A., Thomas, S.M., Ye, R.W., and Wood, T.K. 2004. Gene expression in *Escherichia coli* biofilms. *Appl. Microbiol. Biotechnol.* **64:** 515–524.

Soukup, G.A. and Breaker, R.R. 1999. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5:** 1308–1325.

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

Trotochaud, A.E. and Wassarman, K.M. 2004. 6S RNA function enhances long-term cell survival. *J. Bact.* **186:** 4978–4985.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D.Y., Paulsen, I., Nelson, K.E., Nelson, W., et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

Vogel, D.W., Hartmann, R.K., Struck, J.C.R., Ulbrich, N., and Erdmann, V.A. 1987. The sequence of the 6S RNA gene of *Pseudomonas aeruginosa*. *Nucleic Acids Res.* **15:** 4583–4591.

Wassarman, K.M. 2002. Small RNAs in bacteria: Diverse regulators of gene expression in response to environmental changes. *Cell* **109:** 141–144.

Wassarman, K.M. and Storz, G. 2000. 6S RNA regulates *E. coli* RNA polymerase activity. *Cell* **101:** 613–623.

Watanabe, T., Sugiura, R., and Sugita, M. 1997. A novel small stable RNA, 6Sa RNA, from the cyanobacterium *Synechococcus* sp. strain PCC6301. *FEBS Lett.* **416:** 302–306.

Weinberg, Z. and Ruzzo, W.L. 2004a. Exploiting conserved structure for faster annotation of non-coding RNAs without loss of accuracy. *Bioinformatics* **20 (Suppl. 1):** i334-i341.

———. 2004b. Faster genome annotation of non-coding RNA families without loss of accuracy. In *Proceedings of the Eighth Annual International Conference on Computational Molecular Biology (RECOMB)*3, pp. 243–251. ACM Press, New York.

Wettich, A. and Biebricher, C.K. 2001. RNA species that replicate with DNA-dependent RNA polymerase from *Escherichia coli*. *Biochemistry* **40:** 3308–3315.

Yarnell, W.S. and Roberts, J.W. 1999. Mechanism of intrinsic transcription termination and antitermination. *Science* **284:** 611–615.