Identification of sigma factor 54-regulated small non-coding RNAs by employing genome-wide and transcriptome-based methods in rhizobium strains

Kasthuri Rajendran1 · Vikram Kumar2 · Ilamathi Raja2 · Manoharan Kumariah1 · Jebasingh Tennyson2

Received: 15 September 2021 / Accepted: 12 October 2022 / Published online: 20 October 2022 © King Abdulaziz City for Science and Technology 2022

Abstract

Rhizobium-legume symbiosis is considered as the major contributor of biological nitrogen fixation. Bacterial small non-coding RNAs are crucial regulators in several cellular adaptation processes that occur due to the changes in metabolism, physiology, or the external environment. Identifying and analysing the conditional specific/sigma factor-54 regulated sRNAs provides a better understanding of sRNA regulation/mechanism in symbiotic association. In the present study, we have identified sigma factor 54-regulated sRNAs from the genome of six rhizobium strains and from the RNA-seq data of free-living and symbiotic conditions of Bradyrhizobium diazoefficiens USDA 110 to identify the novel putative sRNAs that are over expressed during the regulation of nitrogen fixation. A total of 1351 sRNAs were predicted from the genome of six rhizobium strains and 1375 sRNAs were predicted from the transcriptome data of B. diazoefficiens USDA 110. Analysis of target mRNA for these novel sRNAs was inferred to target several nodulation and nitrogen fixation genes including nodC, nodJ, nodY, nodJ, nodM, nodW, nodZ, nifD, nifN, nifQ, fixK, fixL, fixD, nolB, and several cytochrome proteins. In addition, sRNAs of B. diazoefficiens USDA 110 which targeted the regulatory genes of nitrogen fixation were confirmed by wet-lab experiments with semi-quantitative reverse transcription polymerase chain reaction. Predicted target mRNAs were functionally classified based on the COG analysis and GO annotations. The genome-wide and transcriptome-based integrated methods have led to the identification of several sRNAs involved in the nodulation and symbiosis. Further validation of the functional role of these sRNAs can help in exploring the role of sRNAs in nitrogen metabolism during free-living and symbiotic association with legumes.

Keywords Rhizobium · Sigma factor-54 · sRNA · Nitrogen fixation

Introduction

sRNAs serve as key regulators in the post-transcriptional/translational process, and their role cannot be neglected in all the three kingdoms of living organisms. Small RNAs excluding tRNAs and rRNAs range from 50 to 500 nts in length and are encoded from intergenic regions (IGRs) with their own promoters or exceptionally transcribed from the promoter of the surrounding genes. Their transcription usually terminates using a strong rho-independent terminator. They regulate the gene expression by perfect or imperfect base pairing with target mRNAs with or without protein partners, such as Hfq and Csr (Moller et al. 2002; Gottesman 2004). sRNAs can be induced and differentially expressed under stress or specific growth conditions, which specify their involvement in many biological process. They are also shown to regulate virulence, nitrogen assimilation, heat stress response, etc. Studying the regulatory role of the sRNAs in the post-transcriptional and translational processes provides a better understanding of the target mRNA-sRNA interaction during important metabolic processes.

In recent times, RNA deep sequencing and several genome-wide computational methods have been widely employed to identify the sRNAs in bacteria. The increasing availability of complete genome sequences has led to the
in silico searches in different bacteria. In the current scenario of sRNA research, there are various bio-computational tools like algorithms and software that are developed for the prediction of putative sRNAs along with techniques and strategies to isolate and to further experimentally validate the sRNAs by Northern analysis. Employment of tools, such as QRNA, RNALfold, RNAz, SIPHT and NAPP, are found to be ideal for detecting the putative sRNAs (Del Val et al. 2007; Livny 2007; Livny and Waldor 2007; Livny et al. 2008; Marchais et al. 2009). Sigma factors are known to initiate transcription of specific set of genes by directing the core RNA polymerase to the promoter binding site. Apart from the sigma factors that are responsible for the expression of housekeeping genes (σ 70-housekeeping sigma factor), yet another sigma factor, σ 54 (RpoN/ntrA) a nitrogen limitation sigma factor, is known to initiate the transcription of nitrogen fixation/regulation genes. RpoN recognizes and binds to the -35/-10-type promoter containing the consensus sequence-5′-TTG GCA GN4TTGCW-3′ (Beynon et al. 1983; Barrios et al. 1999). There are many studies relating the importance of small regulatory RNA molecules in several cellular functions of soil bacteria, but the sRNA-mRNA regulation during free-living and symbiotic association with legumes still remains unanswered. Therefore, in the present study, we have initiated to explore the sigma factor-54 regulated sRNAs and their regulation on mRNAs during the free-living and symbiotic conditions.

The genome of different species of rhizobium harbours multiple replicons and their number varies among the different species. B. diazoefficiens USDA 110 is a slow growing bacterium that harbours a single circular chromosome (9.1 Mb) that can fix molecular nitrogen in symbiosis with the soybean plant (Madhugiri et al. 2012). Sinorhizobium meliloti 1021 is another soil bacterium that can establish an intimate symbiosis with Medicago, Melilotus and Trigonella (Capela et al. 2001). The genome of this strain harbours three replicons, the chromosome (Mb) and two mega plasmids, pSymA (Mb) and pSymB (Mb). Rhizobium etli CFN 42 interacts symbiotically with the common bean Phaseolus vulgaris to form the nitrogen fixing root nodules. The genome of Rhizobium etli organised with seven replicons includes a circular chromosome (6,530,228 bp) and six plasmids: p42a (194,229 bp), p42b (184,338 bp), p42c (250,948 bp), p42d (371,254 bp), p42e (505,334 bp) and p42f (642,517 bp) (González et al. 2006). Comparable to different Rhizobium species, Rhizobium leguminosarum 3841 is known to establish a symbiotic relationship with a legume plants including Pisum sativum (pea) and Vicia sativa (vetch). The genome of R. leguminosarum 3841 is organised with seven replicons including the circular chromosome (5.06 Mb), pRL 10 (0.49 Mb), pRL 11 (0.68 Mb), pRL 12 (0.87 Mb), pRL7 (0.15 Mb), pRL8 (0.15 Mb), pRL9 (0.35 Mb) (Laus et al. 2006; Young et al. 2006). Rhizobium tropici CIAT899 is known to induce the formation of nitrogen-fixing nodules in the roots of the common bean (Phaseolus vulgaris) and the leucaena tree (Leucaena leucocephala). The genome of this species harbours a chromosome (3.84 Mb) and three plasmids pRtrCIAT899a (0.22 Mb), pRtrCIAT899b (0.55 Mb) and pRtrCIAT899c (2.08 Mb) (Ormeño-Orrillo et al. 2012). Sinorhizobium fredii NGR234 can form symbiotic relationship with soybean, cowpea or pigeon pea, and it harbours a chromosome (3.93 Mb) and two plasmids pNGR234a (0.54 Mb) and pNGR234b (2.43 Mb). Earlier studies have shown the existence of sRNAs among different species of Rhizobium such as Sinorhizobium meliloti, Rhizobium etli, Bradyrhizobium japonicum and Mesorhizobium huakii (Del Val et al. 2007; Schluter et al. 2010; Valverde et al. 2008; Vercruyse et al. 2010; López-Leal et al. 2014; Fuli et al. 2017; Raja et al. 2018; Rajendran et al. 2020). However, there is scanty information on the conditional specific sRNAs and its regulation in rhizobium species. In the present study, we have identified sigma factor 54-regulated sRNAs from the genome of six rhizobium species and subsequently sRNAs that are expected to be involved in the regulation of nitrogen fixation were analysed from the free-living and symbiotic specific transcriptome data of the model strain B. diazoefficiens USDA 110.

Methods and materials

Genome-wide prediction of sigma factor 54-regulated sRNAs using improved sRNA scanner

The complete genome sequence and annotation files of all Rhizobium strains were retrieved from the National Center for Biotechnology Information (NCBI). Genome sequences and annotation files were downloaded in fasta nucleic acid (.fna) and protein data file (.ptt) formats, respectively. Accession numbers of various strains with their respective replicons used in our study are listed in the supplementary information 1. In the present study, we employed an improved version of sRNA scanner to predict sigma factor 54-regulated sRNAs. This bioinformatic tool uses positional weigh matrices (PWM) of promoter and rho-independent terminator signals (ST 2), through sliding window-based genome scans, using consensus sequences of sigma factor promoter binding sites: −35 and −10 and rho-independent transcription terminator sequences. This improved sRNA scanner tool is specifically designed for several sigma factors σ24 (extra cytoplasmic/extreme heat shock sigma factor), σ32 (heat shock sigma factor), σ54 (nitrogen limitation sigma factor), and σ70 (housekeeping sigma factor), and their PWM were created based on DNA binding sequences (motifs) of respective sigma factors, whereas in the previous version of
sRNA scanner, PWM was available only for housekeeping sigma factor σ70 (Raja et al. 2018). Sigma factor 54 specific PWM were used to identify the sRNAs from the complete bacterial genome using sRNA scanner. sRNA scanner was used with cumulative sum of score (CSS) of 12 and search length of sRNAs with 50–500 nt. Transcripts with non-coding nature (that does not code for any protein) were considered for further annotation of sRNA. Length and GC content of the putative non-coding transcripts were analysed using customized PERL script as described in our previous publications (Raja et al. 2018; Rajendran et al. 2020). In order to refine the data, other regulatory RNAs which are predicted along with the sRNAs were eliminated by searching against Rfam database. In order to screen the already reported sRNAs, predicted putative non-coding sRNAs were compared against Bacterial Small Regulatory RNA Database (BSRD) (Li et al. 2013). The sRNAs were also compared with previous reports to assess and confirm their novelty. Length and GC content of the putative non-coding transcripts were analysed considering for further annotation of sRNA. Length and GC content of the putative non-coding transcripts were analysed considering for further annotation of sRNA.

**Transcriptome-based sRNA prediction**

The RNA-seq dataset was obtained from the NCBI Gene Expression Omnibus (GEO) (Accession No: GSE69059) (Cuklina et al. 2016). The raw reads of *B. diazoefficiens* USDA 110 under two growth conditions (free-living and symbiotic) were downloaded from the sequence read archive (SRA) database (Accession No.: SRX1033915). The SRA tool kit was used for extracting the transcriptome reads from SRA files in FASTQ format (Leinonen et al. 2010). PolyA, polyT, and Illumina adapters were removed with cutadapt tool (Martin 2011). Sequence quality was analysed using FastQC. Sequence reads having phred score > 20 were used for further analysis. Trimmed reads were aligned to the genome of *B. diazoefficiens* USDA 110 by using Rockhopper tools (https://cs.wellesley.edu/~btjaden/Rockhopper/index.html) for transcriptome read counting (McClureet al. 2013; Tjaden 2015). Based on the alignment data, non-coding transcripts are considered as sRNA. Reads of the coding and non-coding transcripts were separated and aligned to the reference genome. The sRNA sequence was aligned to the genome and visualized using the Integrative genome viewer (IGV). Genomic coordinates of predicted sRNA were extracted from the genome using either samtools or bedtools. Genomic coordinates of these predicted RNAs were extracted from the genome using either samtools or bedtools. Genomic coordinates of these predicted RNAs were considered for further analysis. Clients of free-living and symbiotic association were further functionally annotated. Functional categorization of the predicted target mRNAs was done by clusters of orthologous group (COG) analysis using the Eggnog database (Huerta-Cepas et al. 2019). Gene ontology (GO) annotations and regulatory relationships among the biological process were analysed through the GO regulatory network by using the comparative GO web server (Fruzan-gohar et al. 2015).

**Target and secondary structure prediction for sRNAs**

TargetRNA2 Software was used to predict the mRNA targets for the predicted trans-encoded sRNAs (http://cs.wellesley.edu/~btjaden/TargetRNA2/). TargetRNA2 is a web server that identifies mRNA targets of sRNA involved in regulating gene expression in bacteria. As input, TargetRNA2 takes the sequence of a sRNA and the name of the replicon and it uses a variety of features, including conservation of the sRNA in other bacteria, the secondary structure of the sRNA, each candidate mRNA target and the hybridization energy between the sRNA and mRNA targets (Kery et al. 2014).

RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) was used to predict the secondary structure of sRNAs. FASTA sequence of sRNAs was used for calculating the minimum free energy (ΔG) based on the partition function (default parameter) (Hofacker 2003).

**Prediction of promoter and terminator**

The promoter and rho-independent terminator regions were predicted for the sRNAs of transcriptome data. The promoter and terminator regions of sRNAs were analysed from the region upstream of the transcription start site (TSS) and downstream of the transcription end site (TES), respectively. Genomic co-ordinates of 150-nt sequences upstream of TSS and 150-nt sequences downstream of TES were extracted using ‘Bedtools’ (Quinlan and Hall 2010). After extraction, ‘bprom’ was used to identify the binding sites of σ70 (Sala-mov and Solovyev 2011). Arnold tool was used to identify rho-independent terminators (Nville et al. 2011).

**Functional enrichment analysis of novel putative sRNAs**

The mRNA targets of both free-living and symbiotic association were further functionally annotated. Functional categorization of the predicted target mRNAs was done by clusters of orthologous group (COG) analysis using the Eggnog database (Huerta-Cepas et al. 2019). Gene ontology (GO) annotations and regulatory relationships among the biological process were analysed through the GO regulatory network by using the comparative GO web server (Fruzan-gohar et al. 2015).

**Preparation of plant material and root nodulation**

*B. diazoefficiens* USDA 110 cultures were grown in YEMA broth (Vincent 1970) at 25°C for 48 h on a rotary shaker at 140 rpm till it reached stationary phase. *Glycine max* seeds were surface sterilised, germinated and then infected with *B. diazoefficiens* USDA 110 suspension for 2 h. For control, the same protocol was followed with plain YEMA broth. After infection, seedlings were planted in sterile vermiculate bags and kept in the green house. Matured nodules were harvested on 28th, 29th and 30th days. RNA isolation was performed immediately.
### Semi-quantitative PCR

RNA was isolated from *Rhizobium* as described by Wise et al. (2006). cDNA was synthesized with a first strand cDNA synthesis kit (ABI). Semi-quantitative RT-PCR was performed for selected 10 sRNA candidates with two biological replicates. 16S rRNA was used as a positive control. Semi-quantitative PCR was performed without template and reverse transcriptase as negative controls to rule out the possibility of amplification due to primer dimer formation and DNA contamination, respectively. The following PCR conditions were used for sRNA amplification: denaturation at 95°C for 30 s, annealing at 58°C for 20 s, extension at 72°C for 30 s, for 35 cycles. The densitometry scanning was performed using the software ImageJ, available at [http://rsbweb.nih.gov/ij/download.html](http://rsbweb.nih.gov/ij/download.html).

### Results

#### Genome-wide screening of sigma 54-regulated sRNAs

Prediction of non-coding sRNAs from different nitrogen-fixing rhizobium species was performed by genome-wide computational analysis, based on the PWM matrices of conditional sigma factor 54 (nitrogen limitation sigma factor) by using improved sRNA scanner program (Raja et al. 2018). sRNA scanner demarks the transcription units (TUs) using consensus sequences of sigma factor binding sites (−35 and −10 (SI 2)) and rho-independent transcription terminator sequences. The total number of sRNAs predicted from each of the rhizobium species is graphically represented in Fig. 1a, and the details of the predicted sRNAs are given in Table 1. Among the six species of rhizobium, higher numbers of sRNAs are predicted from the genome of *S. meliloti* followed by *S. fredii, R. etli, R. trifolii, R. leguminosarum*

![Graph](image)

**Fig. 1** Genome-wide prediction of sRNAs using improved sRNA scanner tool. **a** Length distribution and **b** GC% content

#### Table 1  sRNAs identified from the genome of rhizobium strains

| Name of the strain                  | No. of sRNAs predicted | Total number of sRNAs predicted | Homologous identified in Rfam | Homologous identified in BSRD |
|------------------------------------|------------------------|---------------------------------|-------------------------------|------------------------------|
| *Bradyrhizobium diazoefficiens* USDA 110 | 39 45                  | 84                              | 4                             | –                            |
| *Sinorhizobium meliloti* 1021      | 132 160                | 292                             | 61                            | 8                            |
| *Rhizobium etli* CFN 42            | 120 120                | 240                             | 46                            | 1                            |
| *Rhizobium leguminosarum* 3841     | 144 109                | 253                             | 48                            | 1                            |
| *Rhizobium tropici* CIAT899        | 112 108                | 220                             | 53                            | –                            |
| *Sinorhizobium fredii* NGR234      | 130 132                | 262                             | 33                            | –                            |
and *B. diazoefficiens* USDA 110. The genome size and the number of replicons are found to vary among the different species of rhizobium. sRNAs of all the species of rhizobium were equally transcribed from both positive and negative strands, whereas in the *R. leguminosarum*, sRNAs were mostly transcribed from the negative strand (Table 1). Maximum numbers of sRNAs were predicted in chromosomes for all the species of rhizobium. To find out the novel putative sRNAs, predicted sRNAs were searched against Rfam database and BSRD database to eliminate the conserved homologs.

**Length and GC% content distribution of sRNAs**

Generally, sRNAs are found to have length in the range of 50–500 bp. sRNAs that are < 50 and > 500 bp were removed, and rest of them were used for further analysis. Most of the sRNAs predicted from rhizobium genomes are found to have length of 50–150 bp (Fig. 1a). GC content of sRNAs ranges from 30 to 80%. Majority of them were found to have 50–60% content (Fig. 1b).

**sRNA comparison between the species**

It has been reported that sRNAs are conserved and contain conserved interacting sites in their target mRNAs to regulate the physiological process. However, in many cases, sRNAs are known to have their unique regulatory role in bacteria. Therefore, the predicted sRNAs were compared with various rhizobium species to analyse their conservation and to find out the common sRNAs. The present study identified 8 sRNAs are found to be conserved between *R. etli* and *R. leguminosarum*, 1 sRNA between *R. leguminosarum* and *R. tropici* and 1 sRNA between *R. tropici* and *S. fredii* and interestingly, 8 sRNA candidates of *S. fredii* conserved with sRNAs of *S. meliloti*. In order to find out and identify the novel putative sRNAs, the sRNAs predicted from the genome of the rhizobium species were compared with previously reported sRNAs. 70 sRNAs that were predicted from the genome of *S. meliloti* are found to be conserved in sRNAs reported by Del Val et al. (2008) and 30 sRNAs of *S. meliloti* overlapped with the sRNAs reported by Schulter et al. (2010). Interestingly, 3 sRNA candidates were found to be conserved with virulence specific (sigma factor-32 regulated) sRNAs of *Agrobacterium fabrum* (Raja et al. 2018). A total of 6 sRNA candidates of *R. etli* are conserved with sRNAs reported by Vercruysse et al. (2010) and López-Leal et al. (2015) (SI 3a). Since the bacterial cells harbour different types of regulatory RNAs, the predicted sRNAs were analysed in order to find out whether they belong to the category of small regulatory RNAs or any other group of regulatory RNAs. In order to work out the non-coding RNAs, the predicted sRNAs were searched with the database, such as, Rfam and BSRD which are considered to be comprehensive for the requirement of the present study (Table 1). The identified conserved homologs were eliminated, and other novel sRNAs were taken for further analysis. From the above analysis, some more sRNA homologs were identified from the Rfam database.

**Transcriptome-based sRNAs prediction**

The high-quality RNA sequence reads of the bacterium under free-living and symbiotic conditions specific for *B. diazoefficiens* USDA 110 (SI 4) were aligned to respective genome using Rockhopper. Subsequent to alignment, transcripts from the intergenic regions and antisense regions from the complementary strand of the protein-coding genes were identified. sRNAs that are < 50 and > 500 bp were removed and the intergenic sRNAs (trans-encoded) having a length in the range of 50–500 nt were taken for further analysis. As an outcome, a total of 1375 trans-encoded sRNAs were predicted. The lengths of the predicted sRN As are found to vary in length between 50 and 400 nt, and most of the sRNA candidates have the length of 50–100 nt in length. GC content of the sRNAs was found to be in the range of 31–80%, and most of the sRNAs ranged from 51 to 60%. In the present study, predicted sRNAs were searched against Rfam database and BSRD databases in order to eliminate the conserved homologs. Only one sRNA of *B. diazoefficiens* USDA 110 was found to be conserved against sRNAs in the Rfam database (chrB RNA). In order to screen the novel sRNAs, the predicted sRNAs were searched against the previously reported sRNAs. And the results indicated that 8 sRNA candidates of *B. diazoefficiens* USDA 110 are conserved in comparison with sRNAs reported by Hahn et al. 2016 (SI 3b). sRNAs predicted from the transcriptome were searched against the sRNAs generated based on the PWM metrics of sigma factor 54 of *B. diazoefficiens* USDA 110 in order to evaluate their regulatory role in the nitrogen fixation and nitrogen metabolism. A total of 21 sRNAs of *B. diazoefficiens* USDA 110 were found to be conserved with the sigma factor 54-based predicted sRNAs. In order to study the regulatory role of the sRNAs in the regulation of nitrogen fixation and symbiotic association, sRNAs which were found to be regulated in both free-living and symbiotic conditions (based on the expression level from the rockhopper output file) were selected for further analysis. Since a single sRNA can regulate multiple mRNA targets post-transcriptionally, in order to screen the sRNAs and their role in nitrogen fixation post-transcriptionally, their mRNA targets were predicted using the TargetRNA2 tool and the flanking genes for these sRNAs were predicted using IGV (Table 2).
Table 2  Targets and secondary structures of sRNAs identified from the transcriptome data

| sRNA no | sRNA coordinates | Target rpoN/ntrA gene mRNA | Interaction energy with gene mRNA | P-value | Other predicted target genes | Secondary structure | Minimum free energy |
|---------|------------------|----------------------------|----------------------------------|---------|-----------------------------|---------------------|---------------------|
| BD1     | 1927274-1927558  | Transcriptional regulator (fixK) | -10.42                           | 0.019   | imidazole glycerol phosphate synthase subunit (hisH), gamma-glutamyltranspeptidase (gggT) | ![Secondary structure](image) | -76.10 kcal/mol      |
| BD2     | 2174239-2174361  | nitrogen fixation protein (nifQ) | -13.41                           | 0.003   | preprotein translocase subunit (secA), lipoyltransferase (lipB) | ![Secondary structure](image) | -41.50 kcal/mol      |
| BD3     | 2226897-2226997  | nitrogenase molybdenum-iron protein subunit alpha (nifD) | -12.91                           | 0.004   | phosphatidylglycerophosphate synthase, cobalamin synthesis protein (cobW), beta-ketoacyl CoA thiolase (pcaF) | ![Secondary structure](image) | -39.70 kcal/mol      |
| BD4     | 2066092-2066221  | ferrodoxin (fdx) heme exporter protein (ccmC), ABC transporter molybdenum-binding protein (modA) | -8.55 -18.18                      | 0.045   | glycogen debranching protein (glgX), thioredoxin reductase (trxB), succinate dehydrogenase cytochrome (sdhC), histidinol-phosphate aminotransferase (hisC), glyceraldehyde-3-phosphate dehydrogenase (gapA), | ![Secondary structure](image) | -31.70 kcal/mol      |
# Table 2 (continued)

|    | 4506199-4506295 | cbb3 oxidase subunit III – (fixP), cytochrome C oxidase subunit III (coxC), coproporphyrinogen III oxidase (hemN), two component regulator (nodW) | -11.69 | -11.61 | -8.68 | -12.35 | 0.009 | 0.010 | 0.043 | 0.006 | Flagellar basal body rod protein (flgG), glycogen synthase (glgA), phosphoribosylamine–glycine ligase (purD), biopolymer transport protein (exbB) | -37.00 kcal/mol |
|----|----------------|---------------------------------------------------------------------------------|--------|--------|-------|--------|-------|-------|-------|-------|------------|----------------|
| BD6| 1741298-1741451 | nitrate ABC transporter permease (nrtB), cytochrome P-450 CYP117                | -9.35  | -12.99 | 0.032 | 0.004 | glutamyl-tRNA synthetase (glnS), components of type IV pilus pilin subunit (ctpA), flagellar basal-body rod protein (flgB) | -37.80 kcal/mol |
| BD7| 1214552-1214648 | nitrogenase molybdenum-cofactor biosynthesis protein (nifN), nodulation protein (nolV), cytochrome oxidase assembly factor (cox15) | -10.5  | -10.62 | -10.73 | 0.018 | 0.017 | 0.016 | co-chaperonin GroES | -32.20 kcal/mol |
| BD8| 5185970-5186073 | RNA polymerase sigma-54 factor (rpoN1)                                         | -13.04 | 0.004 | L-idonate 5-dehydrogenase, N-(5'-phosphoribosyl) anthranilate isomerase (trpF) | -32.20 kcal/mol |
| BD9| 7340538-7340630 | Oxygen-dependent coproporphyrinogen III oxidase-hemF, nodulation protein nolB | -23.94 | -8.45  | 0.000 | 0.047 | chromosome partitioning protein parB, large ATP-dependent helicase-related protein lhr | -41.60 kcal/mol |
Target prediction

Generally, sRNAs act by short perfect or imperfect base-pairing with complementary sequence stretches of the multiple target mRNAs to mediate post-transcriptional gene regulation. Therefore, it is necessary to identify the relevant mRNA targets for the sRNA. To screen the sRNAs, which post-transcriptionally influence the rhizobium nitrogen fixation/regulation, putative targets of the sRNAs were predicted in all the rhizobium species that were employed in the present study with TargetRNA2 tool. Target mRNAs were predicted for sRNAs derived from genome-wide and transcriptome-based analysis. Target prediction revealed that 13 sRNAs of *S. meliloti*, 30 sRNAs of *S. fredii*, 18 sRNAs of *R. leguminosarum*, 15 sRNAs of *R. etli* and 5 sRNAs of *R. tropici* were found to target nitrogen fixation/regulation genes. sRNA targets of all the rhizobium species employed in the present study were mostly complemented with the binding site of various nodulation genes, such as, *nodI*, *nodD2*, *nodC*, *nodF*, *nodQ1 nodS*, *nopB*, *nopC*, *nopT*, *nopX*, *nopM*, *nopP*, *nolO*, *noel*, *fdxB*, and *fdxN* and nitrogen fixation genes, such as, *nifA*, *nifB*, *nifE*, *nifD1*, *nifD2*, *nifH*, *nifK*, *nifN*, *nifS*, *nifT*, *nifQ*, *fixA*, *fixB*, *fixC*, *fixF*, *fixG*, *fixX*, *fixS*, and *fixU* and several members of the family of cytochrome proteins. All these mRNA targets exhibited significant complementarity with the sRNAs and showed significant \( p \) value (< 0.05). In *B. diazoefficiens* USDA 110, nearly 100 sRNAs which showed high expression from the transcriptome were selected for the target prediction. Among 100 sRNAs, 43 sRNAs from the transcriptome and 6 sRNAs of genome (SI 5) were found to target nitrogen regulation/fixation related genes. Based on the thermodynamic interaction energy (kcal/mol) of hybridization between the sRNA and mRNA targets and significant \( p \) value (< 0.05), sRNAs which regulate nitrogen fixation genes of *B. diazoefficiens* USDA 110 (Table 2) were taken for further analysis and were validated by performing wet-lab experiments.

Prediction of promoter, terminator and secondary structure

Promoter and rho-independent terminator sequences were predicted for the identified putative novel sRNAs of transcriptome data (Table 3). Secondary structure was predicted for the select set of sRNAs using RNAfold server. The predicted minimum free energy for the majority of the sRNAs was in the ranges of \(-20\) to \(-70\) kcal/mol. The colour of the structure represents the base-pairing probabilities (Table 2, SI 5).

Functional categorization of sRNA target genes

In order to study the regulation of target mRNAs involved in nitrogen metabolism and nitrogen fixation by the sRNAs (Table 2), the involvement of the sRNAs in these bioprocess was functionally annotated by COG and GO analysis. The mRNA targets of both free-living and symbiotic association were enriched in COG categories of energy production and conversion, amino acid transport and metabolism, coenzyme biosynthesis and metabolism and post-translation modification, protein turn over and the bioavailability of chaperons (Fig. 2). In GO analysis, the target mRNAs were found to be enriched in 3 categories, viz., biological process, molecular functions and cellular components. Accordingly, the targets were categorized under biological process including the genes involved in cellular and metabolic process and localization, molecular functions, binding, catalytic activity and transporter activity (Figs. 3 and SI6).
GO regulatory network analysis was constructed for the target mRNAs of free-living and symbiotic association conditions (from the final sets of sRNA (Table 2). The target mRNAs of sRNAs identified under the free-living condition is shown in Fig. 4. Signal transduction is known to be the central node in the GRN, governed by nodW and ctpA (GO ID: 7165) and showed interaction with other GO terms, such as, regulation of transcription, positive regulation of sporulation, nodulation, carbohydrate metabolism, protoporphyrinogen IX biosynthetic process, bacterial-type flagellum-dependent swarming motility, protein folding and aerobic respiration. In the case of the target mRNAs of symbiotic association, regulation of transcription is found to be the central node in the network (Fig. 5) governed by ccmC (GO ID: 17004) which showed interaction with nitrogen fixation, aerobic respiration, transmembrane transport and nodulation.

Table 3 Promoter, terminator and flanking genes of sRNAs identified from the transcriptome data

| sRNA | sRNA coordinates − 10 sequence | − 35 sequence | Terminators | Upstream gene | Downstream gene |
|------|-------------------------------|---------------|-------------|---------------|----------------|
| BD1  | 1,927,274–1,927,558            | CTTTATAGT     | TGCGGA      | Outer membrane protein | Hypothetical protein |
| BD2  | 2,174,239–2,174,361            | GGTCATTCT     | TCGATG      | Hypothetical protein | Hypothetical protein |
| BD3  | 2,226,897–2,226,997            | GTGATGCT      | GTGCCA      | Phenolhydroxylase-like protein | Hypothetical protein |
| BD4  | 2,066,092–2,066,221            | GTTGAGATC     | CCCATAAC    | Hypothetical protein | Transposase |
| BD5  | 4,506,199–4,506,295            | CGGGATTC      | GTGAAG      | Alanine racemase | Replicative DNA helicase |
| BD6  | 1,741,298–1,741,451            | GATTAGATG     | TTGGCA      | Hypothetical protein | Site-specific integrase/recombinase |
| BD7  | 1,214,552–1,214,648            | ATCCAGAGT     | TGCA        | Hsp33-like chaperonin | Hypothetical protein |
| BD8  | 5,185,970–5,186,073            | TGGTAGACT     | TCGCA       | Cysteine synthase | Queuine tRNA-ribosyltransferase |
| BD9  | 7,340,538–7,340,630            | –             | –           | Hypothetical protein | Hypothetical protein |
| BD10 | 157,905–158,001                | CCTTAAGCT     | TTGCAGA     | ATP-dependent helicase | Hypothetical protein |

Fig. 2 COG classifications of the target genes of B. diazoefficiens USDA 110. The COG (cluster of orthologous groups) categories are coded as follows: C—energy production and conversion; E—amino acid transport and metabolism; F—nucleotide transport and metabolism; G—carbohydrate transport and metabolism; H—coenzyme metabolism; I—lipid metabolism; J—translation; K—transcription; L—DNA replication, recombination, and repair; M—cell wall/membrane biogenesis; N—cell motility; O—post-translational modification, protein turnover, and chaperones; P—inorganic ion transport and metabolism; Q—secondary metabolite biosynthesis, transport, and catabolism; R—function-unassigned conserved proteins; S—signal transduction; U—intracellular trafficking, secretion, and vesicular transport.

Semi-quantitative PCR analysis

From the combined out of genome and transcriptome data, one common sRNA from the positive strand and nine sRNAs involved in nitrogen fixation from the transcriptome (7 and 2 sRNAs from negative and positive strand, respectively) of B. diazoefficiens USDA 110 were selected. Among the 10 sRNAs, 8 sRNAs (BD1-BD8) showed amplifications.
Semi-quantitative PCR was performed for 8 sRNA, and the primers used to amplify the sRNAs are listed in the SI 7. Densitometric analysis of semi-quantitative PCR revealed that the 16S rRNA expression was constant in both symbiotic and free-living conditions; differential expression was observed between the 8 sRNAs in symbiotic and free-living conditions (Fig. 6).

Discussion

Availability of RNA-seq data, tilling arrays data, genome data and development of several computational software tools has significantly eased out researchers in order to study the existence of various forms of small regulatory RNAs and explore their regulation in cellular process in all domains of life. Several studies have revealed the regulation and importance of sRNA during bacterial stress response conditions. Recently, small non-coding RNA NfeR1, which were involved in regulating nodule formation efficiency, has been reported to have an effect on osmoadaptation and symbiotic efficiency of model rhizobium species, namely, *S. meliloti* (Robledo et al. 2017). In order to understand the functional role of sRNA in rhizobium-legume symbiosis, it is important to study whether it targets the genes involved in nitrogen fixation/assimilation or regulate the gene expression during symbiosis. Hahn et al. (2016) experimentally validated 10 new sRNAs in soybean symbiont *B. diazoefficiens* USDA 110. However, there are only few reports are available related to the conditional specific sRNAs and no reports available on their regulatory role exerted on the target mRNAs. In the present work, we have identified several novel sRNAs by genome-wide and transcriptome-based methods and made an attempt to relate the role of sRNAs in the regulation of nitrogen metabolism during free-living and symbiotic conditions. A total of 1207 sigma factor 54-regulated sRNAs were identified from 6 rhizobium species by employing improved sRNA scanner tool using PWM matrices for the conditional sigma factor 54. Ours is the first report on the existence of sRNA in *R. leguminosarum*, *R. tropici*, and *S. fredii*. Among the 6 species of rhizobium, relatively higher number sRNAs were predicted for *S. meliloti, S. fredii* and
Fig. 4  GO regulatory network based on the mRNA targets of sRNAs predicted from free-living condition

Fig. 5  GO regulatory network based on the mRNA target of sRNAs predicted from symbiotic condition
The conserved homologs of other regulatory RNAs were identified by the batch search against Rfam and BSRD databases. Small non-coding RNAs were searched separately for all the 6 species of rhizobium. Among the 1351 sRNAs, 245 conserved homologs were found in Rfam and 10 homologs were found in BSRD databases. Batch search against Rfam database revealed the presence of 61 other regulatory RNAs including tRNAs, RNAseP, ar7, ar14, ar15 and ar35. Highest numbers of other small regulatory RNA homologs were identified in Rfam database. sRNAs identified in the present study were compared with the previously reported sRNAs. Among the 253 sRNAs of *R. etli*, 6 sRNAs were found to be...
conserved with the earlier reported sRNAs of *R. etli* (Ver-cruysse et al. 2010; López-Leal et al. 2015). Conservation analysis among the rhizobium species revealed that sRNAs of *R. leguminosarum* are highly conserved in comparison with *R. etli*. Both these rhizobium species are known to have seven replicons, which include the symbiotic plasmid pRL10 and p42d, respectively (Table 1). Further, we focussed our study to identify the sRNA target genes involved in the process of rhizobium-legume symbiosis. Interestingly, 84 sRNAs from genome-based method were found to target the genes involved in various nitrogen regulation process especially nodulation and nitrogen assimilation. Some of the sRNAs were found to regulate multiple target genes, and they have been shown to have significant binding sites on *nod* (d1, d2, c, f, q1), *nop* (b, c, m, x, p, t), *nif* (a, b, e, d1, d2, h, k, n, s, t, o, u), *fix* (a, b, c, f, g, x, s, o, u) with minimum interaction energy and *p* value < 0.05. In order to elucidate the sRNAs involved in nitrogen fixation, we have identified differentially expressed sRNAs from the free-living and symbiotic specific transcriptome data. More than 1000 sRNAs were predicted under free-living and symbiotic conditions in *B. diazoefficiens* USDA 110, and from this, 100 sRNAs were selected for target prediction analysis. The results showed that 43 sRNAs are found to have targets on several nitrogen fixation genes. Based on the analysis, the identified sRNAs were classified into three categories, which include sRNAs expressed only under free-living condition, under symbiotic condition and under both free living and symbiotic conditions. sRNAs expressed under symbiotic condition has a significant binding site on *nifQ, nifD, nodJ, fixK, fixL, Fdx, nolB, cytochrome proteins, ABC transporter molybdenum binding protein and heme exporter protein*. 

sRNAs from the bacterium under free-living condition were found to interact with targets related to several nodulation genes which include *nodC, nodY, nodJ, nodM, nodW, nodZ, nifD, fixP, fixK, fixL, nolB, nolV, hemN, groEL, groES, ccmC*, and several cytochrome proteins. The sRNAs which are inferred to target the nitrogen regulation fixation genes were selected for the wet-lab experimental analysis. A total of 10 sRNAs of *B. diazoefficiens* USDA 110 identified from the transcriptome data were selected for semi-quantitative-PCR analysis (Table 2). Semi-quantitative PCR analysis confirmed the presence of 8 sRNAs (BD1-BD8). While analysing the expression pattern of sRNAs, it was found that the sRNA BD1, BD2, and BD3 (Fig. 6a: lane 1, 2, and 3) were highly expressed under symbiotic condition as compared to the expression level under free-living condition. In addition, their targets were found to include nitrogen fixation gene *fixK* (transcriptional regulator), *nifD* (Nitrogenase molybdenum-iron protein subunit α), an essential enzyme required for biological nitrogen fixation and sRNA BD7 (Fig. 6a: lane 7) showed significantly higher level of expression in free-living bacteria as compared to the symbiotic condition which are found to target nitrogen fixation gene *nifN* (Nitrogenase molybdenum-iron co-factor biosynthesis protein) and *nolV* (nodulation protein). sRNAs BD2 and BD5 exhibited lower level of expression under conditions of both free-living and symbiotic conditions. Under free-living condition, notable expression was found in sRNAs BD3, BD6, and BD7.

Since the sRNAs are encoded in the intergenic region, they get transcribed from the orphan promoter and terminate with a *rho*-independent terminator. The promoter and *rho*-independent terminator were predicted for the above 10 sRNAs and found the absence of promoter for BD9 sRNA, and one terminator was predicted for BD1 sRNA (Table 3). In the present study, putative novel sRNA targets of *B. diazoefficiens* USDA 110 were functionally categorised by subjecting to COG and GO analysis. COG analysis revealed that most of the target mRNAs of sRNAs were involved in energy production and conversion, amino acid transport, and co-enzyme metabolism. In the GO enrichment analysis, most of the target genes were associated with cellular and metabolic process; catalytic and binding activity in molecular process; and biomembrane mediated regulation of cellular process. Further, we have also constructed the GRN for the predicted target mRNAs of *B. diazoefficiens* USDA 110 using the biological process as specified by GO terms for both free-living and symbiotic conditions. Network analysis revealed that many target genes are mainly involved in root nodulation and carbohydrate metabolism in the free-living condition, whereas in symbiotic condition, the target genes are mainly involved in nitrogen fixation, transmembrane transport, and aerobic respiration. Thus, the present study not only identified the sigma 54-regulated sRNAs but also related the role of sRNAs in the regulation of nitrogen metabolism and fixation during the free-living and symbiotic conditions. Further characterization of these sRNAs would throw light in the elucidation of their functional role in regulating the nodulation and nitrogen fixation during *Rhizobium*-legume symbiosis and identification of the sRNAs involved in positive regulation of nitrogen fixation would be helpful to increase the nitrogen content which would augment growth and bioproductivity of the host plant. Thus, understanding the role of sRNAs in the post-transcriptional regulation of target mRNAs would contribute to enhancing the yield potential of legume crops.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s13205-022-03394-x.

**Acknowledgements** We thank UGC-BSR for the financial support of KR.

**Author contributions** JT and MK conceived the idea. KR planned and performed the experiments. VK and IR created PWM matrix of improved sRNA scanner.

**Funding** Not applicable.
Availability of data and materials All data generated or analysed during this study are included in this published article.

Code availability Not applicable.

Declarations

Conflict of interest All authors declare that they have no competing interests.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

Consent for publication Not applicable.

References

Barrios H, Valderrama B, Morett E (1999) Compilation and analysis of σ54-dependent promoter sequences. Nucleic Acids Res 27:4305–4313. https://doi.org/10.1093/nar/27.22.4305

Beynon J, Cannon M, Buchanan-Wollaston V, Cannon F (1983) The nif promoters of Klebsiella pneumoniae have a characteristic primary structure. Cell 34:665–671. https://doi.org/10.1016/0092-8674(83)90399-9

Capela D, Barloy-Hubler F, Guoy J, Both G, Ampe F, Batut J, Galibert F (2001) Analysis of the chromosome sequence of the legume symbiont Sinorhizobium meliloti strain 1021, Proc Natl Acad Sci USA 98(17):9877–9882. https://doi.org/10.1073/pnas.161294398

Cukлина J, Hahn J, Imakaye M, Omasits U, Forstner KU, Ljubimov N, Gelfand MS (2016) Genome-wide transcription start site mapping of Bradyrhizobium japonicum grown free-living or in symbiosis—a rich resource to identify new transcripts, proteins and to study gene regulation. BMC Genom 17:302. https://doi.org/10.1186/s12864-016-2602-9

Del Val C, Rivas E, Torres Quesada O, Toro N, Jiménez Zurdo JI (2007) Identification of differentially expressed small non-coding RNAs in the legume endosymbiont Sinorhizobium meliloti by comparative genomics. Mol Microbiol 66:1080–1091. https://doi.org/10.1111/j.1365-2958.2007.05978.x

Fruzangohar M, Ebrahimie E, Gounniiy AD, Mahdi LK, Paton JC, Adelson DL (2015) Correction: comparative GO: a web application for comparative gene ontology and gene ontology-based gene selection in bacteria. BMC Genom 10:e0125537. https://doi.org/10.1001/journal.pone.0125537

Fuli X, Wenlong Z, Xiao W, Jing Z, Zhengzheng Z, Youguo L (2017) A genome-wide prediction and identification of intergenic small RNAs by comparative analysis in Mesorhizobium huakuii 7653R. Front Microbiol 8:1730. https://doi.org/10.3389/fmicb.2017.01730

González V, Santamaría RI, Bustos P, Hernández-González I, Medrano-Soto A, Moreno-Hagelsieb G, Dávila G (2006) The partitioned Rhizobium etli genome: genetic and metabolic redundancy in seven interacting replicons. Proc Natl Acad Sci 103:3834–3839. https://doi.org/10.1073/pnas.0508502103

Gottesman S (2004) The small RNA regulators of Escherichia coli: roles and mechanisms. Annu Rev Microbiol 58:303–328. https://doi.org/10.1146/annurev.micro.58.030603.123841

Hahn J, Tsoy OV, Thalmann S, Cuklina J, Gelfand MS, Evgueniava-Hackenberg E (2016) Small open reading frames, non-coding RNAs and repetitive elements in Bradyrhizobium japonicum USDA 110. PLoS ONE 11:e0165429. https://doi.org/10.1371/journal.pone.0165429

Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31:3429–3431. https://doi.org/10.1093/nar/gkg599

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, von Mering C (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309–D314. https://doi.org/10.1093/nar/gky1085

Kerry MB, Feldman M, Livny J, Tjaden B (2014) TargetRNA2: identifying targets of small regulatory RNAs in bacteria. Nucleic Acids Res 42:W124–W129. https://doi.org/10.1093/nar/gku317

Laus MC, Logman TJ, Lamers GE, Van Brussel AA, Carlson RW, Kijne JW (2006) A novel polar surface polysaccharide from Rhizobium leguminosarum binds host plant lectin. Mol Microbiol 59(6):1704–1713. https://doi.org/10.1111/j.1365-2989.2006.05057.x

Li L, Huang D, Cheung MK, Nong W, Huang Q, Kwan HS (2013) BSRD: a repository for bacterial small regulatory RNA. Nucleic Acids Res 41:D233–D238. https://doi.org/10.1093/nar/gks1264

Livny J, Waldor MK (2007) Identification of small RNAs in diverse bacterial species. Curr Opin Microbiol 10:96–101. https://doi.org/10.1016/j.mib.2007.03.005

Livny J, Teonadi H, Livny M, Waldor MK (2008) High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs. PLoS ONE 3:e3197. https://doi.org/10.1371/annotation/a03e1870-1dd7-4c16-8c46-2268eeb2a50a

López-Leal G, Tabche ML, Castillo-Ramírez S, Mendoza-Vargas A, Ramírez-Romero MA, Dávila G (2014) RNA-Seq analysis of the multipartite genome of Rhizobium etli CE3 shows different replication contributions under heat and saline shock. BMC Genom 15:1–15. https://doi.org/10.1186/1471-2164-15-770

Madhugiri R, Pessi G, Voss B, Hahn J, Sharma CM, Reinhardt R, Evgueniava-Hackenberg E (2012) Small RNAs of the Bradyrhizobium/Rhodopseudomonas lineage and their analysis. RNA Biol 9:47–58. https://doi.org/10.4161/rna.9.1.18008

Marchais A, Naville M, Bohn C, Boulouc P, Gautheret D (2009) Single-pass classification of all noncoding sequences in a bacterial genome using phylogenetic profiles. Genome Res 19:1084–1092. https://doi.org/10.1101/gr.089714.108

McClure R, Balasubramanian D, Sun Y, Bobrovskyk M, Sumby P, Genco CA, Tjaden B (2013) Computational analysis of bacterial RNA-Seq data. Nucleic Acids Res 41(14):e140–e140. https://doi.org/10.1093/nar/gkt444

Ormeño-Orrillo E, Menne P, Almeida L, Ollero FJ, Nicolás MF, Rodrigues EP, Vasconcelos ATR (2012) Genomic basis of broad host range and environmental adaptability of Rhizobium tropici CIAT 899 and Rhizobium sp. PRF 81 which are used in inoculants for common bean (Phaseolus vulgaris L.). BMC Genom 13:1–26. https://doi.org/10.1186/1471-2164-13-735

Rajendran K, Kumar V, Raja I, Kumariam M, Tennyson J (2020) Identification of small non-coding RNAs from Rhizobium etli by integrated genome-wide and transcriptome-based methods. ExRNA 2:1–11. https://doi.org/10.1186/s13059-020-00054-1

Schlüter JP, Reinkensmeier J, Daschkey S, Evgueniava-Hackenberg E, Janssen S, Jänicke S, Becker A (2010) A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium Sinorhizobium meliloti. BMC Genom 11:245. https://doi.org/10.1186/1471-2164-11-245

Tjaden B (2015) De novo assembly of bacterial transcriptomes from RNA-seq data. Genome Biol 16(1):1–10. https://doi.org/10.1186/s13059-014-0572-2

Valverde C, Livny J, Schlüter JP, Reinkensmeier J, Becker A, Parisi G (2008) Prediction of Sinorhizobium meliloti sRNA genes and experimental detection in strain 3111. BMC Genom 9:416. https://doi.org/10.1186/1471-2164-9-416

Vercruysse M, Fauvart M, Cloots L, Engelen K, Thijs IM, Mariscal C, Michiels J (2010) Genome-wide detection of predicted
non-coding RNAs in *Rhizobium etli* expressed during free-living and host-associated growth using a high-resolution tiling array. BMC Genom 11:1–12. https://doi.org/10.1186/gb-2011-12-2-r17

Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration (2010) The sequence read archive. Nucl Acids Res 39:D19–D21. https://doi.org/10.1093/nar/gkq1019

Møller T, Franch T, Udese C, Gerdes K, Valentin-hansen P (2002) Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon 16:1696–1706. Doi: https://doi.org/10.1101/gad.231702.

Raja I, Kumar V, Sabapathy H, Kumariah M, Rajendran K, Tennyson J (2018) Prediction and identification of novel sRNAs involved in *Agrobacterium* strains by integrated genome-wide and transcriptome-based methods. FEMS Microbiol Lett 365:fny247. https://doi.org/10.1093/femsle/fny247

Salamov VS, Solovyevand A (2011) Automatic annotation of microbial genomes and metagenomic sequences. Metagenomics and its applications in agriculture, Nova Science Publishers, Hauppauge, pp 61–78

Wise AA, Liu Z, Binns AN (2006) Nucleic acid extraction from *Agrobacterium* strains. In: *Agrobacterium protocols*. Humana Press, pp 67–76. https://doi.org/10.1385/1-59745-130-4:67

Vincent JM (1970) A manual for the practical study of the root-nodule bacteria. A manual for the practical study of the root-nodule bacteria

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.