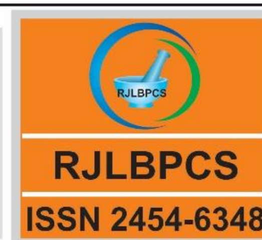

Original Research Article **DOI: 10.26479/2018.0405.58**

COMPREHENSIVE STUDY OF COMPOSITE tRNA GENES IN ARCHAEAL GENOME

Uttam Roy Mandal[1*], Shib Sankar Das[2], Brajadulal Chattopadhyay[3], Satyabrata Sahoo[4],

1.Department of Mathematics, Raidighi College, Raidighi, South 24 Parganas, W.B., India.

2.Department of Mathematics, Uluberia College, Uluberia, Howrah, W.B, India.

3.Department of Physics. Jadavpur University, 188 Raja S.C. Mullick Road, Kolkata, W.B., India.

4.Department of Physics, Dhruba Chand Halder College, Dakshin Barasat, South 24 Parganas, W.B., India.

ABSTRACT: The analysis of archaeal tRNA genes is gradually becoming more important in the field of bioinformatics for its essential role in evaluating the origin and evolution of tRNA molecule. tRNA genes in archaea often have introns intervening between exon sequences. The structural motif at the boundary between exon and intron is the bulge-helix-bulge. Precise insilico identification of the splice-sites on the bulges at the exon-intron boundaries conduce us to infer that a single intron-containing composite tRNA gene can give rise to more than one gene product. A thorough search of different tRNA molecules in archaeal genomes revealed that $tRNA^{Ala}$ (TGC) and $tRNA^{Arg}$ (GCG), and $tRNA^{Gln}$ (TTG) and $tRNA^{Pro}$ (TGG) in *Aciduliprofundum boonei* and $tRNA^{Lys}$ (CTT) and $tRNA^{Gln}$ (TTG) in *Acidilobus Saccharovonans* were co-located. Thus, assuming alternate splicing mechanism we speculated a number of composite tRNA genes giving rise to tRNA isodecoders in archaeal genomes which required another layer of gene arrangement to settle their tRNA gene repertoires.

KEYWORDS: tRNA gene**,** bulge-helix-bulge, alternate splicing, overlapping tRNA, canonical intron.

Corresponding Author: Uttam Roy Mandal* M.Sc, M.Tech

Department of Mathematics, Raidighi College, Raidighi,

South 24 Parganas, West Bengal, India.

Email Address: urmandal77@raidighicollege.in

© 2018 Life Science Informatics Publication All rights reserved

1.INTRODUCTION

The origin and evolution of tRNA is one of the most important subject of discussion in the field of molecular evolution, with varying hypotheses being proposed [1-6]. tRNA plays a key role to link the genetic information carried by mRNA codons and the corresponding amino acids necessary for protein synthesis. Although tRNA is one of the most conversed RNA molecules in structure and function, it is architecturally diverse. In archaca, there have been three types of tRNA genes identified to date: intron-less, intron-containing, and split tRNA [7-9]. Split tRNA genes are encoded on two or three separate regions on genome and then these are processed into single tRNA gene [10, 11]. Intron-containing and split tRNA genes share a common structural motif called bulge-helix-bulge(BHB), or more relaxed bulge-helix-loop (BHL),at the intron/ leader- exon boundary and these are suggested to be evolutionarily interrelated [11-13].There are many hidden patterns around in genome sequences. Many sophisticated non linear algorithms [14] exist for pattern formation and recognition [15, 16]. For instance, there are several computational approaches to detect tRNA genes from a genome [17]. To identify these on the sequences, there are number of algorithms extensively used in this field. Notable amongst these are tRNAScan-SE [18] and ARAGORN [19]. Most of these tRNA- search programs key on primary sequence patterns and /or secondary structures specific to tRNAs. Some of the tRNA genes are either misidentified or missed by existing search algorithms. In the present work, we discuss some of these unidentified tDNA sequence lie embedded in other tDNA sequence. We observed that both of these tDNAs located within the particular region of the respective genomes with same range were identified by our home- algorithm where we modified our algorithm that searched for two halves of cloverleaf, 1-37 and 38-72, from $5^{/}$ and $3^{/}$ end respectively, allowed upto one non-watson –crick base pairing and the possibility of a single putative intron with the standard BHB structure at the exon- intron boundary. We allowed for small variation in the length of D-loop. In this communication we argued that bioinformatics evidence pointed towards one adenine being shifted in the D-loop, the corresponding anticodon shifted accordingly. This anticodon also specifies the particular amino acid that the tRNA carries in these embedded tDNAs. One composite tDNA gives rise to two different tRNAs with different amino acids [20-24]. The single-stranded primary tDNA nucleotide chain folds back into itself to form the cloverleaf secondary structure. This structure has: i) Acceptor or A-arm: In this arm the $5^{/}$ and $3^{/}$ ends of tRNA are base paired into a stem of 7 bp. ii) D-arm or DHU, structurally a stem-loop: D-arm frequently contains the modified base dihydrouracil. It acts as a recognition site for aminoacyl-tRNA synthetases (AARSs), an enzyme involve in the aminoacylation of tRNA molecule. It is highly variable region and is notable for its unusual conformation due to the over-crowding on one of the guanosine residues. It appends to play a large role in the stabilization to the tRNA tertiary structure. It is composed of 7 to 11 base and is closed by a Watson Crick base pair. iii) Anticodon or AC –arm made of a stem and a loop containing the anticodon: At $5^{/}$ end of this anticodon-loop is a pyrimidine

© 2018 Life Science Informatics Publication All rights reserved
Peer review under responsibility of Life Science Informatics Publications

base at 32, followed by an invariant U at 33. The anticodon triplet is located at 34, 35 and 36 in the exposed loop region. iv) An Extra arm or V –arm: This arm is not always present. It is of variable length and largely responsible for the variation in lengths of tRNAs. tRNA classification into types I and II depends on length of V-arm, and T- Ψ-C-arm or T-arm: This has conserved sequence of three ribonucleotides, ribothymidine, pseudouridine and cytosine. T-arm has stem-loop structure. Finally, tRNA terminates with CCA at $3^{/}$ end. Some tDNA sequences may or may not have CCA at $3^{/}$ end. Then, it is added during tRNA maturation. In developing our algorithm to circumvent some of the difficulties in searching missing tRNA sequences by existing programs, we found evidence that tRNA genes overlapped in archaeal genome through introns located at canonical or noncanonical positions [23]. Earlier, in mitochondrial genome we have observed some tRNA sequences overlapped by one to six nucleotides with downstream genes on the same strand [20]. For instance, $tRNA^{Tyr}$ and $tRNA^{Cys}$ of human mitochondrial genome overlap with one another by one nucleotide at the first base of $tRNA^{Cys.}$ This nucleotide is the discriminator base of $tRNA^{Tyr}$ [21, 22]. But tDNA overlaps in archaea are altogether new phenomena. In the genome of *Aciduliprofundum boonei* and *Acidilobus Saccharovonans*, we found a number of tDNAs composite with one another, resulting tRNA isodecoders.

2. MATERIALS AND METHODS

The entire genome sequences were obtained from National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). In the present communication, we focused our study on two archaeal genomes namely, *Aciduliprofundum boonei* (GenBank accession no. NC_013926) and Acidilobus *Saccharovonans* (GenBank accession no. NC_014374). We developed a computational approach which was aimed at searching for any type of tRNA genes and was especially focused on intron containing tRNAs (and split tRNAs) not identified by tRNAscan-SE (Lowe and Eddy 1997) and ARAGORN (Laslett and Canback 2004). We adopted the standard cloverleaf model for studying the secondary structure of predicted tDNAs with salient features that i) T8 ,G18,R19,R53,Y55 and A58 were considered as conserved bases for archaeal tRNAs, ii) The constraints of lengths of stems of regular tRNA A-arm,D-arm,AC-arm and T-arm were 7,4, 5 and 5 bp respectively . We considered at most one non-watson-crick base pairing in any of the four arms or in the 3D-base pairs. In few cases, the constraints on lengths of D arm and AC arm were relaxed. iv) Promoter sequence ahead of the $5^{/}$-end was looked for. v) Base positions optionally occupied in D-loop were 17,17a, 17b, 20a, 20b and 20c. vi) Extra arm or V-arm was taken into consideration. The constraint on length of V-arm was restricted to be less than 21 base vii) Canonical intron (and split) was considered at 37/38 position. The introns and split position constrained to harbor the Bulge-Helix-Bulge (BHB) secondary structure for splicing out during tRNA maturation. With these features in our algorithm, we extracted duplicate copy of $tRNA^{Ala}$ (TGC), $tRNA^{Gln}$ (TTG) and $tRNA^{Lys}$ (CTT). We annotated $tRNA^{Ala}$ (TGC) embedded with $tRNA^{Arg}$ (GCG) and $tRNA^{Gln}$ (TTG) embedded with $tRNA^{Pro}$

© 2018 Life Science Informatics Publication All rights reserved

(TGG) in Aciduliprofundum boonei and $tRNA^{Lys}$ (CTT) embedded with $tRNA^{Gln}$ (TTG) in Acidilobus Saccharovonans.

3. RESULTS AND DISCUSSION

tRNA (ALA^{TGC}) / (ARG^{GCG}) Composite Gene of *Aciduliprofundum boonei*

We investigated the complete genome of *Aciduliprofundum boonei* and found composite tRNA gene of ALA^{TGC} / ARG^{GCG}. In figure 1A we illustrate the construction of composite tRNA genes of ALA^{TGC} / ARG^{GCG}.

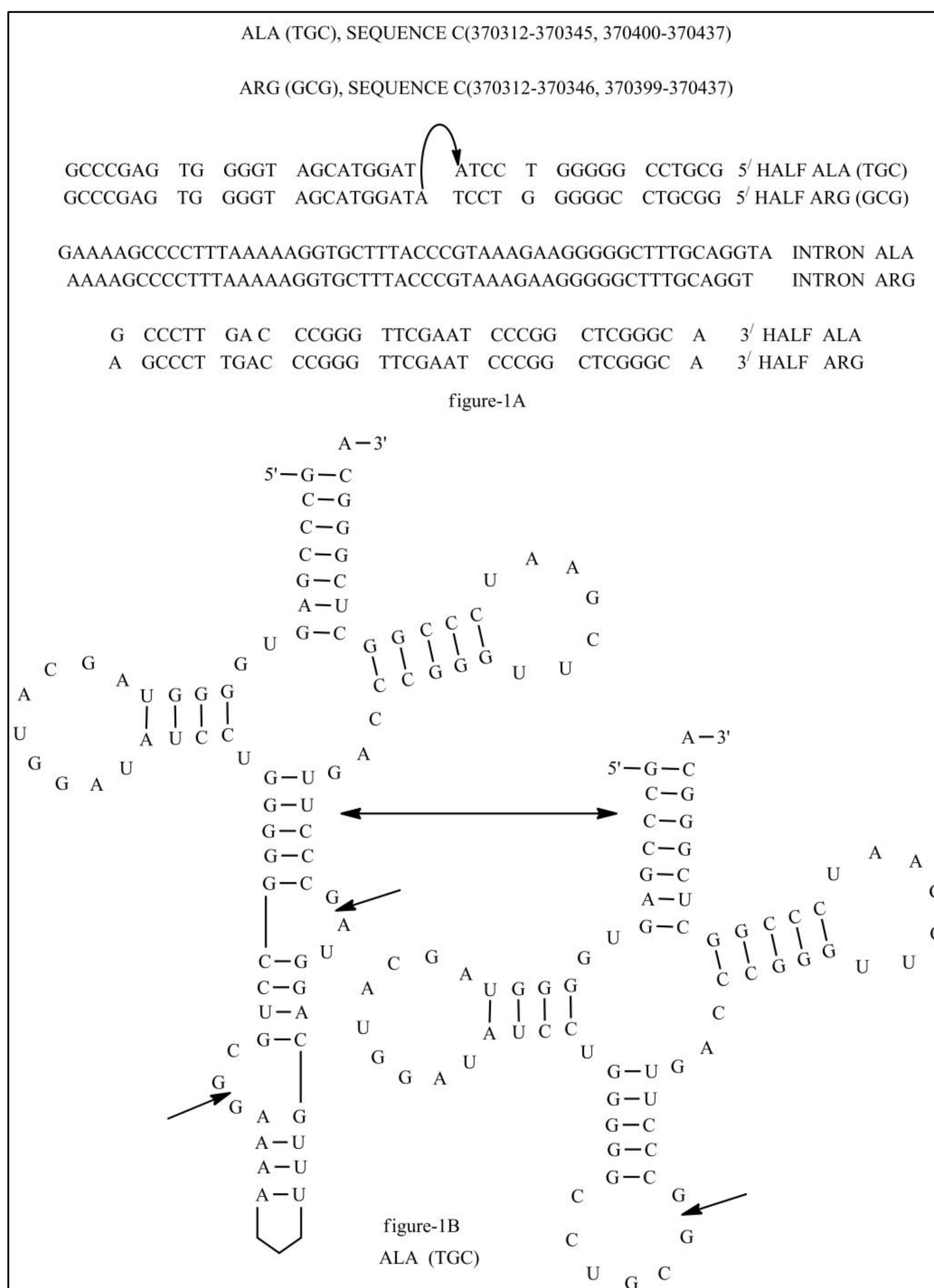

Figure 1A: The locations and the gene sequences of the composite tDNA genes of ALA^{TGC} / ARG^{GCG} in *Aciduliprofundum boonei* genome.

Figure 1B: The schematic diagrams of the second copy of tDNA (ALA^{TGC}) identified in

© 2018 Life Science Informatics Publication All rights reserved

***Aciduliprofundum boonei* genome that split at 37/38 position along with the structures of the BHB motifs. Splice sites are indicated by arrows.**

Using our algorithm, we annotated tRNA (ALA^{TGC}) gene in (860949-860876), which could be easily identified by tRNA Scan-SE but a second copy we identified in the range c (370312-370437). This second copy of tRNA (ALA^{TGC}) gene illustrated in figure 1B was remained unidentified in Aciduliprofundum boonei genome. This alanine tRNA has the important bases A73, G35 and C36 necessary for aminoacylation by AARS (aminoacyl tRNA synthetase), but stronger watson-crick base pairing than existing one. It has the conserved bases and base-pairs like other archaeal analine tRNAs. In this tDNA, presumably, the 54 base long intron was located at 37/38. We also observed that one of the elongator ARG^{GCG} tRNA gene identified to be embedded in this range c(370312-370437). In this tDNA, the 52 base long intron was located at 37/38. Our observations led us to propose two alternate intron splice sites giving rise two different gene products from this single composite tRNA gene. We presented two optimized secondary structures arising from this gene, after the introns were spliced out [Figure 1A]. One of these correspond to tRNA-alanine; the other one is tRNA-argenine. We now analyze insilico the secondary structure at the exon intron boundary – the bulge-helix-bulge (BHB). It is the conformational structure most easily recognized and processed by archaeal splicing mechanism. We explored that there might have been another tRNA alanine gene having the same anticodon TGC but at different location in the genome with different tDNA sequence. Interestingly, new $tRNA^{Ala}$ gene is co-located with $tRNA^{Arg}$. In that case, the base A21 in the D-loop of the ARG^{GCG} tRNA has been shifted to the base A22 in the D-arm of the ALA^{TGC} tRNA. As a result of maintaining the secondary structure of the tRNA alanine and decoding a messenger RNA (mRNA) sequence into protein, the anticodon bases G34, C35 and G36 of arginine (ARG^{GCG}) tRNA gene sequence has been shifted to T34,G35 and C36 of alanine (ALA^{TGC}) tRNA gene sequence. After the comparative analysis in the archaeal genome of *Aciduliprofundum boonei*, we concluded that, the location indentified by our algorithm where same sequence acted as composite tRNA gene was actively involved to transport different amino acids to the ribosome for protein production and tRNA alanine (ALA^{TGC}) was the isodecoder gene.

tRNA (GLN^{TTG}) / (PRO^{TGG}) Composite Gene of *Aciduliprofundum boonei*

In search of missing tRNA genes by our algorithm, we extracted another composite tRNA gene of (GLN^{TTG}) / (PRO^{TGG}) in *Aciduliprofundum boonei* in between (380193-380313). In this range tRNA (PRO^{TGG}) was previously annotated with 47 base long intron (380232-320278) located at 37/38 also identified by us. Using our algorithm we identified tRNA (GLN^{TTG}) isodecoder gene in this range which was remain unidentified in Aciduliprofundum boonei. The sequence was presented in figure 2A. This glutamine tRNA has the important bases A73, U35 and G36 necessary for aminoacylation by AARS (aminoacyl tRNA synthetase) and the watson-crick base pairing in this case was found to be stronger than existing one. It has the conserved bases and base-pairs of other archaeal glutamine

© 2018 Life Science Informatics Publication All rights reserved

tRNAs. In this tDNA, presumably, the 49 base long intron (380231-380279) located at 37/38 has consensus BHB motif (shown in figure 2B) at the exon-intron boundary. In figure 2A-2B we illustrated the construction of composite tRNAs along with tRNA (GLN^{TTG}) gene which was already identified by tRNAScan-SE in between c (367458-367578).

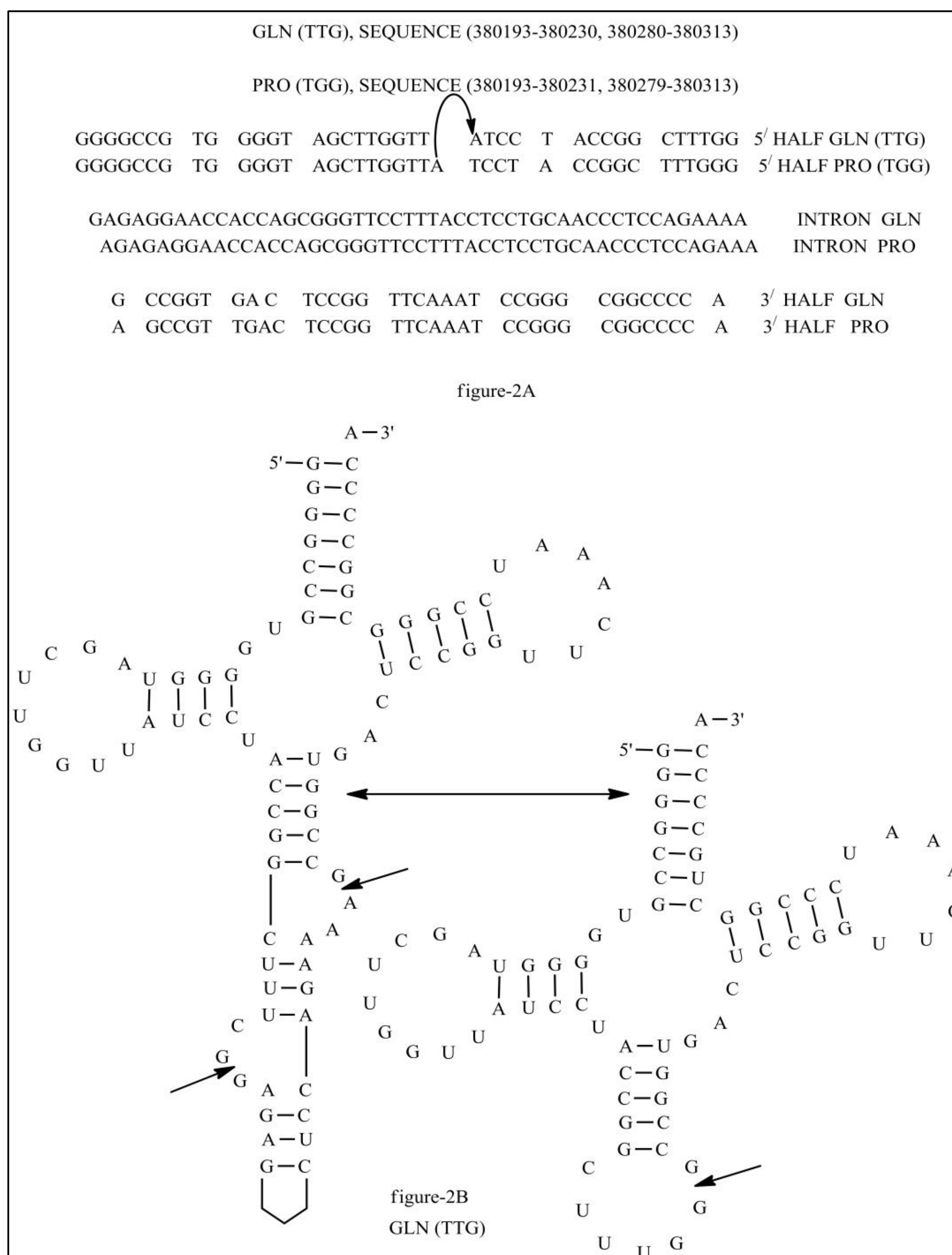

Figure 2A: The locations and the gene sequences of the composite tDNA genes of GLN^{TTG} / PRO^{TGG} in *Aciduliprofundum boonei* genome.

Figure 2B: The schematic diagrams of the second copy of tDNA (GLN^{TTG}) identified in *Aciduliprofundum boonei* genome that split at 37/38 position along with the structures of the BHB motifs. Splice sites are indicated by arrows.

Thus we concluded that tRNA glutamine gene having the same anticodon TTG annotated at different

© 2018 Life Science Informatics Publication All rights reserved

location (380193-380313) in *Aciduliprofundum boonei* genome with different tRNA sequence was to be identified as tRNA isodecoder gene. Surprisingly here also, the base A21 in the D-loop of the tRNA proline ($PRO^{(TGG)}$) was shifted to the base A22 in the D-arm of the tRNA glutamine (GLN^{TTG}), Consequently, to maintain the secondary structure of the $tRNA^{Gln}$ and to decode a messenger RNA (mRNA) sequence into protein some base changes occurred in between 21 to 44 in the tRNA proline molecule. As a result the anticodon loop of $tRNA^{Pro}$ has been transformed to decode the anticodon TTG instead of TGG. The experimental tRNA glutamine (GLN^{TTG}) is ready for protein synthesis. This insilico evidence lead us to conclude that a single intron containing composite gene can give rise two tRNA products i.e. tRNA (GLN^{TTG}) and tRNA (PRO^{TGG}) and tRNA (GLN^{TTG}) is the isodecoder gene of Aciduliprofundum boonei in archaea.

tRNA (LYS^{CTT})/ (GLN^{TTG}) Composite Gene of *Acidilobus Saccharovorans*

In the study of unidentified tRNA genes in the complete genome of *Acidilobus Saccharovorans,* we annotated composite tRNA gene of LYS^{CTT}/ GLN^{TTG} in between c (1034907-1034998).

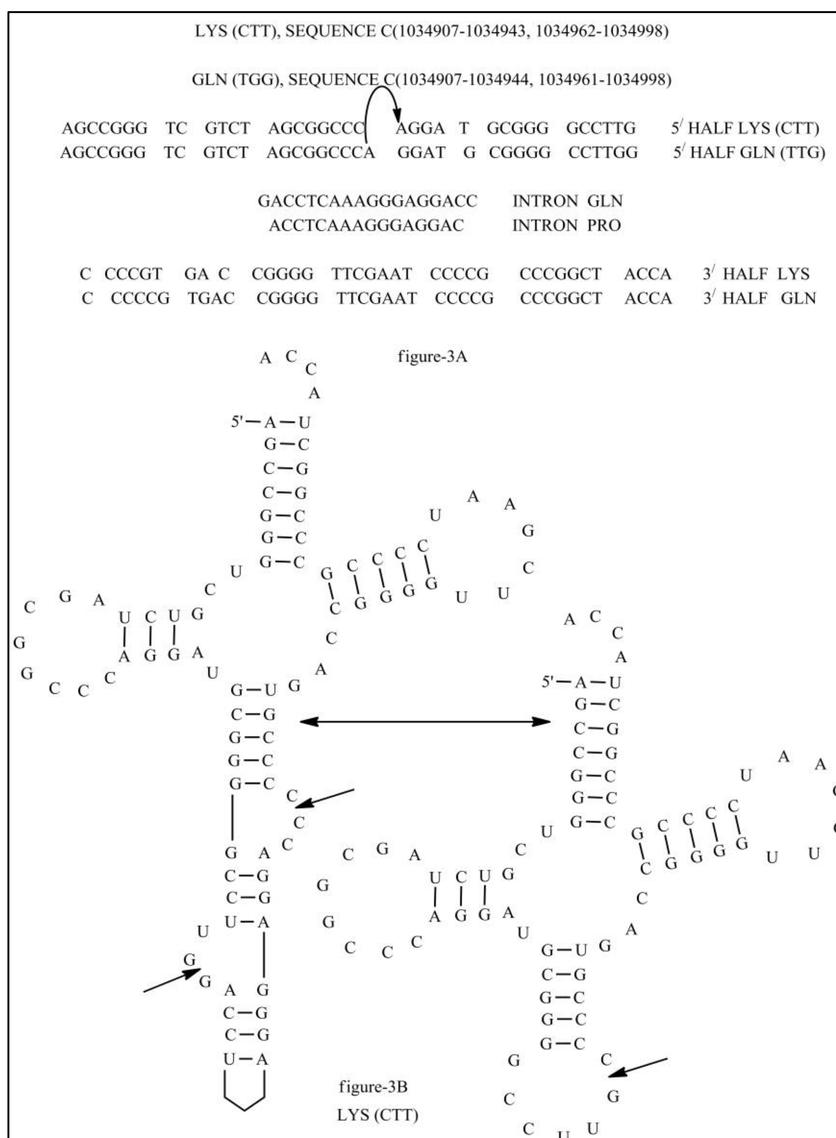

Figure 3A: The locations and the gene sequences of the composite tDNA genes of LYS^{CTT}/

© 2018 Life Science Informatics Publication All rights reserved

GLN^{TTG} in Acidilobus *Saccharovorans* genome.

Figure 3B: The schematic diagrams of the second copy of tDNA (LYS^{CTT}) identified in *Acidilobus Saccharovorans* genome that split at 37/38 position along with the structures of the BHB motifs. Splice sites are indicated by arrows.

In figure 3A-3B, we illustrated the construction of composite tRNA gene of LYS^{CTT}/ GLN^{TTG}. We also extracted another tRNA (LYS^{CTT}) between c (1038698-1038800) in the genome of *Acidilobus Saccharovorans*. This was already annotated by existing search algorithms and could be easily identified by tRNAScan-SE. Using our algorithm; we identified lysine (CTT) between c (1034907-1034998) in the genome of *Acidilobus Saccharovorans*. The sequence was presented in figure 3A-3B. This lysine tRNA gene with anticodon CTT was not annotated earlier in this genome and was found to have stronger base pairing than the existing one. This lysine tRNA also terminates with CCA at $3^{/}$ end. In this tDNA, presumably, the 18 base long intron was located at 37/38. It has consensus BHB motif (illustrated in figure 3B). We also observed that one of the glutamine (GLN^{TTG}) tRNA gene lies exactly embedded within this rage. This was previously identified by tRNAScan-SE. In this tDNA, the 16 base long intron was located at 37/38 having BHB motif at the exon intron boundaries. Remarkably we observed that i) $tRNA^{Lys}$ gene having the same anticodon CTT but annotated at different location c(1034907-1034998) in the *Acidilobus Saccharovorans* genome with different tDNA sequence may be identified as lysine tRNA isodecoder gene. ii) Secondly, the base A21 in the D-loop of the glutamine (GLN^{TTG}) tRNA is shifted to the base A22 in the D-arm of the tRNA (LYS^{CTT}) and consequently, to maintain the secondary structure of the $tRNA^{Lys}$ and to decode a messenger RNA (mRNA) sequence into protein, some base changes occurred from base 21 to 44 in the tRNA glutamine molecule. We have seen that the changes in the anticodon loop of glutamine happens to decode the antcodon CTT instead of UUG. The newly annotated tRNA lysine (LYS^{CTT}) and the existing tRNA glutamine (GLN^{TTG}) is a composite gene in the genome of *Acidilobus Saccharovorans* and (LYS^{CTT}) is the isodecoder gene of *Acidilobus Saccharovorans*.

4. CONCLUSION

In archaea, we studied the availability of all tRNAs over the entire set of sequenced genomes. We found a number of composite tRNAs giving rise to tRNA isodecoers and thus the increase in the tRNA gene number and subsequent divergence is likely the reason of archaeal genome expansion. The evolved variations in the resulting tRNA isodecoders may confer some unique role yet to be determined. We demonstrate that tRNA isodecoders have similar stability and exhibit similar levels of aminoacylation in vivo. We believe that one of the tRNAs is correctly processed in some transcripts, the other in other transcripts, potentially producing both complete transcripts. The sequence diversity among isodecoder tRNA genes may reflect functional variability. The naturally occurring tRNA isodecoders can have large functional variations and suggest that some tRNA

© 2018 Life Science Informatics Publication All rights reserved

isodecoders may perform a function distinct from translation. Our comprehensive study of composite tRNA genes will provide a new molecular basis for upcoming tRNA studies.

ACKNOWLEDGEMENT

Computational facilities Support provided by Raidighi College, Raidighi, South 24 Parganas, West Bengal, is duly acknowledged.

CONFLICT OF INTEREST

We have no conflicts of interest to disclose.

REFERENCES

1. Giulio MD.On the origin of the transfer RNA molecule. J Theor Biol.1992; 159(2):199-214.
2. Dick TP and Schamel WA. Molecular evolution of transfer RNA from two precursor hairpins: implications for the origin of protein synthesis. J Mol Evol.2005; 41(10): 1-9.
3. Maizels N, Weiner AM, Yue D and Shi PY. New evidence for the genomic tag hypothesis: archaeal CCA-adding enzymes and tDNA substrates. Bio Bull. 1999; 196(3): 331-333; discussion 333-334.
4. Nagaswamy U and Fox GE. RNA ligation and the origin of tRNA. Orig. Life Evol Biosph. 2003; 33(2):199-209.
5. Schimmel P, Giege R, Moras D and Yokoyama S. An operational RNA code for amino acids and possible relationship to genetic code. Proc Nat Acad Sc. USA. 1993; 90(19):8763-8768.
6. Tanaka T and Kikuchi Y.Origin of the cloverleaf shape of transfer RNA –the double-hairpin model: implication for the role of tRNA intron and the long extra loop. Viva Orig.2001; 29:134-142.
7. Randau L and Söll D. Transfer RNA genes in pieces. EMBO *Rep*. 2008; 9: 623-628.
8. Sugahara J, Fujishima K, Morita K, Tomita M and Kanai A. Disrupted tRNA gene diversity and possible evolutionary scenarios. J Mol Evol.2009; 69(5): 497-504.
9. Sugahara J, Kikuta K, Fujishima K, Yachie N, Tomita M and Kanai A. Comprehensive analysis of archaeal tRNA genes reveals rapid increase of tRNA introns in the order termoproteales. Mol Bio Evol.2008; 25(12): 2709-2716.
10. Randau L, Munch R, Hohn MJ, Jahn D and Söll D. Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5′- and 3′-halves. Nature. 2005; 433(7025):537–541.
11. Fujishimaa K, Sugaharaa J, Kikuta K, Hirano R, Sato A, Tomita M and Kana A. Tri-split tRNA is a transfer RNA made from 3 transcripts that provides insight into the evolution of fragmented tRNAs in archaea. Proc Natl Acad Sci .USA.2009; 106(8): 2683-2687.
12. Randau L, Calvin K, Hall M, Yuan J, Podar M, Li H and Söll D.The heteromeric Nanoarchaeum equitans splicing endonuclease cleaves noncanonical bulge-helix-bulge motifs of joined tRNA halves. Proc Natl Acad Sci. USA. 2005; 102(50): 17934–17939.

© 2018 Life Science Informatics Publication All rights reserved
Peer review under responsibility of Life Science Informatics Publications

13. Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP. Coevolution of tRNA intron motifs and tRNA endonuclease architecture in Archaea. Proc Natl Acad Sci. USA.2005; 102(43): 15418–15422.
14. Murray JD. Mathematical Biology: An Introduction (3rdedn.) (Berlin:Springer Verlag), 2002
15. Cross MC and Hohenberg P. Pattern formation outside of equilibrium. Rev Mod Phys.1993; 65(3): 851-1112.
16. Koch AJ and Meinhardt H. Biologocal pattern formation: from basic mechanisms to complex structures. Rev Mod Phys.1994; 66(4): 1481-1507.
17. el-Mabrouk N and Lisacek F. Very fast identification of RNA motifs in genomic DNA: application to tRNA search in the yeast genome. J Mol Biol.1996; 264: 46–55.
18. Lowe TM and Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res.1997; 25: 955–964.
19. Laslett D and Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004; 32(1): 11–16.
20. Yokobori S and Paabo S. Transfer RNA editing in land snail mitochondria. Proc Natl Acid Sci. USA.1995; 92: 10432-10435.
21. Reichert A, Rothbaure U and Morl M. Processing and editing of overlapping tRNAs in human mitochondria. J Bio Chem.1998; 273(48): 31977-31984.
22. Binder S, Marchfelder A and Brennicke A. RNA editing of $tRNA^{Phe}$ and $tRNA^{Cys}$ in mitochondria of Oenothera berteriana is initiated in precursor molecules, Mol Genet Genomics. 1994; 244(1): 67-54.
23. Ghosh Z, Chakrabarti J,Mallick B, Das S, Sahoo S and Sethi HS. tRNA-isoleucine-tryptophan composite gene. Biochemical and Biophysical Research Communications. 2006; 339(1): 37–40.
24. Chakrabarti J, Sahoo S, Mallick B, Das S and Ghosh Z. Algorithm for pattern recognition in nano-sized archaea. Indian J Phys.2005; 79 (6): 559-662.

© 2018 Life Science Informatics Publication All rights reserved
Peer review under responsibility of Life Science Informatics Publications