Prediction of Non-coding RNA and Their Targets in *Spirulina platensis* Genome

Tanawut Srisuk[1,2], Natapol Pornputtapong[2], Supapon Cheevadhanarak[2,3], and Chinae Thammarongtham[4]

[1] Bioinformatics and Systems Biology Program, King Mongkut's University of Technology Thonburi, Bangkok, Thailand
[2] Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bangkhuntien, Bangkok, 10150, Thailand
[3] School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bangkhuntien, Bangkok, 10150, Thailand
[4] Biochemical Engineering and Pilot Plant Research and Development Unit, National Center for Genetic Engineering and Biotechnology at King Mongkut's University of Technology Thonburi, Bangkhuntien, Bangkok, 10150, Thailand
tanawut.cool@gmail.com

Abstract. Non-coding RNAs (ncRNAs), transcripts that have function without being translated to protein, have a number of roles in the cell including important regulatory roles. Efforts to identify the whole set of ncRNAs and then to elucidate their functions would gain better biological understanding. Although ncRNA is another type of genome constituent, most of the genes for ncRNA are overlooked by standard genome annotation of genome sequencing projects. This also happens in *Spirulina platensis* genome sequencing project. It is because gene finding tools generally are able to identify only protein-coding genes but not non-protein-coding ones. In this study, *S. platensis* ncRNAs were detected by comparative genomics approach using computational tools, together with RNA secondary structure prediction. It was found that more than 100 predicted ncRNA loci matched with known ncRNAs for example cobalamin riboswitch, RNaseP, Signal Recognition Particle RNA, Group II intron RNA and Yfr1. It has been reported that Yfr1 has been found in most cyanobacterial genomes sequenced. The result showed that more than 70 putative loci were similar to Group II intron RNAs. In addition, approximately 100 predicted ncRNA loci were not matched with any known ncRNAs. The predicted targets for some putative ncRNAs are also proposed.

Keywords: non-coding RNA prediction, non-coding RNA target prediction.

1 Introduction

Besides protein-coding genes, the genes coded for these RNAs have also been recognized as genome constituents since a large fraction of the transcriptome consists of non-protein-coding RNAs [1]. They are involved in many biological processes such

J.H. Chan, Y.-S. Ong, and S.-B. Cho (Eds.): CSBio 2010, CCIS 115, pp. 106–117, 2010.

as transcriptional regulation, chromosome replication, RNA stability and translational regulation, and even proteins stability and translocation [2]. Some of them act as catalytic molecules. Consequently, efforts to identify the whole set of ncRNAs and then to elucidate their functions for better biological understanding are more and more prominent. Availabilities of complete genome sequence data have made it possible to computationally identify ncRNAs in sequenced genomes using bioinformatics approach. Although experimental verification is necessary, it has been recognized that computational identification may be an effective approach to first detect ncRNAs candidates, including novel ncRNA species, followed by biochemical assessment. RNA classification is a step involving in RNA gene annotation. Since ncRNAs are conserved in structures rather than their primary sequences, using a simple homology search is not efficient enough in RNA classification. Comparing structures of unknown ncRNAs, along with similarity at primary sequence level, to known ncRNAs is more powerful. To perform this strategy, a statistical model so-called covariance-model (CM) is a method of choice since CMs are integrated with both primary sequence and secondary structure information of known ncRNAs. CMs will be trained and then be used to search for the region in given sequences which are similar to the feature it has learned [3]. Infernal package [4] is a suite of tools for these tasks ranging from CM construction to searching for the region which is similar to CM in given sequences.

ncRNAs are a heterogeneous group of functional RNAs and showing up in all kingdoms, including prokaryotic domain. In bacteria, ncRNAs mostly function as coordinators of adaptation processes in response to environmental changes, integrating environmental signals and controlling target gene expression. For cyanobacteria, including *Spirulina,* regulatory circuits involving ncRNAs can be expected as well. Identification of ncRNAs will facilitate investigation of another level of controls in *Spirulina,* in addition to regulation by protein mediators, and providing new insights into growth and adaptation to stresses of this cyanobacterium. *Spirulina* genome sequencing project has been established by Thai research consortium. Availabilities of such cyanobacterial genome information together with bioinformatics approach make it possible to computationally identify putative ncRNA genes. In this study, the *Spirulina* genome were computationally analyzed and screened for ncRNAs. Their putative targets were also predicted.

2 Materials and Methods

Sequences of intergenic regions (IG) of *S. platensis* were extracted from genome sequence data and compare with sequences of 34 related species genomes, including *Nostoc punctiforme* ATCC 29133 [5], *Prochlorococcus marinus* subsp. marinus str. SS120 [6], *Synechococcus* sp. JA-3-3Ab, *S.* sp. JA-2-3B'a(2-13), *S.* sp. WH 8102, *S.* sp. CC9902, *S.* sp. CC9605, *S.* sp. CC9311, *S.* sp. WH 7803, *S.* sp. RCC307, *S.* sp. PCC 7002, *S. elongatus* PCC 6301, *S. elongatus* PCC 7942, *Synechocystis* sp. PCC 6803, *Chlorobium tepidum* TLS, *N.* sp., *Thermosynechococcus elongatus* BP-1,

P. marinus str. MIT 9211, *P. marinus* str. MIT 9215, *P. marinus* str. MIT 9301, *P. marinus* str. MIT 9303, *P. marinus* str. MIT 9312, *P. marinus* str. MIT 9313, *P. marinus* str. MIT 9515, *P. marinus* str. NATL1A, *P. marinus* str. NATL2A, *P. marinus* str. AS9601, *P. marinus* subsp. pastoris str. CCMP1986, *Gloeobacter violaceus* PCC 7421, *Anabaena variabilis* ATCC 29413, *Trichodesmium erythraeum* IMS101, *Acaryochloris marina* MBIC11017, *Microcystis aeruginosa* NIES-843, and *Cyanothece* sp. ATCC 51142 [7]. Multiple sequence alignments between *S. platensis* IGs and corresponding regions from other genomes were constructed. These alignments were scores, by ncRNA prediction tool, for possibility of being ncRNAs. This procedure was also applied to *Arthrospira maxima* CS-328 and *Lyngbya* sp. PCC 8106 [8]. The methodology of this work is outlined in the work flow shown on Fig. 1.

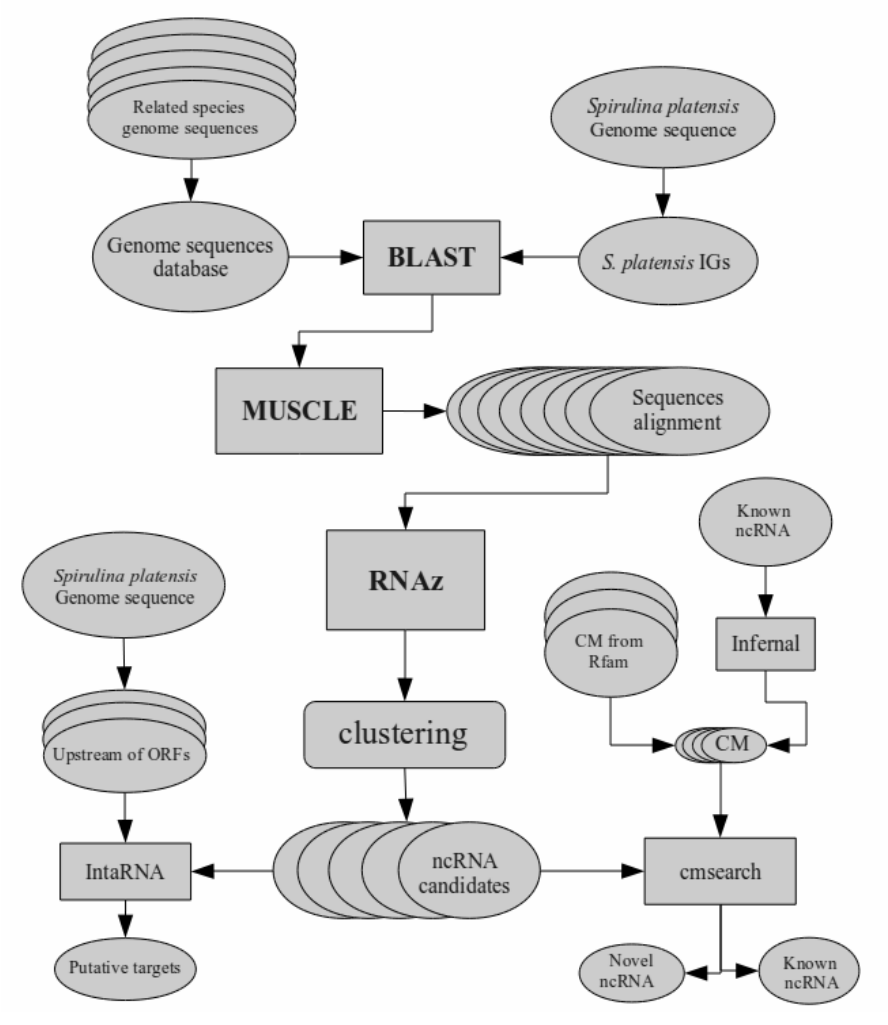

Fig. 1. Work flow of the proposed methodology

2.1 BLAST Based Alignment Generating

To prepare multiple sequence alignment as input for ncRNA prediction program, IG sequences (include 50 base from preceding and consequence ORFs) were searched against cyanobacteria genome database for similar regions in other species using ncbi BLASTN [9] with relaxed parameter (-q -1 -r 1 -F F -e 1e-10) because ncRNAs tend to conserved in their structures rather than primary sequences. Similar regions of each *S. platensis* IGs were grouped by their lengths and positions on *S. platensis* genome then aligned with each IG by MUSCLE [10] with increased gap penalty to avoid gap insertion in distantly related sequences.

2.2 RNAz Scoring and Result Clustering

RNAz [11] package, an ncRNA prediction tool, was used to calculate probability of being ncRNA for each alignment input. Before scoring, each alignment was pre-processed, by scripts in RNAz package, into appropriate form. After each input was scored, results with ncRNA probability score (RNAz P-value) > 0.9 (the closer to 1 the more likely to be ncRNA) were kept. To recover some part of long ncRNA which may lost in prediction due to sliding windows of 40 nt (nucleotides) in preprocessing step, they were also joined into a single ncRNA locus if their locations were over-lapped or located within 40 nt from each other.

2.3 ncRNA Candidate Classification

In order to classify which individual ncRNAs belong to known ncRNAs, each candidate was compared with covariance model (CM) from Rfam database [12] and several our-owned constructed CMs of known ncRNA using cmsearch from INFERNAL package. The resulting matches with E-value lower than 1e-5 were reported as true known ncRNA homologues.

2.4 ncRNA Targets Prediction

To predict targets of each ncRNA candidate, in term of translational regulation, an interaction score between each ncRNA candidate and 5' upstream region (250 nt before start codon to 150 nt after start codon) of predicted ORFs were calculated by IntaRNA [13] then top rank scores from each ncRNA candidate were reported.

3 Results and Discussion

3.1 Predicted ncRNA in *S. Platensis* Genome

A set of 3,976 IGs were searched by BLAST against 34 related species genomes then 10,639 of BLAST hits were selected and grouped before aligning into 2003 alignment. Using RNAz, 334 putative ncRNA loci were computationally identified from *S. platensis* genome. The lengths of these loci are varied between 52 and 1482 nt. Some of particular loci predicted may be only partial, not full length, ncRNAs. Some of them can be merged together into a single larger locus for example RNaseP RNAs and Group II intron RNAs. Based on the method used in this work, the transcription direction of the predicted loci were not determined. Localization of promoter and

terminator of each locus will reveal transcription direction and also increase reliability of putative ncRNA candidates. However experimental identification of promoters is relatively complicated and computational prediction is still challenging. Notably this work primarily focused on ncRNAs located on IGs. Antisense transcripts were not detected by the method used. Such anti-sense ncRNAs can be predicted by CM searching or other methods. The statistical value of the predicted loci is RNAz P-value which is between 0 and 1. The default P-value for a locus to be reported as RNA is 0.5 or higher. The higher P-value the more significant the locus is, inferring a plausible secondary structure forming locus. For all 334 predicted loci, their P-values are 0.9 or above. Therefore they are considered to be putative ncRNA candidates.

3.2 Classification of ncRNA Candidates

By CM searching, 129 ncRNA candidates were matched with CM obtained from Rfam and classified into 12 known RNA families including of 1 5.8S rRNA, 2 Cobalamin riboswitches, 1 CRISPR-DR57, 79 group II introns, 1 mir-598, 2 PK-G12 rRNAs (23S rRNA pseudoknot), 1 bacterial RNaseP type A, 1 bacterial signal recognition particle (SRP), 3 SSU-5s, 37 tRNAs, 1 Yfr1 and 1 Yfr2b. Several long length RNA loci matched to more than 1 family. Interestingly, RNA candidates in a large group containing 79 loci are classified as Group II intron RNA.

3.2.1 Yfr1

Yfr1 is an ncRNA which exists in many cyanobacteria and locates between *trxA* and *guaB*. It is approximately 50-70 nt in length. It has been found that Yfr1 consists of approximately 10-nt conserved unpaired region (5'-ACUCCUCACAC-3') between two stem loop structures [14]. It has been reported that Yfr1 is required for growth under stress condition and target to SbtA mRNA which plays an important role for sodium-dependent bicarbonate transport. In addition, estimated abundance of Yfr1 is about 18,000 molecules per cell [15]. Furthermore, Yfr1 inhibit translation of two outer membrane protein genes by direct base pairing mechanism at ribosome binding site [16]. In *S. platensis* and *A. maxima* genomes, the putative Yfr1 was predicted as a locus resided between thioredoxin and IMP dehydrogenase gene in the genome. This agrees with the report when the Yfr1 was first identified. Fig. 2 represents cmsearch result which indicates matched ncRNA candidate of Yfr1 CM.

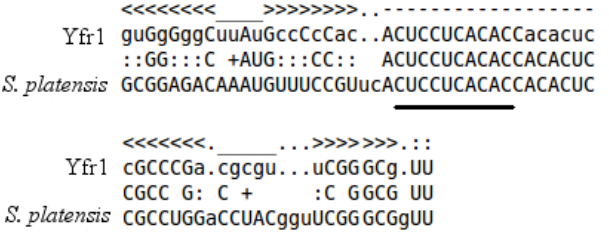

Fig. 2. Alignment from "cmsearch" represents matching between Yfr1 CM from Rfam and ncRNA candidate in *S. platensis*. Middle line between Yfr1 and ncRNA represent matching, either in primary sequence or secondary structure as described in [17]. Conserved region according to previous report is indicated by underlining.

3.2.2 Group-II intron

Bacterial group II intron RNAs are mobile retro-element and catalytic unit which is spliced by lariat intermediate mechanism. These introns transferred to target site by ribonucleprotein complexes assembled from intron-encoded proteins and excised intron RNA lariat. Group II intron RNAs are a large biomolecule consisting of highly structured RNAs with six distinct double-helical domains and reverse transcriptase ORF in fourth domain [18]. There are 79 and 89 ncRNA candidates, which are identified as group II intron RNA, in *S. platensis* and *A. maxima*, respectively. Many loci of group II intron RNAs in *S. platensis* are located near reverse transcriptase or transposase coding sequences. Since the CM for group II intron RNAs is constructed from partial group II intron RNA sequences which is conserved within this RNA family, the hits returned from CM searching are reported as partial sequences matched with the CM. The full length of Group II intron RNA can be traced by investigated neighboring loci predicted as ncRNAs and shared similarity with other regions of group II intron RNA. Fig. 3 represents alignment of group II intron CM to match locus on *S. platensis* genome.

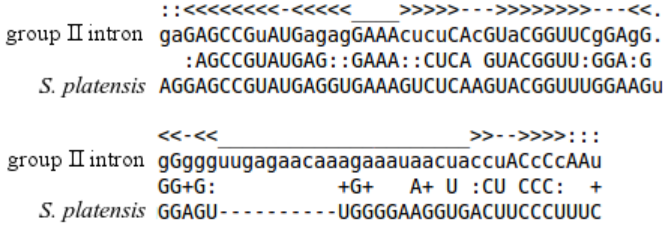

Fig. 3. Alignment from"cmsearch" result represents matching between an ncRNA candidate in *S. platensis* and CM of a group II intron partial structure from Rfam.

3.2.3 Cobalamin riboswitch

Cobalamin riboswitch is a conserved regulatory element located at 5' untranslated region (UTR) of vitamin B_{12} related genes [19]. In *S. platensis* and *A. maxima* genomes, there are three loci of putative cobalamin riboswitchs. A locus of *S. platensis* cobalamin riboswitch locates on 5' UTR of 5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase (MetE) homologous gene. In *Mycobacterium tuberculosis*, cobalamin riboswitch resided on 5' UTR of MetE gene and involved in transcriptional control of the gene [20]. Another predicted locus in *S. platensis* genome matched with cobalamin riboswitch located upstream of cobalamin biosynthesis protein CobW gene [21]. In addition, a possible cobalamin riboswitch was predicted in *S. platensis* genome as it matched with CM of the one reported to be located on 5' UTR of cobalamin biosynthesis protein CbiM gene [22]. However, this locus matched to the CM with E-value higher than 1e-5 which was lower than the cut-off for this work. Fig. 4 represents cmsearch result, indicated how the cobalamin riboswitch CM matched to ncRNA candidate.

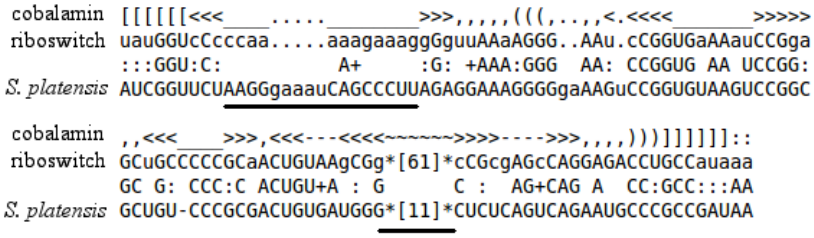

Fig. 4. Alignment from cmsearch result represents matching between Cobalamin riboswitch CM and ncRNA candidate in S. platensis, miss-match regions are indicated by underlining.

3.2.4 Signal Recognition Particle RNA (SRP)

SRP is a ribonuclecoprotein complex, composed of an RNA component and one or more protein domain(s). SRP is required for secretion or integration to membrane for proteins those consisted of signal peptide [23]. Single copy of SRP RNA was classified in each genome of *S. platensis*, *A. maxima, Nostco* sp. PCC 7120, and *Lyngbya* sp. PCC 8160. In *S. platensis* and *A. maxima*, SRP RNA candidate are adjacent to downstream of YCII-related gene and these loci from both species are identical. According to genome annotation, homologues of signal recognition particle protein subunit Ffh and signal recognition particle-docking protein FtsY which are involved in proteins translocation machinery exist in *S. platensis* genome. Fig. 5 represents "cmsearch" result, indicates how ncRNA candidate matches to SRP CM.

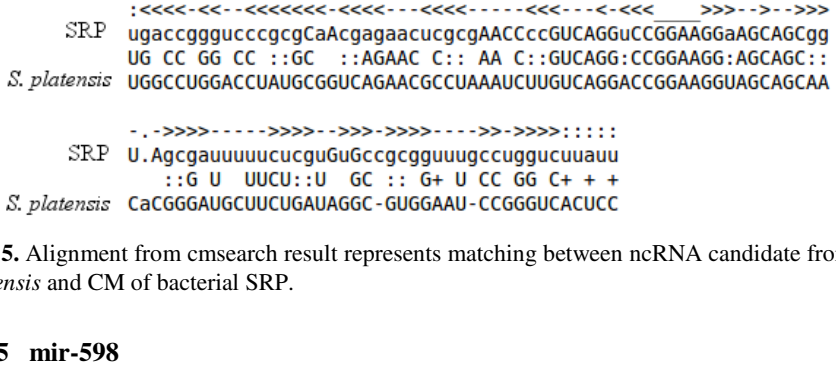

Fig. 5. Alignment from cmsearch result represents matching between ncRNA candidate from *S. platensis* and CM of bacterial SRP.

3.2.5 mir-598

mir-598 is a type of microRNA which is widely observed in animals and plants. Characteristic of RNAs in this family is approximately 20-nt long stretch single stranded RNA which is processed from approximately 100 nt-stem loop RNA by proteins complex. These microRNAs regulate translation by base-pairing mechanism to mRNA after incorporated with an argonaute protein complex [24]. In *S. platensis*, mir-598 candidate locates between predicted hydrogenase accessory protein HypB gene and predicted hydrogenase nickel insertion protein HypA gene. Although microRNAs have been reported only in mammalian and plants but argonaute protein homologues have been reported in several strains of cyanobacteria [25]. This may be considered as an RNA interference system in cyanobacteria. Fig. 6 shows "cmsearch" results which indicate how an ncRNA candidate is matched to mir-598 CM.

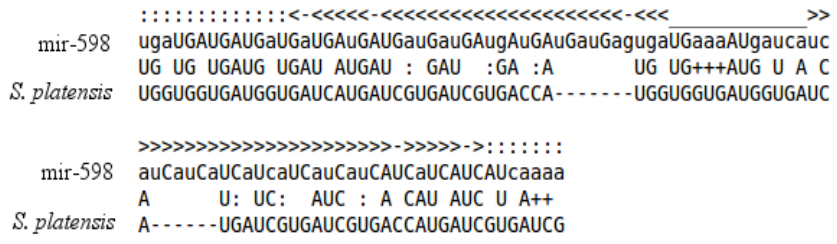

Fig. 6. Alignment from cmsearch result represents matching between ncRNA candidate in *S. platensis* and mir-598 CM.

3.2.6 Bacterial Rnase P Type A

A predicted locus, located on the upstream region of predicted adenylate/guanylate cyclase coding gene, matched with the CM of RNaseP constructed from conserved region of RNase P RNA. Bacterial endoribonuclease RNase P which composed of catalytic RNA domain and protein subunits is essential to generate mature 5' region of tRNA [26]. RNase P cleaves 5' element of pre-tRNA through hydrolysis at specific phosphodiester bond. In addition, other RNase P substrates have been reported also which are polycistronic mRNA [27], tmRNA [28], bacteriophage C4 antisense RNA precursor [29] and SRP RNA [30]. Fig. 7 represents "cmsearch" result which shows how ncRNA candidate is matched to CM of Bacterial RNaseP type A.

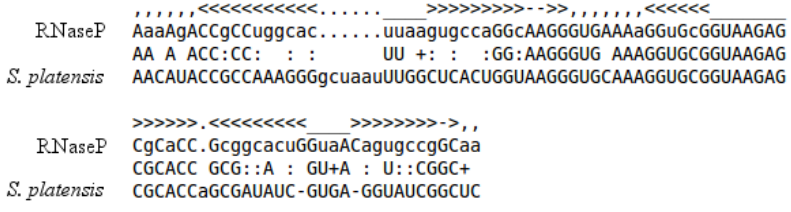

Fig. 7. Alignment from cmsearch result represents matching between ncRNA candidate in *S. platensis* and CM of bacterial RNaseP type A.

3.3 Non-coding RNA Target Prediction

After excluding putative tRNA loci, interaction scores between each ncRNA candidate and all protein-coding sequences in *S. platensis* were calculated by IntaRNA. Candidates for ncRNA targets were extracted from upstream regions of predicted protein-coding genes of *S. platensis* (250 bases upstream of each start codon and 150 bases downstream of each start codon for each protein-coding sequence).

3.3.1 Yfr1 Predicted Targets

Prediction result for Yfr1 indicates that Yfr1 interacts with its target by conserved region (5'-UCACUCCUCACACCACACU-3') located between two stem loops. The conserved region forms base pairing to putative 5' UTR of predicted ORFs listed in Table 1. Many interactions were predicted to occur around -20 to -6 nt (refer to ORF start codon) and some were predicted to occur within ORF region. Yfr1 may regulate

Table 1. Top rank of predicted targets for Yfr1 homolog

gene name	target on mRNA	target on ncRNA	Interaction score
DegT/DnrJ/EryC1/StrS aminotransferase	-20 .. -6	39 -- 52	-16.84
Undecaprenyl-phosphate galactose phosphotransferase	-100 .. -89	41 -- 52	-15.35
phosphoenolpyruvate synthase	32 .. 43	38 -- 50	-15.35
glutamate racemase	42 .. 56	35 -- 50	-15.33
sulfotransferase	-80 .. -65	38 -- 52	-15.18
chromosome partitioning protein, ParB family	-226 .. -216	39 -- 50	-14.45
DNA-cytosine methyltransferase	73 .. 84	36 -- 48	-14.19
Amine oxidase	-117 .. -102	38 -- 52	-14
ATP-dependent metalloprotease FtsH	-178 .. -164	38 -- 52	-13.93
catalytic domain of components of various dehydrogenase complexes	-224 .. -209	38 -- 52	-13.7

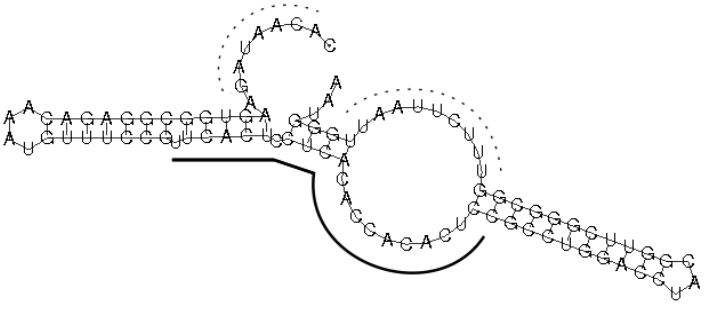

Fig. 8. Predicted structure of putative Yfr1 in *S. platensis* using RNAfold [32]. Thick line represents interaction region and dot line represents A-U rich regions which are Hfq binding motif.

translation of these genes since the predicted interaction regions are at putative ribosome binding sites. Furthermore, predicted Yfr1 structure at interaction site was closed by complementary base pairing from 5' and 3' tails which were adjacent to Hfq binding motif, indicated by A-U rich region and stem loop (Fig. 8) [31]. This suggests that Hfq may involve in Yfr1 target regulation by binding to the motifs to prevent an interaction site from pairing with those 5' and 3' tails.

3.3.2 mir-598 Homolog Predicted Targets

Predicted targets of mir-598 homologue are varied in term of functions and interaction sites while predicted interacting site on mir-598 homolog are likely to be at around the 1st nt to the 50th nt and the 1st nt to the 14th nt as shown in Table 2. Considering to predicting interactions which occur at a half of stem loop on ncRNA (Fig. 9)

which is similar to interaction of micro RNA regulation in higher organisms, this may be RNA interference in *S. platensis*.

Table 2. Top rank of predicted targets for mir-598 homolog

gene name	target on mRNA	target on ncRNA	Interaction score
putative transposase	-244 .. -197	1 -- 50	-24.24
Hemolysin-type calcium-binding region	-30 .. 24	3 -- 50	-22.09
4-hydroxyphenylpyruvate dioxygenase	96 .. 140	28 -- 71	-19.89
short-chain dehydrogenase/reductase SDR	-136 .. -121	1 -- 16	-19.42
SCP-like extracellular	-21 .. 22	4 -- 50	-19.40
glycosyl transferase family 2	-193 .. -132	1 -- 50	-19.37
Polypeptide-transport-associated domain protein ShlB-type	134 .. 147	1 -- 14	-17.95
ribose-phosphate pyrophosphokinase	-193 .. -180	3 -- 14	-17.47
anaerobic ribonucleoside-triphosphate reductase activating	-182 .. -133	3 -- 50	-17.28
restriction endonuclease	134 .. 149	1 -- 14	-17.21
homoserine kinase	-19 .. 34	1 -- 50	-17

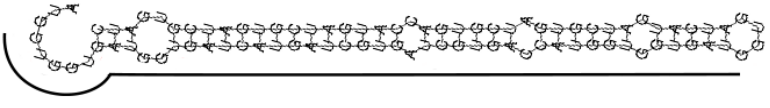

Fig. 9. Predicted structure of mir-598 homolog using RNAfold. Thick line represent interaction region which is predicted to interact with targets.

4 Conclusion and Further Works

This work provided a set of 334 putative ncRNAs including of 129 known ncRNAs and 205 unkown loci for verification and further analysis. In addition, a list of predicted targets for these ncRNAs was also acquired. To investigate reliability and role(s) of predicted ncRNAs in *S. platensis*, an integrating information from "cmsearch" result, ncRNA target prediction, location of transcription regulatory site around ncRNA loci and ncRNA-protein(s) interaction will be performed before analyzing by experimental approach.

Acknowledgments

Srisuk, T. would like to thank National Center for Genetic Engineering and Biotechnology, Thailand (BIOTEC), and King Mongkut's University of Technology Thonburi for the scholarship of Bioinformatics program.

References

1. Washietl, S., Hofacker, I.L., Lukasser, M., Hüttenhofer, A., Stadler, P.F.: Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. Nat. Biotechnol. 23, 1383–1389 (2005)
2. Storz, G.: An expanding universe of noncoding RNAs. Science 296(5571), 1260–1263 (2002)
3. Meyer, I.M.: A practical guide to the art of RNA gene prediction. Brief Bioinform. 8(6), 396–414 (2007)
4. Nawrocki, E.P., Kolbe, D.L., Eddy, S.R.: Infernal 1.0: Inference of RNA alignments. Bioinformatics 25(10), 1335–1337 (2009)
5. http://genome.jgi-psf.org/nospu/nospu.download.html
6. http://www.sb-roscoff.fr/Phyto/Genome_Cyanos/ProSS120/
7. ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria/
8. http://www.ncbi.nlm.nih.gov
9. Alteschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J.: Basic Local alignment search tool. J. Mol. Biol. 215, 403–410 (1990)
10. Edgar, R.C.: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5), 1792–1797 (2004)
11. Washietl, S., Hofacker, I.L., Stadler, P.F.: Fast and reliable prediction of noncoding RNAs. Proc. Natl. Acad. Sci. USA 102(7), 2454–2459 (2005)
12. Griffith-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., Bateman, A.: Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 33, 121–124 (2005)
13. Busch, A., Richer, A.S., Backofen, R.: IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics 24(24), 2849–2856 (2008)
14. Voß, B., Gierga, G., Axmann, I.M., Hess, W.R.: A motif-based search in bacterial genomes identifies the ortholog of the small RNA Yfr1 in all lineages of cyanobacteria. BMC Genomics 8, 375 (2007)
15. Nakamura, T., Naito, K., Yokota, N., Sugita, C., Sugita, M.: A Cyanobacterial Non-coding RNA, Yfr1, is required for Growth Under Multiple Stress Conditions. Plant Cell Physiol. 48(9), 1309–1318 (2007)
16. Richter, A.S., Schleberger, C., Backofen, R., Steglich, C.: Seed-based INTARNA prediction combined with GFP-reporter system identifies mRNA targets of the small RNA Yfr1. Bioinformatics 26(1), 1–5 (2010)
17. Eddy, S.R.: The Infernal user's guide, August 15 (2009), http://infernal.janelia.org/
18. Toro, N., Jimenez-Zurdo, J.I., Gracia-Rodriguez, F.M.: Bacterial group II introns: not just splicing. FEMS Microbiol. Rev. 31, 342–358 (2007)
19. Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., Gelfand, M.S.: Comparative genomics of the vitamin B_{12} metabolism and reglation in prokaryotes. J. Biol. Chem. 278(42), 41148–41159 (2003)
20. Warner, D.F., Savvi, S., Mizrahi, V., Dawes, S.S.: A riboswitch regulates expression of the Coenzyme B_{12}-Independent methionine synthase in *Mycobacterium tuberculosis*: Implications for differential methionine synthase function in strain H37Rv and CDC1551. J. Bacteriol. 189(9), 3655–3659 (2007)
21. Kazanov, M.D., Vitreschak, A.G., Gelfand, M.S.: Abundance and functional diversity of riboswitches in microbial communities. BMC Genomics 8, 347 (2007)

22. Rodinov, D.A., Hebbeln, P., Gelfand, M.S., Eitinger, T.: Comparative and functional genomics analysis of prokaryotic Nickel and Cobalt uptake transporters: Evidence for a novel group of ATP-binding cassette transporters. J. Bacteriol. 188(1), S1 (2006)
23. Nagai, K., Oubridge, C., Kuglstatter, A., Menichelli, E., Isel, C., Jovine, L.: Structure, function and evolution of the signal recognition particle. EMBO J. 22(14), 3479–3485 (2003)
24. Bartel, D.P.: MicroRNAs: target Recognition and Regulatory Functions. Cell 136(2), 215–233 (2009)
25. Makarova, K.S., Wolf, Y.I., van der Oost, J., Koonin, E.V.: Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol. Direct. 4, 29 (2009)
26. Frank, D., Pace, N.: Ribonuclease P: Unity and diversity in a tRNA processing ribozyme. Annu. Rev. Biochem. 67, 153–180 (1998)
27. Alifano, P., Rivellini, F., Piscitelli, C., Arraiano, C.M., Bruni, C.B., Carlomagno, M.S.: Ribonuclease E provides substrates for ribonuclease P-dependeent processing of a ploycistronic mRNA. Genes Dev. 8(24), 3021–3031 (1994)
28. Komine, Y., Kitabatake, M., Yokogawa, T., Nishikawa, K., Inokuchi, H.: A tRNA-like structure is present in 10Sa RNA, a small stable RNA from *Escherichia coli*. Proc. Natl. Acad. Sci. USA 91(20), 9223–9227 (1994)
29. Hartmann, R.K., Heinrich, J., Schlegl, J., Schuster, H.: Precursor of C4 antisense RNA of bacteriophages P1 and P7 is a substrate for RNase P of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 92(13), 5822–5826 (1995)
30. Peck-Miller, K.A., Altman, S.: Kinetics of the processing of the precursor to 4·5 S RNA, a naturally occurring substrate for RNase P from *Escherichia coli*. J. Mol. Biol. 221(1), 1–5 (1990)
31. Moll, I., Afonyushkin, T., Vytvytska, O., Kaberdin, V., And Blasi, U.: Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. RNA 9(11), 1308–1314 (2003)
32. Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, S., Tacker, M., Schuster, P.: Fast Folding and Comparison of RNA Secondary Structures. Monatshefte für Chemie. 125(2), 167–188 (1994)