Journal of Computational Intelligence in Bioinformatics
ISSN 0973-385X Volume 5, Number 2 (2012) pp. 67-74

http://www.ripublication.com/jcib.htm

Computational Analysis of Noncoding Leucine tRNA Genes in Pathogenic Strains of Gammaproteobacteria by Comparative Genomics

K.V. Rajesh[1*], P. Jawahar Babu[1], P. Swathi[1], A. Chaithanya Kumar[1], P. Sanysi Naidu[1] and G.V. Ravi[2]

[1]Department of Biotechnology, Bapatla Engineering College, Bapatla-522101, A.P., India
[2]Nireekshana Research Institute, Raj Mohalla, Near Shalimar Theatre, Hyderabad-500029, A.P., India
**Corresponding Author E-mail: rajeshbiosc@gmail.com*

Abstract

Non coding RNA (nc RNA) is any molecule that is not translated into protein. Non coding RNA includes Infrastructural, Prokaryotic and Eukaryotic RNA. The Transfer RNA one of the Infrastructural RNA is a small RNA chain (73-93 bp) that transfers a specific amino acid to a growing polypeptide chain at the ribosomal site of protein synthesis during translation. In this study, we analyzed Noncoding tRNA Genes *Escherichia coli O157:H7 Leu by* Comparative genomics. The *Escherichia coli O157:H7 Leu* compared with family of Gammaproteobacteria of pathogenic strains as *Salmonella enterica subsp. enterica serovar Typhi Ty2, Vibrio cholerae tRNA I operon, Aeromonas hydrophila sub sp hydrophila ATCC 7966* analyzed by using the BLAST. The selected tRNA gene sequences were compared by using Clustal W and T coffee. Often the conserved sequence was used for PETscan search and identified certain patterns which are highly conserved for certain organisms. The identified tRNAs of selected organisms from tRNAscan-SE used for the prediction of secondary structure through the Mfold server and observed that the conserved sequence fall into the D-arm region of tRNA is act as a recognition site for aminoacyl-tRNA synthetase. Therefore in future, these nc tRNA genes might useful in Antisense technology for therapeutic purposes and as a excellent research tool.

Keywords: Non coding RNA, Infrastructural RNA, Comparative genomics, Secondary structure, Recognition site, Anti sense technology.

Introduction

In addition to protein-coding genes that produce messenger RNAs, there are also genes for non-coding RNAs (ncRNAs) that function directly as structural, regulatory or even catalytic molecules in cells[1,13].Historically, both experimental gene discovery efforts and computational genome sequence annotation have been biased against non-coding genes, since ncRNAs lack open-reading frames (ORFs) and often lack other features such as poly (A) tails. More recently, a number of computational and experimental efforts have been undertaken to systematically identify new ncRNA genes, especially in model systems amenable to genetic and biochemical analysis[5].ThencRNAsare including Infrastructural, Prokaryotic and Eukaryotic RNA. Transfer RNAis one of the Infrastructural RNA, a small RNA chain (73-93 bp) that transfers a specific amino acid to a growing polypeptide chain at the ribosomal site of protein synthesis during translation[3,4]. It has a terminal site for amino acid attachment. This covalent linkage is catalyzed by an amino acyl tRNA synthetase. It also contains a three base region called anticodon that can base pair to the 3 base coding regions on mRNA.

A recent computational screen estimated the number of small regulatory RNAs, which form an important class of non-coding RNAs, in Arabidopsis thaliana to be in the order of 75,000.Among small RNAs, two subclasses form the bulk of all regulatory RNAs as microRNAs (miRNAs) and small interfering RNAs (siRNAs) which are of similar length (21 to 25 nt) and composition but different by origin[5,6]. It is predicted that these two subclasses regulate at least one-third of all human genes. Even for these well-studied RNAs, their precise mode of function remains poorly understood. In addition to such endogenous ncRNAs, antisense oligo nucleotides have been used as exogenous inhibitors of gene expression. Antisense technology is now commonly used for therapeutic purposes and as a research tool. The therapeutic objective of antisense technology is to block the production of disease-causing proteins. In principle, these artificial regulatory RNA molecules could be employed as drugs for the treatment of a variety of human diseases including various types of cancer, rheumatoid arthritis, brain diseases, and viral infections. As a research tool, antisense nucleic acids may be used to study metabolic networks by controlling or interfering with the dynamics and function of various modules in the network.

In present study, we analyzed Noncoding Leucine tRNA genes of *Escherichia coli O157:H7 Leu* against in pathogenic strains of Gammaproteobacteria by comparative genomics. Computational analysis of genes for nc tRNA is to predict RNA transcript initiation, termination, and processing, and find all predicted transcripts that do not have open reading frames[7,8,9].The successful comparative screens for genes of non-coding leucine tRNAs in *E.coli* used as starting material of comparative genome analysis to identify conserved sequences innon-coding regionsin Gammaproteobacteria from related bacteria such as *Escherichia coli O157:H7* (Food poisoning),*Salmonella enterica subsp. enterica serovar Typhi Ty2* (Enteritisand Typhoidfever), *Vibrio cholerae tRNA I operon* (Cholera)and*Aeromonas hydrophila sub sp. hydrophila ATCC 7966 (*Gastroenteritis)[10,11,12].We assembled draft gene sequences of four Gammaproteobacteria species for comparative genomics purposes and often to identify new ncRNA genes. We used these comparative

Gammaproteobacteria species genomic data to perform a tRNAscan-SE screen for new structural non coding tRNA genes in these species[14,15].

Methodology

Computational screening

Non coding Leucine tRNA of*Escherichia coli* 0157:H7 tRNA 45-leu genome sequence and annotation were downloaded from t-RNA database (http://lowelab.ucsc.edu).This sequence was used as query input against nucleotide collection database and sequence comparisons were performed with *BLAST (***http://www.ncbi.nlm.nih.gov/blast/**). There was significant similarity in 226 organisms from which the four pathogenic organisms of Gammaproteobacteria were extracted for the comparative screening.

Identification of conserved intergenic sequence alignments

ClustalW2

The identified leucine tRNA genomic sequences were compared by multiple sequence analysis tool using CLUSTALW2 (http://www.ebi.ac.uk/Tools/clustalw/index.html).Itcalculates the best match for the selected sequences, and lines so that the identities, similarities and differences can be seen. Further sequences were compared to CLUSTALW2withPAM matrix and open gap5.PAM matrices were developed by Dayhoff et al (1978). Point accepted mutation unit corresponds to one amino acid change per 100 residues and Gaps are introduced to improve the alignment between 2 sequences. The insertion of no more than one gap per 20 residues is a reasonable rule of thumb.

T-Coffee

T-Coffee *(www.ebi.ac.uk/t-coffee/*) is Tree-based Consistency Objective Function For alignment Evaluation. It was used for tRNA assessment and could allow combining results obtained with several alignment methods. For instance an alignment coming from ClustalW2, another alignment coming from Dialign, and a structural alignment of some of our sequences. T-Coffee has combined all that information and produces a new multiple sequences having the best agreement with all these methods. By default, T-Coffee was compared all input sequences two by two, producing a global alignment and a series of local alignments.

PET Scan& tRNA Scan search

PETScan ***(www-****unix.mcs.anl.gov/compbio/PatScan)* is Portable, Extensible Toolkit for Scientific Computation, a pattern matcher which searches tRNA sequence archives for instances of a pattern with inputRNA specific Pattern rule isr1={au,ua,gc,cg,gu,ug,ga,ag} and p1=4...8 ggtagac r1 ~p1. It**reports the secondary structure and** Pseudo-knots, long-range interactions between a loop and another portion of the RNA molecule. The PET Scan gave 2000 reports of the conserved pattern GGTAGAC were used to EMBL Nucleotide Sequence Database maintained by EBI .The sequences obtained from EBI were used to do tRNAscan-SE Searches to

find tRNAs present in the sequence. TRNAscan-SE is a program for improved detection of transfer RNA genes in genomic sequence. The tRNAscan-SE servers are accessed via the Lowe Lab Web server Interface at (http://lowelab.ucsc.edu/tRNAscan-SE/) and gives count of total number of tRNAs, number of tRNA pseudogenes, tRNAs with introns and the anticodons that were detected.

Secondary structure prediction

Thensequences obtained from EBI were used to do tRNAscan-SE searches to find tRNAs present in the Bacteriagenomic sequence. The tRNAs were identified from the organisms further used for the prediction of tRNA secondary structure(folding) by energy minimization parameters from DNA MfoldServer (***http://mfold.bioinfo.rpi.edu***).

Results

Computational screening and Identification of conserved sequence alignments

BLAST search gave significant similarity in 226 organisms of which the following were selected to do further analysis.

- Escherichia_coli_O157:H7
- Salmonella enterica subsp. enterica serovar Typhi Ty2
- Vibrio cholerae tRNA I operon
- Aeromonas hydrophila subsp hydrophila ATCC 7966

The conserved regions of tRNA sequences shown as GGTAGAC, GGGTTC, AGTCCC was analyzed through the ClustalW2, ClustalW2with pam25 & gap open 5, T-coffee are shown in fig.1, 2.PetScan is pattern matcher which extracted tRNA sequence archives for instances of a pattern with input RNA specific Pattern rule and gave 2000 reports of the pattern GGTAGAC were used to EMBL Nucleotide Sequence Database are shown in fig.3.It also **reported the secondary structure and** Pseudo-knots, long-range interactions between a loop and another portion of the RNA (not shown).The accession numbers obtained from PetScan were analyzed using EBIdatabase. Then sequence obtained from EBI were used to do tRNAscan-SE Searches and identified the tRNAs present in the genome sequence from the organisms like*Bacillus halodurans, Bacillus subtilis, Thermus thermophilus, Thermofilum pendens*are shown in fig.4 and listed in Table.1.The tRNA from organisms were involved in similar molecular functions of DNA-binding, transposase activity, transcription activity and as a part of transcription start and transcription stop loop are listed in table1.

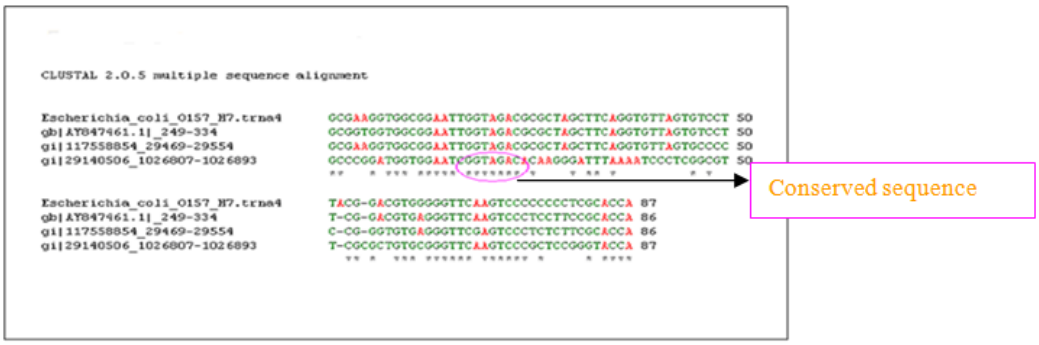


Figure 1: shows the identification conserved regions from multiple sequence alignment of tRNA gene sequences using clustalW2.

Figure 2: shows the identification conserved regions from multiple sequence alignment of tRNAgene sequences using TCoffee.

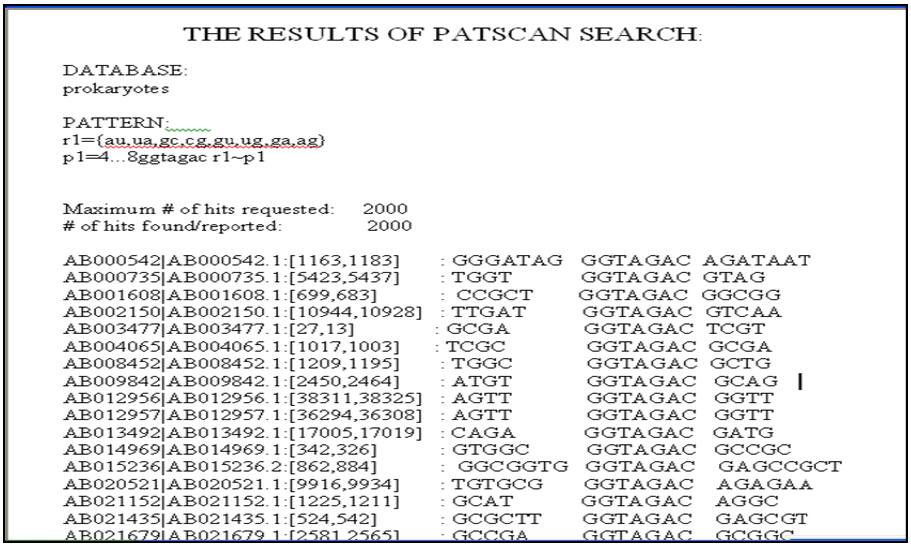


Figure 3: represents the pattern matches through the PetScan search tool which extract tRNA Sequence archives and 2000 hits found with the patterns of GGTAGAC.

Sequence Name	tRNA #	tRNA Begin	Bounds End		
embl|AB002150|AB002150	1	12931	13003	Val	TAC
embl|AB002150|AB002150	2	13011	13083	Thr	TGT
embl|AB002150|AB002150	3	13106	13186	Tyr	GTA
embl|AB002150|AB002150	4	13194	13265	Gln	TTG

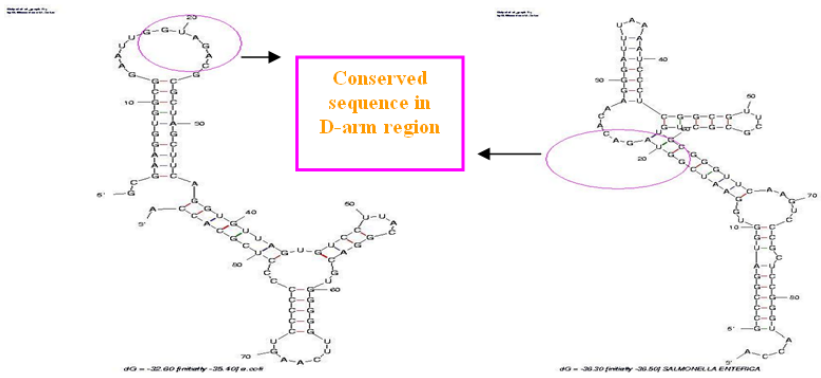
Figure 4: shows the identification of the tRNAs sequence obtained from EBI were used to perform tRNAscan-SE Searches.

Table 1: The list of organisms that showed the tRNA from tRNAscan-SE searches.

Accession number	Organism name	Number andName of tRNAs
AB031212	*Bacillus halodurans*	2, tRNA-Asn,Thr
AB002150	*Bacillus subtilis*	4,tRNA-Val,Thr,Tyr,Gln
X07394	*Thermus thermophilus*	1,tRNA-Ser
X14835	*Thermofilum pendens*	2,tRNA-Met,Gly
AE008707	*Salmonella typhimurium*	1,tRNA-Asp

Prediction of structural tRNA by MfoldServer

The secondary structure ofthe tRNA from organismswas viewed in *Mfold*Server and the conserved sequence was found to occur in D-arm region. The D loop's main function is that of recognition. It is widely believed that it will act as a recognition site for aminoacyl-tRNA synthetase which is an enzyme involved in the aminoacylation of the tRNAmolecule. The secondary structures of all the non-coding tRNA of are shown in fig.5.



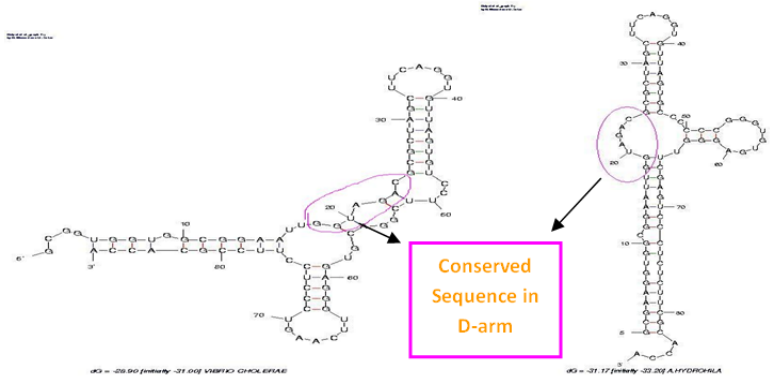


Figure 5: The conserved sequence shown in thein D-arm region of secondary structure in Non-coding tRNA of organisms such as a) *Escherichia_coli_O157:H7*b) *Salmonella enterica subsp. enterica serovar Typhi Ty2* c) *Vibrio cholerae tRNA I operon* d) *Aeromonas hydrophila sub sp hydrophila ATCC 7966.*

Discussion

The results from PATSCAN and tRNAscan-SE showed that most of the tRNA containing organisms such as *Bacillus halodurans AB 031212* and *Bacillus subtilis AB 002150* are involves part of molecular functions such as DNA-binding,trascription factor activity, sigma factor activity, DNA-binding transposase activity. The organisms such as *Bacillus subtilis, Themus Thermophilus X0739, Thermofilum pendens X1483*etc. containing tRNA formed part of transcription start and transcription stop loop.

Conclusion

We identifiednon coding tRNAsfrom Gammaproteobacteria A of *Escherichia Coli, Salmonella,Vibrio,and Aeromonas Hydrophila*are pathogenic organisms and identified certain pattern which is highly conserved in these organisms are GGTAGAC, GGGTTC and AGTCCC from computation screening. These conserved Patterns similarly pragmatic in D-arm region of tRNA. They are involved in similar molecular functions as DNA- binding transposase activity, transcription activity, and part of transcription start and transcription stop loop. Hence in future, thesenc tRNA genes could beused as antisense oligonucleotides in therapeutic purposes for the treatment of a variety of human diseases and as a research tool.

Acknowledgement

Authors are acknowledges to Head of the Department of the Biotechnology Division, Bapatla Engineering College, Bapatla and Centre for Biotechnology, Acharya Nagarjuna University, Guntur for providing facilities to carry out work. The authors also thankful to moral support by faculties and research scholars in both the institutes.

Bibliography

[1] Eddy, S.R. (2001) Non-coding RNA genes and the modern RNA world. Nature Rev. Genet., 2, 919-929.
[2] ENCODE Project Consortium, Nature 447, 799 (2007).
[3] Erdmann, V.A., Barciszewska, M.Z., Szymanski, M., Hochberg, A., de Groot, N., and Barciszewski, J. (2001). Nucleic Acids Res. *29*,189–193.
[4] Fichant, G.A. and Burks, C. (1991) Identifying potential tRNA genes ingenomic DNA sequences. J. Mol. Biol., 220, 659-671.
[5] Gesteland, R.F., Cech, T.R., and Atkins, J.F., eds (1999). The RNAWorld, Second Edition. (Cold Spring Harbor Laboratory Press: ColdSpring Harbor, NY).
[6] Hüttenhofer A, Schattner P, Polacek N: Non-coding RNAs: hope or hype?Trends Genet 2005, 21(5), 289-97.
[7] Rivas,E., Klein,R.J., Jones,T.A. and Eddy,S.R. (2001) Computationalidentification of noncoding RNAs in E. coli by comparative genomics.Curr. Biol., 11, 1369-1373.
[8] Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: a program for improveddetection of transfer RNA genes in genomic sequence. Nucleic AcidsRes., 25, 955-964.
[9] M. W. Vaughn and R. Martienssen, Science 309, 1525 (2005).regions. Nuc.Acids Res. 32: 4925-4936.
[10] Pedersen, JS et al. (2006) “Identification and Classification of Conserved RNA Secondary Structures in the Human Genome” PLOS Computational Biology 2.4.251-262.
[11] Prakash A. and Tompa M. (2005) Discovery of regulatory elements in vertebrates through comparative genomics. Nat. Biotechnol. 23:1249-1256.
[12] Rivas,E. and Eddy,S.R. (2001) Noncoding RNA gene detection usingcomparative sequence analysis. BMC Bioinformatics, 2, 8.
[13] Storz, G. (2002) An expanding universe of noncoding RNAs. Science, 296, 1260-1263.
[14] Torarinsson E, Yao Z, Wiklund ED, Bramsen JB, Hansen C, Kjems J,Tommerup N, Ruzzo WL, Gorodkin J: Comparative genomics beyondsequence-based alignments: RNA structures in the ENCODE regions.Genome Res 2008, 18(2):242-51.
[15] Wang AX, Ruzzo WL, Tompa M: How accurately is ncRNA aligned withinwhole-genome multiple alignments? BMC Bioinformatics 2007, 8:41.