Mining regulatory 5′UTRs from cDNA deep sequencing datasets

Jonathan Livny[1,2,*] and Matthew K. Waldor[2,3]

[1]The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, [2]Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, 181 Longwood Avenue, Boston MA 02115, USA and [3]Howard Hughes Medical Institute

ABSTRACT

Regulatory 5′ untranslated regions (r5′UTRs) of mRNAs such as riboswitches modulate the expression of genes involved in varied biological processes in both bacteria and eukaryotes. New high-throughput sequencing technologies could provide powerful tools for discovery of novel r5′UTRs, but the size and complexity of the datasets generated by these technologies makes it difficult to differentiate r5′UTRs from the multitude of other types of RNAs detected. Here, we developed and implemented a bioinformatic approach to identify putative r5′UTRs from within large datasets of RNAs recently identified by pyrosequencing of the *Vibrio cholerae* small transcriptome. This screen yielded only ~1% of all non-overlapping RNAs along with 75% of previously annotated r5′UTRs and 69 candidate *V. cholerae* r5′UTRs. These candidates include several putative functional homologues of diverse r5′UTRs characterized in other species as well as numerous candidates upstream of genes involved in pathways not known to be regulated by r5′UTRs, such as fatty acid oxidation and peptidoglycan catabolism. Two of these novel r5′UTRs were experimentally validated using a GFP reporter-based approach. Our findings suggest that the number and diversity of pathways regulated by r5′UTRs has been underestimated and that deep sequencing-based transcriptomics will be extremely valuable in the search for novel r5′UTRs.

INTRODUCTION

Non-coding RNAs are now known to regulate gene expression in species from all kingdoms of life. Regulatory RNAs in bacteria, which have been identified in diverse species, fall into two main classes: *trans*-acting RNAs (sRNAs) and regulatory 5′ untranslated regions (r5′UTRs) [reviewed in ref. (1)]. sRNAs are transcribed independently from their target genes and, in most cases, hybridize to cognate mRNAs over short regions of imperfect complimentarity thereby modulating mRNA stability and/or availability for translation. In contrast, r5′UTRs are encoded as part of the mRNA and regulate transcription elongation, translation initiation, or message stability by switching between alternative structures in response to a specific stimulus.

r5′UTRs participate in the regulation of a variety of cellular functions, including the biosynthesis, metabolism, and transport of amino acids, small metabolites, and vitamins, the heat- and cold-shock responses, and the autoregulation of ribosomal protein expression (2–6). While all r5′UTRs mediate regulation of gene expression after transcription initiation, the mechanisms by which they act vary considerably. Riboswitches are the most diverse and well-studied class of r5′UTRs. Binding of a cognate metabolite to a riboswitch alters its conformation and thereby affects the stability of a transcription terminator or alters the accessibility of the ribosome binding site (7). A similar mechanism is employed by T-boxes, r5′UTRs identified mainly in Gram-positive species, whose interaction with uncharged tRNAs leads to destabilization of a Rho-independent terminator 5′ of aminoacyl-tRNA synthtetase genes as well as genes involved in amino acid biosynthesis and transport (8–10). Leader peptides, which have been identified predominantly in Gram-negative species, also regulate amino acid biosynthesis operons by modulating transcription elongation (4). These r5′UTRs encode small ORFs with clusters of codons for a specific amino acid(s) followed by a Rho-independent terminator. When the cellular level of their cognate amino acid(s) is low, ribosomal stalling at these clusters destabilizes the adjacent terminator, leading to increased expression of the downstream operon. r5′UTRs can also mediate post- or co-transcriptional autoregulation of gene expression

---

*To whom correspondence should be addressed. Tel: +617 714 7315; Fax: 617 525 4660; Email: livny@broadinstitute.org

through direct interactions with proteins, a mechanism common in regulating the expression of ribosomal proteins and proteins mediating the cold shock response (3,5). Finally, r5′UTRs known as thermosensors undergo a conformational shift following changes in temperature that affects transcription or translation of the downstream gene (6).

r5′UTRs are relatively short and usually do not encode proteins and thus functional homologues of known r5′UTRs are difficult to identify based on primary sequence conservation. However, since secondary-structure plays a central role in r5′UTR function, covariance models that identify predicted RNA structure conservation have proven useful in identifying functional homologues of characterized r5′UTRs. Kingdom-wide searches using covariance models have lead to the identification of many putative homologues of known r5′UTRs in diverse species and provided important insights into the evolution of r5′UTRs (11,12). However, several recent studies using bioinformatic approaches not based on homology to known r5′UTRs have yielded novel classes of r5′UTRs regulating biological pathways not previously known to be regulated by r5′UTRs (13–16). These observations suggest that current annotations of r5′UTRs represent only a partial catalogue, particularly for Gram-negative species where fewer r5′UTRs are known.

High-throughput DNA sequencing technologies have recently been used to profile bacterial transcriptomes with unprecedented sensitivity (17–20). These studies have generated very large datasets that contain a great diversity of transcripts, including primary transcripts, processed derivatives, and degradation intermediates of messenger, structural, catalytic and regulatory RNAs. These new methodologies hold great potential for discovery of novel regulatory RNAs. However, to date, methods to distinguish r5′UTRs from the large number of other functional transcripts or from 'transcriptional noise' have not been reported. Here, we developed and implemented a bioinformatic approach to mine *Vibrio cholerae* cDNA deep sequencing datasets for r5′UTR-encoding loci. The results of this screen validate the sensitivity and specificity of our approach in distinguishing r5′UTRs from other types of transcripts, including catalytic, structural and trans-acting regulatory RNAs. Subsequent analyses of the RNAs identified in our screen revealed several putative *V. cholerae* functional homologues of known r5′UTRs that had been missed in previous annotations, including one that had been misannotated as an sRNA. We also identified dozens of candidates for novel r5′UTRs, two of which were shown to regulate expression of their downstream genes.

MATERIALS AND METHODS

Bioinformatic analyses

Alignment of the 454 reads [Supplementary Table S2 in ref. (17)] to the *V. cholerae* N16961 genome was conducted with BLASTN version 2.2.17 from NCBI. For each read, only the top hit was kept and only if the percent identity of the alignment was ≥90. Filtering of the 454 dataset for putative r5′UTRs was done using a variety of filters and parameters; the reported combination of filters and parameters was chosen as it yielded the highest ratio of known or putative r5′UTRs to candidate r5′UTRs. For step IV in our filtering of the 454 datasets, reads in different samples were considered to be overlapping if both the 5′- and 3′-ends of one read were within 40 nucleotides of the corresponding ends of the other read. Genome sequences and ORF and COG annotations were obtained from NCBI (accession numbers NC_002505, NC_002506). Gene Ontology (GO) Role Category designations were obtained from TIGR. Rfam annotations were based on version 9.1. Transcription terminators were predicted by RNAMotif, TransTerm and FindTerm as described (21). The Artemis Comparison Tool release 9 (22) was used to visualize 454 read abundances superimposed on genome annotations. Information on the function of *Escherichia coli* proteins was obtained from EcoCyc (23).

Construction of GFP reporter strains

5′UTRs were amplified using the oligos listed in Supplementary Table S1 and cloned into the NsiI and NheI sites of plasmid pXG10 (24). The respective 5′-end of each 5′UTR insert was determined based on the 5′-end of the corresponding cDNAs detected by 454. Each r5′UTR inserts included the start codon of its 3′ ORF along with up to 19 additional codons.

GFP reporter assays

Escherichia coli DH5a or *V. cholerae* NI6961 strains carrying the indicated plasmids as well as those carrying a control plasmid [pXG0 (24)] expressing luciferase instead of GFP were grown overnight in LB + 2.5 µg/ml chloramphenicol (Cm2.5). For the experiments shown in Figure 3, these overnight cultures were subcultured 1:50 in 96-well plates in M63 medium (0.2% glucose, 1 mM MgSO4, 1 µg/ml B1, Cm2.5) supplemented with 16 L-amino acids (Sigma Aldrich) (excluding L-leucine, glycine, L-histidine and L-lysine) each added to a final concentration of 20–200 µM. Where indicated L-leucine and glycine (Sigma Aldrich) were added to a final concentration of 2.5 mM, respectively. For experiments described in Figure 4, overnight cultures were sub-cultured 1:250 in LB Cm2.5 and grown to $OD_{600} \sim 0.8$–1. Aliquots of these cultures were washed twice with one volume M63 medium and diluted to an OD_{600} of ~0.4 in M6 medium supplemented with casamino acids (Dibco) to the final concentrations indicated. Total culture fluorescence was measured using a Synergy HT Multi-Mode Microplate Reader (BioTek) with 485/20 nm optical excitation filter, 528/20 nm emission filter and measurement height of 8.0 mm. GFP fluorescence was calculated by subtracting the fluorescence of strains carrying the GFP fusions from those carrying the luciferase control plasmid. For experiments described in Figure 4, GFP fluorescence was normalized to OD_{600}.

Table 1. Candidate r5′UTRs sharing conserved genomic context with known families of r5′UTRs or with putative RibEx RLEsa

No./ name[a]	ORF No.	ORF product	3′CGC (Rfam)[b]	3′CGC and/or putative motif (RibEx)[b,c]
1	*VC0326*	*50S ribosomal protein L10*	*L10 leader(117)*	*RLE0035(35)*
2	**VC0570**	**50S ribosomal protein L13**	**L13 leader(64)**	**RLE0227(17)**
3	VC0647	*Polynucleotide phosphorylase/polyadenylase*	*S15(1)*	*RLE01541(8)*
4*	**VC0705**	***Chorismate mutase/prephenate dehydratase***	***T-box(11), T-box(1)***	–
5	**VC0875**	**Prolyl-tRNA synthetase**	**T-box(15)**	**RLE018(5)**
6*	VC0894	Thiamine biosynthesis protein ThiI	SAM-IV(1)	RLE0079(7)
7	VC1091	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein	T-box(4), SAM(8)	RLE0210(6)
8*	**VC1623**	**Carboxynorspermidine decarboxylase**	**speF(25), Lysine(17)**	–
9	*VC2030*	*Ribonuclease E*	*rne5(32)*	–
10*	VC2108	Erythronate-4-phosphate dehydrogenase	T-box(3), SAM(1)	–
11*	VC2334	Hypothetical protein	ykoK(3)	RLE0310(6)
12	**VC2356**	**Sodium/alanine symporter**	**Glycine(118)**	–
13	**VC2439**	**Methyl-accepting chemotaxis protein**	**GEMM RNA motif(20), SAM(1)**	–
14	VC2522	Hypothetical protein	yybP-ykoY(9)	RLE0334(5)
15	VC2645	Acetylornithine deacetylase	TPP(4)	–
16	*VC2712*	*Xanthine/uracil permease family protein*	*PyrR(32), Purine(17), TPP(4)*	–
17	VCA0142	C4-dicarboxylate transport transcriptional regulatory protein	MOCO RNA motif(1), GEMM RNA motif(2)	RLE0123(2)
18	**VCA0179**	**NupC family protein**	**Purine(11)**	–
19	VCA0278	Serine hydroxymethyltransferase	Glycine(8)	RLE0085(7)
20	*VCA0287*	*Threonyl-tRNA synthetase*	*T-box(60)*	*RLE020(5)*
C1*	*VC2490*	*2-isopropylmalate synthase*	*T-box(11), Leu leader(28), ydaO-yuaA(1)*	–
21	*VC0007*	*50S ribosomal protein L34*	–	–
22	*VC0218*	*Ribosomal protein L28*	–	*RLE0348*
23	*VC0324*	*50S ribosomal protein L11*	–	*RLE0241, RLE0148(6)*
24	*VC2597*	*30S ribosomal protein S10*	–	*RLE0110(25)*
25	*VC2679*	*50S ribosomal protein L31*	–	*RLE0089(8)*
26	*VCA0166*	*Cold-shock transcriptional regulator CspA*	–	*RLE0357(6)*
27	*VCA0184*	*Cold-shock DNA-binding domain-containing protein*	–	*RLE0357(6)*
28	*VCA0933*	*Cold-shock domain-contain protein*	–	*RLE0357(6)*
29	VCA0819	Co-chaperonin GroES	–	RLE0003(75)
30	VCA1075	Hypothetical protein	–	RLE0037
31	VCA0518	Bifunctional fructose-specific PTS protein	–	RLE0062(8)
32	VC2431	DNA topoisomerase IV subunit B	–	RLE0226, SAM
33	VC2738	Phosphoenolpyruvate carboxykinase	–	RLE0239(5)
34*	VC1923	Trigger factor	–	RLE0241(7)
35*	VC1046	3-ketoacyl-CoA thiolase	–	RLE0244(2)
36	VC1258	DNA gyrase, subunit A	–	RLE0300(4)
37	VC0633	Outer membrane protein OmpU	–	RLE0331(5)
38	VC0972	Porin, putative	–	RLE0331(5)
39	VC1130	DNA-binding protein H-NS	–	RLE0337(5)

[a]Asterisk denote candidates containing or directly upstream of a putative transcription terminator.
[b]The number in parentheses denotes the number of r5′UTR in each family or of RLEs found to share 3′CGC with the candidate r5′UTR.
[c]RLEs predicted to be encoded by the candidate are underlined, RLEs with 3′CGC are not.

RESULTS AND DISCUSSION

Summary of *V. cholerae* small transcriptome datasets used in this study

The datasets used in our analyses were obtained by 454 pyrosequencing of DNA libraries complementary to primary transcripts and transcript fragments (from here on referred to collectively as transcripts) 14–200 nucleotides in length isolated from four independent cultures of *V. cholerae* (17). Of the 681 205 total reads in these datasets, 362 345 align with 100% identity to the *V. cholerae* genome, corresponding to 37 494 non-identical transcripts and 6208 sets of non-overlapping transcripts. Transcripts overlapping 17 of 19 (90%) and 20 of 22 (91%) previously annotated or characterized sRNAs and r5′UTRs, respectively, were identified. Initial analysis of these datasets yielded numerous candidates for novel sRNAs, several of which were confirmed by northern analysis (17). The identification of so many hitherto unannotated putative sRNAs along with the sensitivity with which previously annotated r5′UTRs were identified suggested to us that there would likely be unannotated *V. cholerae* r5′UTRs among the transcripts detected by deep sequencing. However, identifying unknown r5′UTRs

Table 2. Candidate for novel r5′UTR lacking conserved genomic context with Rfam r5′UTRs or RibEx RLEs

Name	ORF	ORF product	Gene Ontology (GO) role category
40	VC2656	Fumarate reductase flavoprotein subunit	Anaerobic respiration
41	VCA0013	Maltodextrin phosphorylase	Carbohydrate metabolism
42	VC2188	Flagellin	Ciliary or flagellar motility
43	VC2678	Primosome assembly protein PriA	DNA replication, synthesis of RNA primer
44	VC1442	Cytochrome *c* oxidase, subunit CcoN	Electron transport
45	VC2295	Na(+)-translocating NADH-quinone reductase subunit A	Electron transport
46*	VC2701	Thiol:disulfide interchange protein precursor	Electron transport, protein thiol-disulfide exchange, cytochrome complex assembly
47	VCA1067	Aldehyde dehydrogenase	Fermentation
48	VC0374	Glucose-6-phosphate isomerase	Glycolysis
49	VCA0843	Glyceraldehyde-3-phosphate dehydrogenase	Glycolysis
50	VC0986	Adenylate kinase	Nucleobase, nucleoside and nucleotide interconversion
51	VCA0623	Transaldolase B	Pentose-phosphate shunt
52	VCA0870	D-alanyl-D-alanine endopeptidase	Peptidoglycan biosynthetic
53	VC2421	N-acetyl-anhydromuranmyl-L-alanine amidase	Peptidoglycan metabolism
54	VC0322	Preprotein translocase subunit SecE	Protein secretion
55	VC2748	Nitrogen regulation protein NR(II)	Regulation of nitrogen utilization
56*	VC1796	Middle operon regulator-related protein	Regulation of transcription, DNA-dependent
57	VC1901	Sodium/proton antiporter	Sodium ion transport, hydrogen transport
58	VC0347	RNA-binding protein Hfq	Targeting of mRNA for destruction, involved in RNA interference
59	VC0038	Hypothetical protein	–
60	VC0381	Hypothetical protein	–
61	VC1576	Hypothetical protein	–
62	VC1613	Hypothetical protein	–
63	VC1891	Hypothetical protein	–
64	VC2002	Hypothetical protein	–
65	VC2264	Hypothetical protein	–
66	VC2647	Hypothetical protein	–
67	VCA0327	Hypothetical protein	–
68	VCA0363	Hypothetical protein	–
69	VCA0743	Hypothetical protein	–

Asterisk denotes candidates containing or directly upstream of a putative transcription terminator.

within these large datasets required an effective way to distinguish such transcripts from the great number and diversity of other types of transcripts.

Filtering the datasets for r5′UTRs

We took several steps to filter our 454 datasets for transcripts derived from r5′UTRs. First, we discarded all transcripts that did not overlap putative 5′UTRs, defined as regions 100 bp upstream of annotated start codons and not overlapping other annotated genes. This filter led to a large reduction in the numbers of total transcripts and known sRNAs but to only a small decrease in the number of previously annotated r5′UTRs (Figure 1, filter I). However, we found that nearly half of previously annotated *trans*-acting regulatory, structural, and catalytic RNAs and nearly 20% of all unique transcripts detected overlapped putative 5′UTRs. Indeed, transcripts overlapping the 5′UTRs of 1048 (27%) *V. cholerae* ORFs were identified. Most of these 5′UTR transcripts are likely the result of aborted transcription or incomplete mRNA degradation and do not represent r5′UTRs. Transcripts produced by r5′UTR-mediated regulation usually do not extend into the coding region of the mRNA. Thus, we next removed all transcripts overlapping annotated ORFs. This step also led to a significant
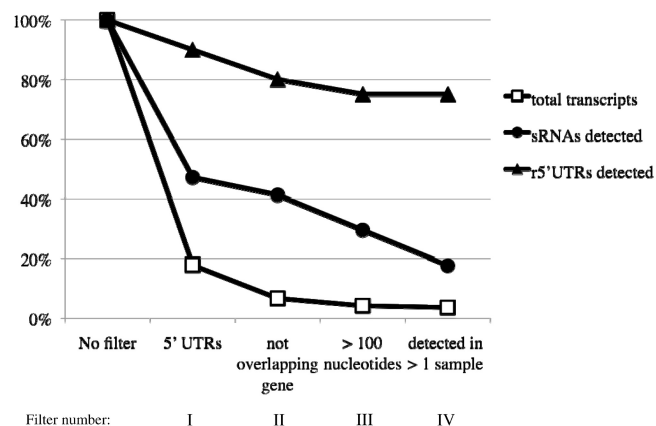
Figure 1. Results of *in silico* mining of *V. cholerae* cDNA datasets for r5′UTRs. The values shown correspond to the percentage of all transcripts in the 454 datasets remaining after addition of each filter.

decrease in the number of total transcripts but to only a modest reduction in the numbers of known r5′UTRs (Figure 1, filter II). In an effort to enrich our datasets for r5′UTRs relative to sRNAs, we next eliminated all transcripts shorter than 100 nucleotides. This was based on previous observations that transcripts associated with

characterized r5′UTRs are almost always longer than 100 nucleotides, whereas a number of known sRNAs are <100 nucleotides in length. Finally, from the set of transcripts fulfilling the above criteria, we filtered out those that did not overlap another transcript in at least one of the three other independent samples (Figure 1, filter IV), since we reasoned that transcripts detected in only one of the four independent samples were less likely to correspond to real r5′UTRs.

As shown in Figure 1, application of filters I–IV removed the vast majority of the total transcripts and of the sRNAs in the dataset but left behind most of the annotated r5′UTRs. Specifically, 96% of the total unique transcripts and 82% of the annotated sRNAs were eliminated, whereas only 25% of the annotated r5′UTRs were removed. The three sRNAs that remained included C1, msr and 6S RNA. C1 is an uncharacterized small intergenic transcript that was discovered by our group in a bioinformatic screen and annotated as a putative sRNA (25). As described below, C1 actually corresponds to an r5′UTR rather than to an sRNA. msr is a non-coding RNA gene found in Retron elements in diverse bacteria whose biological function is poorly understood (26,27). 6S RNA has been well characterized in *E. coli* where it acts as *trans*-acting regulatory RNA; however, unlike the vast majority of other characterized sRNAs, 6S does not target mRNAs but rather interacts with and modulates the activity of RNA polymerase (28). Thus, none of the sRNAs remaining in our filtered dataset corresponds to a canonical mRNA-targeting *V. cholerae* sRNA such as RyhB, Qrr1-4, VrrA, or MicX. Taken together, these observation suggest that our approach was effective in sensitively and specifically distinguishing known r5′UTRs from other transcripts, including other regulatory or catalytic non-coding RNAs.

The above analysis yielded transcripts corresponding to 15 characterized or putative r5′UTRs previously annotated in the Rfam database. Rfam is a collection of multiple sequence alignments, consensus secondary structures, and covariance models (CMs) representing families of non-coding RNAs. New members of these families are identified by Rfam based on predicted secondary-structure conservation using sensitive BLAST filters in combination with CMs (12). We were initially surprised that previously annotated putative *V. cholerae* TPP riboswitches and a putative r5′UTR of ribosomal protein S15 were identified in our screen. Based on characterization of their homologues in other species, these r5′UTRs are thought to act through sequestration of ribosome binding sites (29,30). Thus, these putative r5′UTRs, unlike r5′UTRs that regulate expression through transcription termination, were not expected to produce discrete short transcripts. However, it has been shown that r5′UTRs that do not employ Rho-independent transcriptional termination elicit formation of stem loop structures at specific locations in the 5′UTR. We therefore postulate that even in the absence of a strong termination signal, these structured elements act as sites for transcription termination, RNA processing, and/or boundaries for RNA degradation to reproducibly yield short transcripts

terminating within the 5′UTR that are detectable by high-throughput sequencing.

Five previously annotated r5′UTRs were eliminated by our filtering. Three were eliminated because their corresponding transcripts overlapped ORFs while the other two were lost because their transcript terminated >100 bp upstream of the annotated start codon of their respective 3′ORFs. Two of the r5′UTRs removed by our filtering, LR-PK1 and mini-ykkC, are putative structured motifs identified in computational screens that have yet to be experimentally validated and thus may not correspond to functional r5′UTRs. However, one of the r5′UTRs missed has been experimentally characterized and the other two belong to well-characterized families of r5′UTRs and thus also likely correspond to bona fide r5′UTRs. It is therefore likely that a significant number of unannotated r5′UTRs are also missing from our list of candidate r5′UTRs.

Identification of candidate *V. cholerae* functional homologues of previously characterized or predicted r5′UTR based on conserved genomic context

In addition to transcripts overlapping previously characterized or putative regRNAs, we identified transcripts corresponding to 69 r5′UTRs that were not identified in Rfam (from here on referred to as candidate r5′UTRs). Since r5′UTRs are co-transcribed with their target mRNAs, we reasoned that comparing the respective genomic location of candidate and known r5′UTRs vis-à-vis their 3′ genes might enable us to identify functional homologues of known r5′UTRs missed in previous annotations. Thus, we compared the Clusters of Orthologous Groups (COG) designations of genes downstream of all loci we identified to those of genes downstream of known or putative r5′UTRs in Rfam. We also conducted similar comparisons with genes downstream of putative Riboswitch-like elements (RLEs) found in the RibEx database (13). Unlike Rfam, RibEx identifies putative r5′UTRs based on conserved primary sequence upstream of orthologous groups of genes in multiple genera and has been used to identify several hundred putative RLE families in addition to those annotated in Rfam.

All genes 3′ of previously annotated *V. cholerae* r5′UTRs shared COG designations with genes 3′ of numerous Rfam r5′UTRs in other species. We also found many candidate r5′UTRs that share 3′ conserved genomic context (3′CGC) with known or putative r5′UTRs annotated in the Rfam database and/or with putative RLEs in RibEx (Table 1). In some cases, candidates shared 3′CGC with only a few annotated Rfam r5′UTRs and/or with seemingly functionally unrelated families of r5′UTRs, suggesting that the apparent conservation in genomic context might be coincidental and unlikely to reflect a functional or evolutionary relationship between the candidate and previously characterized r5′UTRs. However, 12 candidates (bold in Table 1) were found to share 3′CGC with more than 10 known or putative Rfam r5′UTRs in the same family, strongly suggesting they correspond to bona fide but previously
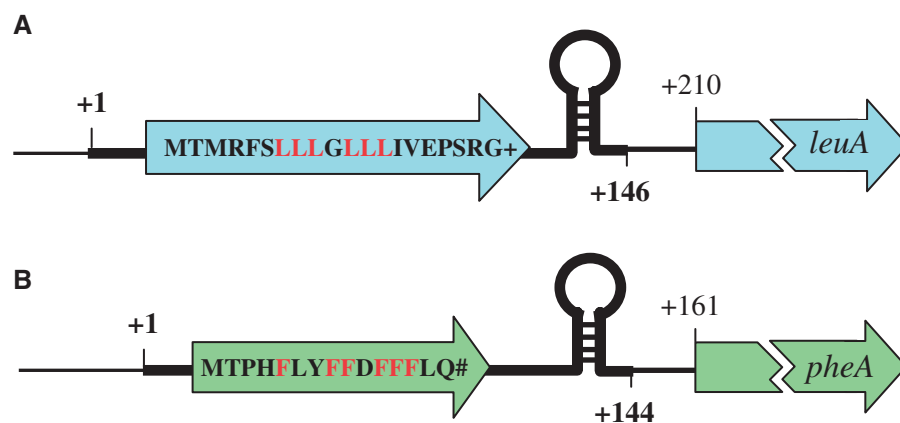
Figure 2. Features of putative *V. cholerae* (**A**) LeuL and (**B**) PheL leader peptides. The two numbers in bold denote the relative positions of the 5′- and 3′-ends of each transcript based on the 454 data. The third number indicates the relative position of the downstream ORF as annotated by NCBI. Cognate clusters of codons for each leader peptide are shown in red. The '+' and '#' symbols denote stop codons.

unannotated *V. cholerae* r5′UTRs. We were surprised to find that C1 shared 3′CGC with numerous r5′UTRs in Rfam. For C1 and 14 candidate r5′UTRs in Table 1 (italicized) there is additional experimental and/or bioinformatic evidence suggesting they correspond to bona fide r5′UTR. Seven of these candidates are discussed in more detail below.

(i) *C1*—The gene 3′ of C1, 2-isopropylmalate synthase, is annotated as the first gene in a putative leucine biosynthesis operon homologous to the *leuABCD* operon found in *E. coli* and many other Enterobacteriacea. In *E. coli*, this operon is regulated by the leader peptide LeuL (31). As shown in Figure 2A, we found that the C1 transcript overlaps a short open reading frame of 20 residues that encodes two clusters of three leucine-encoding codons. Importantly, half of these are CUA codons, which represent only 8% of all annotated leucine codons in *V. cholerae*. This over-representation of rare codons is a hallmark of characterized leader peptides and is important for their function (32). Consistent with other leader peptides, the short ORF encoded by C1 is followed by a Rho-independent terminator, which presumably mediates transcription termination in the absence of ribosome stalling. These observations suggest that C1 was misannotated as an sRNA; instead, it likely corresponds to the *V. cholerae* LeuL leader peptide.

(ii) *Candidate No. 4*: *vc0705* is a homologue of the gene encoding the *E. coli* PheA which is subject to co-transcriptional regulation through a leader peptide (33). Indeed, as shown in Figure 2B, candidate No. 4 possesses all the features of a phenylalanine-regulated leader peptide, as it encodes a 45 bp open reading frame that contains a cluster of six phenylalanine residues directly upstream of a putative Rho-independent terminator.

(iii) *Candidate No. 9*: The *E. coli* RNase E has been shown to reduce the stability of its own transcript through its interactions with the *rne* 5′UTR (34).

Putative homologues of this RNase E regulated motif have been identified by Rfam in several genera of Gamma-proteobacteria including *Salmonella* and *Yersinia* sp.

(iv) *Candidates No. 20*: In *E. coli*, binding of the threonyl-tRNA to the 5′UTR of its own mRNA has been shown to prevent initiation of translation (35).

(v) *Candidates No. 3*: Expression of *E. coli* polynucleotide phosphorylase (Pnp) is post-transcriptionally auto-regulated through degradation of a double-stranded structure in the *pnp* mRNA leader (36–38).

(vi) *Candidate No. 1*: In *E. coli*, L10 binds the 5′UTR of its own transcript to modulate its translation (39).

(vii) *Candidate No. 16*: *In silico* annotations for both purine and PyrR-dependent r5′UTRs suggest they are found almost exclusively in Gram-positive species, with only five of the 357 purine or PyrR r5′UTRs in the Rfam database predicted in Gram-negative strains. However, variants of purine-sensing riboswitches have recently been discovered in *Mesoplasma florum* that share conserved structure with sequences upstream of putative xanthine/uracil permease genes in *Vibrio* sp. (40).

Interestingly, only one of the seven putative r5′UTR homologues described above correspond to putative riboswitches, suggesting covariance models may be more effective in identifying functional homologues of known riboswitches compared to those of other types of r5′UTRs. This may reflect the fact that the function of riboswitches dictates a relatively higher level of structure conservation. Specifically, the secondary structures of riboswitches are constrained both at the expression platform and aptamer region, the latter needing to maintain a very specific conformation to preserve ligand specificity. In contrast, the only structural constraint on leader peptide function is in their terminator/anti-terminator region. Similarly, auto-regulatory r5′UTRs need only to maintain structures that modulate RBS accessibility and/or affect transcript stability in response
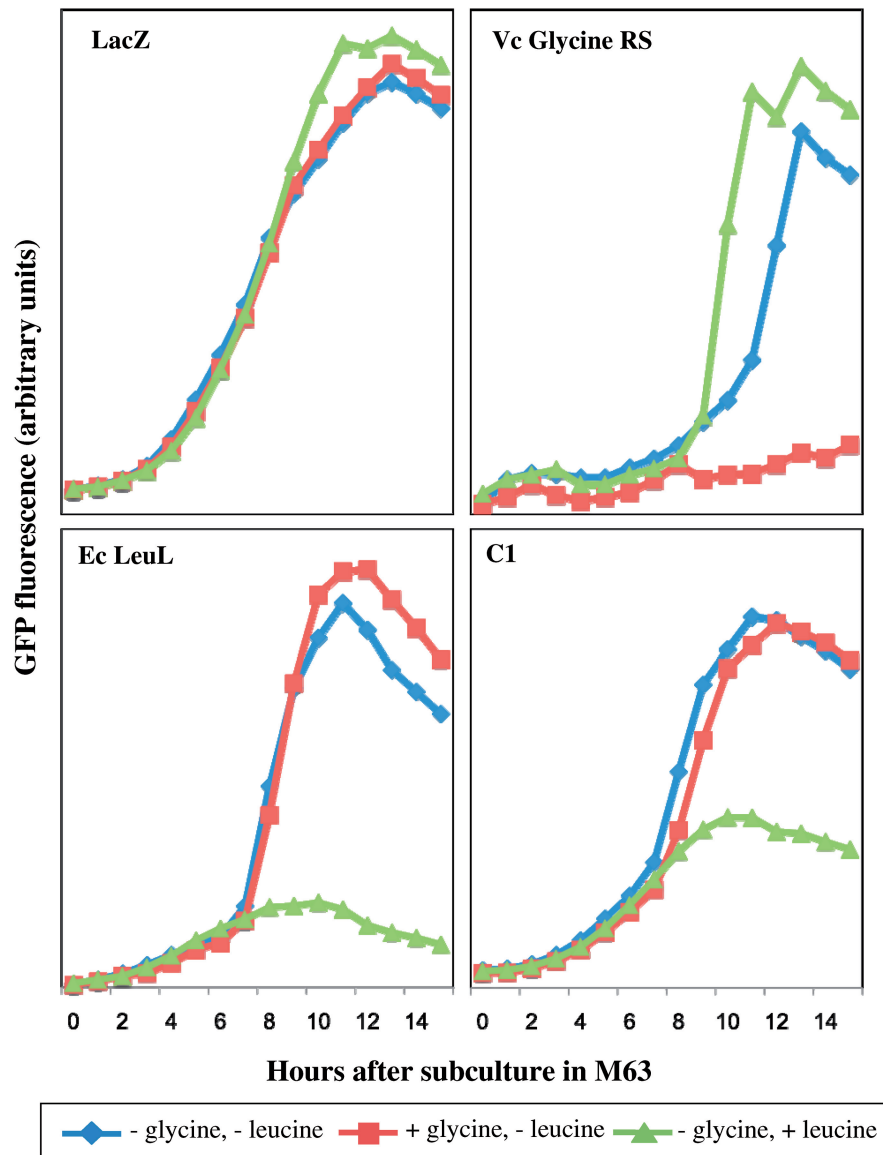
Figure 3. Expression of GFP fused to indicated 5′UTRs. *Escherichia coli* strains carrying the indicated fusions were grown in defined media lacking both glycine and L-leucine (blue diamonds) or supplemented with either glycine (red squares) or L-leucine (green triangles). Results from a representative experiment are shown.

to protein binding. However, while covariance models such as those used by Rfam may be effective in identifying well-conserved functional homologues of known riboswitches, transcriptome mining may be more effective in identifying significantly diverged variants of known riboswitch families, such as functional homologues of the purine-sensing riboswitches in *Mesoplasma florum*.

In addition to the seven loci described above, we found a number of candidate r5′UTRs that do not share 3′CGC with r5′UTR families annotated in Rfam but whose genomic context strongly suggests they are r5′UTRs nonetheless (candidates 21–28 in Table 1). Five of these candidates are upstream of genes encoding ribosomal proteins. These observations are consistent with several studies showing that post-transcriptional or co-transcriptional

auto-regulation are common mechanisms for modulating the expression of ribosomal proteins (41). Indeed, transcription of the *V. cholerae* S10 operon has been shown to be regulated by an attenuator in the 5′UTR (42). We also identified candidate loci upstream of genes encoding putative *V. cholerae* homologues of the *E. coli* cold-shock proteins CspA and CspE. In *E. coli*, several genes involved in the cold-shock response, including *cspA* and *cspE*, are subject to auto-regulation mediated by structural changes in their 5′UTRs (5). Finally, candidate no. 29 was identified 5′ of the gene encoding the protein chaperone GroES; a putative 'fourU' thermosensor has been identified in the 5′UTR of GroES in *Salmonella* sp. (43). Taken together, our findings suggest that our mining was effective in identifying previously unannotated functional homologues of characterized 5′UTRs.

Identification of candidate r5′UTR that do not share conserved genomic context with known or putative r5′UTRs

Of the 69 previously unannotated putative r5′UTRs we identified in our screen, a total of 30 do not share 3′CGC with known families of r5′UTRs (Table 2). These candidate r5′UTRs were found upstream of genes implicated in a variety of cellular processes and of 12 ORFs encoding hypothetical proteins. Multiple candidates were identified upstream of genes implicated in glycolysis, electron transport and peptidoglycan biosynthesis/metabolism, suggesting these may represent members of novel r5′UTRs families that, like other families of r5′UTRs such as TPP riboswitches, are responsible for regulating different steps in the same pathway or process.

Using a GFP reporter approach to measure r5′UTR-mediated regulation

To experimentally test and begin to characterize a few of the candidate r5′UTRs identified in our screen, we adapted an approach that was developed by Urban *et al.* (24) to study sRNA-mediated regulation of mRNAs in trans. Urban and colleagues constructed a plasmid into which 5′UTRs of interest can be introduced directly downstream of a constitutive promoter to create translational fusions with a gene encoding GFP. The fluorescence generated from GFP is used as a means to gauge GFP expression from different fusions; a control fusion of GFP to the *E. coli* LacZ 5′UTR serves as a negative control for these assays. Since the identical constitutive promoter is present in all fusions, these constructs can be used to measure regulation of gene expression that is not mediated by changes in transcription initiation.

To test the efficacy of this approach for measuring r5′UTR-mediated regulation, we compared expression of GFP fused to two characterized r5′UTRs, *E. coli* LeuL and the *V. cholerae* Glycine riboswitch, to that of GFP fused to the *E. coli* LacZ 5′UTR. As shown in Figure 3, when cultures were grown in minimal media supplemented with 16 amino acids excluding leucine and glycine, expression of all three fusions increased significantly. In the control (LacZ) fusion, GFP expression was similar when this medium was supplemented with either glycine or leucine (Figure 3). In contrast, expression of GFP fused to the *V. cholerae* Glycine riboswitch was almost completely repressed when glycine was added to the media; inhibition of GFP expression by glycine appears to be fairly specific, since addition of leucine did not repress GFP expression (Figure 3). Amino acid specificity was also observed with GFP expression from the LeuL fusion. However, in this case, expression was markedly decreased when leucine was added; the addition of glycine did not inhibit expression (Figure 3). Taken together, these observations suggest that monitoring GFP expression with this reporter system is a useful technique for investigating r5′UTR-mediated regulation. Finally, similar to the LeuL fusion, expression of a C1-GFP fusion was also repressed by leucine but not

glycine, providing strong support for the bioinformatic evidence implicating C1 as the *V. cholerae* LeuL.

In the case of LeuL, our results with the GFP fusion are consistent with previous studies showing that the LeuL leader down-regulates expression of the *leuABCD* operon in response to leucine (31,32). However, the glycine-mediated repression of GFP expression by the *V. cholerae* Glycine riboswitch was surprising as Mandal *et al.* (44) found that glycine had the opposite effect on expression of a reporter gene fused to the *B. subtilis* Glycine riboswitch. Interestingly, even though the *B. subtilis* and *V. cholerae* Glycine riboswitches share significant structure conservation (44), they are encoded 5′ of unrelated genes, the former upstream of *gcvT*, an aminomethyltransferase that mediates conversion of glycine to serine, and the latter upstream of *vc1422*, a putative sodium/alanine symporter. VC1422 is a homologue of *E. coli* CycA, an APC family transporter of glycine, serine and alanine (45), as well as several other gene products annotated as glycine symporters in both Gram-positive and Gram-negative species. Our findings suggest that even though the aptamer regions of the *V. cholerae* and *B. subtilis* Glycine riboswitches share significant structural conservation that has maintained their specificity for glycine, the two riboswitches elicit opposite regulatory responses on their respective 3′ genes. The *V. cholerae* Glycine riboswitch appears to have evolved to up-regulate glycine uptake in the absence of glycine, whereas the *B. subtilis* Glycine riboswitch has evolved to up-regulate glycine catabolism when glycine is abundant. The mechanisms that account for how these similar riboswitches elicit opposite effects on the expression of their respective 3′ genes warrants further investigation.

Two candidates for novel r5′UTRs down-regulate expression of their downstream gene in response to increased amino acid concentration

In the experiments described above, the cognate signals for the r5′UTRs of interest were known based on previous studies. However, for the candidate r5′UTRs that do not share 3′CGC with well-characterized classes of r5′UTRs, *a priori* determination of these signals is difficult. Since many r5′UTRs are known to be regulated by amino acids, we constructed fusions of GFP with several candidate r5′UTRs and measured their expression in minimal media supplemented with either 1 or 0.1% casamino acids (CAA). As shown in Figure 4, a construct carrying the *E. coli* LacZ 5′UTR produced more GFP in the presence of 1% casamino acids (CAA) than in 0.1% CAA, presumably due to an increase in translation efficiency. In contrast, fusions of GFP with *E. coli* or *V. cholerae* LeuL or the *V. cholerae* Glycine riboswitch exhibited less GFP expression in high CAA. Similar patterns of GFP expression (higher in 0.1% than 1% CAA) were also observed when two candidate r5′UTRs were fused to GFP (Figure 4, ppbG and thiI). One of these candidates is upstream of *vca0870* encoding the *V. cholerae* homologue of penicillin-binding protein 7 (*pbpG*), a protein involved in peptidoglycan metabolism (46). The other candidate is upstream of a gene annotated
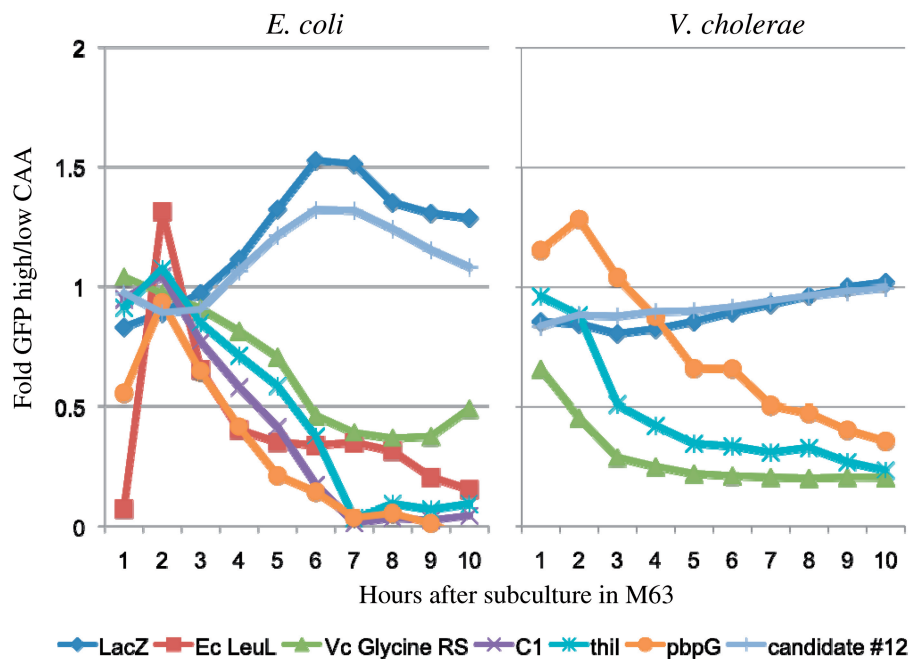
Figure 4. Effect of increased amino acid concentration on the expression of GFP fused to known or candidate r5′UTRs. Results from representative experiments in *E. coli* and *V. cholerae* are shown.

as *thiI*. ThiI has been implicated in thiamine biosynthesis and tRNA modification; in *Salmonella typhimurium,* ThiI is the only component of the thiamine biosynthesis pathway whose expression is not regulated by TPP riboswitches (47). These observations suggest that the *pbpG* and *thiI* UTRs mediate co- or post-transcriptional repression of their respective downstream genes when amino acid concentrations increase. However, it is not clear from these data whether the *pbpG* and *thiI* UTRs influence on gene expression is triggered by their direct interaction with amino acids or through the participation of other factor(s). As shown in Figure 4, both of these candidate r5′UTRs exhibited more GFP expression in 0.1% than 1% CAA in *V. cholerae* as well as in *E. coli.* Thus, if additional factors are required for the regulatory effects of these *V. cholerae* UTRs, these factors appear to be conserved in *E. coli.* The relative expression of GFP from the reporter construct carrying the 5′UTR of candidate No. 12 in low and high CAA was similar to the control LacZ construct (Figure 4), suggesting that this candidate r5′UTR is not sensitive to changes in amino acid concentration; alternatively this candidate may not correspond to a r5′UTR.

As shown in Table 1, the *thiI* r5′UTR shares conserved 3′CGC with one SAM-IV riboswitch and with the RLE0079 motif. The latter motif was identified upstream of *thiI* homologues in seven Gram-negative species (13). We identified a canonical Rho-independent terminator near the 3′ end of the *thiI* 5′UTR, suggesting that the regulatory effects of this UTR on *thiI* expression may be achieved through a terminator/antiterminator switch. Indeed, in Northern analyses, the abundance of a small transcript overlapping the *thiI* 5′UTR was markedly

increased in high versus low CAA (data not shown). Putative Rho-independent terminators were also identified within 100 bps of the *thiI* start codon in several *E. coli* strains, *Shewanella* sp., *Streptococcus* sp. and *Vibrio* sp. (21), suggesting that the *thiI* homologues in these strains may be regulated by a similar mechanism.

The *pbpG* r5′UTR lacks 3′CGC with any known or putative r5′UTRs. Thus, it is not clear if this motif is conserved in other species. Since no terminator was predicted in the *pbpG* 5′UTR, the mechanism by which this r5′UTR mediates regulation of its downstream message is not clear.

CONCLUSIONS

Taken together our findings suggest that transcriptome profiles acquired through new deep sequencing techniques will be a rich source of information about r5′UTRs. We developed a simple set of filters to mine the *V. cholerae* small transcriptome acquired by pyrosequencing of cDNA libraries. Our approach appears to be effective, as we identified most of the previously annotated, though in most cases not experimentally verified, r5′UTRs but relatively few of the total transcripts or trans-acting regulatory RNAs found in the original datasets. We also identified numerous candidate r5′UTRs not annotated in previous computational screens that share conserved genomic context with known r5′UTRs. Finally, we identified candidate r5′UTRs upstream of several classes of genes whose expression has not been previously shown to be subject to regulation by r5′UTRs. Thus, our findings highlight the utility of mining deep-sequencing transcriptome data as a complementary

approach to computational screens for identifying r5′UTRs. Overall, our observations suggest that the distribution of known classes of r5′UTRs and the diversity of functions regulated by r5′UTRs are much greater than what has been suggested by previous in silico genomics-based annotations.

Although conservation-based computational approaches such as Rfam are invaluable for identification of r5′UTRs, their reliance on homology to known r5′UTRs is an inherent limitation which preclude the identification of new classes of r5′UTRs. Also, since these approaches often rely on seed alignments of r5′UTRs from closely related species, identification of functional homologues of known r5′UTRs in species that are highly diverged from those represented in the seed is often not possible. Thus, using high-throughput transcriptomics to identify novel r5′UTRs and/or functional homologues of known r5′UTRs in less well-studied bacterial species and then integrating these loci into kingdom-wide bioinformatic screens could significantly improve annotations for r5′UTRs, particularly outside well-studied genera.

Several recent studies have revealed that the diversity of ligands and environmental cues that elicit r5′UTR-mediated regulation is greater than previously thought. Thus, as more families of r5′UTRs are identified using a variety of approaches, the task of identifying each of their specific cognate signals will become increasingly daunting. The GFP reporter approach we have implemented for validating r5′UTR-mediated regulation here should be useful in addressing this challenge, providing an efficient way to screen a large number of candidate r5′UTRs in a wide variety of conditions.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Conflict of interest statement. None declared.

REFERENCES

1. Waters,L.S. and Storz,G. (2009) Regulatory RNAs in bacteria. *Cell*, **136**, 615–628.
2. Dambach,M.D. and Winkler,W.C. (2009) Expanding roles for metabolite-sensing regulatory RNAs. *Curr. Opin. Microbiol.*, **12**, 161–169.
3. Paul,B.J., Ross,W., Gaal,T. and Gourse,R.L. (2004) rRNA transcription in Escherichia coli. *Annu. Rev. Genet.*, **38**, 749–770.
4. Henkin,T.M. and Yanofsky,C. (2002) Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *Bioessays*, **24**, 700–707.
5. Gualerzi,C.O., Giuliodori,A.M. and Pon,C.L. (2003) Transcriptional and post-transcriptional control of cold-shock genes. *J. Mol. Biol.*, **331**, 527–539.
6. Klinkert,B. and Narberhaus,F. (2009) Microbial thermosensors. *Cell Mol. Life Sci.*, **66**, 2661–2676.
7. Tucker,B.J. and Breaker,R.R. (2005) Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.*, **15**, 342–348.
8. Grundy,F.J. and Henkin,T.M. (1993) tRNA as a positive regulator of transcription antitermination in B. subtilis. *Cell*, **74**, 475–482.
9. Gutierrez-Preciado,A., Henkin,T.M., Grundy,F.J., Yanofsky,C. and Merino,E. (2009) Biochemical features and functional implications of the RNA-based T-box regulatory mechanism. *Microbiol. Mol. Biol. Rev.*, **73**, 36–61.
10. Merino,E., Jensen,R.A. and Yanofsky,C. (2008) Evolution of bacterial trp operons and their regulation. *Curr. Opin. Microbiol.*, **11**, 78–86.
11. Gardner,P.P., Daub,J., Tate,J.G., Nawrocki,E.P., Kolbe,D.L., Lindgreen,S., Wilkinson,A.C., Finn,R.D., Griffiths-Jones,S. *et al.* (2009) Rfam: updates to the RNA families database. *Nucleic Acids Res.*, **37**, D136–D140.
12. Griffiths-Jones,S., Moxon,S., Marshall,M., Khanna,A., Eddy,S.R. and Bateman,A. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.*, **33**, D121–D124.
13. Abreu-Goodger,C. and Merino,E. (2005) RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res.*, **33**, W690–W692.
14. Weinberg,Z., Barrick,J.E., Yao,Z., Roth,A., Kim,J.N., Gore,J., Wang,J.X., Lee,E.R., Block,K.F., Sudarsan,N. *et al.* (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.*, **35**, 4809–4819.
15. Regulski,E.E., Moy,R.H., Weinberg,Z., Barrick,J.E., Yao,Z., Ruzzo,W.L. and Breaker,R.R. (2008) A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Mol. Microbiol.*, **68**, 918–932.
16. Sudarsan,N., Lee,E.R., Weinberg,Z., Moy,R.H., Kim,J.N., Link,K.H. and Breaker,R.R. (2008) Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science*, **321**, 411–413.
17. Liu,J.M., Livny,J., Lawrence,M.S., Kimball,M.D., Waldor,M.K. and Camilli,A. (2009) Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res.*, **37**, e46.
18. Passalacqua,K.D., Varadarajan,A., Ondov,B.D., Okou,D.T., Zwick,M.E. and Bergman,N.H. (2009) Structure and complexity of a bacterial transcriptome. *J. Bacteriol.*, **191**, 3203–3211.
19. Perkins,T.T., Kingsley,R.A., Fookes,M.C., Gardner,P.P., James,K.D., Yu,L., Assefa,S.A., He,M., Croucher,N.J., Pickard,D.J. *et al.* (2009) A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *PLoS Genet.*, **5**, e1000569.
20. Yoder-Himes,D.R., Chain,P.S., Zhu,Y., Wurtzel,O., Rubin,E.M., Tiedje,J.M. and Sorek,R. (2009) Mapping the Burkholderia cenocepacia niche response via high-throughput sequencing. *Proc. Natl Acad. Sci. USA*, **106**, 3976–3981.
21. Livny,J., Teonadi,H., Livny,M. and Waldor,M.K. (2008) High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs. *PLoS ONE*, **3**, e3197.
22. Carver,T.J., Rutherford,K.M., Berriman,M., Rajandream,M.A., Barrell,B.G. and Parkhill,J. (2005) ACT: the Artemis Comparison Tool. *Bioinformatics*, **21**, 3422–3423.
23. Keseler,I.M., Bonavides-Martinez,C., Collado-Vides,J., Gama-Castro,S., Gunsalus,R.P., Johnson,D.A., Krummenacker,M., Nolan,L.M., Paley,S., Paulsen,I.T. *et al.* (2009) EcoCyc: a comprehensive view of Escherichia coli biology. *Nucleic Acids Res.*, **37**, D464–D470.
24. Urban,J.H. and Vogel,J. (2007) Translational control and target recognition by *Escherichia coli* small RNAs in vivo. *Nucleic Acids Res.*, **35**, 1018–1037.
25. Livny,J., Fogel,M.A., Davis,B.M. and Waldor,M.K. (2005) sRNAPredict: an integrative computational approach to identify sRNAs in bacterial genomes. *Nucleic Acids Res.*, **33**, 4096–4105.

26. Ahmed,A.M. and Shimamoto,T. (2003) msDNA-St85, a multicopy single-stranded DNA isolated from Salmonella enterica serovar Typhimurium LT2 with the genomic analysis of its retron. *FEMS Microbiol. Lett.*, **224**, 291–297.
27. Shimamoto,T., Kobayashi,M., Tsuchiya,T., Shinoda,S., Kawakami,H., Inouye,S. and Inouye,M. (1999) A retroelement in *Vibrio cholerae*. *Mol. Microbiol.*, **34**, 631–632.
28. Wassarman,K.M. (2007) 6S RNA: a regulator of transcription. *Mol. Microbiol.*, **65**, 1425–1431.
29. Benard,L., Philippe,C., Ehresmann,B., Ehresmann,C. and Portier,C. (1996) Pseudoknot and translational control in the expression of the S15 ribosomal protein. *Biochimie*, **78**, 568–576.
30. Winkler,W., Nahvi,A. and Breaker,R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952–956.
31. Wessler,S.R. and Calvo,J.M. (1981) Control of leu operon expression in *Escherichia coli* by a transcription attenuation mechanism. *J. Mol. Biol.*, **149**, 579–597.
32. Carter,P.W., Bartkus,J.M. and Calvo,J.M. (1986) Transcription attenuation in Salmonella typhimurium: the significance of rare leucine codons in the leu leader. *Proc. Natl Acad. Sci. USA*, **83**, 8127–8131.
33. Gavini,N. and Davidson,B.E. (1991) Regulation of pheA expression by the pheR product in *Escherichia coli* is mediated through attenuation of transcription. *J. Biol. Chem.*, **266**, 7750–7753.
34. Jain,C. and Belasco,J.G. (1995) RNase E autoregulates its synthesis by controlling the degradation rate of its own mRNA in *Escherichia coli*: unusual sensitivity of the rne transcript to RNase E activity. *Genes Dev.*, **9**, 84–96.
35. Schlax,P.J. and Worhunsky,D.J. (2003) Translational repression mechanisms in prokaryotes. *Mol. Microbiol.*, **48**, 1157–1169.
36. Jarrige,A.C., Mathy,N. and Portier,C. (2001) PNPase autocontrols its expression by degrading a double-stranded structure in the pnp mRNA leader. *EMBO J.*, **20**, 6845–6855.
37. Robert-Le Meur,M. and Portier,C. (1992) *E. coli* polynucleotide phosphorylase expression is autoregulated through an RNase III-dependent mechanism. *EMBO J.*, **11**, 2633–2641.
38. Robert-Le Meur,M. and Portier,C. (1994) Polynucleotide phosphorylase of *Escherichia coli* induces the degradation of its RNase III processed messenger by preventing its translation. *Nucleic Acids Res.*, **22**, 397–403.
39. Johnsen,M., Christensen,T., Dennis,P.P. and Fiil,N.P. (1982) Autogenous control: ribosomal protein L10-L12 complex binds to the leader sequence of its mRNA. *EMBO J.*, **1**, 999–1004.
40. Kim,J.N., Roth,A. and Breaker,R.R. (2007) Guanine riboswitch variants from Mesoplasma florum selectively recognize 2′-deoxyguanosine. *Proc. Natl Acad. Sci. USA*, **104**, 16092–16097.
41. Zengel,J.M. and Lindahl,L. (1994) Diverse mechanisms for regulating ribosomal protein synthesis in *Escherichia coli*. *Prog. Nucleic Acid Res. Mol. Biol.*, **47**, 331–370.
42. Allen,T.D., Watkins,T., Lindahl,L. and Zengel,J.M. (2004) Regulation of ribosomal protein synthesis in Vibrio cholerae. *J. Bacteriol.*, **186**, 5933–5937.
43. Waldminghaus,T., Heidrich,N., Brantl,S. and Narberhaus,F. (2007) FourU: a novel type of RNA thermometer in Salmonella. *Mol. Microbiol.*, **65**, 413–424.
44. 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.
45. Robbins,J.C. and Oxender,D.L. (1973) Transport systems for alanine, serine, and glycine in *Escherichia coli* K-12. *J. Bacteriol.*, **116**, 12–18.
46. Romeis,T. and Holtje,J.V. (1994) Penicillin-binding protein 7/8 of Escherichia coli is a DD-endopeptidase. *Eur. J. Biochem.*, **224**, 597–604.
47. Webb,E., Claas,K. and Downs,D.M. (1997) Characterization of thiI, a new gene involved in thiazole biosynthesis in *Salmonella typhimurium*. *J. Bacteriol.*, **179**, 4399–4402.