RESEARCH ARTICLE

Riboswitch regulation in cyanobacteria is independent of their habitat adaptations

Payal Singh¹ · Nilesh Kumar¹ · Minesh Jethva¹ · Saurabh Yadav² ·
Pragati Kumari³ · Archana Thakur⁴ · Hemant Ritturaj Kushwaha⁵

Received: 7 March 2017 / Revised: 13 December 2017 / Accepted: 2 January 2018 / Published online: 8 February 2018
© Prof. H.S. Srivastava Foundation for Science and Society 2018

Abstract Cyanobacteria are one of the ancient bacterial species occupying a variety of habitats with diverse metabolic preferences. RNA regulators like riboswitches play significant role in controlling the gene expression in prokaryotes. The taxonomic distribution of riboswitches suggests that they might be one of the oldest mechanisms of gene control system. In this paper, we analyzed the distribution of different riboswitch families in various cyanobacterial genomes. It was observed that only four riboswitch classes were abundant in cyanobacteria, B₁₂-element (Cob)/AdoCbl/AdoCbl-variant riboswitch being the most abundant. The analysis suggests that riboswitch mode of regulation is present in cyanobacterial species irrespective of their habitat types. A large number of unidentified genes regulated by riboswitches listed in this analysis indicate the wide range of targets for these riboswitch families. The analysis revealed a large number of genes regulated by riboswitches which may assist in elaborating the diversity among the cyanobacterial species.

Keywords Cyanobacteria · Riboswitch · Thiamine · Cobalamin · Regulation

Introduction

Cyanobacteria are oxygenic photosynthetic bacteria and one of the most ancient in the evolutionary lineages (Schopf and Walter 1982; Charpy et al. 2012). They occupy diverse niche and possess a wide range of metabolic capabilities (Ma et al. 2014). Over the years, availability of a large number of bacterial genomes have made it possible to analyze their crucial complex processes like photosynthesis, photo-regulation of genetic expression, nitrogen fixation, and molecular evolution.

RNAs are considered as one of the major components of the bacterial regulatory mechanism (Erdmann et al. 2001). Unique folding of each RNA class is responsible for its biochemical action (Doty et al. 1959; Fresco et al. 1960). Riboswitches are highly structured domains that sense and bind cellular metabolites to regulate various metabolic pathways (Breaker 2011). These are *cis*-encoding RNAs present in the leader region of various mRNAs (Nahvi et al. 2002). The structure of a riboswitch is composed of an aptamer domain which binds to ligand with high affinity, and an expression platform which undergoes structural change on metabolite binding at the aptamer domain to modulate gene expression (Nudler 2006). Often riboswitch classes are named on the basis of cellular metabolite they bind (Regulski et al. 2008; Barrick et al. 2007; Sudarsan et al. 2008). Their widespread taxonomic distribution and ability to regulate highly conserved metabolic pathways indicates that they are one of the oldest forms of gene regulatory mechanism (Vitreschak et al. 2004). Riboswitches play a role in the regulation of several crucial

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12298-018-0504-9>) contains supplementary material, which is available to authorized users.

✉ Hemant Ritturaj Kushwaha
hemantkushwaha@jnu.ac.in

¹ Synthetic Biology and Biofuel, International Center for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India

² Department of Biotechnology, Hemwati Nandan Bahuguna Garhwal University, Srinagar Garhwal, Uttarakhand, India

³ Indian Agriculture Research Institute, New Delhi, India

⁴ University Grants Commission, New Delhi, India

⁵ School of Biotechnology, Jawaharlal Nehru University, New Delhi, India

biochemical pathways including cofactors, amino acids and nucleotide biosynthesis (Vitreschak et al. 2004). Riboswitch regulate the downstream gene mainly by transcription termination or by inhibition of translation initiation. However, a few not so common forms of riboswitch action have also been reported. For example, recently, the B₁₂-riboswitch has been identified downstream and antisense to the adjacent gene, which shows another unique way of regulation exerted through riboswitches (Mellin et al. 2013). Such riboswitches reported in *Clostridium acetobutylicum* (Rodionov et al. 2003) and *Listeria monocytogenes* (Mellin et al. 2013) are known as “Marooned” riboswitches as they appeared to be ‘marooned’ in the genomes with no associated gene to be regulated (Mellin and Cossart 2015).

Engineered riboswitches in their regulatory mode can be a promising tool for various molecular biology studies (Wittmann and Suess 2012). In the current analysis, we have analyzed the distribution of riboswitch families in various cyanobacterial genomes, with cyanobacterial species belonging to diverse niche and some possessing nitrogen-fixing capabilities. Also, we have identified probable ‘marooned’ riboswitches and their associated genes. The analysis presented here would also be helpful in underlining those genes which are regulated by riboswitch mechanism and their probable role in cyanobacteria.

Materials and methods

Phylogenetic tree of cyanobacterial species

The 16S ribosomal sequences of 317 cyanobacterial species (Supplementary Table 1) analyzed were aligned using MUSCLE (Edgar 2004), multiple alignment software with default parameters. The alignment thus obtained was used to generate maximum likelihood (ML) tree with 100 bootstrap value using PhyML (version 3.0) (Guindon et al. 2010). The tree was analyzed and visualized using Mega (version 6) (Tamura et al. 2013).

Analysis of riboswitches present in cyanobacteria

A total of 317 completely sequenced and annotated cyanobacterial genomes were obtained from (Joint Genome Institute) JGI database (<http://genome.jgi.doe.gov/>) (Supplementary Table 1). These organisms were classified on the basis of their habitat and capabilities to fix nitrogen. The covariance models of twenty-four riboswitch families were obtained from Rfam database (version 12.0) (<http://rfam.xfam.org/>) (Burge et al. 2013). The covariance models are the probabilistic models of conserved sequence and secondary structure of RNA family. These covariance

models were used to scan the cyanobacterial genomes using Infernal (version 1.1) (Nawrocki and Eddy 2013) software with a score cut-off obtained from Rfam database (Supplementary Table 3) and inclusion thresholds of E-value 0.01 or less. Finally, in-house perl scripts were used for parsing and analysis of the results obtained from the Infernal.

Identification of ‘Marooned’ riboswitches

‘Marooned’ riboswitches are found downstream and antisense of the adjacent genes (Mellin et al. 2013). In order to find the ‘marooned’ riboswitches in 317 cyanobacterial species, the genes at a distance < 1500 bp from the riboswitch end (downstream and antisense of the adjacent genes), in the sense as well as the antisense strands were considered in the category. Further, the organisms possessing such type of riboswitches were classified on the basis of their nitrogen-fixing capabilities.

Results

Distribution of riboswitches in cyanobacteria

The genomes of 317 cyanobacterial species were obtained from JGI database (<http://img.jgi.doe.gov/>). These genomes were classified according to their habitat and nitrogen fixing capabilities (Supplementary Table 1) (Fig. 1). The riboswitch occurrences were mapped onto these genomes. Of 317 genomes under study, 302 species were found to be under riboswitch mode of regulation. Twenty four riboswitch families were searched in all genomes, out of which twenty families were found in cyanobacteria. Four of the riboswitch families which were found missing were SAM-V (RF01826), SAM-SAH (RF01727), MFR (RF01510) and SAM-IV (RF00634). The analysis of riboswitch occurrence in different cyanobacterial genomes revealed that *Prochlorococcus* sp. scB243_498P3 has maximum number of genes under riboswitch regulation (Zhang and Gladyshev 2008). A few other species which employ riboswitch mode of regulation are *Oscillatoriales cyanobacterium* JSC-12 (22 genes), *Mastigocoleus testarum* BC008 (20 genes), *Calothrix* PCC 7103 (16 genes), *Oscillatoria* PCC 6407 (15 genes), *Oscillatoria* PCC 6506 (15 genes) and *Hassallia byssoidea* VB512170 (13 genes). Of all the species under riboswitch regulation, 32 were nitrogen fixing with *Oscillatoria* PCC 6407 having the highest number of genes regulated by riboswitches (15 genes). Other nitrogen fixing species showing riboswitch regulation were *Cyanothece* sp. PCC 7425 (12 genes), *Calothrix* PCC 6303 (10 genes), *Microcoleus vaginatus* FGP-2 (10 genes) and *Nostoc* sp. PCC 7120 (9 genes).

Fig. 1 Distribution of various Riboswitch families in various cyanobacterial species. The 16S rRNA sequences were used to plot the phylogenetic tree highlighting their evolutionary relationships. The riboswitch families were shown in various rings marked from

We found that out of twenty riboswitch families found in cyanobacteria, only four riboswitch families namely cobalamin (Cob): RF00174, adenosylcobalamin (AdoCbl): RF01482, adenosylcobalamin (AdoCbl)-variant: RF01689 and thiamine pyrophosphate (TPP): RF00059 (Fig. 2) were in abundance while others have only a few instances. The Cob riboswitches were in the upstream of 462 genes across various cyanobacterial genomes (including nitrogen-fixing species). AdoCbl riboswitches were observed in the upstream of 338 genes while TPP and AdoCbl-variant (RF01689) riboswitch regulated 308 and 173 genes respectively. Among the nitrogen-fixing cyanobacteria species, the Cob riboswitches were upstream of 69 genes followed by AdoCbl-variant riboswitches (57 genes). Other riboswitch families such as Moco, cyclic di-GMP-I and II, SAM-I, SAM-II, SAM-I-IV variant, PreQ1-I, PreQ1-II, Glycine, SMK, Purine, FMN, glmS, lysine, SAH and THF

inside towards the outside. The numbers of genes upstream of which various riboswitches are present are shown in various colors. The highlighted tree branches are specific for cyanobacterial species which possess nitrogen fixing capabilities

were identified only in a few cyanobacterial genomes (Fig. 2).

B₁₂-element (Cob)/AdoCbl/AdoCbl-variant riboswitch

Cobalamin riboswitch is known to regulate genes involved in the vitamin B₁₂ pathway. This riboswitch class is distributed widely across prokaryotes. In cyanobacteria, cobalamin riboswitches are the most abundant with a total of 463 occurrences spanning across 218 cyanobacterial genomes, out of which 31 species are nitrogen fixing. The maximum number of genes under riboswitch regulation are observed in *Mastigocoleus testarum* BC008 followed by *Prochlorococcus* sp. scB243_498P3 (Supplementary Table 2) (Fig. 1). The cobalamin riboswitch in cyanobacteria regulate a variety of genes and/or operons belonging

Fig. 2 Analysis of number of the genes possessing various riboswitch elements showing large presence of Cobalamin (RF00174), AdoCbl (RF01482), TPP (RF00059) and AdoCbl-var (RF01689) riboswitch in the upstream of various genes in various cyanobacterial species

to cobalamin biosynthesis pathway (*cobG* and *cobW*) and transport (*cbiM*). A few genes annotated as metal binding proteins and hydrogenase/urease accessory protein encoding genes (*hupE*) are also found to be regulated by riboswitches. Earlier analysis showed that *hupE* gene is part of B₁₂ regulon in *Prochlorococcus marinus*, *Synechococcus* sp. and *Synechocystis* sp. (Rodionov et al. 2003). Some genes with cobalamin riboswitch are involved in various tRNA-charging pathways, peptidoglycan, and chorismate biosynthesis. The analysis of habitat of cyanobacteria possessing cobalamin riboswitches showed that they belonged to aquatic fresh water and marine habitats. A large number of genes under cobalamin riboswitch regulation in cyanobacteria were annotated as hypothetical (128 genes).

Another family of riboswitch found in abundance in cyanobacteria species genomes was Adenosylcobalamin (AdoCbl) riboswitch (RF01482) with a total of 339 occurrences (Supplementary Table 2) (Fig. 1). Also known as coenzyme B₁₂, AdoCbl is a cofactor required for catalysis of a number of isomerization reactions (Roth et al. 1996). A total of 190 cyanobacterial genomes contained AdoCbl riboswitches that were located upstream of a variety of genes. Out of 102 cyanobacterial species, 29

were involved in nitrogen fixation. The maximum number of occurrences (5) of AdoCbl riboswitch were observed in *Oscillatoria* (PCC 6506, PCC 6407 and PCC 7112) and *Calothrix* (PCC 7103) followed by four in *Pleurocapsa* (PCC 7327), *Arthrospira* sp. PCC 8005, *Arthrospira platensis* (C1 and Paraca), *Calothrix* (PCC 7507), *Chamaesiphon* (PCC 6605), *Leptolyngbya boryana* (PCC 6306), *Mastigocoleus testarum* (BC008), *Microcoleus vaginatus* (FGP-2 and PCC 9802) and *Pseudanabaena* (PCC 6802 and PCC 7429). Similar to B₁₂-element, AdoCbl riboswitch were found mostly in the upstream region of cobalamin biosynthesis genes (*cobG* and *cobW*), cobalt transporter *CbiM* gene, and hydrogenase/urease accessory protein encoding gene (*hupE*).

The AdoCbl-variant riboswitch (RF01689) are observed in 131 cyanobacterial species of which 17 species were nitrogen fixing (Supplementary Table 2) (Fig. 1). The maximum number of AdoCbl-variant riboswitch controlled genes are present in *Calothrix* PCC 7103 (5 genes) followed by *Oscillatoria* PCC 6407, *Oscillatoria* PCC 6506, *Calothrix desertica* PCC 7102 and *Scytonema tolypothrichoides* VB-61278 (each with three riboswitch controlled genes). Similar to cobalamin and AdoCbl riboswitch, AdoCbl-variant riboswitch are mainly present upstream of

cobalamin biosynthesis genes (*cobG* and *cobW*), cobalt transporter *CbiM* gene and hydrogenase accessory protein encoding gene (*hupE*).

THI-element (TPP-riboswitch)

The TPP riboswitch regulate various biosynthetic enzymes and predicted transporters involved in the salvage and synthesis of thiamine and its metabolic precursors, hydroxymethylpyrimidine and hydroxyethylthiazole (Rodionov et al. 2002). The TPP riboswitches (RF00059) are distributed across 290 cyanobacterial genomes, out of which 31 are involved in nitrogen fixation, regulating a total of 308 genes (Supplementary Table 2) (Fig. 1). The maximum number of TPP regulated genes are present in *Prochlorococcus* sp. scB243_498P3 (5 genes) followed by *Hassallia bysoidea* VB512170, *Mastigocoleus testarum* BC008, *Prochlorococcus marinus* bv. (HNLC1) and *Scytonema millei* VB511283 (3 genes each). TPP riboswitches regulate mainly *thiC* gene in cyanobacteria. Previously, TPP riboswitch has been reported to be present upstream of *thiC*, which encodes an essential thiamine biosynthesis enzyme (Rodionov et al. 2002). Many hypothetical genes are also riboswitch regulated for this family. The cyanobacteria species which showed the presence of THI-element in the upstream of various genes were found mainly in aquatic marine and aquatic fresh water habitat.

Moco (molybdenum cofactor) riboswitch

The regulation and uptake of molybdenum and tungsten cofactors, which essentially participate in the oxygen atom transfer in nitrogen, sulfur, and carbon metabolism is performed by Moco (molybdenum cofactor) riboswitch (Zhang and Gladyshev 2008). This riboswitch family is shown to have restricted phylogenetic distribution, present mostly in α -proteobacteria, *Clostridiaceae*, and *Thermotogaceae* (Sun et al. 2013). Surprisingly, this limited riboswitch family is observed in 11 cyanobacterial species, none of which is nitrogen fixing (Supplementary Table 2) (Fig. 1). The species which possess Moco riboswitch belong to aquatic fresh water habitat. The highest number of genes under Moco riboswitches regulation are observed in *Oscillatoriales* JSC-12 (15 genes) followed by *Dactylococcopsis salina* PCC 8305 and *Spirulina major* PCC 6313 in which three genes are regulated by the riboswitch. Few hypothetical genes are also riboswitch regulated in this class.

Cyclic di-GMP-I and II riboswitch (GEMM RNA motifs)

In bacteria, the genes controlled by cyclic di-GMP responsive riboswitches are involved in motility, virulence,

and biofilm formation (Reader et al. 2004). The cyclic di-GMP-I riboswitch (RF01051) is observed in ten cyanobacterial genomes with a maximum of seven genes under riboswitch regulation in *Cyanothece* sp. PCC 7425, a nitrogen fixing bacteria. Cyclic di-GMP-II (RF01786) riboswitch are present in twelve genomes regulating maximum of two genes in *Leptolyngbya* PCC 6406, *Leptolyngbya* PCC 7376, *Prochlorococcus* sp. scB243_498P3 and *Cyanobium* sp. (PCC 7001). Among the species under cyclic di-GMP-II riboswitch regulation, only one cyanobacterial species namely *Synechococcus* sp. PCC 7335 is a nitrogen fixing species (Supplementary Table 2) (Fig. 1). For both cyclic di-GMP-I and II riboswitch, majority of riboswitch regulated genes belong to hypothetical protein encoding genes.

S-box (SAM), SAM-II (α -proteobacteria) and SAM-I/IV-variant riboswitch

The genes regulated by SAM (S-adenosylmethionine) riboswitch are largely involved in the cysteine and methionine metabolic pathways which affect the intracellular balance of SAM cofactor (Sun et al. 2013). In cyanobacteria, the distribution of this riboswitch family is sparse with only twelve cyanobacterial species possessing SAM riboswitch regulated genes (Supplementary Table 1) (Fig. 1), the largest number being present in *Prochlorococcus* sp. scB243_498P3. The habitat of the species possessing SAM-riboswitch regulated genes is not known. The SAM-II riboswitch (RF00521) which is usually found in α -proteobacteria, is present in only five cyanobacterial species namely *Fischerella* PCC 9605, *Mastigocoleus testarum* BC008, *Obscuribacter phosphatis* Mle1 12, *ThermoSynechococcus* sp. NK55a and *Scytonema millei* VB511283 (Supplementary Table 2)(Fig. 1). The SAM-I/IV-variant riboswitch (RF01725) is observed in only two cyanobacterial species namely *Gastranaerophilaceae* (Zag_1) and *Gastranaerophilaceae* (Zag_111) which inhabit animal digesta (Supplementary Table 2) (Fig. 1). One of the genes under SAM-I/IV-variant riboswitch regulation belongs to the CBS domain family protein predicted to have a role in fermentation.

PreQ1-I and PreQ1-II (pre queuosine) riboswitch

The biosynthesis and uptake of queuosine nucleoside is regulated by PreQ1-I riboswitches (Reader et al. 2004; Eitinger et al. 2011). The PreQ1-I riboswitches (RF00522) in cyanobacteria are thinly dispersed with only fourteen genomes showing their presence, of which only two species namely *Trichodesmium erythraeum* IMS 101 and *Anabaena* sp. 90 are nitrogen-fixing (Supplementary Table 2) (Fig. 1). The largest number of riboswitch regulated genes

are present in *Microcystis aeruginosa* NIES 843 and *Trichodesmium erythraeum* IMS 101. Only single gene is under PreQ1-II (pre-queuosine) riboswitch (RF01054) regulation in *Leptolyngbya* sp. (Heron Island J). The PreQ1-II riboswitch is present upstream of soluble [2Fe-2S] ferredoxin gene in *Leptolyngbya* sp. (Heron Island J).

Glycine riboswitch

The glycine riboswitches are known to carry two aptamers which bind to the glycine cooperatively (Tucker and Breaker 2005). The glycine riboswitch (RF00504) is observed in seven cyanobacterial genomes namely *Leptolyngbya* PCC 7375, *Mastigocoleus testarum* BC008, *Halotheca* PCC 7418, *Myxosarcina* sp. G11, *Prochlorococcus* sp. scB243_498P3, *Scytonema millei* VB511283 and *Tolypothrix campylonemoides* VB511288 (Supplementary Table 2) (Fig. 1), of which *Leptolyngbya* PCC 7375 is a nitrogen-fixing bacteria. Aspartate racemase, aspartyl/asparaginyl beta-hydroxylase encoding genes are glycine riboswitch regulated in cyanobacteria.

SMK box translational riboswitch

The SMK box translational riboswitches (RF01767) show sporadic presence, being observed in only eight cyanobacterial species namely *Cyanothece* sp. ATCC 51472, *Cyanothece* sp. PCC 7822 (both nitrogen fixing bacteria), *Prochlorococcus marinus* sub. *pastoris* CCMP 1986, *Gastranaerophilaceae* (MH 37 and Zag_1) and *Prochlorococcus* sp. (EQPAC1, MIT9201 and W7) (Supplementary Table 2) (Fig. 1). Most of the cyanobacterial species possessing SMK box translational riboswitch belong to aquatic marine habitat.

Purine riboswitch

The purine riboswitches (RF00167) are also observed in limited cyanobacterial genomes with only three species namely *Prochlorococcus* sp. (scB243_498P3, scB241_526B17 and scB241_527L15) showing their presence (Fig. 1). It controls the genes that are relevant for purine biosynthesis and salvage. The genes with upstream purine riboswitch are present in the pathway of purine de novo biosynthesis.

FMN riboswitch (RFN element)

The FMN riboswitches (RF00050) were observed in only five cyanobacterial genomes namely *Prochlorococcus* sp. (scB243_498P3 and PAC1), *Planktothrix* sp. (st147 and 585) and *Planktothrix agardhii* (NIVA-CYA 126/8) (Supplementary Table 2) (Fig. 1). The gene preceding which

FMN riboswitch are observed mainly belong to PIN domain superfamily, tryptophanyl-tRNA synthetase and tRNA 2-selenouridine synthase.

glmS (glucosamine-6-phosphate activated) ribozyme

The glmS (glucosamine-6-phosphate activated) ribozyme (RF00234) which acts as a ribozyme (upon binding to its native ligand GlcN6P) as well as riboswitch was found in six cyanobacterial genomes namely *Calothrix* (PCC 6303), *Prochlorococcus* sp. (scB243_498P3, MIT 9313 and MIT0604) and *Oscillatoria* (PCC 6407 and PCC 6506), with each of them consisting of only a single gene under the riboswitch control (Supplementary Table 2) (Fig. 1). This riboswitch regulates the expression of *glmS* genes which is responsible for the production of glucosamine-6-phosphate (GlcN6P) (Weinberg et al. 2007). Nitrogen fixing species with glmS riboswitch regulation are *Calothrix* (PCC 6303) and *Oscillatoria* (PCC 6407). The cyanobacterial species having glmS ribozyme belong to aquatic marine and fresh water habitat. Genes under glmS riboswitch regulation code for amidohydrolase, thioredoxin and cytochrome c biogenesis protein (CcdA).

Lysine riboswitch

Lysine riboswitch regulate genes involved in the biosynthesis and transport of amino acid lysine. Cyanobacterial species namely *Prochlorococcus* sp. (scB243_498P3) possess lysine riboswitch (RF00168) preceding its gene (Supplementary Table 2) (Fig. 1). The genes upstream of which lysine riboswitches are present encoded for D-hydantoinase, Holliday junction DNA helicase subunit (RuvA) and an amino acid carrier protein.

SAH (S-adenosyl-L-homocysteine) riboswitch

The SAH (S-adenosyl-L-homocysteine) riboswitch regulates the genes responsible for methionine metabolism (Nudler 2006). The SAH riboswitches (RF01057) are present only in *Lyngbya confervoides* BDU141951 and *Tolypothrix campylonemoides* VB511288 (Supplementary Table 2) (Fig. 1). Most proteins observed under SAH riboswitch regulation are hypothetical proteins.

THF (Tetrahydrofolate) riboswitch

The THF (Tetrahydrofolate) riboswitch is known for binding broad spectrum of ligands (Ames et al. 2010; Trausch et al. 2011). Only a single gene in *Prochlorococcus* sp. (scB243_498P3) is observed to have THF riboswitch (RF01831) in its upstream (Supplementary Table 2) (Fig. 1).

Analysis of sense/antisense and ‘marooned’ riboswitches

The analysis of riboswitches orientation (sense or antisense) in various cyanobacterial species showed half of the genes in sense and half in antisense strand in all the major riboswitch families (Supplementary Table 4). Three riboswitch families namely SAM-var, SAH, and PreQ1-II are present on the antisense strand in many cyanobacterial species.

The identification of ‘marooned’ riboswitches indicates their common occurrence in various cyanobacterial species (Fig. 3, Supplementary Table 5). In various cyanobacterial species, a total of 270 cobalamine riboswitches are ‘marooned’ riboswitches followed by 197 AdoC1 and 191 TPP riboswitches (Fig. 3). Further, 109 AdoCbl-var riboswitches are found in the same category. Among nitrogen-fixing cyanobacterial species, maximum ‘marooned’ riboswitches are present in *Cyanothece* sp. (PCC 7425) followed by eight in *Microcoleus vaginatus* (FGP-2). For the cobalamin riboswitch, the maximum ‘marooned’ riboswitch instances are found in *Mastigocoleus testarum* (BC008) (total 8). In case of AdoCbl riboswitches, maximum ‘marooned’ riboswitches are in *Oscillatoria* sp. (PCC 6506). For the TPP riboswitches, the maximum number of ‘marooned’ riboswitches (3) were found in *Hassallia bysoidea* (VB512170), *Mastigocoleus testarum* (BC008), *Prochlorococcus marinus* bv. (HNLC1), *Prochlorococcus* sp. (scB243_498P3) and *Scytonema millei* (VB511283).

Discussion

Cyanobacteria belong to the group of ancient bacteria which are most ubiquitous throughout the world inhabiting a wide range of habitats. The hypothesis of the RNA-centric world suggests the existence of RNA as major gene regulators in ancient bacteria, providing all the necessary catalytic and genomic functions of the earliest organisms. One of the gene regulators is RNA riboswitch, a genetic switch which acts as a receptor for specific metabolites without any protein aid, thus controlling the gene expression via allosteric structural changes (Charpy et al. 2012). Due to the conserved features, some of the riboswitches are considered as the reminiscent of the ancient sensory and regulatory system (Vitreschak et al. 2004; Nahvi et al. 2004; Breaker 2006). These regulatory mechanisms have been well known in bacterial genomes. In order to analyze the diversity and distribution of riboswitches in cyanobacteria with respect to their habitats and nitrogen-fixing capabilities, genomes of 317 species were scanned for the presence of twenty-four known riboswitch families. Interestingly, despite cyanobacteria being considered as one of the early life forms is not using the riboswitch mode of regulation abundantly. Only 302 of the total 317 cyanobacterial genome possess one or the other type of riboswitch family in their respective genomes. As bacterial species grow and evolve rapidly, the low presence of the variety of riboswitch in cyanobacterial genomes show that riboswitches are not primitive but are highly refined sensors and switches that are counters to the protein factors (Breaker 2011). The absence of riboswitches in 15

Fig. 3 ‘Marooned’ riboswitches in various cyanobacterial species. The distance (< 1500 bp) between the 3′ end of the riboswitch and adjacent ORF was plotted. The boxes denote all the riboswitch classes observed in various cyanobacterial species

cyanobacteria species shows that some species may have regulatory proteins that fulfill the same role or there might be other extreme structural variants of riboswitch classes which are not known yet.

In cyanobacteria, only four riboswitch families namely Cob (RF00174), AdoCbl (RF01482), TPP (RF00059) and AdoCbl-variant (RF01689) are observed to be abundant. The cobalamin specific riboswitches (or Cob riboswitch) are the most common throughout diverse bacterial species (Winkler and Breaker 2005). Earlier, the phylogenetic analysis suggested that cobalamin riboswitch is widely distributed amongst both Gram-positive and Gram-negative prokaryotes (Winkler and Breaker 2005; Rodionov et al. 2003; Vitreschak et al. 2003). In several bacterial genomes, duplications of B₁₂-elements have led to the expansion of the vitamin B₁₂ regulon (Vitreschak et al. 2003). The abundance of B₁₂-elements and THI-element in cyanobacteria shows that ligand-binding properties of riboswitch aptamer domains have been maintained throughout expansive evolutionary timescales. Some of the genes containing the riboswitch binding sites for cobalamin in various cyanobacterial species are the cobalt transporter gene *hupE*. Various *hupE* homologs are the membrane proteins predicted to contain at least six transmembrane helices (Hoffmann et al. 2006). The *hupE* genes in *Rhizobium leguminosarum* encode nickel transporter (Ruiz-Argüeso et al. 2000; Brito et al. 2010), a mutant of which was found to be incapable of growing in the presence of metal chelator, nitrilotriacetate (Hoffmann et al. 2006). Our results also predicted the presence of vitamin B₁₂ dependent riboswitch in the upstream of the *hupE* gene which corroborated with the earlier evidence. The TPP riboswitch which is the most common amongst the major bacterial taxonomic group, is also present in abundance in various cyanobacterial species genome. Thiamine (also known as Vitamin B1), is an essential co-factor in all the living systems. De novo synthesis of thiamine, regulated via RNA sequences known as riboswitches is well established in prokaryotes, algae, plants, and certain fungi. They have been shown to control splicing in plants and fungi (Winkler and Breaker 2005; Sudarsan et al. 2003). The direct binding of metabolites, thiamine or its pyrophosphate derivative to the pre-mRNA in the absence of any cofactors, leads to the change in RNA secondary structure that interferes with gene expression (Sudarsan et al. 2003; Winkler et al. 2002). Other widely present riboswitch families in cyanobacteria genomes, AdoCbl, and AdoCbl-variant are known to inhibit cobalamin biosynthesis by promoting transcription termination (Fox et al. 2009). In *E. coli* and other related species of bacteria, AdoCbl does not contribute to the growth under standard laboratory conditions but they contribute to the metabolism of compounds that are

important sources of carbon, nitrogen, and energy in specific environments (Fowler et al. 2010).

Out of total 317 cyanobacterial species considered for analysis, the number of *Prochlorococcus* species genomes was the highest (total 97 species). This is due to the fact that among the known oxygenic phototrophs, genus *Prochlorococcus* is most abundant in the marine conditions (Partensky et al. 1999; Kettler et al. 2007). The genus is known to have smallest genomes (~ 1.6–2.7 Mbp) and contributes significantly to the oceanic productivity. Earlier analyses have suggested extensive streamlining of the *Prochlorococcus* genomes during the evolution. Our analysis shows that large number of genes in *Prochlorococcus* sp. (scB243_498P3) were under the riboswitch regulation. Other *Prochlorococcus* genomes had comparatively less number of genes under riboswitch regulation which shows probable signs of streamlining of the genomes during the evolution. The TPP riboswitch was found to be most abundant in various *Prochlorococcus* genomes.

Some of the riboswitch classes are extremely rare, even though their metabolite sensing and gene control system are very efficient. This might be due to the competitive affinity of near-identical riboswitch classes (Breaker 2011). Among these competing class of riboswitches, cyclic-di-GMP, SAM, and PreQ1 were observed in the upstream of various genes but their presence was comparatively very low in various cyanobacterial species. Like other bacterial candidates, various riboswitches are present upstream of many genes encoding proteins of unknown function. This phenomenon was observed in almost all the families of the riboswitches found in various cyanobacteria. The earlier analysis suggests that riboswitches can also regulate RNAs other than mRNAs which was evident from the various studies in which the riboswitches located at the 3' end of one gene and positioned antisense to the 5' adjacent open reading frame (ORF) (Mellin et al. 2013; Rodionov et al. 2003, 2004). Due to their distant presence, these riboswitches are unable to regulate the expression of the adjacent genes in a classical manner, that is, by controlling transcription or translation of the downstream ORF(s). These 'marooned' riboswitches are hypothesized to be regulating the expression of antisense RNAs (asRNAs) (Mellin et al. 2013). Evidence of such riboswitches is found in *Clostridium acetobutylicum*, with SAM riboswitch identified antisense to the three gene *ubiG mcbB mcbA* operon (Rodionov et al. 2004). Earlier identified in *Clostridium acetobutylicum*, 'marooned' riboswitch are hypothesized to function as independent sRNAs that regulates the expression of a gene in *trans* by base-pairing with mRNAs or titrating RNA-binding proteins (De Lay and Garsin 2016). The analysis of such riboswitches specifically of Cob, AdoCbl, TPP, and AdoCbl-var family showed their large presence in the

cyanobacteria species irrespective of their nitrogen-fixing capabilities. Thus, in light of the previous hypothesis these ‘marooned’ riboswitches are present in cyanobacteria species, which further needs to be experimentally validated.

Conclusion

Riboswitch-based sensing is considered as an important platform for screening of a large number of environmental or genetic conditions. Due to the ability to sense and modulate the gene expression in response to the ever-changing environmental conditions, riboswitches are considered as smart and efficient regulators. In the light of RNA world hypothesis, 317 finished cyanobacterial genomes inhabiting varied habitats and nitrogen-fixing capabilities are analyzed for the presence of the riboswitch mechanism. The information about the genes present and their regulation is very meagre till date in many of the considered cyanobacterial genomes. The analysis showed that the riboswitch dependent control of metabolic pathways is not ubiquitous in cyanobacterial species except for pathway genes involving B₁₂ and thiamine, which shows the conserved nature of the two pathways in cyanobacterial species. RNA motifs regulons have been hypothesized to evolve and adapt to the changes in habitats and lifestyles in microbial lineages but the current analysis has shown that in cyanobacteria, riboswitch regulation is not entirely dependent on the habitat and their metabolic preferences. The current analysis has revealed a large number of hypothetical genes, which showed presence of various riboswitches to its upstream. The regulatory information can be used for the annotation of these genes as well as enable more detailed studies on the co-evolution of the regulatory RNAs, their downstream genes, and their cognate metabolic pathways. Further, the identification of riboswitch-regulated non-coding RNAs in various cyanobacterial genomes adds to the list which can be analyzed for their functional involvement. The number and type of riboswitch reported in the analyses in various cyanobacterial species will assist in developing single gene deletion or over expression libraries which can reveal various genetic factors that affect metabolite levels, thus identifying genes directly involved in the related pathways. The current analysis can be used for various synthetic biology applications leading to the production of industrially important products.

Funding This work is supported by grants received from Department of Science and Technology, Government of India as INSPIRE Faculty award.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

References

- Ames TD, Rodionov DA, Weinberg Z, Breaker RR (2010) A eubacterial riboswitch class that senses the coenzyme tetrahydrofolate. *Chem Biol* 17(7):681–685. <https://doi.org/10.1016/j.chembiol.2010.05.020>
- Barrick JE, Breaker RR (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol* 8(11):1. <https://doi.org/10.1186/gb-2007-8-11-r239>
- Breaker RR (2006) Riboswitches and the RNA world. *Cold Spring Harb Monogr Arch* 43:89–107. <https://doi.org/10.1101/cshperspect.a003566>
- Breaker RR (2011) Prospects for riboswitch discovery and analysis. *Mol Cell* 43(6):867–879. <https://doi.org/10.1016/j.molcel.2011.08.024>
- Brito B, Prieto RI, Cabrera E, Mandrand-Berthelot MA, Imperial J, Ruiz-Argüeso T, Palacios JM (2010) *Rhizobium leguminosarum hupE* encodes a nickel transporter required for hydrogenase activity. *J Bacteriol* 192(4):925–935. <https://doi.org/10.1128/JB.01045-09>
- Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP et al (2013) Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res* 41:D226–D232. <https://doi.org/10.1093/nar/gks1005>
- Charpy L, Casareto BE, Langlade MJ, Suzuki Y (2012) Cyanobacteria in coral reef ecosystems: a review. *J Mar Biol*. <https://doi.org/10.1155/2012/259571>
- De Lay NR, Garsin DA (2016) The unmasking of ‘junk’ RNA reveals novel sRNAs: from processed RNA fragments to marooned riboswitches. *Curr Opin Microbiol* 30:16–21. <https://doi.org/10.1016/j.str.2011.06.019>
- Doty P, Boedtker H, Fresco JR, Haselkorn R, Litt M (1959) Secondary structure in ribonucleic acids. *Proc Natl Acad Sci USA* 45(4):482–499
- Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform* 5(1):1. <https://doi.org/10.1186/1471-2105-5-113>
- Eitinger T, Rodionov DA, Grote M, Schneider E (2011) Canonical and ECF-type ATP-binding cassette importers in prokaryotes: diversity in modular organization and cellular functions. *FEMS Microbiol Rev* 35(1):3–67. <https://doi.org/10.1111/j.1574-6976.2010.00230.x>
- Erdmann VA, Barciszewska MZ, Szymanski M, Hochberg A, de Groot N, Barciszewski J (2001) The non-coding RNAs as riboregulators. *Nucleic Acids Res* 29(1):189–193. <https://doi.org/10.1093/nar/29.1.189>
- Fowler CC, Brown ED, Li Y (2010) Using a riboswitch sensor to examine coenzyme B(12) metabolism and transport in *E. coli*. *Chem Biol* 17(7):756–765. <https://doi.org/10.1016/j.chembiol.2010.05.025>
- Fox KA, Ramesh A, Stearns JE, Bourgoigne A, Reyes-Jara A, Winkler WC, Garsin DA (2009) Multiple posttranscriptional regulatory mechanisms partner to control ethanolamine utilization in *Enterococcus faecalis*. *Proc Natl Acad Sci USA* 106(11):4435–4440. <https://doi.org/10.1073/pnas.0812194106>
- Fresco JR, Alberts BM, Doty P (1960) Some molecular details of the secondary structure of ribonucleic acid. *Nature* 188:98–101. <https://doi.org/10.1038/188098a0>

- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321. <https://doi.org/10.1093/sysbio/syq010>
- Hoffmann D, Gutekunst K, Klissenbauer M, Schulz-Friedrich R, Appel J (2006) Mutagenesis of hydrogenase accessory genes of *Synechocystis* sp. PCC 6803. *FEBS J* 273(19):4516–4527. <https://doi.org/10.1111/j.1742-4658.2006.05460.x>
- Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML et al (2007) Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* 3(12):e231
- Ma AT, Schmidt CM, Golden JW (2014) Regulation of gene expression in diverse cyanobacterial species by using theophylline-responsive riboswitches. *Appl Environ Microbiol* 80(21):6704–6713. <https://doi.org/10.1128/AEM.01697-14>
- Mellin JR, Cossart P (2015) Unexpected versatility in bacterial riboswitches. *Trends Genet* 31(3):150–156. <https://doi.org/10.1016/j.tig.2015.01.005>
- Mellin JR, Tiensuu T, Bécavin C, Gouin E, Johansson J, Cossart P (2013) A riboswitch-regulated antisense RNA in *Listeria monocytogenes*. *Proc Natl Acad Sci USA* 110(32):13132–13137. <https://doi.org/10.1073/pnas.1304795110>
- Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR (2002) Genetic control by a metabolite binding mRNA. *Chem Biol* 9(9):1043–1049. [https://doi.org/10.1016/S1074-5521\(02\)00224-7](https://doi.org/10.1016/S1074-5521(02)00224-7)
- Nahvi A, Barrick JE, Breaker RR (2004) Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res* 32(1):143–150. <https://doi.org/10.1093/nar/gkh167>
- Nawrocki EP, Eddy SR (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29(22):2933–2935. <https://doi.org/10.1093/bioinformatics/btt509>
- Nudler E (2006) Flipping riboswitches. *Cell* 126(1):19–22. <https://doi.org/10.1016/j.cell.2006.06.024>
- Partensky F, Hess WR, Vaulot D (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol* 63:106–127
- Reader JS, Metzgar D, Schimmel P, de Crécy-Lagard V (2004) Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine. *J Biol Chem* 279(8):6280–6285. <https://doi.org/10.1074/jbc.M310858200>
- Regulski EE, Moy RH, Weinberg Z, Barrick JE, Yao Z, Ruzzo WL, Breaker RR (2008) A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Mol Microbiol* 68(4):918–932. <https://doi.org/10.1111/j.1365-2958.2008.06208.x>
- Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS (2002) Comparative genomics of thiamin biosynthesis in prokaryotes. NEW genes and regulatory mechanisms. *J Biol Chem* 277(50):48949–48959. <https://doi.org/10.1074/jbc.M208965200>
- Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS (2003) Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. *J Biol Chem* 278(42):41148–41159. <https://doi.org/10.1074/jbc.M305837200>
- Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS (2004) Comparative genomics of the methionine metabolism in gram-positive bacteria: a variety of regulatory systems. *Nucleic Acids Res* 32(11):3340–3353. <https://doi.org/10.1016/j.chembiol.2010.05.025>
- Roth JR, Lawrence JG, Bobik TA (1996) Cobalamin (coenzyme B12): synthesis and biological significance. *Annu Rev Microbiol* 50(1):137–181. <https://doi.org/10.1146/annurev.micro.50.1.137>
- Ruiz-Argüeso T, Imperial J, Palacios JM (2000) Uptake hydrogenases in root nodule bacteria. In: Triplett EW (ed) *Prokaryotic nitrogen fixation: a model system for the analysis of a biological process*. Horizon Scientific Press, Norfolk, pp 489–507
- Schopf JW, Walter MR (1982) Origin and early evolution of cyanobacteria: the geological evidence, chapter 21. University of California Press, p 543
- Sudarsan N, Barrick JE, Breaker RR (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* 9(6):644–647. <https://doi.org/10.1261/rna.5090103>
- Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH et al (2008) Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* 321(5887):411–413. <https://doi.org/10.1126/science.1159519>
- Sun EI, Leyn SA, Kazanov MD, Saier MH, Novichkov PS, Rodionov DA (2013) Comparative genomics of metabolic capacities of regulons controlled by cis-regulatory RNA motifs in bacteria. *BMC Genom* 14(1):1. <https://doi.org/10.1186/1471-2164-14-597>
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30(12):2725–2729. <https://doi.org/10.1186/1471-2105-5-113>
- Trausch JJ, Ceres P, Reyes FE, Batey RT (2011) The structure of a tetrahydrofolate-sensing riboswitch reveals two ligand binding sites in a single aptamer. *Structure* 19(10):1413–1423. <https://doi.org/10.1016/j.str.2011.06.019>
- Tucker BJ, Breaker RR (2005) Riboswitches as versatile gene control elements. *Curr Opin Struct Biol* 15(3):342–348. <https://doi.org/10.1016/j.sbi.2005.05.003>
- Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS (2003) Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. *RNA* 9(9):1084–1097. <https://doi.org/10.1261/rna.5710303>
- Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS (2004) Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet* 20(1):44–50. <https://doi.org/10.1016/j.tig.2003.11.008>
- Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J et al (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* 35(14):4809–4819. <https://doi.org/10.1093/nar/gkm487>
- Winkler WC, Breaker RR (2005) Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol* 59:487–517. <https://doi.org/10.1146/annurev.micro.59.030804.121336>
- Winkler W, Nahvi A, Breaker RR (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419(6910):952–956. <https://doi.org/10.1038/nature01145>
- Wittmann A, Suess B (2012) Engineered riboswitches: expanding researchers' toolbox with synthetic RNA regulators. *FEBS Lett* 586(15):2076–2083. <https://doi.org/10.1016/j.febslet.2012.02.038>
- Zhang Y, Gladyshev VN (2008) Molybdoproteomes and evolution of molybdenum utilization. *J Mol Biol* 379(4):881–899. <https://doi.org/10.1016/j.jmb.2008.03.051>