

Suppressive Variants of Selenocysteine tRNA in the Complete Genome of *Methanopyrus kandleri* AV19

Uttam Roymandal¹, Shib Sankar Das², Riya Rakshit³ and Satyabrata Sahoo^{4*}

¹Department of Mathematics, Raidighi College, Raidighi, South 24 Parganas, WB, India

²Department of Mathematics, Uluberia College, Uluberia, Howrah, WB, India

³Department of Botany, Dhruva Chand Halder College, Dakshin Barasat, South 24 Parganas, WB, India

⁴Department of Physics, Dhruva Chand Halder College, Dakshin Barasat, South 24 Parganas, WB, India

*Corresponding author: Satyabrata Sahoo, Department of Physics, Dhruva Chand Halder College, Dakshin Barasat, South 24 Parganas, WB, India, Tel: +91 9437447274 ; E-mail: dr_s_sahoo@yahoo.com

Received date: March 23, 2018; Accepted date: May 09, 2018; Published date: May 19, 2018

Copyright: © 2018 Roymandal U, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

In Archaea, previous studies have revealed the presence of multiple intron-containing tRNAs and split tRNAs. The full unexpurgated analysis of archaeal tRNA genes remains a challenging task in the field of bioinformatics, because of the presence of various types of disrupted tRNA genes in archaea. Here we suggested a computational method that searched for widely separated genes encoding tRNA halves to generate suppressive variants of missing tRNAs. Considering the existence of split tRNA genes widely separated throughout the genome, we developed our tRNA search algorithm to predict such separated tRNA genes by searching both a conserved terminal 5'- and 3'-motif of tRNA in agreement with the split hypothesis on the basis of cloverleaf prediction and precise *in silico* determination of bulge-helix-bulge secondary structure at the splice sites. By a comprehensive search for missing tRNA genes, we characterized new variants of selenocysteine tRNA that had been found inserted in *Methanopyrus kandleri* AV19. Analysis of the complete genome sequence of *M. kandleri* AV19 revealed the multiple mode of transcription of a non-coding RNA decoding UGA to read selenocysteine (Sec) and suggested further study of the disrupted tRNA genes 002E.

Keywords: *M. Kandleri* AV19; Selenocysteine tRNA; Fragmented tRNA s; tRNA search algorithm; Split hypothesis

Introduction

Transfer RNA (tRNA) is a central genetic element in the decoding of genome information for all types of life form in the earth [1]. Since its discovery in 1950's, the search for this molecule becomes the most important subject of discussion in the field of molecular evolution [2]. The proper identification of transfer RNAs (tRNAs) is critical for a detailed understanding of its essential role in the decoding operation by which the base sequence in a messenger RNA (mRNA) molecule becomes translated to an amino acid sequence of a protein [3]. *In-silico* prediction of all tRNAs in a complete genome of an organism still remains a unsolved problem in the field of bioinformatics [4-12]. Although, tRNAs have common secondary and tertiary structures that govern the important role in controlling their activity [13,14] some tRNAs are difficult to recognize due to their unusual sub-structures and may result in the detection of the wrong anticodon. Therefore, the detection of all tRNA genes in a genome remains an important challenge.

Notably, there have been three types of tRNA genes previously identified in microbial genomes: a) Non-intronic tRNA which is a single uninterrupted gene with no intervening intron sequence, b) Intron-containing tRNA in which case a number of introns that vary in length may be present at various locations in a single gene [3,15-18] and c) Trans-spliced tRNA which comprises of 5' and 3' halves that are located on two separate genes in the genome sequence [3,19,20]. A number of studies have revealed that many missing tRNAs have been

left unexplored [1] in Archaea. In search of missing tRNA genes in the nuclear genome of the red algae *Cyanidioschyzon merolae*, another type of fragmented tRNA genes was deciphered in 2007 [21]. In some cases, the distant 5' and 3' halves of split tRNAs are found inverted within a single gene [22]. The organization of these fragmented intron-containing tRNAs that require processing to form mature structure is surprising. The introns intervening the fragmented or disrupted tRNAs representing different amino acids are unrelated, but they are not entirely irrelevant. An endonuclease cleaves the tRNA precursors at both ends of the intron and it is a well-established fact that archaeal tRNA splicing endonuclease recognizes a bulge-helix-bulge (BHB) structural motif around the intron-exon boundaries [23,24]. An interesting insight into the evolution of tRNA splicing is provided by the archaeal enzymes which recognize their substrates in the relaxed form of the BHB motif denoted as HBh or BHB like (BHL) motif [3,15,18,25].

A number of previous studies have revealed the presence of various types of disrupted and fragmented intron-containing tRNAs and split tRNAs in Archaea [3]. A comprehensive prediction of archaeal tRNA genes, thus, remains a challenging task in the field of bioinformatics [3]. In the present communication, we have developed a computational methodology towards genome analysis that searches for fragmented parts of a tRNA gene in widely separated coding sequences. The search algorithm is formulated on the basis of conserved cloverleaf structure of a tRNA with invariant or semi invariant base pairs so that predicted tRNA halves could form mature tRNA molecules [19]. We know that the genetic code is very much universal, but some exceptions are also found to exist in certain species. These deviations from the standard genetic code may affect the codons involved in initiation and

termination of protein synthesis. A striking example is the addition of Selenocysteine (Sec) as twenty first amino acid to the genetic code [26-28]. The incorporation of the unusual amino acid selenocysteine at certain UGA codons within the genes that code for selenoproteins are found in both prokaryotes and eukaryotes to provide an essential nutrient for their normal growth and development [29]. The specific changes that recode UGA from stop codon might be accomplished by the surrounding bases and is determined by the local secondary structure of mRNA, in particular by the presence of a hairpin loop downstream of the UGA. We focused our study to identify potential tRNA genes in the complete genome of *M. kandleri* AV19 that are spliced to generate suppressive variants of a non-coding RNA decoding UGA to read selenocysteine (Sec).

Materials and Methods

The whole genome sequence of *M. kandleri* AV19 (GenBank accession no: NC-003551) along with gene annotations were retrieved from NCBI GenBank. All sequences marked as CDS were considered. It also includes those sequences which were annotated as hypothetical or predicted by computational methods. The genome sequence of *M. kandleri* AV19 contains 1,694,969 base pairs. The GC content of the genome is about 61% which is unusually high compared to other methanogens. *M. kandleri* is among the most heat tolerant organisms with an optimum growth temperature near 98°C. In terms of maintaining stability with the extreme nature of its environment, it exhibits an unusual and diverse pattern of post transcriptional modification of tRNA nucleosides [30]. In this study, we developed a computational approach to characterize suppressive variants of selenocysteine tRNA gene that had been found inserted in the complete genome of *M. kandleri* AV19.

We adopted the standard cloverleaf model for studying the secondary structure of predicted tDNAs in our computational approach. The salient features included in our search algorithm considered T8, G18, R19, T44, R53, Y55 and A58 as conserved bases for archaeal tRNAs. The constraints on lengths of stems of regular tRNA A-arm, D-arm, AC-arm and T-arm were considered to be 7 bp, 4 bp, 5 bp and 5 bp respectively. We considered at most one mismatch in every of the four arms and in the 3D-base pairs. In few cases the constraints on lengths of D arm and AC arm were relaxed. Base positions optionally occupied in D-loop were considered to be 17, 17a, 17b, 20a, 20b and 20c. Moreover, extra arm or V-arm was also considered in our search- algorithm. The constraint on length of V-arm was restricted to be less than 21 bases. Our search algorithm looked for the promoter sequence ahead of the 5' -end. Splicing of tRNA was considered at 37/38 or 36/37 position. The introns and split position were constrained to harbour BHB secondary structure for splicing out during tRNA maturation.

To provide a complete picture of tRNA diversity, we based our search for the prediction of nuclear tRNA genes on two features which were characteristic of tRNA structures [5]. These were local potential hairpin and stem structures consistent with the cloverleaf structural motif of the typical tRNA sequences [5] and the presence of several invariant or semi-invariant bases that define two conserved regions [5,31] T, Y, C and D regions. Our program searches sequences which can define the highly conserved T-loop region and the adjacent 3' -acceptor stem sequence so that the reverse complementary sequence (defining the 5' -acceptor stem sequence) plus a D-stem identifies the corresponding 5' -half [19]. The length of V loop can be adjusted as necessary for structuring anticodon stem. Finally, predicted tRNA-

halves that recognizes the canonical or relaxed form of bulge-helix-bulge motif around the boundaries are re-joined *in-silico* and analysed by tRNA Scan [8,19]. In addition to identifying the set of tRNAs predicted by the tRNA Scan-SE, our algorithm predicted several other tRNA halves spread throughout the genome. Surprisingly, many of these tRNA halves that follow the split hypothesis at the exon-intron boundaries could be joined to predict the missing tRNAs [19]. Here we reported that analysis of the complete genome sequence of *M. kandleri* AV19 revealed a good number of genes encoding the selenocysteine transfer RNA species.

Results and Discussion

Archaea, an intermediate between Eukarya and Bacteria, have tRNAs that share many similarities with those of either or both of these domains [32]. In the present study, we developed an algorithm to search for widely separated tRNA signature sequences in the archeal genome. Since its development in 1997, tRNA scan-SE has become a tool of choice for tRNA prediction. This program is sensitive and highly selective, and all currently known tRNAs could be found using its 'maximum sensitivity' mode. However, a number of unusual tRNAs [33] containing long introns remain unexplored. Later, ARAGON [7] is added to tRNA prediction software packages. All known tRNAs with and without the presence of canonical tRNA introns are predicted with high accuracy by using algorithms such as tRNA scan-SE and ARAGORN. More recently, modified algorithms, which allow the prediction of tRNAs containing non-canonical introns, multiple introns, and split tRNA genes, for annotation of tRNAs at the genome level have been developed [17-20,23,32]. Considering the existence of split tRNA genes widely separated throughout the genome, application of our tRNA prediction software to the completely genome sequence of *M. Kandleri* AV19 predicted a number of missing tRNAs. Algorithm has been developed to detect fragmented tRNA genes in the genome sequence by searching consensus 5' - and 3' - halves of tRNA genes. Searching a pairing stretch of 7 nucleotides of 3'-region to a reverse complementary sequence in the 5' - region for the tRNA's acceptor stem, matching pairs of tRNA-half genes were identified by our computational approach.

In a first step, we remove all matching pairs to from cloverleaf structure of un-split tRNAs and the remaining 5' and 3' regions are used to predict possible split tRNAs by searching matching pairs of their anticodon stem sequences. Finally, the respective tRNA halves are joined and the resulting sequence have been analysed by using tRNA scan-SE. The split tRNA genes are widely separated throughout the genome. Most of these tRNA genes are found to be split into two tRNA-halves after the anticodon-adjacent position 37, the normal location of tRNA introns. Aside from usual split at 37, a good number of conjugated halves may also be split after the anticodon position. Each of the split tRNA pairs possessed a flanking leader sequence at the 5' and 3' -end on two separate genes sharing a common BHB motif or its relaxed form for recognition and cleavage by tRNA endonucleases. It is well investigated that split tRNA genes in prokaryote genomes are usually expressed through RNA splicing [34]. It has been reported that both split and permuted tRNA genes are much common in *Nanoarchaeum equitans* [35]. More recently, a computational analysis of the *N. equitans* genome revealed that a number of tRNA halves spread throughout the genome [19] could be joined *in silico* producing tRNA genes encoding the glutamate (tRNA^{Glu}), histidine (tRNA^{His}), tryptophan, and methionine [36].

Although many potential missing tRNAs may be detected by our split tRNA search programme, we have been interested in search of selenocysteine tRNA in the present study. To our knowledge selenoproteins have been found to be widely distributed in all sphere of life [29]. In this context, the existence of a number of tRNA sec half genes in *M. kandleri* AV19 is most exciting. A good number of 5' tRNA halves of tRNA (sec) were identified by our programme that differed mainly in anticodon loop/arm or in V loop for a single 3' half. With a relaxation of single nucleotide mismatch in the acceptor arm, a critical analysis has revealed a good number of genes encoding the selenocysteine transfer RNA genes in *M. kandleri* AV19. The sequences also reveal characteristic nucleotides in the respective tRNA species needed for recognition by the cognate aminoacyl-tRNA synthetase [19]. For example, the tRNA_{Sec} half genes encode the unique G-1:C73 base pair required for aminoacylation of tRNA_{Sec} (by histidyl-tRNA synthetase), and the tRNA_{Sec} isoacceptors contain the characteristic D-loop nucleotides 20a and 20b [19]. The locations of the fragmented part of the isotypes of the predicted sec tDNAs are summarized in Figures 1-3.

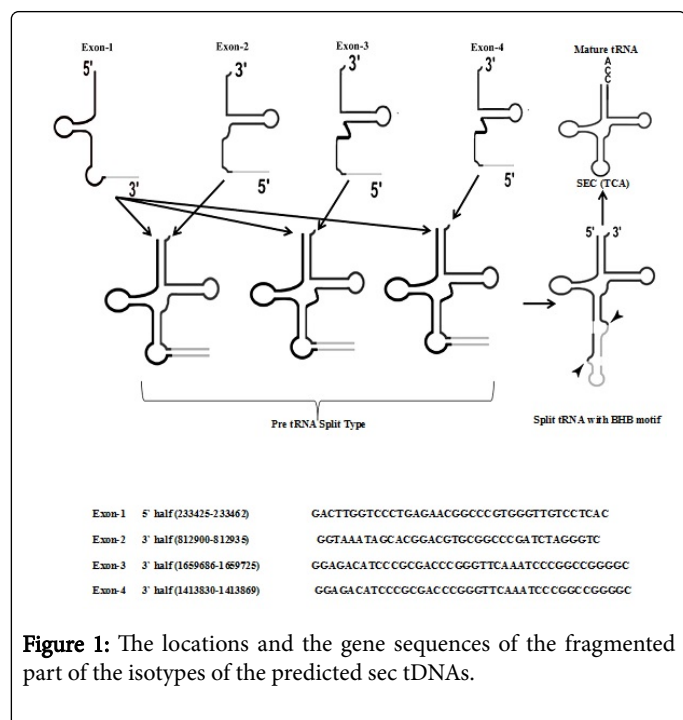


Figure 1: The locations and the gene sequences of the fragmented part of the isotypes of the predicted sec tDNAs.

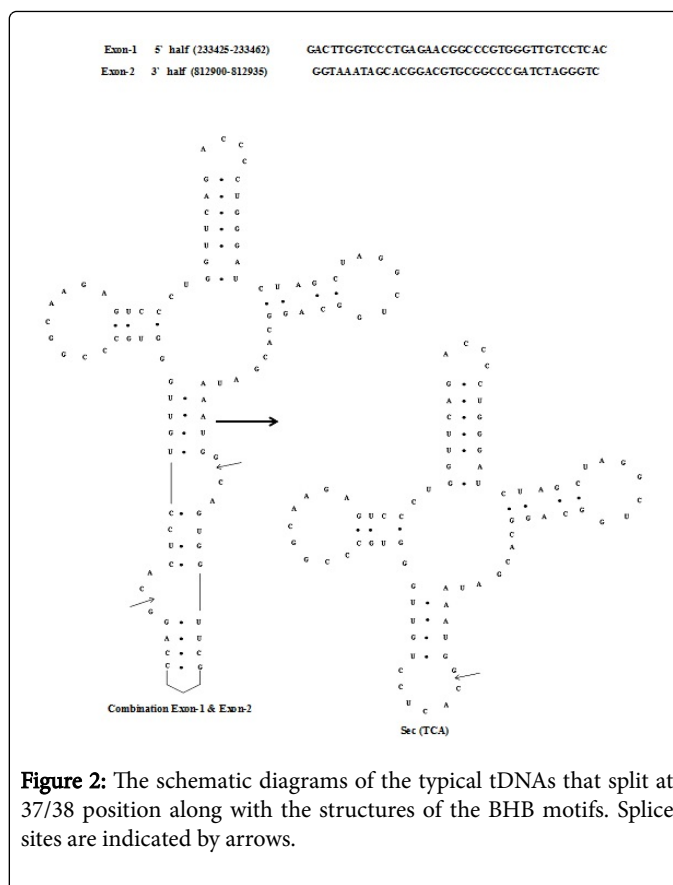


Figure 2: The schematic diagrams of the typical tDNAs that split at 37/38 position along with the structures of the BHB motifs. Splice sites are indicated by arrows.

Also we remarkably observed in Figure 1, that exon-1 combined with the exon-2, exon-3 and exon-4 respectively. These fragmented parts are located in different regions within the genome. The schematic diagrams of the typical tDNAs that split at 37/38 or 36/37 position are displayed in Figures 2 and 3 respectively along with the structures of the BHB motifs. The analysis of exon-intron junctions of the newly found split tRNA halves has revealed that canonical or relaxed BHB structural motifs exist at splicing junctions [3]. In some cases, an unusual single-nucleotide mismatch was also found within the central H helix [3]. It is well known that the heterotetrameric enzyme is able to process both canonical and relaxed form of BHB motifs at various splice sites in the tRNA structure [3,24] and is even capable of cleaving each strand at a bulge in a bulge-helix-bulge motifs with variable bulge length [3,23]. Finally, based on the hypothesis that split tRNA genes encoding tRNA_{sec} should contain the consensus bases of all archaeal elongator tRNA_{sec}, namely U8, A14, G18, G19 and the T-loop GTTCAAATC [19], we characterize new variant of selenocysteine tRNA gene (Figure 3) that had been found inserted in the complete genome of *M. kandleri* AV19 with its 1st half located at (674560-674600) and 2nd half at (1146698-1146735).

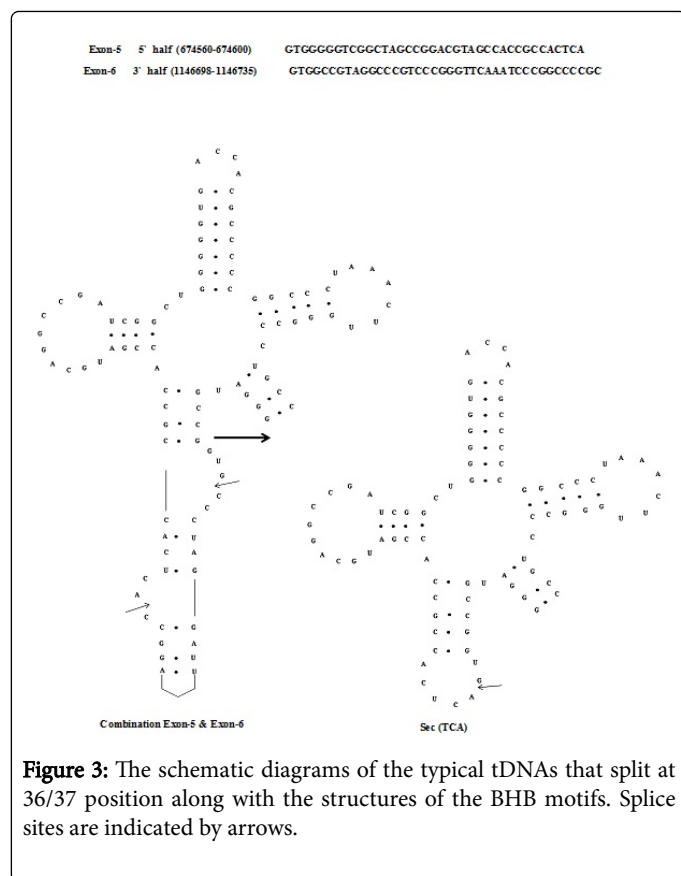


Figure 3: The schematic diagrams of the typical tDNAs that split at 36/37 position along with the structures of the BHB motifs. Splice sites are indicated by arrows.

In summary, we have developed a new split tRNA search program for detecting missing tDNAs by integration of fragmented tDNAs which are widely separated throughout the genome and have been able to identify a novel tDNA candidate complementing missing tRNAsec in *M. kandleri* AV19. The present study reveals that the integration of tRNA fragments widely separated throughout the genome might provide a mechanism for the generation of the suppressive tRNA genes. Our program, however, should prove useful in examining newly sequenced genomes for the presence of nonstandard amino acids within the genetic code.

References

- Maruyama S, Sugahara J, Kanai A, Nozaki H (2010) Permuted tRNA Genes in the Nuclear and Nucleomorph Genomes of Photosynthetic Eukaryotes. *Mol Biol Evol* 27: 1070-1076.
- Sugahara J, Fujishima K, Morita K, Tomita M, Kanai A (2009) Disrupted tRNA Gene Diversity and Possible Evolutionary Scenarios. *J Molecular Evolution* 69: 497-504.
- Fujishima K, Sugahara J, Kikuta K, Reiko HR, Sato A, et al. (2009) Tri-split tRNA is a transfer RNA made from 3 transcripts that provides insight into the evolution of fragmented tRNAs in archaea. *PNAS* 106: 2683-2687.
- el-Mabrouk N, Lisace F (1996) Very fast identification of RNA motifs in genomic DNA: application to tRNA search in the yeast genome. *J Mol Biol* 264: 46-55.
- Fichant GA, Burks C (1991) Identifying potential tRNA genes in genomic DNA sequences. *J Mol Biol* 220: 659-671.
- Gautheret D, Lambert A (2001) Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J Mol Biol* 313: 1003-1011.
- Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32: 11-16.
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25: 955-964.
- Pavesi A, Conterio F, Bolchi A, Dieci G, Ottonello S (1994) Identification of new eukaryotic tRNA genes in genomic DNA databases by a multi-step weight matrix analysis of transcriptional control regions. *Nucleic Acids Res* 22: 1247-1256.
- Sagara J, Shimizu S, Kawabata T, Nakamura S, Ikeguchi M, et al. (1998) The use of sequence comparison to detect 'identities' in tRNA genes. *Nucleic Acids Res* 26: 1974-1979.
- Sprinzel M, Horn C, Brown M, Ioudovitch A, Steinberg S (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* 26: 148-153.
- Steinberg S, Misch A, Sprinzel M (1993) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* 21: 3011-3015.
- Goodenbour JM, Pan T (2006) Diversity of tRNA genes in eukaryotes. *Nucleic Acid Research* 34: 6137-6146.
- Randau L, Soll D (2008) Transfer RNA genes in pieces. *EMBO Reports* 9: 623-628.
- Marck C, Grosjean H (2003) Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea. Evolutionary implications. *RNA* 9: 1516-1531.
- Marck C, Grosjean H (2002) tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea and Bacteria reveals anticodon-sparing strategies and domain specific features. *RNA* 8: 1189-1232.
- Sugahara J, Yachie N, Sekine Y, Soma A, Matsui M, et al. (2006) SPLITS: a new program for predicting split and intron containing tRNA genes at the genome level. *In Silico Biol* 6: 411-418.
- Sugahara J, Yachie N, Arakawa K, Tomita M (2007) In silico screening of archaeal tRNA-encoding genes having multiple introns with bulge-helix-bulge splicing motifs. *RNA* 13: 671-681.
- Randau L, Munch R, Hohn MJ, Jahn D, Soll D (2005) Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5' -and 3' -halves. *Nature* 433: 537-541.
- Randau L, Pearson M, Soll D (2005) The complete set of tRNA species in Nanoarchaeum equitans. *FEBS Lett* 579: 2945-2947.
- Soma A, Onodera A, Sugahara J, Kanai A, Yachie N, et al. (2007) Permuted tRNA genes expressed via a circular RNA intermediate in *Cyanidioschyzon merolae*. *Science* 318: 450-453.
- Heinemann IU, Soll D, Randau L (2010) Transfer RNA processing in archaea: Unusual pathways and enzymes. *FEBS Letters* 584: 303-309.
- Randau L, Calvin K, Hall M, Yuan J, Podar M, et al. (2005) The heteromeric Nanoarchaeum equitans splicing endonuclease cleaves noncanonical bulge-helix-bulge motifs of joined tRNA halves. *Proc Natl Acad Sci* 102: 17934-17939.
- Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP (2005) Coevolution of tRNA intron motifs and tRNA endonuclease architecture in Archaea. *Proc Natl Acad Sci* 102: 15418-15422.
- Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP (2009) Processing of multiple-intron-containing pre-tRNA. *Proc Natl Acad Sci USA* 106: 20246-20251.
- Steinberg SV, Ioudovitch A, Cedergren R (1998) The secondary structure of eukaryotic selenocysteine tRNA: 7/5 versus 9/4 RNA 4: 241-245.
- Sturchler C, Westhof E, Carbon P, Krol A (1993) Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA^{Sec}. *Nucleic Acids Research* 21: 1073-1079.
- Takaharu MT, Goto C (2000) Eukaryotic selenocysteine tRNA has the 9/4 secondary structure. *FEBS Letters* 466: 359-362.
- Chiba S, Itoh Y, Sekine Y, Yokoyama S (2010) Structural Basis for the Major Role of O-Phosphoryl-tRNA Kinase in the UGA-Specific Encoding of Selenocysteine. *Molecular Cell* 39: 410-420.

-
30. Sauerwald A, Sitaramaiah D, James A, McCloskey JA, Soll D, et al. (2005) N6-Acetyladenosine: A new modified nucleoside from *Methanopyrus kandleri* tRNA. FEBS Letters 579: 2807-2810.
 31. Rich A, RajBhandary UL (1976) Transfer RNA: Molecular structure, sequence, and properties. Annu Rev Biochem 45: 805-860.
 32. Ghosh Z, Chakrabarti J, Mallick B, Das S, Sahoo S, et al. (2006) tRNA-isoleucine-tryptophan composite gene. Biochemical and Biophysical Research Communications 339: 37-40.
 33. Lobanov AV, Kryukov GV, Hatfield DL, Gladyshev VN (2006) Is there a twenty third amino acid in the genetic code? Trends Genet 22: 357-360.
 34. Martinez-Abarca F, Toro N (2000) Group II introns in the bacterial world. Mol Microbiol 38: 917-926.
 35. Randau L (2012) RNA processing in the minimal organism *Nanoarchaeum equitans*. Genome Biology 13: 1-11.
 36. Cobucci-Ponzano B, Rossi M, Moracci M (2006) Interrupted genes in extremophilic archaea: mechanisms of gene expression in early organisms. Orig Life Evol Biosph 36: 487-492.