Identification of Sigma Factor 54-Regulated Small Non-Coding RNAs by Employing Genome-Wide and Transcriptome-Based Methods in *Rhizobium* Strains

Kasthuri Rajendran
Madurai Kamaraj University
Vikram Kumar
Madurai Kamaraj University
Ilamathi Raja
Madurai Kamaraj University
Manoharan Kumariah
Madurai Kamaraj University
Jebasingh Tennyson (jebasinghs@gmail.com)
Madurai Kamaraj University https://orcid.org/0000-0001-9703-6090

Research Article

Keywords: Rhizobium, Sigma factor 54, sRNA, nitrogen fixation

Posted Date: June 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-624650/v1

Abstract

Rhizobium-legume symbiosis is considered as the major contributor of biological nitrogen fixation. In the present study, we have identified sigma factor 54-regulated sRNAs from the genome of five *Rhizobium* strains and integrated with the free-living and symbiotic specific transcriptome data to identify the novel putative sRNAs that are over expressed during the regulation of nitrogen fixation. A total of 1059 sRNAs were predicted from each genome of the select set of *Rhizobium* strains and 1,375 sRNAs were predicted from the transcriptome data of *Bradyrhizobium japonicum*. Target mRNA analysis revealed the functional role of putative novel sRNAs from different free-living and symbiotic strains. Those novel sRNAs were inferred to target several nodulation and nitrogen fixation genes including *nodC*, *nodJ*, *nodY*, *nodJ*, *nodM*, *nodW*, *nodZ*, *nifD*, *nifN*, *nifQ*, *fixK*, *fixL*, *Fdx*, *nolB*, and several cytochrome proteins. Further, sRNAs of *Bradyrhizobium japonicum* which targeted the regulatory genes of nitrogen fixation were experimentally confirmed with semi-quantitative reverse transcription polymerase chain reaction. Predicted target mRNAs were functionally classified based on the COG analysis and GO annotations. Studies on this sigma factor 54-regulated sRNA identification could be a better method to relate the role of sRNAs in nitrogen metabolism during free-living and symbiotic association with legumes.

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. sRNAs other than tRNAs and rRNAs which range from 50 to 500 nts in length, encoded from intergenic regions (IGRs) having their own promoters or exceptionally transcribed from the promoter of the surrounding genes. Their transcription usually terminates at a strong *rho*-independent terminator. They regulate the gene expression by perfect or imperfect base pairing with target mRNAs with or without protein partners like Hfq, Csr, *etc* (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*. 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’ -TTGGCAG- N4 -TTGCW- 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 regulatory role of non-coding RNA (ncRNA) in free-living and symbiotic association with legumes, still remain unanswered. The regulatory role of the sRNAs in the biological process need to be addressed clearly. It is pertinent to mention that there are no reports on the sigma factor 54-regulated sRNAs.

The genome of *Rhizobium* harbours multiple replicons and their number varies among the strains. *Bradyrhizobium japonicum* USDA 110 is a slow growing bacterium that harbours a single circular chromosome (9.1 Mb) can fix molecular nitrogen in symbiosis with the soybean plant (Madhugiri et al., 2012). *Rhizobium etli* interacts symbiotically with the common bean *Phaseolus vulgaris* to form the nitrogen fixing root nodules. The genome of *R. etli* organised with seven replicons includes circular chromosome (6,530,228 bp) and six plasmids: p42a (194,229 bp), p42b (184,338bp), p42c (250,948 bp), p42d (371,254 bp), p42e (505,334 bp) and p42f (642,517 bp) (Gonzalez *et al*., 2006). *Rhizobium leguminosarum* like other Rhizobia establishes a symbiotic relationship with a legume plant. The genome of *R. leguminosarum* 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) (Young *et al*., 2006). *Rhizobium tropici* is another strain which can induce the formation of nitrogen-fixing nodules on the roots of the common bean (*Phaseolus vulgaris*) and the leucaena tree (*Leucaena leucocephala*). The genome of this strain harbours 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 fredi* 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 (2.43Mb). Earlier studies have reported the existence of sRNAs from different *Rhizobium* strains 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 the present study, we have identified sigma factor 54-regulated sRNAs in 5 different *Rhizobium* strains and then analyzed the sRNAs involved in regulation of nitrogen fixation during free-living and symbiotic conditions.

Methods And Materials

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

The complete genome sequence and annotation files of five different strains namely *Bradyrhizobium japonicum*, *Rhizobium etli*, *Rhizobium leguminosarum*, *Sinorhizobium fredii*, *Rhizobium tropici* 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 have employed the improved version of the sRNA scanner to predict sigma factor 54-regulated sRNAs. This bioinformatic tool uses positional weigh matrices (PWM) of promoter and *rho*-independent terminator signals (SI 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 (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 true non-coding nature were considered for further annotation of sRNA. Length and GC content of the putative non-coding transcripts were analyzed 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 searched 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. Filtered, putative non-coding RNAs (sRNAs) were used for subsequent analysis.

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 *Bradyrhizobium japonicum* under two 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, Sugawara and Shumway, 2010). PolyA, polyT and Illumina adapters were removed with cutadapt tool (Martin 2011). Sequence quality was analyzed using FastQC. Sequence reads having phred score > 20 were used for further analysis. Trimmed reads were aligned to the genome of *R. etli* by using Rockhopper tools for transcriptome read counting (McClureet *et al*., 2013; Tjaden, 2015). Based on the alignment data, non-coding transcripts are considered as sRNA. The reads per kilobase of transcript per million mapped reads (RPKM) values of experimental conditions (free-living and symbiotic) were compared with control to calculate the fold change. 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 provided in the Rockhopper output file.

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 the 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, the secondary structure of 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 of sRNAs were analyzed for the predicted sRNAs. BPROM was used to identify the binding sites of σ70 (Salamov and Solovyev 2011) and Arnold to identify *rho*-independent terminators (Naville *et a*l., 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 analyzed through the GO regulatory network by using the comparative GO web server (Fruzangohar et al., 2015).

Preparation Of Plant Material And Root Nodulation

Bradyrhizobium japonicum cultures were grown to stationary phase in YEMA medium (Werner et al., 1975) at 25°C on a rotary shaker at 140 rpm. *Glycine max* seeds were surface sterilised, germinated and then infected with *Bradyrhizobium japonicum* suspension for 2 hrs. For control, the same protocol was followed with plain YEMA medium. 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, 60°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.

Results

Genome-wide screening of sigma54-regulated sRNAs

Prediction of non-coding sRNAs from the various nitrogen-fixing *Rhizobium* strains were 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 strain of *Rhizobium* is graphically represented in Fig. 1a and the details of the predicted sRNAs are given in Table 1. Among the six *Rhizobium* strains, higher numbers of sRNAs are predicted from the genome of *S. fredii* followed by *R. etli*, *R. trifoli*, *R. leguminosarum*, *and B. japonicum.* The genome size and the number of replicons are found to vary among the *Rhizobium* strains. sRNAs of all *Rhizobium* strains were equally transcribed from both positive and negative strands, whereas in the *R. leguminosarum* sRNAs are mostly transcribed from the negative strand (Fig. 1b). Maximum numbers of sRNAs are predicted in chromosomes for all the *Rhizobium* strains. To find out the novel putative sRNAs, predicted sRNAs were searched against Rfam database and BSRD database to eliminate the conserved homologs.

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
	Negative strand	Positive strand			
Bradyrhizobium japonicum USDA 110	39	45	84	4	-
Rhizobium etli CFN 42	120	120	240	46	1
Rhizobium leguminosarum 3841	144	109	253	48	1
Rhizobium tropici CIAT 899	112	108	220	53	-
Sinorhizobiium fredii NGR234	130	132	262	33	-

Length And Gc% Content Distribution Of Srnas

Generally, sRNAs are found to have length of 50 − 500bp. sRNAs that are < 50 and > 500bp 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. 1c). GC content of sRNAs ranges from 30–80%. Majority of them were found to have 50–60 % content (Fig. 1d).

Srna Conservation And Comparison Analysis

sRNAs are highly conserved in nature. The predicted sRNA candidates were compared to identify the conservation between the Rhizobium strains. It is inferred that 8 sRNAs are found to be conserved between *R. etli* and *R. leguminosarum*, 1 sRNA between *R. leguminosarum* and *R.tropoci* and 1 sRNA between *R. tropoci* and *S. fredii* and interestingly 8 sRNA candidates of *S. fredii* conserved with the sRNAs of *S. meliloti* (unpublished data). In order to find out and identify the novel putative sRNAs, the sRNAs predicted from the genome of the Rhizobium strains were compared with previous reported sRNAs from the literature. A total of 6 sRNA candidates of *R. etli* are conserved with the 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 validated to find out whether they are truely sRNAs or any other regulatory RNAs. In order to validate and functionally characterize the non-coding RNAs, the predicted sRNAs were searched against the most comprehensive database like Rfam and BSRD repositories (Table 1). The identified conserved homologs were eliminated and other novel sRNAs were taken for further analysis.

Transcriptome-based Srnas Prediction

The high quality RNA sequence reads of the bacterium under free-living and symbiotic conditions specific of *B. japonicum* (SI 4) were aligned to respective genome using Rockhopper. After 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 > 500bp were removed and the intergenic sRNAs having a length of 50-500nt were taken for further analysis. A total of 1375 *trans*-encoded sRNAs are predicted. The lengths of the predicted sRNAs are found to vary in length between 50 and 400 nt and most of the sRNA candidates have the length of 50 to 100 nt in length. GC content of the sRNAs ranged from 31–80% and most of the sRNAs ranged between 51 and 60%. Predicted sRNAs were searched against Rfam database and BSRD databases to eliminate the conserved homologs. Only one sRNA of *B. japonicum* was found to be conserved against predicted 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 indicate that 8 sRNAs candidates of *B. japonicum* are conserved and are compable 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 matrics of sigma factor 54-based from the genome of *B. japonicum* to relate their role in the nitrogen regulation. A total of 21 sRNAs of *B. japonicum* 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 are highly regulated (based on the expression level) were selected for further analysis. Since a single sRNA can regulate

multiple mRNA targets, in order to screen the sRNAs and its 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
Bf1	1927274-1927558	Transcriptional regulator fixK	-10.42	0.019	imidazole glycerol phosphate synthase subunit hisH, gamma-glutamyltranspeptidase ggt		**-76.10** kcal/mol
Bf2	2174239-2174561	nitrogen fixation protein nifQ	-13.41	0.003	preprotein translocase subunit secA, lipoyltransferase lipB		**-41.50** kcal/mol
Bf3	2226897-2226997	nitrogenase molybdenum-iron protein subunit alpha nifD	-12.91	0.004	phosphatidylglycerophosphate synthase, cobalamin synthesis protein cobW, beta-ketoadipyl CoA thiolase pcaF		**-39.70** kcal/mol
Bf4	2066092-2066221	ferredoxin fdx heme exporter protein C ccmC, ABC transporter molybdenum-binding protein modA	-8.55 -18.18	0.045 0.000	glycogen debranching protein glgX, thioredoxin reductase trxB, succinate dehydrogenase cytochrome sdhC, histidinol-phosphate aminotransferase hisC, glyceraldehyde-3-phosphate dehydrogenase gapA,		**-31.70** kcal/mol
Bf5	4506199-4506295	cbb3 oxidase subunit III - fixP, cytochrome C oxidase subunitIII ccoG, coproporphyrinogen III oxidase-hemN, two component regulator ncrdV	-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		**-37.60** kcal/mol
Bf6	1741298-1741451	nitrate ABC transporter permease nrtB, cytochrome P-450 CYP117	-9.35 -12.99	0.032 0.004	glutaminyl-tRNA synthetase glnS, components of type IV pilus pilin subunit cupA, flagellar basal-body rod protein flgB		**-37.80** kcal/mol
Bf7	1214553-1214648	nitrogenase molybdenum-cofactor biosynthesis protein nifN, nodulation protein nodV, cytochrome oxidase assembly factor coxI5	-10.5 -10.62 -10.73	0.018 0.017 0.016	gglgB, co-chaperonin GroES		**-32.20** kcal/mol
Bf8	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
Bf9	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
Bf10	157905-158001	N-acetylglucosaminyltransferase nodC, molecular chaperone groEL	-16.86 -10.61	0.000 0.017	glycerol kinase glpK , pyridoxine 5'-phosphate synthase pdxJ, nicotinate phosphoribosyltransferase pncB, phosphoserine phosphatase serB		**-48.90** 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 directly influence the Rhizobium nitrogen fixation/regulation post-transcriptionally; putative targets of the sRNAs were predicted in all the Rhizobium strains with TargetRNA2 tool. Target mRNAs were predicted for both genome- and transcriptome-derived sRNAs. Target prediction revealed that 30 sRNAs of *Sinorhizobium 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* strains employed in the present study were mostly complemented with the binding site of various nodulation genes like *nodI*, *nodD2*, *nodC*, *nodF*, *nodQ1 nodS*, *nopB*, *nopC*, *nopT*, *nopX*, *nopM*, *nopP*, *nolO*, *noeI*, *fdxB*, and *fdxN* and nitrogen fixation genes like *nifA*, *nifB*, *nifE*, *nifD1*, *nifD2*, *nifH*, *nifK*, *nifN*, *nifS*, *nifT*, *nifQ*, *nifZ*, *fixA*, *fixB*, *fixC*, *fixF*, *fixG*, *fixX*, *fixS*, and *fixU* and several cytochrome proteins. All these mRNA targets exhibit significant complementarity with the sRNAs and show significant P-value (< 0.05). In *B. japonicum*, nearly 100 sRNAs 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 target interaction energy and high scores, sRNAs which targets more nitrogen fixation genes of *B. japonicum* (Table 2) were taken for further analysis and were experimentally validated.

Promoter, Terminator, Secondary Structure Prediction

Promoter and *rho*-independent terminator sequences were predicted for the identified putative novel sRNAs (Table 3). Secondary structure was predicted for the selected sRNAs using RNAfold server. The predicted minimum free energy for the majority of the sRNAs ranges from − 20 to -70 kcal/mol. The color of the structure represents the base-pairing probabilities (Table 2, SI 5).

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

sRNA	sRNA cordinates	-10 sequence	-35 sequence	Terminators	Upstream gene	Downstrean
BJ1	1927274–1927558	CTTTATAGT	TTGCGG	TGACAGAGACCTTGCGCGGCTTCTCGCGCGACTTTTGGAATGGA	outer membrane protein	hypothetica
BJ2	2174239–2174361	GGTCATTCT	TCGATG	-	hypothetical protein	hypothetica
BJ3	2226897–2226997	GTGTATACT	GTGCCA	-	phenolhydroxylase-like protein	hypothetica
BJ4	2066092–2066221	GGTTAGCAT CCCCATAA	CTCAGT CCCATAAC	-	hypothetical protein	transposase
BJ5	4506199–4506295	CCGGATTCT	GTGAAG	-	alanine racemase	replicative D helicase
BJ6	1741298–1741451	GATTAGAGT	TTGCCC	-	hypothetical protein	site-specific integrase/re
BJ7	1214552–1214648	ATCCAGAGT	TTGCCA	-	Hsp33-like chaperonin	hypothetica
BJ8	5185970–5186073	TGTTAGACT	TTCGCA	-	cysteine synthase	queuine tRN ribosyltrans
BJ9	7340538–7340630	-	-	-	hypothetical protein	hypothetica
BJ10	157905–158001	CCTTAAGCT	TTGCGA	-	ATP-dependent helicase	hypothetica

Experimental Validation

From the combined genome and transcriptome data, one sRNA from the positive strand of sigma 54-regulated sRNAs, nine sRNAs from the transcriptome (7 and 2 sRNAs from negative and positive strand, respectively) of *B. japonicum* were selected. Among the 10 sRNAs, 8 sRNAs (BJ1-BJ8) have shown amplifications in semi-quantitative PCR (Fig. 2).

Semi-quantitative Pcr Analysis

Semi-quantitative PCR was performed for 8 sRNA and the primers used to amplify the sRNAs are listed in the SI 6. 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. 2).

Functional Categorization Of Srna Target Genes

In order to study the role of target mRNAs of sRNAs, select set of target mRNAs were 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. 3). In GO analysis, the target mRNAs were enriched into 3 categories, *viz*., biological process, molecular functions and cellular components. The targets are categorized under biological process including the genes involved in cellular and metabolic process and localization, molecular functions, binding, catalytic activity and transporter activity (Fig. 4).

Go Regulatory Network

GO regulatory network analysis were constructed for the target mRNAs of free-living and symbiotic association conditions. The target mRNAs of sRNAs identified under the free-living condition is shown in Fig. 5. 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, protophyrinogen 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. 6) governed by ccmc (GO ID: 17004) which showed interaction with nitrogen fixation, aerobic respiration, transmembrane transport and nodulation.

Discussion

sRNAs play crucial role and they regulate diverse cellular process in all the three kingdoms. There are several bio-computational techniques including genome-wide and transcriptome based methods which have been extensively employed in order to identify the existence of sRNAs in both gram-positive and gram-negative bacteria. However, there are only few reports available related to the conditional specific sRNAs and their regulatory role exerted on the target mRNAs. In this present work, we have identified several novel sRNAs by genome-wide and transcriptome based methods and then made an attempt to relate the role of sRNAs in the regulation of nitrogen metabolism during the free-living and symbiotic conditions. A total of 1,059 sigma factor 54-regulated sRNAs were identified from 5 *Rhizobium* strains 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 5 strains of *Rhizobium*, higher number sRNAs were predicted for *S. fredii* and *R. leguminosarum*. The length of the sRNAs varied between 50 and 500nt, most of the sRNAs were 50-200nt in length. The conserved homologs of other regulatory RNAs were identified by the batch search against Rfam and BSRD databases. sRNAs were searched separately for all the 5 strains of *Rhizobium*. Among the 1059 sRNAs, 184 conserved homologs were found in Rfam and 2 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. sRNAs which were compared for their molecular and structural features with the previously reported sRNAs available from the literature. Among the 253 sRNAs of *R. etli*, 6 sRNAs were found to be conserved with the earlier reported sRNAs of *R. etli* (Vercruysse et al., 2010; López-Leal *et al*., 2015). Conservation analysis within the *Rhizobium* strains revealed that sRNAs of *R. leguminosarum* are highly conserved with the *R. etli*. Both these strains of the *Rhizobium* are known to have seven replicons, which include the symbiotic plasmid pRL10 and p42d, respectively (Table 1). Target prediction analysis revealed that sigma factor 54-regulated sRNAs can regulate multiple target genes involved in various nitrogen regulation process especially the nodulation and nitrogen assimilation. A total of 68 sRNAs 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, Q, Z*), *fix* (*A, B, C, F, G, X, S*,*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 role of sigma 54-regulated sRNAs 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 the *B. japonicum* and from this, 100 sRNAs were selected for target prediction analysis. The results showed that 43 sRNAs found to have targets on several nitrogen fixation genes. The identified sRNAs are classified into three categories, sRNAs expressed only under free-living condition, symbiotic condition and in both 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 has targets on several nodulation genes which include *nodC*, *nodY*, *nodJ*, *nodM*, *nodW*, *nodZ*, *nifD*, f*ixP*, *fixK*, *fixL nolB*, *nolV*, *fdx*, *hemN*, *groEL*, *groES*, *ccmC*, and several cytochrome proteins. sRNAs screened from the genome by using conditional sigma factor 54 were integrated with sRNAs derived from the free-living and symbiotic conditions transcriptome data to identify the differentially expressed sRNAs under conditions of regulation of nitrogen metabolism. Based on the comparative analysis, 21 sRNAs of *B. japonicum* are found to overlap between the genome and transcriptome data. 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. japonicum* were selected for performing semi-quantitative-PCR analysis (Table 2). Semi-quantitative PCR analysis confirmed the presence of 8 sRNAs (BJ1-BJ8). While analyzing the expression pattern of sRNAs, it was found that the sRNA BJ3 (Fig. 2a: lane 6) was highly expressed under symbiotic condition than in free-living condition, which targets the nitrogen fixation gene *nifD* (Nitrogenase molybdenum-iron protein subunit α), an essential enzyme required for biological nitrogen fixation. Another sRNA, i.e. BJ4 (Fig. 2a: lane 7) also has significant level of expression, which targets the gene *fdx* (Ferredoxin) electron transport protein which mediates electron transfer to dinitrogenase reductase for the bioproduction of ammonia. sRNA, BJ7 (Fig. 2a: lane 10) showed significantly higher level of expression in free-living bacteria as compared to the bacteria grown under symbiotic conditions which are found to target nitrogen fixation gene *nifN* (Nitrogenase molybdenum-iron co-factor biosynthesis protein) and *nolV* (nodulation protein).

The putative novel sRNA targets of *B. japonicum* were functionally categorised by 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. japonicum* using the biological process GO terms for both free-living and symbiotic conditions. Network analysis revealed that many target genes are mainly involved in 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. From the present study, we have identified and experimentally quantified the expression of 8 sRNAs (BJ1-BJ8) under free-living and symbiotic conditions for *B. japonicum*. A significant differential expression was noted for all the 8 sRNAs under free-living and symbiotic conditions. 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 by integrating genome-wide and transcriptome based methods.

Declarations

Acknowledgement

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

Funding

Not applicable.

Authors' information

[1]Department of Plant Morphology and Algology, School of Biological Sciences, Madurai Kamaraj University, Madurai-625 021, Tamil Nadu, India

[2]Department of Plant Sciences, School of Biological Sciences, Madurai Kamaraj University, Madurai-625 021, Tamil Nadu, India

Authors' contributions

Jebasingh Tennyson and Manoharan Kumariah conceived the idea. Kasthuri Rajendran planned and performed the experiments. Vikram Kumar and Ilamathi Raja created PWM matrix of improved sRNA scanner.

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Code availability

Not applicable

Ethical approval

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

Competing interests

All authors declare that they have no competing interests.

Consent for publication

Not applicable.

References

1. 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
2. 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
3. Čuklina J, Hahn J, Imakaev M, Omasits U, Förstner K, U, 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
4. 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
5. Fruzangohar M, Ebrahimie E, Ogunniyi 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 PLoS ONE 10: e0125537. https://doi.org/10.1371/journal.pone.0125537
6. Fuli X, Wenlong Z, Xiao W, Jing Z, Baohai H, Zhengzheng Z, Youguo L (2017) A genome-wide prediction and identification of intergenic small RNAs by comparative analysis in *Mesorhizobium huakuii* 7653R. Frontiers in microbiology 8:1730. https://doi.org/10.3389/fmicb.2017.01730
7. 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. Proceedings of the National Academy of Sciences 103: 3834–3839. https://doi.org/10.1073/pnas.0508502103
8. 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
9. Hahn J, Tsoy OV, Thalmann S, Čuklina J, Gelfand MS, Evguenieva-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
10. 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
11. Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic acids research 31:3429–3431. https://doi.org/10.1093/nar/gkg599
12. Kery MB, Feldman M, Livny J, Tjaden B (2014) TargetRNA2: identifying targets of small regulatory RNAs in bacteria. Nucleic acids research 42:W124–W129. https://doi.org/10.1093/nar/gku317
13. 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
14. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration (2010) The sequence read archive. Nucleic acids research 39: D19-D21. https://doi.org/10.1093/nar/gkq1019
15. Li L, Huang D, Cheung MK, Nong W, Huang Q, Kwan HS (2013) BSRD: a repository for bacterial small regulatory RNA. Nucleic acids research 41:D233–D238. https://doi.org/10.1093/nar/gks1264
16. 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

17. 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 replicon contributions under heat and saline shock. BMC Genom 15:1–15. https://doi.org/10.1186/1471-2164-15-770
18. Madhugiri R, Pessi G, Voss B, Hahn J, Sharma CM, Reinhardt R, Evguenieva-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
19. Marchais A, Naville M, Bohn C, Bouloc P, Gautheret D (2009) Single-pass classification of all noncoding sequences in a bacterial genome using phylogenetic profiles. Genome Res, 19:1084–1092. http://www.genome.org/cgi/doi/10.1101/gr.089714.108
20. 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. http://www.genesdev.org/cgi/doi/10.1101/gad.231702
21. Ormeño-Orrillo E, Menna P, Almeida LGP, 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
22. 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
23. Rajendran K, Kumar V, Raja I, Kumariah 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/s41544-020-00054-1
24. Salamov VS, Solovyevand A (2011) Automatic annotation of microbial genomes and metagenomic sequences. Metagenomics and its applications in agriculture. Nova Science Publishers Hauppauge, NY, pp 61–78
25. Schlüter JP, Reinkensmeier J, Daschkey S, Evguenieva-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
26. Werner D, Wilcockson J, Zimmermann E (1975) Adsorption and selection of rhizobia with ion-exchange papers. Arch Microbiol 105:27–32. https://doi.org/10.1007/BF00447108
27. Wise AA, Liu Z, Binns AN (2006) Nucleic acid extraction from *Agrobacterium* strains. In *Agrobacterium protocols* (pp. 67–76). Humana Press. https://doi.org/10.1385/1-59745-130-4:67
28. Valverde C, Livny J, Schlüter JP, Reinkensmeier J, Becker A, Parisi G (2008) Prediction of *Sinorhizobium meliloti* sRNA genes and experimental detection in strain 2011. BMC Genom 9:416. https://doi.org/10.1186/1471-2164-9-416
29. Vercruysse M, Fauvart M, Cloots L, Engelen K, Thijs IM, Marchal K, 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

Figures

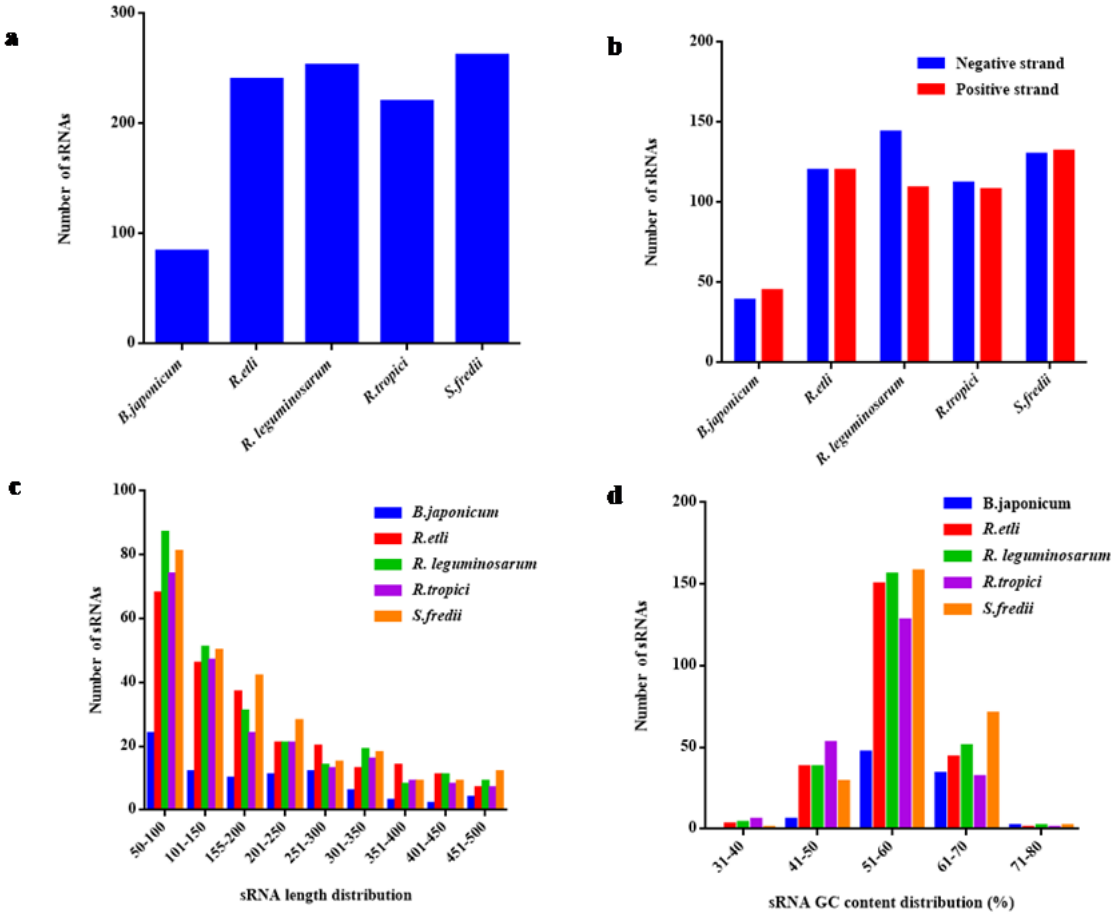

Figure 1

Genome-wide prediction of sRNAs using improved sRNA scanner tool. (a) Number of sRNAs predicted in various Rhizobium srains: Bradyrhizobium japonicum, Rhizobium etli, Rhizobium leguminosarum, Sinorhizobium fredii, Rhizobium tropici; (b) from positive and negative strands; (c) length distribution and (d) GC% content.

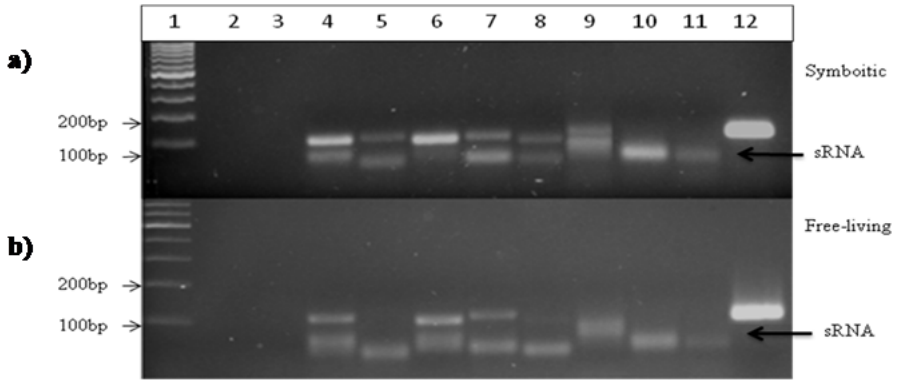

Figure 2

Semi-quantitative PCR analysis of sRNAs of Bradyrhizobium japonicum at symbiotic and free-living conditions. Gel picture shows amplification of sRNAs in the symbiotic (a) and free-living condition (b). Lane 1: 100 bp DNA marker, lane 2: non-template control, lane 3: no RT control, 4–11: sRNAs BJ1, BJ2, BJ3, BJ4, BJ5, BJ6, BJ7, and BJ8, respectively, lane 12: 16S rRNA.

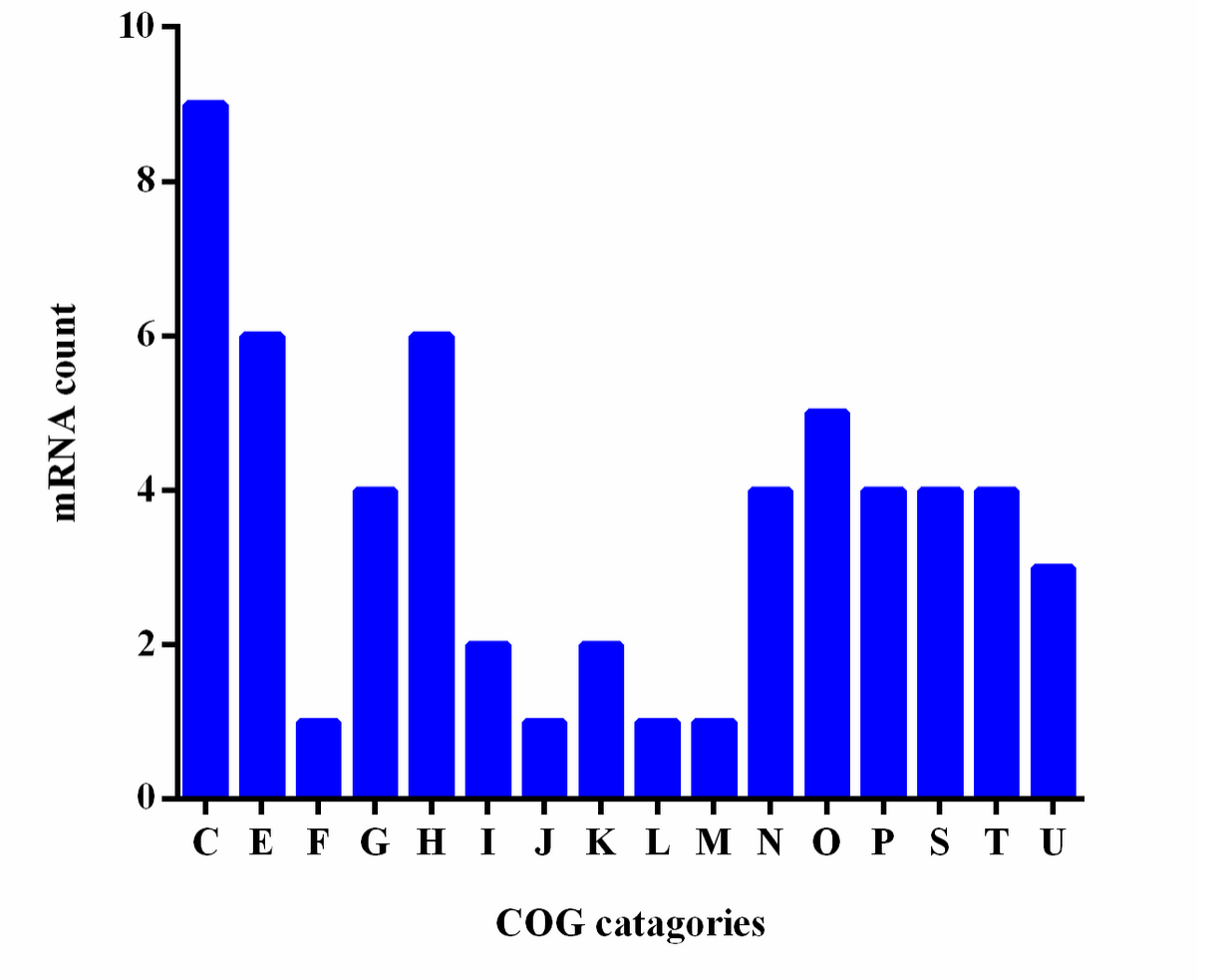

Figure 3

COG classifications of the target genes of B. japonicum. 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; S- function-unassigned conserved proteins; T- signal transduction; U- intracellular trafficking, secretion, and vesicular transport; and V- defence mechanisms.

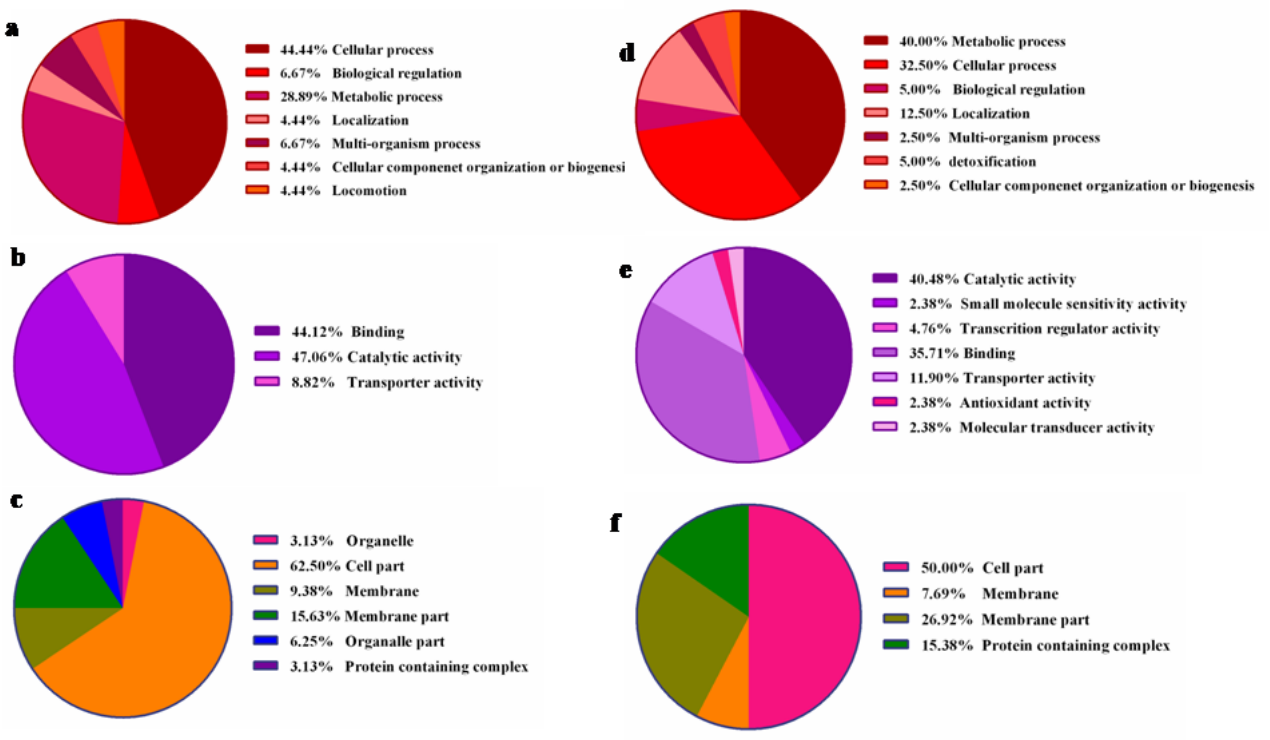

Figure 4

Gene Ontology analysis of predicted target genes for sRNAs of Bradyrhizobium japonicum. GO analysis of target genes that are predicted to be involved in (a) biological processes, (b) molecular functions and (c) cellular components of free-living specific sRNAs; (d) biological processes, (e) molecular functions and (f) cellular components of symbiotic specific sRNAs.

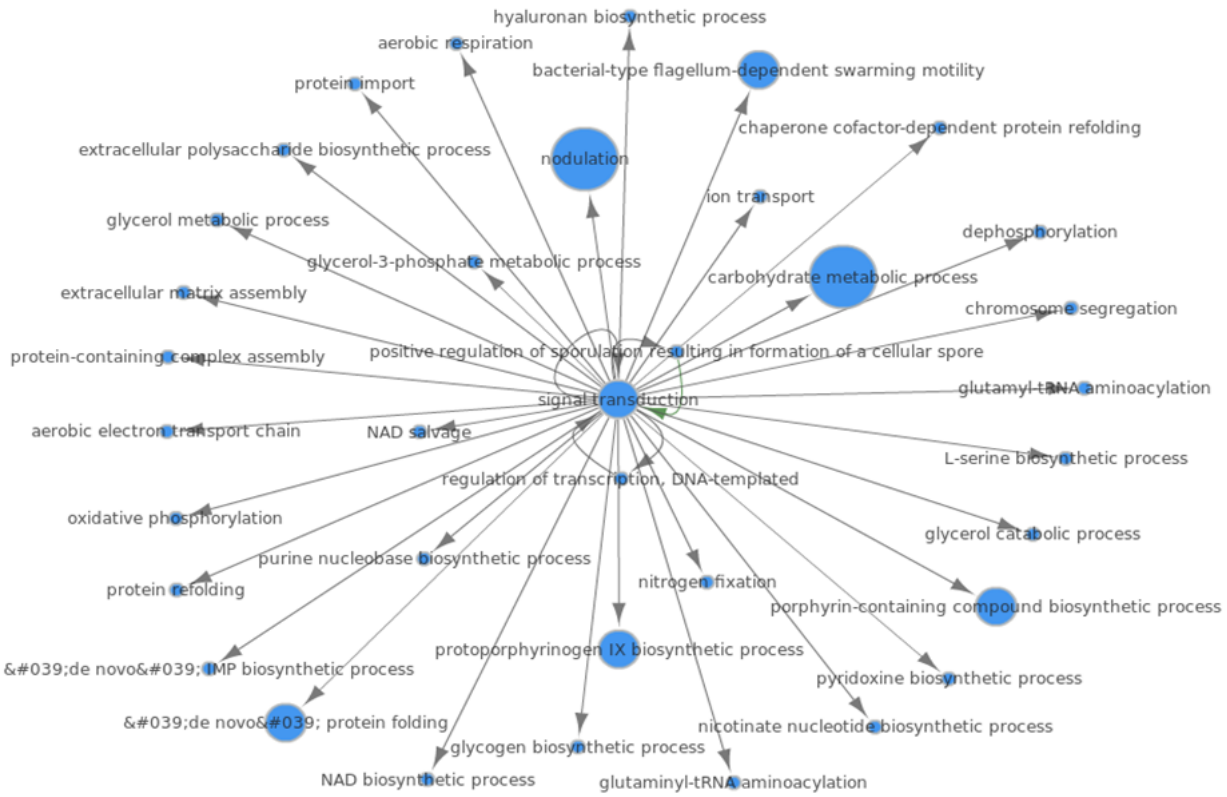

Figure 5

GO regulatory network based on the mRNA targets of sRNAs predicted from free-living condition.

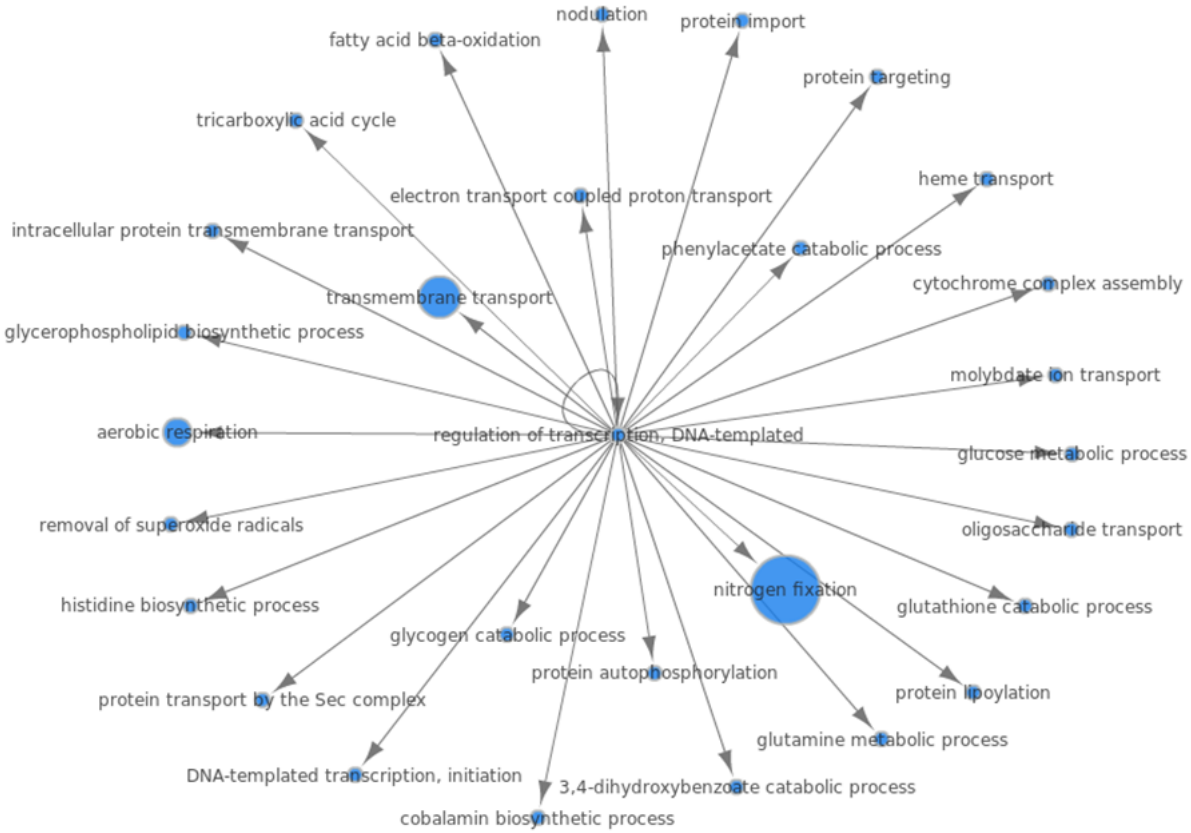

Figure 6

GO regulatory network based on the mRNA target of sRNAs predicted from symbiotic condition.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfiles.doc