



Novel Genomic and Evolutionary Perspective of Cyanobacterial tRNAs

Tapan K. Mohanta^{1*}, Asad S. Syed², Fuad Ameen² and Hanhong Bae^{1*}

¹ School of Biotechnology, Yeungnam University, Gyeongsan, South Korea, ² Department of Botany and Microbiology, College of Science, King Saud University, Riyadh, Saudi Arabia

Transfer RNA (tRNA) plays a central role in protein synthesis and acts as an adaptor molecule between an mRNA and an amino acid. A tRNA has an L-shaped clover leaf-like structure and contains an acceptor arm, D-arm, D-loop, anti-codon arm, anti-codon loop, variable loop, Ψ -arm and Ψ -loop. All of these arms and loops are important in protein translation. Here, we aimed to delineate the genomic architecture of these arms and loops in cyanobacterial tRNA. Studies from tRNA sequences from 61 cyanobacterial species showed that, except for few tRNAs (tRNA^{Asn}, tRNA^{Leu}, tRNA^{Gln}, and tRNA^{Met}), all contained a G nucleotide at the 1st position in the acceptor arm. tRNA^{Leu} and tRNA^{Met} did not contain any conserved nucleotides at the 1st position whereas tRNA^{Asn} and tRNA^{Gln} contained a conserved U¹ nucleotide. In several tRNA families, the variable region also contained conserved nucleotides. Except for tRNA^{Met} and tRNA^{Glu}, all other tRNAs contained a conserved A nucleotide at the 1st position in the D-loop. The Ψ -loop contained a conserved U¹-U²-C³-x-A⁵-x-U⁷ sequence, except for tRNA^{Gly}, tRNA^{Ala}, tRNA^{Val}, tRNA^{Phe}, tRNA^{Thr}, and tRNA^{Gln} in which the U⁷ nucleotide was not conserved. However, in tRNA^{Asp}, the U⁷ nucleotide was substituted with a C⁷ nucleotide. Additionally, tRNA^{Arg}, tRNA^{Gly}, and tRNA^{Lys} of cyanobacteria contained a group I intron within the anti-codon loop region. Maximum composite likelihood study on the transition/transversion of cyanobacterial tRNA revealed that the rate of transition was higher than the rate of transversion. An evolutionary tree was constructed to understand the evolution of cyanobacterial tRNA and analyses revealed that cyanobacterial tRNA may have evolved polyphyletically with high rate of gene loss.

Keywords: cyanobacteria, tRNA, evolution, intron, transition, transversion

INTRODUCTION

Transfer RNAs (tRNA) are short non-coding RNAs comprised of 75–95 nucleotides, universally present in all organisms from the prokaryotes to the eukaryotes. The 75–95 residues polynucleotide sequences of tRNAs fold back upon themselves and form hydrogen-bonded clover leaf-like structures that fold into L-shaped tertiary structure (Holley et al., 1965; Wilusz, 2015). The tRNA is organized into a linear double stranded helix. Determination of the three-dimensional structure of a tRNA was a landmark discovery for modern molecular biology research. The major function of the tRNA is to bridge between an amino acid and an mRNA during protein synthesis where the tRNA transfers the cognate amino acid to the translating polypeptide chain. The clover leaf-like structure of the tRNA is comprised of an acceptor arm, D-arm, D-loop, anti-codon arm, anti-codon loop, variable loop, Ψ -arm, and Ψ -loop (**Figure 1**; Kirchner and Ignatova, 2014; Mohanta and Bae, 2017).

OPEN ACCESS

Edited by:

Ashok Sharma,
Augusta University, United States

Reviewed by:

Jian Lu,
National Institutes of Health (NIH),
United States
Liang Liu,
University of Georgia, United States

*Correspondence:

Tapan K. Mohanta
nostoc.tapan@gmail.com
Hanhong Bae
hanhongbae@ynu.ac.kr

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Genetics

Received: 12 June 2017

Accepted: 21 November 2017

Published: 13 December 2017

Citation:

Mohanta TK, Syed AS, Ameen F and
Bae H (2017) Novel Genomic and
Evolutionary Perspective of
Cyanobacterial tRNAs.
Front. Genet. 8:200.
doi: 10.3389/fgene.2017.00200

The acceptor arm is 7 base pairs long, the D-arm is 3–4 base pairs, the D-loop is 4–12 nucleotides, the anti-codon arm is 5 base pairs, the anti-codon loop is 7 nucleotides, the variable loop is 4–23 nucleotides, the Ψ -arm is 5 base pairs and the Ψ -loop is 7 nucleotides long (**Figure 1**; Kirchner and Ignatova, 2014). The heterogeneity in the length of tRNAs is caused by variable number of bases in D-loop and the variable loop. The most important functional parts of a tRNA are: the anti-codon triplet, which reads the codons of a messenger RNA; a 3'-CCA tail where the cognate amino acid is attached and the Ψ -arm and Ψ -loop that hold the ribosome machinery.

There are 21 different types of iso-acceptor families for the 20 amino acids, one for each amino acid and one for selenocysteine. In prokaryotes, an isoacceptor family may have one (tRNA^{Trp}, and tRNA^{Met}) to six tRNA members (tRNA^{Ser}, tRNA^{Arg}, and tRNA^{Leu}). To date, four different types of tRNA genes have been reported so far including intron-containing, non-intronic, split and permuted types (Randau et al., 2005; Fujishima et al., 2009; Chan et al., 2011). The intron-containing and non-intronic

tRNAs are encoded in a single gene, whereas the split tRNAs are encoded in two or more separate genes; in permuted tRNAs, the 3' half of the tRNA is located up-stream to the 5' half of the tRNA (Tocchini-Valentini et al., 2009; Maruyama et al., 2010; Soma, 2014). The intron-containing, permuted and split tRNA genes are called “disrupted tRNA genes” whereas the non-intronic tRNA genes are called “continuous tRNA genes” (Sugahara et al., 2009; Kanai, 2015). The disrupted tRNA genes may have evolved earlier than the continuous genes (Di Giulio, 2008a,b; Fujishima et al., 2009). The tRNA mini helix is the most ancient form of tRNA and forms the top half of the tRNA that contains the acceptor arm and the Ψ -arm (Weiner and Maizels, 1987, 1999; Shi et al., 1992, 1998; Sun and Caetano-Anollés, 2008). The tRNA mini-helix is a validated substrate for aminoacyl-tRNA synthetases and the 3'-CCA tail adding enzymes (Shi et al., 1992, 1998). The genomic tag hypothesis suggests that the upper half of the tRNA had evolved earlier than the lower half. Phylogenetic analysis also suggests that the top half of the tRNA is more ancient than the bottom half (Sun and

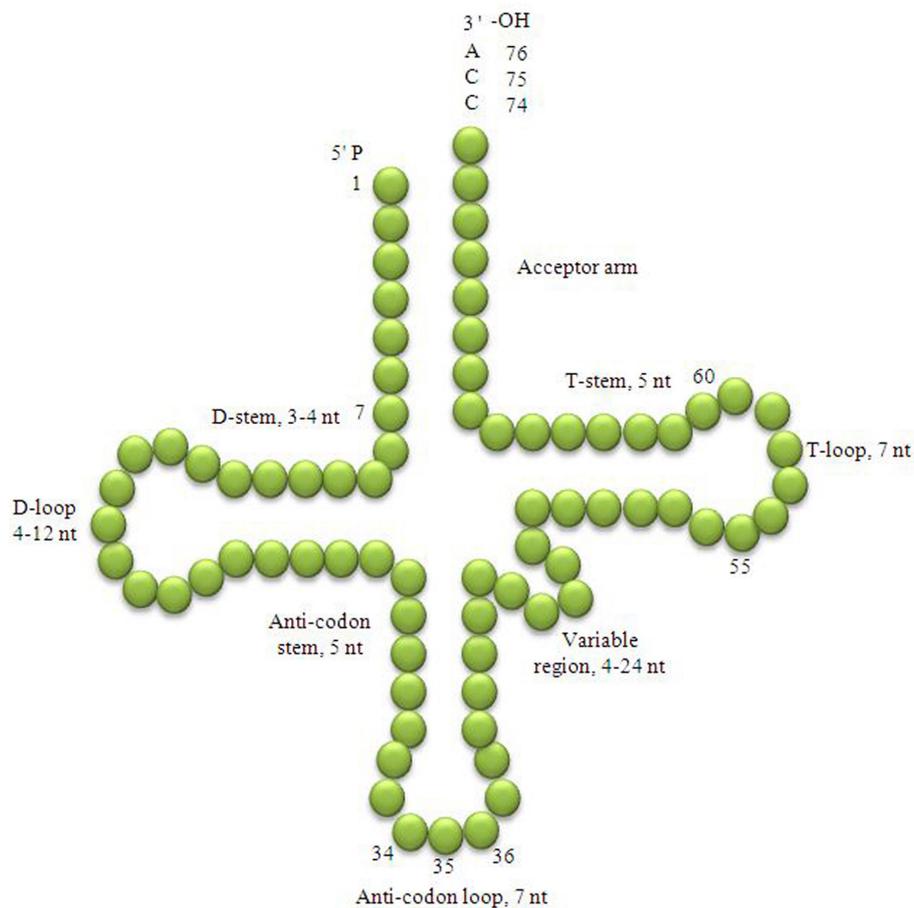


FIGURE 1 | Clover leaf-like structure of tRNA. The tRNA possess the acceptor arm (7 nt), D-arm (3–4 nt), D-loop (4–12 nt), anti-codon arm (5 nt), anti-codon loop (7 nt), variable region (4–23 nt), Ψ -arm (5 nt) and Ψ -loop (7 nt). The D-arm, D-loop, and variable region possess variable number of nucleotides whereas the nucleotide number in the acceptor arm, anti-codon arm, anti-codon loop, Ψ -arm and Ψ -loop is always constant. The anti-codon nucleotides in the anti-codon loop is always numbered 34, 35, and 36 whereas, the nucleotides of C-C-A tail is always numbered with 74, 75, and 76.

Caetano-Anollés, 2008). The evolution of tRNA occurred with respect to its surrounding environment and is far more complex than that of proteins, because tRNA perform multiple roles in the cell (Navarre and Schneewind, 1999; Kirchner and Ignatova, 2014). tRNA interacts with numerous molecules; during protein synthesis, it interacts with amino-acyl tRNA synthetase, ribosome and messenger RNA, and a large number of transcription factors and other enzymes (Nirenberg and Leder, 1964). tRNA^{Gly} acts as a structural element of the peptidoglycan of the bacterial cell wall whereas tRNA^{Lys} and tRNA^{Ala} are used to aminoacylate membrane lipid to permeate cationic antibiotics (Roy and Ibba, 2008). Cyanobacteria are the most ancient living organism whose existence dates back to 3.3–3.5 billion years (Schopf and Packer, 1987; Altermann et al., 2006). Because it is most likely that the tRNA genes date back to the age of cyanobacteria, and studying the genomic and evolutionary perspectives of cyanobacterial tRNA is very important. Given their high level of conservation it is especially important to study the genomics, origin and evolution of tRNAs. Here, a comprehensive study of cyanobacterial tRNA was conducted by analyzing the tRNA sequences of 61 cyanobacterial species.

MATERIALS AND METHODS

Identification and Analysis of Cyanobacterial tRNAs

The genomic tRNA sequences of cyanobacteria were searched from Joint Genome Portal (<http://genome.jgi.doe.gov/>) and cyanobacterial genome database of National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>). In total, tRNA sequences from 61 cyanobacterial species were retrieved and analyzed (**Supplementary data 1**). The downloaded genomic tRNA sequences of cyanobacterial species were analyzed using tRNAScan-SE software (Lowe and Eddy, 1997). All the sequences analyzed during this study were found to encode for tRNA. Different statistical parameters used to run the tRNAScan-SE software were as follows: sequence source, bacterial/mixed; search mode, default; and genetic code for tRNA isotype prediction, universal. The tRNAScan-SE software is the most accurate software and it identifies 99–100% tRNA genes from DNA sequences. It produces less than one false positive per 15 gigabases (Lowe and Eddy, 1997).

Sequence Alignment

Multiple sequence alignment was conducted to identify the conserved consensus sequences of cyanobacterial tRNA. Multalin software was used to conduct the multiple sequence alignment of cyanobacterial tRNA using the default parameters as described previously (Mohanta et al., 2015a,b,c, 2016) with minor modification.

Analysis of Transition/Transversion Rate of Cyanobacterial tRNAs

To analyze the transition and transversion rate of the cyanobacterial tRNA, a clustal file for each of the tRNA gene family was generated using the multiple sequence alignment program MUSCLE (<http://www.ebi.ac.uk/Tools/>

msa/muscle/). The generated clustal file was downloaded and converted to a MEGA file format. The generated MEGA file of the cyanobacterial tRNA was uploaded in MEGA6 software to analyze the transition/transversion rate (Tamura et al., 2013). Following statistical parameters were used to analyze the transition/transversion rate: analysis, estimate transition/transversion bias (MCL); scope, all selected taxa; statistical method, maximum composite likelihood; substitution type, nucleotide; model/method, Tamura-Nei model; and gaps/missing data treatment, pair wise deletion.

Construction of Phylogenetic Trees

A phylogenetic tree was constructed to understand the evolutionary aspects of cyanobacterial tRNAs. All the genomic tRNA sequences of cyanobacterial tRNA were subjected to clustal omega server to construct a clustal file. Generated clustal file of the cyanobacterial tRNA was converted to MEGA file format using MEGA6 software. The MEGA files of the cyanobacterial tRNA were subjected to MEGA6 software to construct the phylogenetic trees (Tamura et al., 2013). Following statistical parameters were used to construct the phylogenetic tree: analysis, phylogeny reconstruction; scope, all selected taxa; statistical method, neighbor-joining; test of phylogeny, bootstrap method; no. of bootstrap replicate, 1,000; substitution type, nucleotide; model/method, maximum composite likelihood; substitution to include, d: transition+transversion; rates among sites, uniform rates; pattern among lineage, same (homogeneous); and gaps/missing data treatment, pair wise deletion. The pairwise distances and substitution model parameters were estimated by maximizing the composite likelihood because maximum likelihood approach can accurately estimate the parameters that drive the evolutionary process (Akaike, 1998; Varin et al., 2011; Xu and Reid, 2011; Zhao et al., 2015). Phylogenetic trees with low bootstrap values were collapsed with 50% cutoff values.

Gene Duplication and Loss

To understand the duplication and loss event of cyanobacterial tRNA, all the clustal files generated to study the transition/transversion rate of individual tRNA family were used to construct the phylogenetic tree. Twenty phylogenetic trees were generated separately for each tRNA gene family using the similar statistical parameters as mentioned previously. A species tree of studied cyanobacterial species was downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>). The phylogenetic tree of the cyanobacterial tRNA and species tree of cyanobacterial species were subjected to Notung2.6 software and further reconciliation of gene tree with the species tree led to finding of duplicated and loss genes of cyanobacterial tRNA (Chen et al., 2000; Vernet et al., 2008).

RESULTS

The Conservation of tRNA Sequences Are Family Specific

To understand the basic genomic and evolutionary aspects of cyanobacterial tRNAs, genomic tRNA sequences of cyanobacterial species were downloaded from the Joint

TABLE 1 | Cyanobacterial species and the number of tRNA genes used during this study.

Sl. No.	Name of the species	No. of tRNAs
1	<i>Anabaena cylindrica</i> PCC 7122	61
2	<i>Anabaena</i> sp. PCC 7108	43
3	<i>Calothrix desertica</i> PCC 7102	65
4	<i>Calothrix</i> sp. PCC 6303	40
5	<i>Calothrix</i> sp. PCC 7103	65
6	<i>Calothrix</i> sp. PCC 7507	72
7	<i>Chamaesiphon minutus</i> PCC 6605	54
8	<i>Cyanobacterium</i> PCC 7702	49
9	<i>Chroococciopsis</i> sp. PCC 6712	51
10	<i>Chroococciopsis thermalis</i> PCC 7203	46
11	<i>Crinalium epipsammum</i> PCC 9333	43
12	<i>Cyanobacterium cyanothece</i> sp. BH63E	42
13	<i>Cyanobacterium</i> ESFC-1	72
14	<i>Cyanobacterium aponinum</i> PCC 10605	43
15	<i>Cyanobacterium stanieri</i> PCC 7202	43
16	<i>Cyanobium gracile</i> PCC 6307	44
17	<i>Cylindrospermum stagnale</i> PCC 7417	69
18	<i>Dactylococcopsis salina</i> PCC 8305	42
19	<i>Fischerella</i> sp. PCC 9339	50
20	<i>Fischerella</i> sp. PCC 9431	69
21	<i>Fischerella</i> sp. PCC 9605	42
22	<i>Geitlerinema</i> sp. PCC 7105	50
23	<i>Geitlerinema</i> sp. PCC 7407	46
24	<i>Geminocystis herdmannii</i> PCC 6308	40
25	<i>Gloeocapsa</i> sp. PCC 73106	40
26	<i>Gloeocapsa</i> sp. 7428	41
27	<i>Halothece</i> sp. PCC 7418	46
28	<i>Kamptonema formosum</i>	71
29	<i>Leptolyngbya boryana</i> PCC 6306	67
30	<i>Leptolyngbya</i> sp. PCC 6406	44
31	<i>Leptolyngbya</i> sp. PCC 7375	63
32	<i>Leptolyngbya</i> sp. PCC 7376	44
33	<i>Mastigocladopsis repens</i> PCC 10914	43
34	<i>Microchaete</i> sp. PCC 7126	69
35	<i>Microcoleus</i> sp. PCC 7113	70
36	<i>Microcoleus vaginatus</i> FGP-2	72
37	<i>Nodosilinea nodulosa</i> PCC 7104	43
38	<i>Nostoc</i> sp. PCC 7107	78
39	<i>Nostoc</i> sp. PCC 7524	65
40	<i>Oscillatoria acuminata</i> PCC 6304	66
41	<i>Oscillatoria</i> sp. PCC 10802	65
42	<i>Oscillatoria nigro-viridis</i> PCC 7112	71
43	<i>Oscillatoriales</i> sp. JSC-12	44
44	<i>Pleurocapsa</i> sp. PCC 7319	42
45	<i>Pleurocapsa</i> sp. PCC 7327	42
46	<i>Prochlorothrix hollandica</i> PCC 9006	41
47	<i>Pseudanabaena</i> sp. PCC 6802	75
48	<i>Pseudanabaena</i> sp. PCC 7367	46
49	<i>Pseudanabaena</i> sp. PCC 7429	44
50	<i>Rivularia</i> sp. PCC 7116	52

(Continued)

TABLE 1 | Continued

Sl. No.	Name of the species	No. of tRNAs
51	<i>Scytonema hofmannii</i> UTEx 2349	53
52	<i>Spirulina major</i> PCC 6313	43
53	<i>Spirulina subsalsa</i> PCC 9445	41
54	<i>Stanieria cyanosphaera</i> PCC 7437	43
55	<i>Synechococcus elongatus</i> PCC 7942	44
56	<i>Synechococcus</i> sp. PCC 9616	42
57	<i>Synechococcus</i> sp. PCC 6312	41
58	<i>Synechococcus</i> sp. PCC 7336	44
59	<i>Synechococcus</i> sp. PCC 7502	43
60	<i>Synechocystis</i> sp. PCC 7509	46
61	<i>Xenococcus</i> sp. PCC 7305	36

Genome Institute (JGI) genome portal and National Center for Biotechnology Information (NCBI) (Table 1). In total, the tRNA sequences of 61 cyanobacterial species were downloaded and analyzed for their conserved genomic aspects. A conserved genomic sequence is required for the conserved clover leaf-like structure of a tRNA. Although significant conservation is present in tRNAs at the nucleotide level, occasionally the conserved structure of tRNAs varies at the family level.

In tRNA^{Gly} of the *Pseudanabaena* sp. PCC 7367 (gene id: 2504679288) was found to contain only a conserved C² nucleotide instead of the conserved G¹-C²-G³ (Table 2, Supplementary Figure 1). Similarly, the acceptor arm of *Cylindrospermum stagnale* PCC 7417 (gene id: 2509767604) was found to contain G¹-G²-A³ nucleotides. The nucleotide U⁸ and A⁹ in the acceptor arm of tRNA^{Gly} was found to be conserved (Table 3). The D-arm was found to be unconserved whereas the D-loop has conserved A¹-x₂-G⁴-G⁵ nucleotides followed by the presence of conserved U²-x-C⁴-C⁵-A⁶ nucleotides in anti-codon arm (Table 2). The variable loop was found to be unconserved whereas the Ψ-arm has conserved G⁵ nucleotide and the Ψ-loop has conserved U¹-U²-C³-x-A⁵ nucleotides (Table 2).

In tRNA^{Ala}, the acceptor arm was found to contain conserved G¹-G²-G³ nucleotides at the 5' end (Table 2) whereas tRNA^{Ala} of *Pseudanabaena* sp. PCC 6802 (gene id: 2507088349) and *Microchaete* sp. PCC 7126 (gene id: 2509784201) was found to contain A¹-G²-G³ nucleotides. The nucleotide U⁸ and A⁹ in the acceptor arm were found to be conserved (Table 3) followed by the presence of conserved G¹-C²-U³-C⁴ nucleotides in the D-arm. The D-loop contained conserved A¹-x₂-U⁴-G⁵-G⁶-x-A⁸ nucleotides followed by the presence of conserved U²-x-G⁴-C⁵-A⁶ nucleotides in the anti-codon arm. A G⁵ nucleotide was found to be conserved in the variable region as well as in the Ψ-arm followed by the presence of conserved U¹-U²-C³-x-A⁵ nucleotides in the Ψ-loop (Table 2).

The acceptor arm of tRNA^{Pro} was found to contain conserved C¹-x-G³-G⁴-x₂-G⁷ nucleotides (Table 2). The nucleotide U⁸ and A⁹ in the acceptor arm and C⁴ in the D-arm were found to be conserved followed by the presence of conserved A¹-G²-x₆-A⁹ nucleotides in the D-loop (Tables 2, 3). The anti-codon loop contained conserved U¹-U²-x-G⁴-G⁵-G⁶ followed

TABLE 2 | Conserved nucleotide signature elements of Cyanobacterial tRNAs.

tRNA Isotypes	5' Acceptor Arm (1–7 nt)	D-arm (10–13 nt)	D-loop	AC- arm (27–31 nt)	Anti-codon loop (32–38 nt)	Variable loop	Ψ-arm (49–53 nt)	Ψ-loop (54–60 nt)
Gly	G ¹ -C ² -G ³	*****	A ¹ -x ₂ -G ⁴ -G ⁵	*****	U ² -x-C ⁴ -C ⁵ -A ⁶	*****	G ⁵	U ¹ -U ² -C ³ -x-A ⁵
Ala	G ¹ -G ² -G ³	G ¹ -C ² -U ³ -C ⁴	A ¹ -x ₂ -U ⁴ -G ⁵ -G ⁶ -x-A ⁸	*****	U ² -x-G ⁴ -C ⁵ -A ⁶	G ⁵	G ⁵	U ¹ -U ² -C ³ -x-A ⁵
Pro	C ¹ -x-G ³ -G ⁴ -x ₂ -G ⁷	C ⁴	A ¹ -G ² -x ₆ -A ⁹	*****	U ¹ -U ² -x-G ⁴ -G ⁵ -G ⁶	G ³ -x-C ⁵	G ¹ -x ₂ -G ⁴ -G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -A ⁶ -U ⁷
Val	G ¹ -G ² -x-C ⁴	C ² -x-C ⁴	A ¹ -G ² -x ₃ -G ⁶ -x-U ⁸ -A ⁹	*****	U ¹ -U ² -x-A ⁴ -C ⁵ -A ⁶	G ³ -U ⁴ -C ⁵	G ⁴ -G ⁵	U ¹ -U ² -C ³ -x-A ⁵
Leu	*****	*****	A ¹ -A ² -x ₂ -G ⁵ -G ⁶ -x-A ⁸	*****	U ² -x-A ⁴	*****	G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -x-U ⁷
Ile	G ¹ -G ² -G ³ -C ⁴	G ¹ -C ² -U ³ -C ⁴	A ¹ -x ₄ -G ⁶ -x ₂ -A ⁹	*****	C ¹ -U ² -x-A ⁴ -U ⁵ -A ⁶ -A ⁷	G ³ -U ⁴	G ⁴ -G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -x-U ⁷
Met	*****	G ²	G ⁶	*****	C ¹ -U ² -C ³ -A ⁴ -U ⁵ -A ⁶ -A ⁷	*****	G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -x-U ⁷
Phe	C ¹ /G ¹ -C ² -C ³ -x ₂ -G ⁶	G ¹ -C ² -U ³ -C ⁴	A ¹ -G ² -U ³ -U ⁴ -G ⁵ -G ⁶ -U ⁷	G ⁴	U ² -G ³ -A ⁴ -A ⁵ -x-A ⁷	G ³ -U ⁴ -C ⁵	G ⁵	U ¹ -U ² -C ³ -x-A ⁵
Tyr	G ¹ -G ² -G ³ -U ⁴ -C ⁵	G ¹ -x-C ³ -C ⁴	A ¹ -G ² -U ³ -G ⁴ -G ⁵ -U ⁶ -U ⁷ -A ⁸	*****	U ² -G ³ -U ⁴ -A ⁵	*****	G ⁴ -G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -x-U ⁷
Trp	G ¹	G ¹ -U ²	A ¹ -x ₃ -G ⁵	G ⁴ -U ⁵	C ¹ -U ² -C ³ -C ⁴ -A ⁵ -A ⁶ -A ⁷	*****	G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -x-U ⁷
Ser	G ¹ -G ² -A ³	G ¹ -C ²	A ² -x ₂ -G ⁵ -G ⁶	*****	U ² -x ₃ -A ⁵ -A ⁶	*****	A ⁴ /G ⁴ -G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -x-U ⁷
Thr	G ¹ -C ²	G ¹ -C ² -x-C ⁴	U ² -x-G ⁴ -U ⁵ -A ⁶ -A ⁷	*****	U ² -x-G ⁴ -U ⁵ -A ⁶ -A ⁷	G ³	G ⁵	U ¹ -U ² -C ³ -x-A ⁵
Cys	G ¹	G ¹ -C ² -C ³	A ¹ -A ² -G ³ -x-G ⁵ -G ⁶ -U ⁷	*****	C ¹ -U ² -G ³ -C ⁴ -A ⁵ -A ⁶ -A ⁷	C ⁵	G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -x-U ⁷
Asn	U ¹ -C ² -C ³ -x-C ⁵	G ¹ -C ² -U ³	A ¹ -x ₁₋₂ -G ³ -G ⁴	*****	C ¹ -U ² -G ³ -U ⁴ -U ⁵ -A ⁶ -A ⁷	G ³ -U ⁴	G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -x-U ⁷
Gln	U ¹ -G ² -x ₃₋₄ -G ⁶	****	A ¹ /U ¹ -A ² -x ₂ -G ⁴	*****	U ¹ -U ² -G ³ -U ⁴ -G ⁵	*****	A ³ -G ⁴ -G ⁵	U ¹ -U ² -C ³ -x-A ⁵
Lys	G ¹ -G ² -G ³	C ²	A ¹ -x ₂ -G ⁴ -G ⁵	*****	C ¹ -U ² -U ³ -U ⁴ -U ⁵ -A ⁶ -A ⁷	*****	G ⁴ -G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -x-U ⁷
His	G ¹ -C ² /U ² -G ³ -x ₃ -G ⁷	G ¹ -C ² -C ³	A ¹ -A ² -G ³ -x-G ⁵ -G ⁶ -U ⁷	G ⁴ -G ⁵	U ¹ -U ² -G ³ -U ⁴ -G ⁵	*****	G ¹ -x-G ³ -G ⁴ -G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -x-U ⁷
Arg	G ¹ -G ²	G ¹	A ¹ -x ₂₋₃ -G ⁴ -x ₃ -A ⁸	****	U ¹ /C ¹ -U ² -x-C ⁴	*****	G ⁴ -G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -x-U ⁷
Asp	G ¹ -G ² -G ³	G ¹ -x-U ³ -C ⁴	A ¹ -x ₂₋₃ -G ⁴ -G ⁵ -U ⁶ -x-A ⁸	****	C ¹ -U ² -G ³ -U ⁴ -C ⁵ -A ⁶	A ¹ -x-G ³ -U ⁴	G ¹ -x-G ³ -x-G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -x-C ⁷
Glu	G ¹	****	G ⁵	****	U ² -U ³ -U ⁴ -C ⁵	****	G ³ -x-G ⁵	U ¹ -U ² -C ³ -x-A ⁵ -x-U ⁷

The asterisk (*) mark represent the absence of conserved nucleotide consensus sequence in respective region of the tRNA.

by the presence of conserved G³-x-C⁵ nucleotides in the variable region. The Ψ-arm contained conserved G¹-x₂-G⁴-G⁵ nucleotides followed by the presence of conserved U¹-U²-C³-x-A⁵-A⁶-U⁷ nucleotides in the Ψ-loop (Table 2).

The acceptor arm of tRNA^{Val} was found to contain conserved G¹-G²-x-C⁴ nucleotides (Table 2) except for species *Cyanobacterium* sp. ESFC 1 (gene id: 2517646710), *Microcoleus vaginatus* FGP-2 (gene id: 2506348931), *Nostoc* sp. PCC 7107 (gene id: 2503741640), *Kamptonema formosum* PCC 6407 (gene id: 2508873173), *Pseudanabaena* sp. PCC 6802 (gene id: 2507089028), *Oscillatoria acuminata* PCC 6304 (gene id: 2509421383), *Microchaete* sp. PCC 7126 (gene id: 2509784199), and *Microcoleus* sp. PCC 7113 (gene id: 2509433894). The tRNA^{Val} of *Pleurocapsa* sp. PCC 7319 (gene id: 2509708986). They were found to contain U¹ nucleotide instead of the G¹ or A¹ nucleotides. The U⁸ and A⁹ nucleotides were conserved in tRNA^{Val} (Table 3) followed by the presence of conserved C²-x-C⁴ nucleotides in the D-arm. The D-loop contained conserved A¹-G²-x₃-G⁶-x-U⁸-A⁹ nucleotides followed by the presence of conserved U¹-U²-x-A⁴-C⁵-A⁶ nucleotides in the anti-codon arm (Table 2). The variable region contained conserved G³-U⁴-C⁵ nucleotides, followed by the presence of conserved G⁴-G⁵ nucleotides in the Ψ-arm and U¹-U²-C³-x-A⁵ nucleotides in the Ψ-loop (Table 2).

The acceptor arm of tRNA^{Leu} contained variable nucleotides. However, more than half of the tRNA^{Leu} were contained G¹-C²-G³ nucleotides (Table 2). The U⁸ and U⁹ nucleotide were conserved at the 8th and 9th position, respectively (Table 3), whereas the D-arm did not contain any conserved nucleotide sequences whereas the D-loop contained conserved A¹-A²-x₂-G⁵-G⁶-x-A⁸ nucleotides. The anti-codon loop contained

conserved U²-x-A⁴ nucleotides followed by the presence of G⁵ nucleotide in the Ψ-arm and U¹-U²-C³-x-A⁵-x-U⁷ nucleotides in the Ψ-loop (Table 2). No conserved nucleotides were found in the variable loop.

The acceptor arm of tRNA^{Ile} contained conserved G¹-G²-G³-C⁴ nucleotides (Table 2). The nucleotides U⁸ and A⁹ were also found to be conserved in their respective position (Table 3) followed by the presence of conserved G¹-C²-U³-C⁴ nucleotides in the D-arm and A¹-x₄-G⁶-x₂-A⁹ nucleotides in the D-loop. The anti-codon loop contained conserved C¹-U²-x-A⁴-U⁵-A⁶-A⁷ nucleotides followed by the presence of conserved G³-U⁴ nucleotides in the variable region, G⁴-G⁵ nucleotides in the Ψ-arm and U¹-U²-C³-x-A⁵-x-U⁷ nucleotides in the Ψ-loop (Table 2).

In tRNA^{Met}, no consensus conserved nucleotides were found at the 5' end of the acceptor arm. A few tRNAs were contained G¹-G²-C³ nucleotides whereas some other were contained C¹-G²-C³, or C¹-C²-A³ nucleotides (Table 2). The U nucleotide at the 8th position was conserved in the acceptor arm whereas the 9th position was occupied by either A⁹/G⁹ or U⁹ nucleotide. The D-arm and the D-loop was found to contain conserved G² and G⁶ nucleotide, respectively. The anti-codon loop contained conserved C¹-U²-C³-A⁴-U⁵-A⁶-A⁷ nucleotides followed by the presence of conserved G⁵ nucleotide in the Ψ-arm and U¹-U²-C³-x-A⁵-x-U⁷ nucleotides in the Ψ-loop (Table 2).

The acceptor arm of tRNA^{Phe} contained conserved C¹/G¹-C²-C³-x₂-G⁶ nucleotides (Table 2). The nucleotides U⁸ and A⁹ in the acceptor arm were conserved (Table 3) followed by the presence of conserved G¹-C²-U³-C⁴ nucleotides in the D-arm and A¹-G²-U³-U⁴-G⁵-G⁶-U⁷ nucleotides in the D-loop, respectively. In the majority of the cases, the anti-codon arm contained conserved G⁴

TABLE 3 | Variations of nucleotide conservation at the 8th and 9th position of tRNAs in the 5' end of the acceptor arm in cyanobacteria.

tRNA	8th position	9th position
Glycine (Gly)	U	A
Alanine (Ala)	U	A
Proline (Pro)	U	A
Valine (Val)	U	A
Leucine (Leu)	U	G
Isoleucine (Ile)	U	A
Methionine (Met)	U	A/G/U
Phenylalanine (Phe)	U	A
Tyrosine (Tyr)	U	G
Tryptophan (Trp)	U	A
Serine (Ser)	U	G
Threonine (Thr)	U	A/G
Cysteine (Cys)	U	A/C/G
Asparagine (Asn)	U	A
Glutamine (Gln)	U	A/C/G
Lysine (Lys)	U	A/G
Histidine (His)	U/A	A/C
Arginine (Arg)	U	A/G
Aspartate (Asp)	U	A
Glutamate (Glu)	U/C/G	A/C/U

nucleotide and the anti-codon loop contained conserved U²-G³-A⁴-A⁵-x-A⁷ nucleotides. The variable loop contained conserved G³-U⁴-C⁵ nucleotides followed by the presence of conserved G⁵ nucleotide the Ψ-arm and U¹-U²-C³-x-A⁵ nucleotides in the Ψ-loop (**Table 2**).

In tRNA^{Tyr}, the acceptor arm contained conserved G¹-G²-G³-U⁴-C⁵ nucleotides (**Table 2**). The U⁸ and G⁹ nucleotides in the acceptor arm were conserved (**Table 3**) followed by the presence of conserved G¹-x-C³-C⁴ nucleotides in the D-arm and A¹-G²-U³-G⁴-G⁵-U⁶-U⁷-A⁸ nucleotides in the D-loop, respectively (**Table 2**). The anti-codon loop contained conserved U²-G³-U⁴-A⁵ nucleotides followed by the presence of conserved G⁴-G⁵ nucleotides in the Ψ-arm and U¹-U²-C³-x-A⁵-x-U⁷ nucleotides in the Ψ-loop (**Table 2**).

In the majorities of cases, the acceptor arm of tRNA^{Trp} contained a conserved G¹ nucleotide whereas few others were found to contain A¹ nucleotide. The U⁸ nucleotide of the acceptor arm was conserved whereas A⁹ nucleotide was sometimes substituted with G⁹ nucleotide (**Tables 2, 3**). The D-arm contained conserved G¹-U² nucleotides followed by the presence of conserved A¹-x₃-G⁵ nucleotides in the D-loop. The anti-codon arm contained conserved G⁴-U⁵ nucleotides followed by the presence of conserved C¹-U²-C³-C⁴-A⁵-A⁶-A⁷ nucleotides in the anti-codon loop, a G⁵ nucleotide in the Ψ-arm and U¹-U²-C³-x-A⁵-x-U⁷ nucleotides in the Ψ-loop, respectively (**Table 2**).

In tRNA^{Ser}, the acceptor arm contained conserved G¹-G²-A³ nucleotides (**Table 2**). A few of the tRNA^{Ser} were contained U¹ or G¹ nucleotides instead of G¹-G²-A³ consensus nucleotide sequence. The U⁸ and G⁹ nucleotide were found to be conserved (**Table 3**) followed by the presence of

conserved G¹-C² nucleotides in the D-arm and A²-x₂-G⁵-G⁶ nucleotides in the D-loop, respectively. The anti-codon loop contained conserved U²-x₃-A⁵-A⁶ nucleotides (**Table 2**) followed by the presence of conserved C²¹ nucleotide in variable region, A⁴/G⁴-G⁵ nucleotides in the Ψ-arm and U¹-U²-C³-x-A⁵-x-U⁷ nucleotides in the Ψ-loop, respectively (**Table 2**).

The acceptor arm of tRNA^{Thr} contained conserved G¹-C² nucleotides (**Table 2**). The U⁸ nucleotide was found to be conserved whereas the 9th position was occupied by A⁹ or G⁹ nucleotide (**Table 3**). The D-arm contained conserved G¹-C²-x-C⁴ nucleotides and A¹-x₁₋₄-G³-G⁴-x-A⁶ nucleotides in the D-loop, respectively. The anti-codon loop contained conserved U²-x-G⁴-U⁵-A⁶-A⁷ nucleotides followed by the presence of conserved G³ nucleotide in the variable loop, G⁵ nucleotide in the Ψ-arm and U¹-U²-C³-x-A⁵ nucleotides in the Ψ-loop, respectively (**Table 2**). At the 7th position of the Ψ-loop, the U⁷ nucleotide was sometimes substituted with C⁷ nucleotide.

The acceptor arm of tRNA^{Cys} contained a conserved G¹ nucleotide. The U⁸ and A⁹ nucleotides were found to be conserved while in some cases, A⁹ nucleotide was substituted with either C⁹ or G⁹ nucleotides (**Table 3**). The D-arm contained conserved G¹-C²-C³ nucleotides followed by the presence of conserved A¹-A²-G³-x-G⁵-G⁶-U⁷ and C¹-U²-G³-C⁴-A⁵-A⁶-A⁷ nucleotides in the D-loop and anti-codon loop, respectively. The C⁵ nucleotide was conserved in the variable region of the tRNA^{Cys} followed by the presence of conserved G⁵ and U¹-U²-C³-x-A⁵-x-U⁷ nucleotides in the Ψ-arm and Ψ-loop, respectively (**Table 2**).

The acceptor arm of tRNA^{Asn} contained conserved U¹-C²-C³-x-C⁵ nucleotides (**Table 2**) with a few exceptions. The U⁸ and A⁹ nucleotides of the acceptor arm were conserved (**Table 3**) followed by the presence of conserved G¹-C²-U³ nucleotides in the D-arm and A¹-x₁₋₂-G³-G⁴ nucleotides in the D-loop, respectively (**Table 2**). The anti-codon loop contained conserved C¹-U²-G³-U⁴-U⁵-A⁶-A⁷ nucleotides followed by the presence of conserved G³-U⁴ nucleotides in the variable region, G⁵ in the Ψ-arm and U¹-U²-C³-x-A⁵-x-U⁷ nucleotides in the Ψ-loop, respectively (**Table 2**).

The acceptor arm of tRNA^{Gln} contained conserved U¹-G²-x₃₋₄-G⁶ nucleotides (**Table 2**). The U⁸ nucleotide of the acceptor arm was conserved whereas the 9th position was substituted with either A⁹/C⁹, or G⁹ nucleotides (**Table 3**). The D-loop contained conserved A¹/U¹-A²-x₂-G⁴ nucleotides followed by the presence of conserved U¹-U²-U³-U⁴-G⁵ nucleotides in the anti-codon loop, A³-G⁴-G⁵ nucleotides in the Ψ-arm and U¹-U²-C³-x-A⁵ nucleotides in the Ψ-loop, respectively (**Table 2**).

The acceptor arm at tRNA^{Lys} contained conserved G¹-G²-G³ nucleotides (**Table 2**). In a few cases, the G¹ nucleotide at the 1st position was substituted with either C¹ or U¹ nucleotide. The nucleotide U⁸ and A⁹ in the acceptor arm were conserved (**Table 3**) followed by the presence of conserved C² nucleotide in the D-arm and A¹-x₂-G⁴-G⁵ nucleotides in the D-loop. The anti-codon loop contained conserved C¹-U²-U³-U⁴-U⁵-A⁶-A⁷ nucleotides, followed by the presence of conserved G⁴-G⁵ sequence in the Ψ-arm and U¹-U²-C³-x-A⁵-x-U⁷ nucleotides in the Ψ-loop (**Table 2**). In a few cases, the U⁷ nucleotide in the Ψ-loop was substituted with C⁷ nucleotide (**Table 2**).

In tRNA^{His}, the acceptor arm contained conserved G¹-C²/U²-G³-x₃-G⁷ nucleotides with a few exceptions (Table 2). The U⁸ and A⁹ nucleotides in the acceptor arm were conserved followed by the presence of conserved G¹-C²-C³ nucleotides in the D-arm and A¹-A²-G³-x-G⁵-G⁶-U⁷ nucleotides in the D-loop, respectively (Table 2). The anti-codon arm contained conserved G⁴-G⁵ nucleotides followed by the presence of conserved U¹-U²-G³-U⁴-G⁵ nucleotides in the anti-codon loop, G¹-x-G³-G⁴-G⁵ nucleotides in the Ψ-arm and U¹-U²-C³-x-A⁵-x-U⁷ nucleotides in the Ψ-loop. In a few cases, the U⁷ nucleotide in the Ψ-loop was substituted with C⁷ nucleotide (Table 2).

In tRNA^{Arg}, the acceptor arm at the 5' end contained conserved G¹-G² nucleotides (Table 2). In a few cases, the G¹ nucleotide at the 1st position substituted by A¹ or C¹ nucleotide. The U⁸ nucleotide in the acceptor arm was conserved whereas the 9th position was substituted with A⁹/G⁹ or C⁹ nucleotides (Table 3). The D-arm was found to contain conserved G¹ nucleotide, followed by the presence of conserved A¹-x₂₋₃-G⁴-x₃-A⁸ nucleotide in the D-loop and C¹/U¹-U²-x-C⁴ nucleotides in the anti-codon loop (Table 2). The Ψ-arm and Ψ-loop was found to contain conserved G⁴-G⁵ and U¹-U²-C³-x-A⁵-x-U⁷ nucleotides, respectively (Table 2).

The acceptor arm of tRNA^{Asp} contained conserved G¹-G²-G³ nucleotides (Table 2). A few species were found to contain C¹ nucleotide instead of G¹ nucleotide at the 1st position. The U⁸ and A⁹ nucleotides in the acceptor arm were conserved (Table 3), followed by the presence of conserved G¹-x-U³-C⁴ and A¹-x₂₋₃-G⁴-G⁵-U⁶-x-A⁸ nucleotides in the D-arm and D-loop, respectively (Table 2). The anti-codon loop contained conserved C¹-U²-G³-U⁴-C⁵-A⁶ nucleotides followed by the presence of conserved A¹-x-G³-U⁴ nucleotides in the variable region, G¹-x-G³-x-G⁵ nucleotides in the Ψ-arm and U¹-U²-C³-x-A⁵-x-C⁷ nucleotides in the Ψ-loop, respectively (Table 2). The C⁷ nucleotide in the Ψ-loop sometimes substituted with U⁷ nucleotide.

In tRNA^{Glu}, the majorities of species were contained conserved G¹ nucleotide in the acceptor arm (Table 2). In a few species, G¹ nucleotide was substituted by A¹ nucleotide at the first position. The U⁸ nucleotide in the acceptor arm was conserved whereas the 9th position was substituted by A⁹/G⁹ or C⁹ nucleotides (Table 3). The D-arm was devoid of any conserved nucleotide sequence whereas the D-loop contained conserved G⁵ nucleotide. The anti-codon loop contained conserved U²-U³-U⁴-C⁵ nucleotides, followed by the presence of conserved G³-x-G⁵ and U¹-U²-C³-x-A⁵-x-U⁷ nucleotides in the Ψ-arm and Ψ-loop, respectively. The U⁷ nucleotide in the Ψ-loop sometimes substituted with the C⁷ nucleotide.

Cyanobacterial tRNAs Contain Group I Intron

In our study, cyanobacterial tRNA was found to contain group I introns. The intron was present in the anti-codon loop region of the tRNA in the *Nostoc* sp. PCC 7524 (gene id: 2509813156), *Gloeocapsa* sp. PCC 73106 (gene id: 2508643885), and *Nostoc* sp. PCC 7107 (gene id: 2503742551) (Figure 2). tRNA^{Arg} (ACG) of *Nostoc* sp. PCC 7524 contained an intron from nucleotide

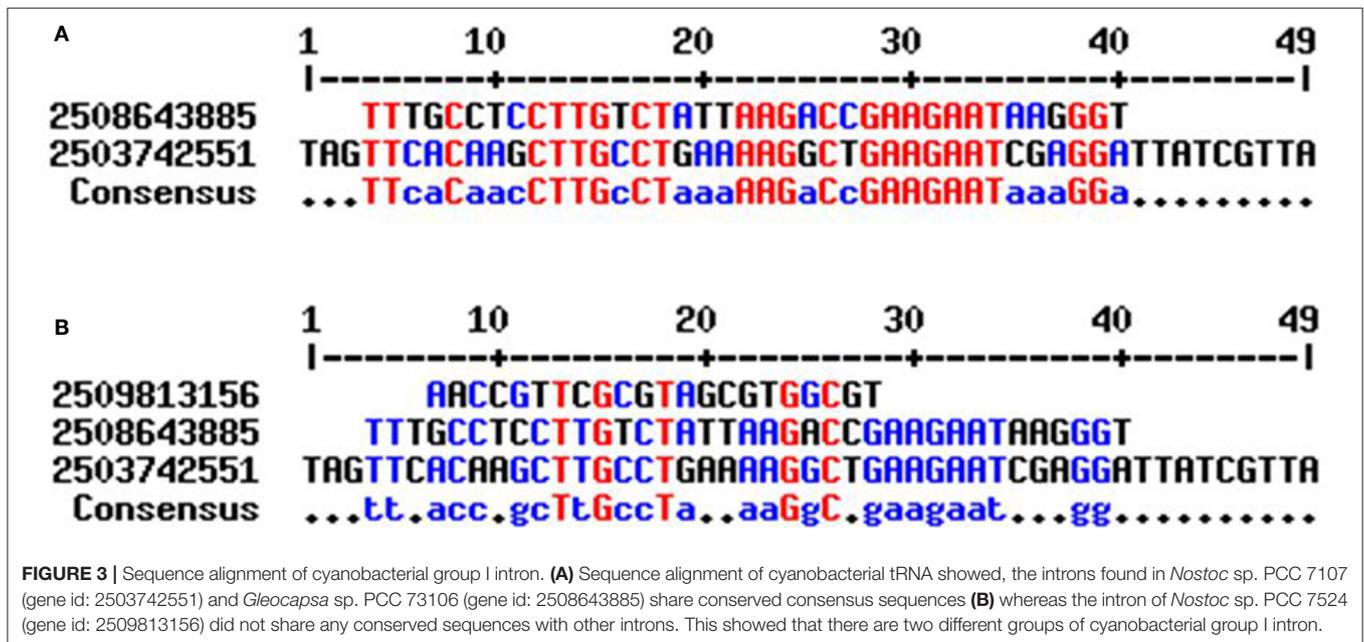
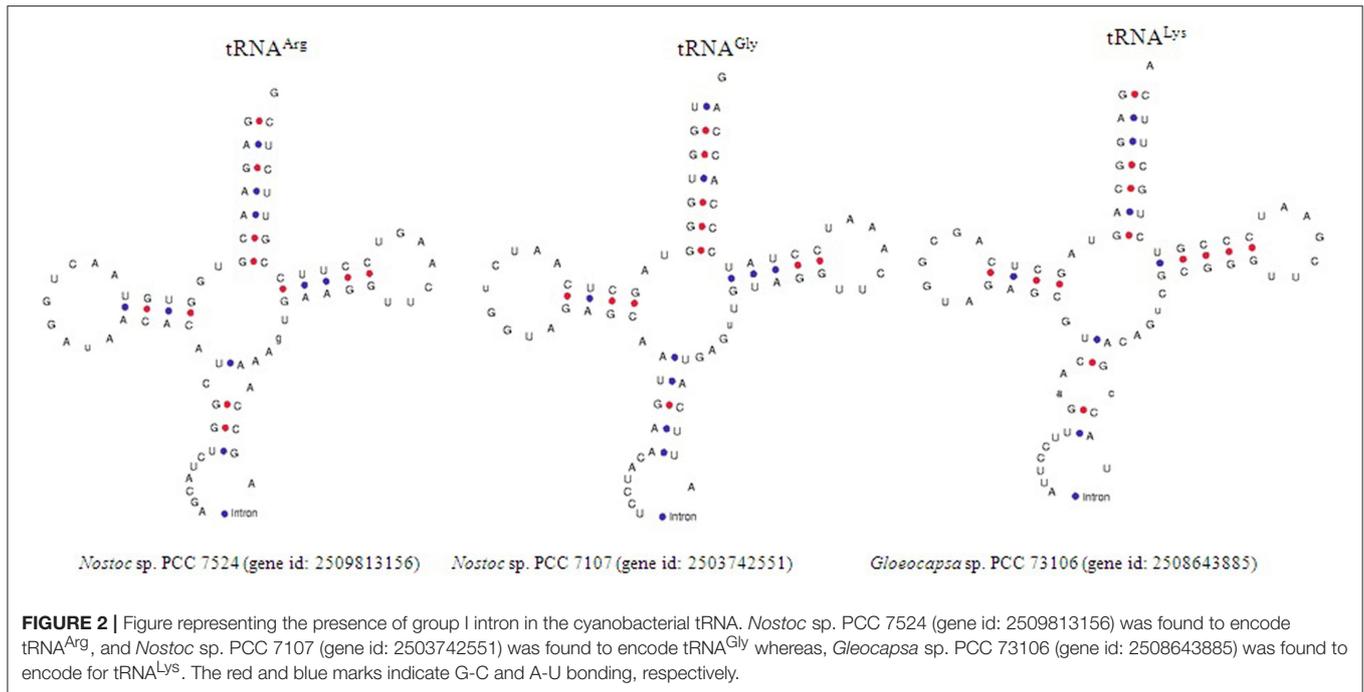
39–60 whereas in tRNA^{Gly} (TCC) of the *Nostoc* sp. PCC 7107 the intron was found from nucleotide 39–87. One of the genes encoding tRNA^{Lys} (CTT) of the *Gloeocapsa* sp. PCC 73106 was found to contain an intron, from nucleotide 38 to 74. Previous studies reported that an intron was present only in tRNA^{Leu} (UAA) and tRNA^{fMet} (UAC) of cyanobacterial tRNA (Paquin et al., 1997; Rudi and Jakobsen, 1999). However in our study, we found that the cyanobacterial group I intron was also present in tRNA^{Arg}, tRNA^{Gly}, and tRNA^{Lys} (Figure 2). Multiple sequence alignment has shown the presence of conserved T-x-G-x₂-T and G-x-C motifs in the introns of cyanobacterial tRNA (Figure 3). Sequence alignment between the introns of *Gloeocapsa* sp. PCC 73106 (tRNA^{Lys}) and *Nostoc* sp. PCC 7107 (tRNA^{Gly}) showed the presence of conserved T-T-x₂-C, C-T-T-G-x-C-T, A-A-G-x-C, G-A-A-G-A-A-T, and G-G motifs. The length of the intron ranged between 22 and 49 nucleotides. The introns of the *Nostoc* sp. PCC 7524 (gene id: 2509813156) and *Nostoc* sp. PCC 7107 (gene id: 2503742551) were found to be 22 and 38 nucleotides, respectively, whereas that of the *Gloeocapsa* sp. PCC 73106 (gene id: 2508643885) was 49 nucleotides long. No introns were found in other parts of the tRNA except for the anti-codon loop.

The Rate of Transition of Cyanobacterial tRNAs Were Higher Than the Rate of Transversion

The genes of cyanobacterial tRNA are found to range from 55 to 96 nucleotides; a tRNA gene possessing only 55 nucleotides is the smallest gene reported thus far. Studying the substitution pattern of mutation is important for understanding the molecular basis of evolution. Therefore, we examined the substitutions (transition/transversion) of cyanobacterial tRNA. The mutation, which changes a purine (A<>G) nucleotide to another purine or a pyrimidine nucleotide to another pyrimidine (T<>C) is called transition, whereas a mutation that changes a purine nucleotide to a pyrimidine and vice versa, is called a transversion. In our study, we found that the rate of transition, in cyanobacterial tRNA, was higher than that of transversion (Table 4). The highest rate of transition was 42.05 in tRNA^{Trp} followed by 25.66 in tRNA^{Cys}, where an adenine was substituted with a guanine (Table 4); the highest rate of transversion was 9.7 in tRNA^{Ser}, where a uracil and a cytosine were substituted with a guanine (Table 4). The highest rate of transition from a guanine to an adenine was found in tRNA^{Trp} (28.89), whereas the highest rate of transition from a uracil to a cytosine was found in tRNA^{Tyr} (Table 4). The highest rate of transition from a cytosine to a uracil was found in tRNA^{Tyr}. The lowest rate of transition was found in tRNA^{Ser} where a guanine was substituted with an adenine. Similarly, the lowest rate of transversion was found in tRNA^{Trp} where a uracil and a cytosine were substituted by an adenine (Table 4). The transition/transversion bias of tRNA^{Trp} was highest, followed by tRNA^{Asp} and tRNA^{Tyr} (Table 4).

Cyanobacteria Species Evolved via the Loss of tRNA Genes

In addition to substitution (transition/transversion), gene duplication and loss play crucial roles in the evolution of a gene.



Duplication study of cyanobacterial tRNA showed that tRNA genes in all 20 tRNAs families, were duplicated by more than 50% (Table 5, Supplementary Figure 1). The highest duplication of 85% was observed in tRNA^{Ile}, whereas the lowest duplication was observed in tRNA^{His} (Table 5). The highest percentage of conditional duplication was found in tRNA^{Asn} (20.75%), whereas the lowest percentage was found in tRNA^{Ile} (5.26%), followed by tRNA^{Gln} (5.68%) and tRNA^{Asp} (5.84%) (Table 5). Unlike duplication and conditional duplication, the highest percentage

of losses with respect to the species tree (Figure 4), was found in tRNA^{Asp} (284.93%) followed by tRNA^{Glu} (280.76%). The low level of loss was found in tRNA^{Arg} (221.45%) followed by tRNA^{Ala} (226.63%).

Cyanobacterial tRNAs May Be Polyphyletic

Understanding the evolution of cyanobacterial tRNA is important for delineating the evolution of tRNA, its evolutionary lineages, and subsequent diversification. In this study, we

TABLE 4 | Maximum composite likelihood transition/transversion rate of cyanobacterial tRNAs.

	A	U	C	G
GLYCINE				
A	–	4.61	4.76	17.77
U	3.43	–	17.54	5.25
C	3.43	16.99	–	5.25
G	11.59	4.61	4.76	–
SERINE				
A	–	6.6	7.18	9.66
U	6.19	–	12.92	9.7
C	6.19	11.88	–	9.7
G	6.17	6.6	7.18	–
THREONINE				
A	–	4.66	4.63	17.7
U	3.93	–	16.07	5.69
C	3.93	16.19	–	5.69
G	12.22	4.66	4.63	–
CYSTEINE				
A	–	3.74	5.37	25.66
U	4.2	–	10.05	5.46
C	4.2	7	–	5.46
G	19.73	3.74	5.37	–
TYROSINE				
A	–	2.19	2.65	17.47
U	1.96	–	28.31	3.1
C	1.96	23.4	–	3.1
G	11.04	2.19	2.65	–
ASPARAGINE				
A	–	4.73	4.39	19.84
U	3.73	–	14.51	6.1
C	3.73	15.63	–	6.1
G	12.12	4.73	4.39	–
GLUTAMINE				
A	–	4.47	5.18	10.59
U	3.72	–	23.69	5.97
C	3.72	20.46	–	5.97
G	6.59	4.47	5.18	–
ALANINE				
A	–	6.84	6.95	13.53
U	5.62	–	10.62	9.1
C	5.62	10.46	–	9.1
G	8.36	6.84	6.95	–
VALINE				
A	–	5.96	6.37	13.83
U	5.53	–	13.07	7.43
C	5.53	12.23	–	7.43
G	10.3	5.96	6.37	–
ASPARTATE				
A	–	2.11	2.66	20.62
U	1.64	–	27.74	3.27
C	1.64	21.99	–	3.27
G	10.33	2.11	2.66	–
LEUCINE				
A	–	5.68	5.92	15.54
U	5.04	–	13.02	7.86
C	5.04	12.5	–	7.86
G	9.96	5.68	5.92	–

(Continued)

TABLE 4 | Continued

	A	U	C	G
ISOLEUCINE				
A	–	5.56	5.97	21.94
U	4.74	–	9.12	7.94
C	4.74	8.04	–	7.94
G	13.09	5.56	5.59	–
PROLINE				
A	–	3.61	4.38	22.18
U	2.82	–	18.54	5.95
C	2.82	15.27	–	5.95
G	10.49	3.61	4.38	–
PHENYLALANINE				
A	–	5.89	6.63	14.24
U	5.96	–	11.59	8.46
C	5.96	10.29	–	8.46
G	10.03	5.89	6.63	–
TRYPTOPHAN				
A	–	1.24	1.34	42.05
U	1.09	–	9.65	1.58
C	1.09	8.91	–	4.58
G	28.89	1.24	1.34	–
METHIONINE				
A	–	4.42	5.9	13.38
U	4.58	–	19.75	6.15
C	4.58	14.79	–	6.15
G	9.97	4.42	5.9	–
LYSINE				
A	–	4.77	5.02	16.76
U	3.97	–	16.71	5.34
C	3.97	15.88	–	5.34
G	12.46	4.77	5.02	–
ARGININE				
A	–	4.86	5.86	19.01
U	4.06	–	14.36	7.29
C	4.06	11.91	–	7.29
G	10.58	4.86	5.86	–
HISTIDINE				
A	–	5.66	5.89	20.01
U	3.86	–	12.2	7.46
C	3.86	11.71	–	7.46
G	10.35	5.66	5.89	–
GLUTAMATE				
A	–	4.97	6.81	19.43
U	4.45	–	12.61	6.49
C	4.45	9.19	–	6.49
G	13.32	4.97	6.81	–
ALL TRNAS				
A	–	5.65	6.28	16.97
U	4.91	–	12.2	7.64
C	4.91	10.98	–	7.64
G	10.9	5.65	6.28	–

Analyses showed that the rate of transition of cyanobacterial tRNAs are higher than the rate of transversion. The values marked bold in the table represent the rate of "transition" whereas italicized are represents rate of "transversion."

constructed a phylogenetic tree by examining all the tRNAs of the studied species (Figure 5). The phylogenetic tree was divided into five distinct clusters (Supplementary Figure 1). Cluster I contained tRNA^{Ala} (red), tRNA^{Val} (blue), tRNA^{Ile} (purple), tRNA^{Arg} (teal), tRNA^{Lys} (lime), tRNA^{Glu} (green), tRNA^{Asp} (maroon), and tRNA^{Gly} (green). Cluster II was found to contain tRNA^{Pro} (red), tRNA^{Thr} (fuchsia), tRNA^{Phe} (blue), tRNA^{Met} (olive), tRNA^{Asn} (navy), and tRNA^{Phe} (blue) (Supplementary Figure 1). Cluster III contained tRNA^{Arg} (teal) and tRNA^{Trp} (fuchsia) whereas cluster IV had tRNA^{Lys} (lime), tRNA^{Leu} (maroon), tRNA^{Gln} (aqua), tRNA^{Leu} (maroon), tRNA^{Val} (blue), tRNA^{Glu} (green), tRNA^{His} (red), tRNA^{Ala} (black), tRNA^{Leu} (maroon), tRNA^{Gly} (green), tRNA^{Trp} (fuschia), tRNA^{Gln} (aqua), tRNA^{His} (red), tRNA^{Cys} (navy), tRNA^{Tyr} (gray), tRNA^{Met} (olive), and tRNA^{Ser} (fuschia) (Supplementary Figure 1). Cluster V contained tRNA^{Arg} (teal), tRNA^{Met} (olive), tRNA^{Ala} (red), tRNA^{Met} (olive), and tRNA^{Gln} (aqua). In cluster I, tRNA^{Ala} and tRNA^{Val} were present together with tRNA^{Val} and tRNA^{Ile}. In cluster I tRNA^{Arg} was shared by tRNA^{Asp}, and tRNA^{Gly}. Cluster II was shared by tRNA^{Pro}, tRNA^{Thr}, tRNA^{Phe}, tRNA^{Met}, and tRNA^{Asn} (Supplementary Figure 1). Cluster III was shared by tRNA^{Arg} and tRNA^{Trp}. Cluster IV contained tRNA^{Lys}, tRNA^{Leu}, tRNA^{Gln}, tRNA^{Val}, tRNA^{His}, tRNA^{Ala}, tRNA^{Gly}, tRNA^{Trp}, tRNA^{Gln}, tRNA^{Cys}, tRNA^{Tyr}, tRNA^{Met}, and tRNA^{Ser}. Cluster V contained tRNA^{Ala}, tRNA^{Met}, tRNA^{Gln}, and tRNA^{Arg}. tRNA^{Ala} was distributed in clusters I, V, and VI whereas tRNA^{Gln} was distributed in clusters IV, V, and VI. tRNA^{Val}, tRNA^{Lys}, and tRNA^{Gly} were distributed in clusters I and IV; tRNA^{Asn} was distributed in cluster II and IV, and tRNA^{Trp} was distributed in cluster III and IV (Supplementary Figure 1). Due to low bootstrap values in few clades, it was difficult to infer the phylogenetic result properly. Therefore, we collapsed the clades with low bootstrap values (cut-off values for condensed tree was 50%) (Figure 5, Supplementary Figure 1). The overall tree architecture remained unchanged post-collapsing the clades with 50% cut-off values. Several potentially novel evolutionary trends were found in the phylogenetic analysis of cyanobacterial tRNA. In between tRNA^{Met} and tRNA^{Asn} of cluster II, several other tRNAs were also found. In cluster V, tRNA^{Asp} and tRNA^{Asn} were found in close to tRNA^{Met}, suggesting their possible evolution from tRNA^{Met}. Similarly, a few tRNAs^{Val} (*Microcoleus* sp. PCC 7113, gene id: 2509433894; *Leptolyngbya boryana* PCC 6306, gene id: 2509801628) were present close to tRNA^{Ala}. This suggests that, tRNA^{Val} has likely evolved from tRNA^{Ala}. In cluster V, tRNA^{Gly} (*Microcoleus* sp. PCC 7113, gene id: 2509433891) was clustered with tRNA^{Met}, reflecting its evolution from tRNA^{Met}.

DISCUSSION

Nucleotide sequence conservation is an important phenomenon that demonstrates conserved functional role. Therefore, understanding the conserved nucleotide consensus sequences in cyanobacterial tRNA was very important. In this study we found that except for tRNA^{Asn}, tRNA^{Leu}, tRNA^{Gln}, and tRNA^{Met}, all the tRNAs were contained conserved G¹ nucleotide at the 1st

TABLE 5 | Duplication, conditional duplications and losses of cyanobacterial tRNAs.

tRNA	k1 (Purines)	k2 (Pyrimidines)	R (Transition/Transversion Bias)	No. of sequences Studied
Alanine	1.494	1.515	0.732	258
Arginine	2.607	2.449	1.207	255
Asparagine	3.389	3.268	1.618	90
Aspartate	6.971	10.501	4.131	84
Cysteine	4.699	1.87	1.65	78
Glutamate	3.951	2.82	1.626	87
Glutamine	1.626	4.461	1.499	94
Glycine	3.2	3.426	1.631	192
Histidine	2.832	2.25	1.199	68
Isoleucine	0.231	1.512	0.39	124
Leucine	1.845	2.114	0.964	343
Lysine	3.149	3.386	1.623	126
Methionine	1.805	3.28	1.231	202
Phenylalanine	1.619	1.661	0.811	89
Proline	3.708	4.158	1.825	204
Threonine	3.133	3.453	1.618	202
Tryptophan	26.77	7.30	8.409	83
Tyrosine	5.114	10.015	3.658	76
Serine	0.914	1.82	0.654	271
Valine	1.959	2.169	1.018	159

Highest transition/transversion bias was found in tRNA^{Trp} whereas the lowest was found in tRNA^{Ile}.

position in the acceptor arm. tRNA^{Leu} and tRNA^{Met} did not contain any conserved nucleotides at the 1st position whereas tRNA^{Asn} and tRNA^{Gln} contained a conserved U¹ nucleotide at the 1st position. With a few exceptions, the G² nucleotide was conserved at the 2nd position in tRNA^{Ala}, tRNA^{Val}, tRNA^{Ile}, tRNA^{Phe}, tRNA^{Tyr}, tRNA^{Ser}, tRNA^{Gln}, tRNA^{Lys}, tRNA^{His}, tRNA^{Arg}, and tRNA^{Asp}. The nucleotide C² was conserved in tRNA^{Gly} and tRNA^{Thr} whereas tRNA^{His} contained either a C² or a U² nucleotide. At the 3rd position, the G³ nucleotide was conserved in tRNA^{Gly}, tRNA^{Ala}, tRNA^{Pro}, tRNA^{Ile}, tRNA^{Tyr}, tRNA^{Lys}, and tRNA^{Asp}, whereas tRNA^{Phe} contained a C³; tRNA^{His} contained C³/U³; whereas tRNA^{Ser} contained an A³ nucleotide. At the 4th position of the acceptor arm, tRNA^{Val} and tRNA^{Ile} contained the conserved C⁴ nucleotide whereas tRNA^{Tyr} contained a U⁴ nucleotide and tRNA^{Pro} contained the G⁴ nucleotide. In tRNA, the C⁵ nucleotide was conserved, whereas tRNA^{Phe} and tRNA^{Gln} contained the conserved G⁶ nucleotide. In all canonical tRNA, 1st, 2nd, and 3rd base of the acceptor arm pairs with the 72nd, 71st, and 70th base, respectively of the acceptor arm. Their conserved structure is highly important to have proper functional clover leaf-like structure of tRNA. At the 7th position, tRNA^{Pro} and tRNA^{His} contained conserved G⁷ nucleotide. In tRNA^{Tyr}, all the nucleotides in the positions 1–5 were conserved across the species. In the 1st and 2nd positions of the acceptor arm of tRNA, the G¹-G² conserved consensus sequence present in tRNA^{Ala}, tRNA^{Val}, tRNA^{Ile}, tRNA^{Tyr}, tRNA^{Ser}, tRNA^{Lys}, tRNA^{Arg}, and tRNA^{Asp} whereas the G¹-C²

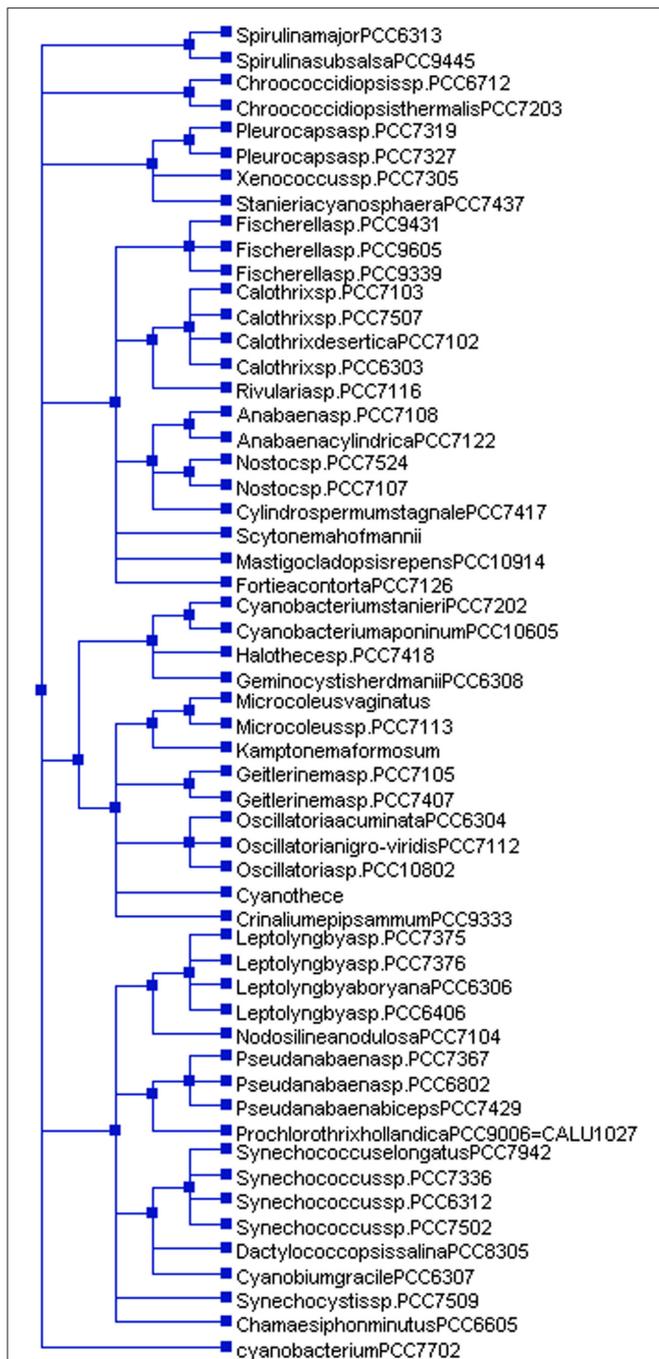


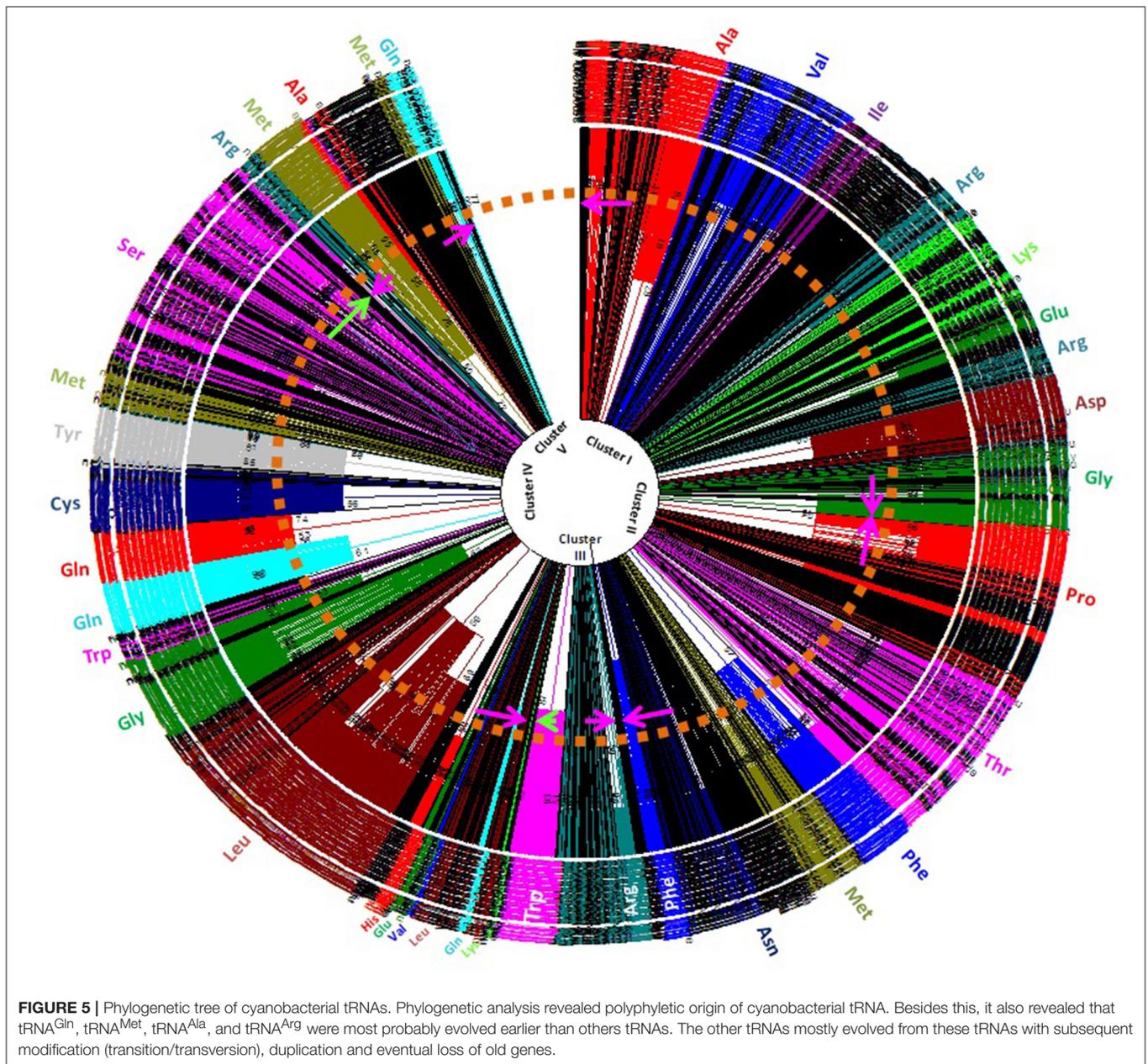
FIGURE 4 | Species tree of cyanobacteria. The species tree was constructed using NCBI taxonomy browser (<https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>). *Cyanobacterium* sp. PCC 7702 in the species tree falls independently with regard to other cyanobacterial species. This shows cyanobacterial groups show polyphyletic origins.

conserved consensus sequence present in tRNA^{Gly}, tRNA^{Phe}, tRNA^{Thr}, and tRNA^{Lys}. The nucleotide, at the 2nd position of the acceptor arm, was substituted with either G² or C², whereas only tRNA^{His} had a U² at the 2nd position. Only tRNA^{Asn} and tRNA^{Gln} contained a U¹ nucleotide at the 1st position

of the acceptor arm whereas tRNA^{Tyr} contained conserved U⁴ nucleotide at the 4th position and tRNA^{His} contained the conserved U² nucleotide at the 2nd position (Table 2). Except the above-mentioned four tRNAs, none of the other tRNAs were found to encode U nucleotide in the acceptor arm. Previously, the 8th position of tRNA between the acceptor arm and the D-arm was thought to contain a conserved U⁸ nucleotide (Table 3). In this study, the majority of tRNAs were found to contain a conserved U⁸ nucleotide. However, tRNA^{His} was found to contain either a U⁸ or an A⁸ whereas tRNA^{Glu} contained a U⁸, C⁸ or a G⁸ nucleotide (Table 3). This showed that the 8th position of cyanobacterial tRNA can be substituted with any nucleotide. Similarly, at the 9th position, the majority of tRNAs were found to contain a conserved A⁹ nucleotide (Table 3). However, tRNA^{Leu}, tRNA^{Tyr} and tRNA^{Ser} were found to contain a G⁹ nucleotide whereas tRNA^{Thr}, tRNA^{Lys}, and tRNA^{Arg} contained an A⁹ or a G⁹ nucleotide (Table 3). tRNA^{His} contained an A⁹ or a C⁹ nucleotide at the 9th position whereas tRNA^{Cys} and tRNA^{Gln} contained either a A⁹, C⁹, or a G⁹ (Table 3). tRNA^{Met} contained either an A⁹, G⁹ or a U⁹ nucleotide whereas tRNA^{Glu} contained an A⁹, C⁹, or a U⁹ nucleotide at the 9th position. This negated the previous assumption about the conservation of the U⁸ nucleotide in cyanobacterial tRNA and showed that it can be prone to variations; the major variation was observed in tRNA^{Glu} which was found to contain varied nucleotides at 8th and 9th position.

In the D-arm, the majority of tRNA contained a conserved G¹ nucleotide at the 1st position except for tRNA^{Gly}, tRNA^{Pro}, tRNA^{Val}, tRNA^{Met}, tRNA^{Gln}, tRNA^{Lys}, and tRNA^{Glu}. In total of 20 tRNA families, 10 families were found to contain a conserved C² nucleotide at the 2nd position (Table 2). These were tRNA^{Ala}, tRNA^{Val}, tRNA^{Ile}, tRNA^{Phe}, tRNA^{Ser}, tRNA^{Thr}, tRNA^{Cys}, tRNA^{Asn}, tRNA^{Lys}, and tRNA^{His} whereas tRNA^{Met} was found to contain a G² and tRNA^{Trp} contained a U² nucleotide at the 2nd position (Table 2). With respect to the first two nucleotides in the D-arm, the G¹-C² conserved consensus sequence can be found in tRNA^{Ala}, tRNA^{Ile}, tRNA^{Phe}, tRNA^{Ser}, tRNA^{Thr}, tRNA^{Cys}, tRNA^{Asn}, and tRNA^{His}. At the 3rd position, the U³ nucleotide conserved in tRNA^{Ala}, tRNA^{Ile}, tRNA^{Phe}, tRNA^{Asn}, and tRNA^{Asp}. At the 4th position of the D-arm, the C⁴ nucleotide was found to be conserved in tRNA^{Ala}, tRNA^{Val}, tRNA^{Ile}, tRNA^{Phe}, tRNA^{Tyr}, tRNA^{Thr}, and tRNA^{Asp} (Table 2). The D-stem expected to have rigid structure due to high G/C content, but due to the presence of more A-U and G-U base pairing, it seems rather weak (Hardt et al., 1993). Previous study reported that D-arm play important role in recognition of aminoacyl-tRNA synthetase and is highly variable and possess unusual conformation due to over-crowding of G residue (Smith and Yarus, 1989; Hardt et al., 1993). However, in our study we found that D-arm is also highly conserved, suggesting its common functional role.

Except for tRNA^{Met} and tRNA^{Glu}, 18 tRNA families were found to contain a conserved A¹ nucleotide at the 1st position in the D-loop. At the 2nd position, the G² nucleotide was conserved in tRNA^{Pro}, tRNA^{Val}, tRNA^{Phe}, and tRNA^{Tyr}, whereas tRNA^{Thr} contained a U² nucleotide at the 2nd position (Table 2). In the 3rd position, the U³ nucleotide was conserved in tRNA^{Phe},



and tRNA^{Tyr} and the G³ nucleotide was conserved in tRNA^{Cys}, tRNA^{Asn}, and tRNA^{His}. In the 4th position of the D-loop, the G⁴ nucleotide was conserved in tRNA^{Gly}, tRNA^{Tyr}, tRNA^{Thr}, tRNA^{Asn}, tRNA^{Gln}, tRNA^{Lys}, tRNA^{Arg}, and tRNA^{Asp} whereas at the same position tRNA^{Phe} contained conserved U⁴ nucleotide (Table 2). At the 5th position, the G⁵ nucleotide was found to be conserved in tRNA^{Gly}, tRNA^{Ala}, tRNA^{Leu}, tRNA^{Phe}, tRNA^{Tyr}, tRNA^{Trp}, tRNA^{Ser}, tRNA^{Cys}, tRNA^{Lys}, tRNA^{His}, tRNA^{Asp} and tRNA^{Glu}. The majority of tRNAs were found to contain a conserved G⁵ nucleotide whereas tRNA^{Thr} contain a U⁵ nucleotide. At the 6th position, tRNA^{Ala}, tRNA^{Val}, tRNA^{Leu}, tRNA^{Ile}, tRNA^{Met}, tRNA^{Phe}, tRNA^{Ser}, tRNA^{Cys}, and tRNA^{His} contained the G⁶ nucleotide whereas tRNA^{Tyr}, and tRNA^{Asp}

contained the U⁶ nucleotide. At the 7th position of the D-loop, the U⁷ nucleotide was found to be conserved in tRNA^{Phe}, tRNA^{Tyr}, tRNA^{Cys}, and tRNA^{His} whereas the 8th position had a conserved A⁸ nucleotide in tRNA^{Ala}, tRNA^{Leu}, tRNA^{Tyr}, tRNA^{Arg} and tRNA^{Asp}. The A⁹ nucleotide was conserved in tRNA^{Val}, and tRNA^{Ile}. This indicates that a great variation was present in the nucleotide composition of the D-loop. However, the A¹ nucleotide was conserved in the majority of the cases, whereas conservation of nucleotides in the other parts of the D-loop was specific to an individual tRNA family. The D-loop interacts with the Ψ-loop via long range interactions where G¹⁸ and G¹⁹ interact with Ψ⁵⁵ and C⁵⁶, respectively (Hanawa-Suetsugu et al., 2001). Although the position is dynamic, except

for tRNA^{Gln}, all tRNAs were contained at least one conserved G nucleotide in the D-loop, suggesting their common functional role.

The anti-codon arm of tRNA was less conserved than other parts. In tRNA^{Phe}, tRNA^{Trp} and tRNA^{His}, the G⁴ nucleotide was found to be conserved, whereas tRNA^{Trp} had a conserved U⁵ nucleotide and in tRNA^{His}, the G⁵ nucleotide was conserved. The anti-codon loop, which reads the codon of an mRNA during protein translation, is the most important part of tRNA. The C¹ nucleotide was found to be conserved at the 1st position of the anti-codon loop in tRNA^{Ile}, tRNA^{Met}, tRNA^{Trp}, tRNA^{Cys}, tRNA^{Asn}, tRNA^{Lys}, and tRNA^{Asp}. Additionally, the U¹ nucleotide at the 1st position of the anti-codon loop, was found to be conserved in tRNA^{Pro}, tRNA^{Val}, tRNA^{Gln}, tRNA^{His}, and tRNA^{Arg} (Table 2). Occasionally, tRNA^{Arg} contained a C¹ nucleotide in the anti-codon loop. All the tRNAs were found to contain a conserved U² nucleotide at the 2nd position of the anti-codon loop. The A⁶ nucleotide, at the 6th position of the anti-codon loop, was conserved in tRNA^{Gly}, tRNA^{Ala}, tRNA^{Val}, tRNA^{Ile}, tRNA^{Met}, tRNA^{Trp}, tRNA^{Ser}, tRNA^{Thr}, tRNA^{Cys}, tRNA^{Asn}, tRNA^{Lys}, and tRNA^{Asp} (Table 2). In the 7th position of the anti-codon loop, the A⁷ nucleotide was conserved in tRNA^{Ile}, tRNA^{Met}, tRNA^{Trp}, tRNA^{Thr}, tRNA^{Cys}, tRNA^{Asn}, and tRNA^{Lys}. The presence of a conserved A⁶ nucleotide at the 6th position suggests the possible entry site for group I intron in the anti-codon loop.

Previous study reported that the variable loop of tRNA is dynamic in nature and does not contain any conserved nucleotides. However, in our study we found that the variable loop contained conserved nucleotides at the specific positions at the level of individual tRNA family (Table 2). tRNA^{Val} and tRNA^{Phe} were found to contain the conserved G³-U⁴-C⁵ nucleotides at the 3rd, 4th, and 5th positions of the variable loop, whereas tRNA^{Pro} contained the conserved G³-x-C⁵ nucleotides at 3rd and 5th position (Table 2). In tRNA^{Ile} and tRNA^{Asn}, the G³-U⁴ nucleotides were found to be conserved at 3rd and 4th positions. At the 5th position, the C⁵ nucleotide was conserved in tRNA^{Pro}, tRNA^{Val}, tRNA^{Phe}, and tRNA^{Cys} (Table 2). This study indicates that the G³, U⁴, and C⁵ nucleotides at 3rd, 4th, and 5th position, respectively, were the most likely to be conserved in the variable loop of the cyanobacterial tRNA.

All the tRNAs were found to contain a conserved G⁵ nucleotide in the Ψ -arm. Several tRNAs contained a conserved G⁴ nucleotide at the 4th position whereas others were found to contain conserved G³ nucleotide. The conserved G nucleotide at the 4th and 5th position was detected as the G⁴-G⁵ configuration in tRNA^{Pro}, tRNA^{Val}, tRNA^{Ile}, tRNA^{Tyr}, tRNA^{Ser}, tRNA^{Gln}, tRNA^{Lys}, tRNA^{His}, and tRNA^{Arg}.

The Ψ -loop of all the tRNA contained the conserved U¹-U²-C³-x-A⁵ nucleotides at the 1st, 2nd, 3rd, and 5th positions, respectively. Few other tRNAs were found to contain the extended conserved U¹-U²-C³-x-A⁵-x-U⁷ sequence from the 1st to the 7th position of the Ψ -loop (Table 2). The tRNA containing the conserved U¹-U²-C³-x-A⁵-x-U⁷ nucleotides were tRNA^{Leu}, tRNA^{Ile}, tRNA^{Met}, tRNA^{Tyr}, tRNA^{Trp}, tRNA^{Ser}, tRNA^{Cys}, tRNA^{Asn}, tRNA^{Lys}, tRNA^{His}, tRNA^{Arg}, and tRNA^{Glu} (Table 2). tRNA^{Pro} was found to contain the conserved

U¹-U²-C³-x-A⁵-A⁶-U⁷ consensus sequence, where the A⁶ nucleotide was found to be conserved at the 6th position. In tRNA^{Asp}, it was found to contain conserved U¹-U²-C³-x-A⁵-x-C⁷ consensus sequence in which the U⁷ nucleotide substituted with a C⁷ nucleotide (Table 2). These analyses show that tRNAs were conserved in the acceptor arm, D-arm, D-loop, anti-codon arm, anti-codon loop, the variable loop and the Ψ -arm at the individual family level with majority of tRNAs showing greater conservation in the Ψ -loop.

In spite of the presence of highly conserved genetic architecture of cyanobacterial tRNA, they also found to possess introns within it. Previous studies reported that an intron was present only in tRNA^{Leu} (UAA) and tRNA^{Met} (UAC) of cyanobacterial tRNA (Paquin et al., 1997; Rudi and Jakobsen, 1999). However in our study, we found that the cyanobacterial group I intron was also present in tRNA^{Arg}, tRNA^{Gly}, and tRNA^{Lys} (Figure 3). Because cyanobacteria are a prokaryotic organism, the intron, present in cyanobacterial tRNA can be compared with the group I intron in the plastid. Previous studies reported that the group I intron, found in the plastid, are immobile (Rudi and Jakobsen, 1999). The group I introns found in various other locations including the mitochondria, nucleus, bacteria, and bacteriophages share conserved sequence motifs and have evolved via lateral gene transfer (Rudi and Jakobsen, 1999). Cyanobacterial group I introns were also possessed conserved motifs. Rudi and Jakobsen (1999) have classified the group I intron as intron element P, Q, R, and S. Cluster analysis revealed that tRNA^{Arg} of the *Nostoc* sp. PCC 7524 were grouped in intron element P, whereas the introns of the *Nostoc* sp. PCC 7107 and *Gloeocapsa* sp. PCC 73106 were clustered independently. None of these were clustered with intron elements Q, R, or S. This suggests that there may be another conserved intron element in the cyanobacteria which is yet to be elucidated. The presence of an intron, in the anti-codon loop of cyanobacterial tRNA agrees with the exon theory of the evolution of tRNA genes where modern tRNA is reported to have evolved by duplication of one half of tRNA (Di Giulio, 1992). Report suggests that the introns of cyanobacterial tRNA moved via lateral gene transfer (Marck and Grosjean, 2002).

Self-splicing group I intron of cyanobacterial tRNA^{Met} in genera *Dermocarpa*, *Scytonema*, and *Synechocystis* possess open reading frame (ORF). The intron of *Synechocystis* possesses ORF of 150 codons. The introns in cyanobacteria distributed sporadically without any correlation between relatedness in intron sequence corroborates with the lateral transfer of intron (Dujon, 1989; Biniszkiwicz et al., 1994). Sporadic occurrence of these group I introns in homologous genes of closely related species show consistent similarity in the mobility. The group I introns are suggested to be either mobile and ancient in origin (Rudi and Jakobsen, 1999). Closely related group I intron are inserted in the UAA anti-codon of tRNA^{Leu} of cyanobacteria and chloroplast (Biniszkiwicz et al., 1994). Study of introns from 22 cyanobacterial strains by Rudi and Jakobsen (1999) demonstrated relatively recent gain and/or loss of intron in some cyanobacterial lineage (Rudi and Jakobsen, 1999). The presence of self-splicing group I intron in tRNA^{Leu} at the same position of the same gene in cyanobacteria and chloroplast indicated

that this intron predates the invasion of eukaryotic cells by endosymbiosis (Reinhold-Hurek and Shub, 1992). The group I intron of tRNA^{Leu} have higher homology with the group I intron of tRNA^{Ile} (CAU) and tRNA^{Arg} (CCU) (Rudi and Jakobsen, 1997, 1999).

Evolution of a gene occurred through random mutation/substitution. The pattern and frequencies of nucleotide substitution is largely depends on the mutational events of the gene (Shimizu et al., 1989; Zhang and Gerstein, 2003; Arnheim and Calabrese, 2009). Although transitions are more frequent in coding genes because they are less likely to result in amino acid substitutions due to a wobble in the genetic code, the presence of a higher rate of transition in cyanobacterial tRNA is intriguing. A universal bias may favor transition over transversion and a diverse rate of transitions/transversions, combined with the bias may help to explain the multiple evolutionary rates and lineages. The rates of transition/transversion bias of tRNA^{Trp}, tRNA^{Asp}, and tRNA^{Tyr} were 8.409, 4.131, and 3.658, respectively; these were quite higher than the rates of other tRNAs (Table 6). At a low level of genetic divergence, the transition/transversions bias remains high, whereas at a high level of genetic divergence, the transition/transversion bias remains low (Yang and Yoder, 1999). This indicates that tRNA^{Trp}, tRNA^{Asp}, and tRNA^{Tyr} have a very low level of genetic divergence, whereas other tRNAs have a very high rate of genetic divergence. Therefore, cyanobacterial species have only a fewer number of tRNA^{Trp}, tRNA^{Asp}, and tRNA^{Tyr} genes in their genome compared to other tRNAs and rest of the tRNAs have a very high rate of divergence. Due to high genetic divergence, the frequency of other tRNAs are more abundant than tRNA^{Trp}, tRNA^{Asp}, and tRNA^{Tyr}. Silent mutation can have significant impact on exhibiting genetic divergence which accompany through selection/genetic drift or novel adaptation (Palumbi, 1994).

Understanding the process of evolution and function of new gene is important in genomics and evolutionary biology; gene duplication is a powerful process in generating novel gene function via sub-functionalization and neo-functionalization, whereas loss of a gene can greatly shape the gene family (Lynch and Conery, 2000; Niimura and Nei, 2007; Rasmussen and Kellis, 2012; Teufel et al., 2016). Gene duplication is a fundamental process in the evolution of novel species, especially in eukaryotes in which it plays major roles in creating novel gene functions (Cotton and Page, 2005; Blomme et al., 2006; Zhao et al., 2015). Similarly, the loss of a gene occurs via segmental deletion, and pseudogenization preserves a minimum number of functional gene copies (Ohno, 1970; Lynch and Conery, 2000; Ting et al., 2004; Cotton and Page, 2005; Blomme et al., 2006; Demuth et al., 2006). Due to the lowest percentage of duplication, the genome of the cyanobacteria encodes only one to two tRNA^{His} genes per genome, whereas the number of tRNA^{Ile} genes is more numerous due to highest percentage of duplication (Table 5). The low level of duplication and conditional duplication may responsible for the low frequency in the occurrence of tRNA^{His}, tRNA^{Asn}, and tRNA^{Asp} in the cyanobacterial genome. Coupled with the low percentage of duplication, conditional duplication and the high percentage of gene loss led to the low frequency of tRNA^{His}, tRNA^{Asn}, tRNA^{Gln}, and tRNA^{Asp} genes in the cyanobacterial

TABLE 6 | Transition/transversion bias of cyanobacterial tRNAs.

tRNA	D/L score	Duplications	Conditional duplications	Losses
Alanine	797.5	163 (66.8%)	15 (6.14%)	553 (226.63%)
Arginine	712.5	131 (56.22%)	18 (7.72%)	516 (221.45%)
Asparagine	260.0	44 (53.01%)	11 (20.75%)	194 (233.73%)
Aspartate	287.5	53 (67.94%)	5 (6.84%)	208 (284.93%)
Cysteine	229.5	41 (59.42%)	4 (5.79%)	168 (243.47%)
Glutamate	304.5	57 (73.07%)	9 (11.53%)	219 (280.76%)
Glutamine	292.0	50 (56.81%)	5 (5.68%)	217 (246.59%)
Glycine	641.5	115 (65.34%)	17 (9.65%)	469 (266.47%)
Histidine	197.0	32 (52.45%)	7 (11.47%)	149 (244.26%)
Isoleucine	453.6	97 (85.08%)	6 (5.26%)	308 (270.17%)
Leucine	1081.5	193 (61.66%)	25 (7.98%)	792 (253.03%)
Lysine	372.5	67 (56.77%)	9 (7.62%)	272 (230.50%)
Methionine	598.0	108 (59.34%)	19 (10.43%)	436 (239.56%)
Phenylalanine	276.5	47 (56.62%)	8 (9.63%)	206 (248.19%)
Proline	699.5	133 (70.74%)	16 (8.51%)	500 (265.95%)
Serine	788.5	139 (55.60%)	23 (9.2%)	580 (232.00%)
Threonine	596.5	107 (57.83%)	15 (8.10%)	436 (235.67%)
Tryptophan	284.0	54 (70.12%)	5 (6.49%)	203 (263.63%)
Tyrosine	212.0	38 (57.57%)	6 (9.09%)	155 (234.84%)
Valine	485.5	85 (57.82%)	19 (12.92%)	358 (243.53%)

Result showed loss of cyanobacterial tRNA gene predominate the duplication and conditional duplication event.

genome. Similarly, low percentage of gene loss, coupled with the high percentage of gene duplication and conditional duplication in tRNA^{Arg} and tRNA^{Ala}, resulted multiple and variable copies of tRNA^{Arg} and tRNA^{Ala} in the cyanobacterial genome (Table 5). Previous studies showed that tRNAs are the classic example of gene families that were continuously duplicated and lost (Withers et al., 2006; Tang et al., 2009; Bermudez-Santana et al., 2010; Rogers et al., 2010). The operon or tRNA clusters are highly dynamic and unstable genomic region. Therefore, it became an easy target for duplication or loss event. However, in cyanobacteria, the loss event was predominant over the duplication and conditional duplication.

Numerous studies have been conducted on the evolution of tRNA. Reports suggest that tRNA hairpin structures were the precursors of tRNA, having evolved from the combination of different sub-structures, where the top half (acceptor arm) of the tRNA is considered older than the bottom half containing the anti-codon arm (Sun and Caetano-Anollés, 2008). According to the “genomic tag” hypothesis, the upper half of the tRNA harbors the ancestral genomic information and the bottom half provides the specificity for the genetic code (Weiner and Maizels, 1987). The hairpin structure having the anti-codon loop may have evolved as an intermediary in the process of protein synthesis. Introns in tRNA are usually present in the anti-codon loop, and the 5' and 3' halves of the tRNA were likely the mini-genes before their merger (Di Giulio, 1992, 2004). The presence of an intron in tRNA suggests that the two modules of tRNA were separated by introns; hence, tRNA is the exon part of the gene

described by the “exonic gene theory” of the origin of tRNA (Di Giulio, 1995). The tRNA molecule is universally present in all cellular organisms and is transferred by “horizontal gene transfer.” Its ubiquitous presence in all cellular lineages strongly supports its early origin, explaining its presence in organisms of early origin. Fossil record suggests that cyanobacteria are older than 2 billion years and indirect evidence suggests that they may be older than 2.45 billion years (Amard and Bertrand-Sarfati, 1997; Bekker et al., 2004; Schirmermeister et al., 2011; Mohanta et al., 2017). The presence of tRNA^{Arg} and tRNA^{Met} in clusters I, III, V, and VI in the phylogenetic tree reflected their ancient origin (Figure 5); hence, their co-distribution in different clusters may have been followed by subsequent evolution of tRNA^{Gln} and tRNA^{Ala}. The presence of tRNA^{Arg} and tRNA^{Met} in multiple clusters vividly illustrates that they have undergone duplication and subsequently diverged into different species. The distribution of tRNA^{Ile}, tRNA^{Glu}, tRNA^{Asp}, tRNA^{Pro}, tRNA^{Thr}, tRNA^{Gly}, tRNA^{Trp}, tRNA^{His}, tRNA^{Cys}, tRNA^{Tyr}, and tRNA^{Ser} was confined to a single cluster only (Figure 5). This suggests that they have evolved recently, compared with tRNA^{Met}, tRNA^{Arg}, tRNA^{Ala}, and tRNA^{Gln} (Figure 5). In cluster I, tRNA^{Arg} grouped twice and present close to tRNA^{Asp} and tRNA^{Gly}, suggesting their evolution from tRNA^{Arg} (Figure 5). In cluster I, tRNA^{Ala} was present close to tRNA^{Val}; this suggests that tRNA^{Val} evolved from tRNA^{Ala}. In cluster II, tRNA^{Pro} was grouped separately whereas tRNA^{Thr}, tRNA^{Phe}, and tRNA^{Asn} were grouped with tRNA^{Met}, suggesting their evolution from tRNA^{Ala} (Figure 5). Cluster III contained tRNA^{Arg} and tRNA^{Trp}, suggesting possible evolution of tRNA^{Trp} from tRNA^{Arg}. In cluster IV, tRNA^{Gln} present close to tRNA^{His}, and tRNA^{Met} present closer to tRNA^{Tyr} and tRNA^{Cys}. This suggests that tRNA^{His} may have evolved from tRNA^{Gln} whereas tRNA^{Tyr} and tRNA^{Cys} may have evolved from tRNA^{Met}. tRNA^{Ser} found in cluster IV shared a branch with tRNA^{Met}, suggesting the evolution of tRNA^{Ser} from tRNA^{Met}. tRNA^{Gln}, tRNA^{Met}, and tRNA^{Ala} were found together in cluster V, suggesting that these tRNAs may have coevolved together, and later diversified and evolved into other tRNAs. These tRNAs were grouped with several pseudo-tRNAs in the cluster of tRNA^{Asn} (Figure 5, Supplementary Figure 1). This suggests that these tRNAs are likely in the verge of becoming pseudo-tRNAs and will be lost from the evolutionary lineage. tRNA^{Phe} was present as sister branch to tRNA^{Asn} in cluster III, suggesting their evolution from tRNA^{Asn}. Several tRNAs^{Lys}, tRNAs^{His}, and tRNAs^{Ala} in cluster IV were grouped with the pseudo-tRNAs, suggesting their possible loss of function and, hence, extinction from the evolutionary lineage by pseudogenization. However, the evolutionary history of cyanobacterial tRNAs may be more complicated. The presence of common secondary and tertiary structures occurring with an invariant number of nucleotide residues, 20 invariant gene families, and similar functional roles

during the translation is the phenomena that are complicated to delineate. It is assumed that for a total of 20 tRNA gene families, there would be 20 branches, corresponding to a group in each tRNA gene family, in the phylogenetic tree. However, this assumption is incorrect. The differences between tRNA families, with regard to their respective amino acid and tRNA specificity were found overlap significantly. The non-isoacceptor tRNAs shared the phylogenetic groups with the isoacceptors, suggesting their evolution from multiple lineages. This shows that tRNAs had evolved by multiple duplication with extensive genetic mutation and subsequent loss of old genes.

CONCLUSION

Sequence analysis revealed that cyanobacterial tRNAs were dynamic and transiently conserved within their respective tRNA families. However, the U¹-U²-C³ conserved consensus sequence in the Ψ -loop was found in all of the studied tRNAs. Several cyanobacterial tRNAs contained a group I intron in the anticodon region; further study is required to discover more introns and to understand their specific genomic and evolutionary aspects. The rate of mutational transition of cyanobacterial tRNAs was higher than that of transversion; cyanobacterial tRNA may have evolved polyphyletically by loss of tRNA genes.

AUTHOR CONTRIBUTIONS

TM: Conceived the idea, performed the study, analyzed data, drafted and revised the manuscript. AS: Revised the manuscript. FA: Revised the manuscript. HB: Revised the manuscript.

ACKNOWLEDGMENTS

This work was carried out with the support of the Next-Generation Biogreen 21 Program (PJ011113), Rural Development Administration, Korea. Authors also like to extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through research group No (RG-1438-078).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2017.00200/full#supplementary-material>

Supplementary Figure 1 | Phylogenetic tree of cyanobacterial tRNAs. The phylogenetic clades with low bootstrap replicates were collapsed with 50% cutoff values.

Supplementary data 1 | Supplementary data showing genomic tRNA sequences of 61 cyanobacterial species.

REFERENCES

Akaike, H. (1998). “Information theory and an extension of the maximum likelihood principle,” in *Selected Papers of Hirotugu Akaike*, eds E. Parzen, K. Tanabe, and G. Kitagawa (New York, NY: Springer), 199–213.

Altermann, W., Kazmierczak, J., Oren, A., and Wright, T. (2006). Cyanobacterial calcification and its rock-building potential during 3.5 billion years of Earth history. *Geobiology* 4, 147–166. doi: 10.1111/j.1472-4669.2006.00076.x

Amard, B., and Bertrand-Sarfati, J. (1997). Microfossils in 2000 Ma old cherty stromatolites of the Franceville Group, Gabon.

- Precambrian Res.* 81, 197–221. doi: 10.1016/S0301-9268(96)00035-6
- Arnheim, N., and Calabrese, P. (2009). Understanding what determines the frequency and pattern of human germline mutations. *Nat. Rev. Genet.* 10, 478–488. doi: 10.1038/nrg2529
- Bekker, A., Holland, H. D., Wang, P. L., Rumble, D., Stein, H. J., Hannah, J. L., et al. (2004). Dating the rise of atmospheric oxygen. *Nature* 427, 117–120. doi: 10.1038/nature02260
- Bermudez-Santana, C., Attolini, C. S., Kirsten, T., Engelhardt, J., Prohaska, S. J., Steigle, S., et al. (2010). Genomic organization of eukaryotic tRNAs. *BMC Genomics* 11:270. doi: 10.1186/1471-2164-11-270
- Binizskiewicz, D., Cesnaviciene, E., and Shub, D. A. (1994). Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. *EMBO J.* 13, 4629–4635.
- Blomme, T., Vandepoele, K., De Bodt, S., Simillion, C., Maere, S., and Van de Peer, Y. (2006). The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol.* 7:R43. doi: 10.1186/gb-2006-7-5-r43
- Chan, P. P., Cozen, A. E., and Lowe, T. M. (2011). Discovery of permuted and recently split transfer RNAs in Archaea. *Genome Biol.* 12, R38–R38. doi: 10.1186/gb-2011-12-4-r38
- Chen, K., Durand, D., and Farach-Colton, M. (2000). NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* 7, 429–447. doi: 10.1089/106652700750050871
- Cotton, J. A., and Page, R. D., (2005). Rates and patterns of gene duplication and loss in the human genome. *Proc. R. Soc. B.* 272, 277–283. doi: 10.1098/rspb.2004.2969
- Demuth, J. P., De Bie, T., Stajich, J. E., Cristianini, N., and Hahn, M. W. (2006). The evolution of mammalian gene families. *PLoS ONE* 1:e85. doi: 10.1371/journal.pone.0000085
- Di Giulio, M. (1992). On the origin of the transfer RNA molecule. *J. Theor. Biol.* 159, 199–214. doi: 10.1016/S0022-5193(05)80702-7
- Di Giulio, M. (1995). Was it an ancient gene codifying for a hairpin RNA that, by means of direct duplication, gave rise to the primitive tRNA molecule? *J. Theor. Biol.* 177, 95–101. doi: 10.1016/S0022-5193(05)80007-4
- Di Giulio, M. (2004). The origin of the tRNA molecule: implications for the origin of protein synthesis. *J. Theor. Biol.* 226, 89–93. doi: 10.1016/j.jtbi.2003.07.001
- Di Giulio, M. (2008a). The split genes of Nanoarchaeum equitans are an ancestral character. *Gene* 421, 20–26. doi: 10.1016/j.gene.2008.06.010
- Di Giulio, M. (2008b). Transfer RNA genes in pieces are an ancestral character. *EMBO Rep.* 9, 820. doi: 10.1038/embor.2008.153
- Dujon, B. (1989). Group I introns as mobile genetic elements: facts and mechanistic speculations—a review. *Gene* 82, 91–114. doi: 10.1016/0378-1119(89)90034-6
- Fujishima, K., Sugahara, J., Kikuta, K., Hirano, R., Sato, A., Tomita, M., 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. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2683–2687. doi: 10.1073/pnas.0808246106
- Hanawa-Suetsugu, K., Bordeau, V., Himeno, H., Muto, A., and Felden, B. (2001). Importance of the conserved nucleotides around the tRNA-like structure of *Escherichia coli* transfer-messenger RNA for protein tagging. *Nucleic Acids Res.* 29, 4663–4673. doi: 10.1093/nar/29.22.4663
- Hardt, W. D., Schlegl, J., Erdmann, V. A., and Hartmann, R. K., (1993). Role of the D arm and the anticodon arm in tRNA recognition by eubacterial and eukaryotic RNase P enzymes. *Biochemistry* 32, 13046–13053. doi: 10.1021/bi00211a014
- Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., et al. (1965). Structure of a ribonucleic acid. *Science* (80-). 147, 1462–1465. doi: 10.1126/science.147.3664.1462
- Kanai, A. (2015). Disrupted tRNA genes and tRNA fragments: a perspective on tRNA gene evolution. *Life* 5, 321–331. doi: 10.3390/life5010321
- Kirchner, S., and Ignatova, Z. (2014). Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat. Rev. Genet.* 16, 98–112. doi: 10.1038/nrg3861
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.0955
- Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* (80-). 290, 1151–1155. doi: 10.1126/science.290.5494.1151
- Marck, C., and 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. doi: 10.1017/S1355838202022021
- Maruyama, S., Sugahara, J., Kanai, A., and Nozaki, H. (2010). Permuted tRNA genes in the nuclear and nucleomorph genomes of photosynthetic eukaryotes. *Mol. Biol. Evol.* 27, 1070–1076. doi: 10.1093/molbev/msp313
- Mohanta, T. K., Arora, P. K., Mohanta, N., Parida, P., and Bae, H. (2015a). Identification of new members of the MAPK gene family in plants shows diverse conserved domains and novel activation loop variants. *BMC Genomics* 16:58. doi: 10.1186/s12864-015-1244-7
- Mohanta, T. K., and Bae, H. (2017). Analyses of genomic tRNA reveal presence of novel tRNAs in *Oryza sativa*. *Front. Genet.* 8:90. doi: 10.3389/fgene.2017.00090
- Mohanta, T. K., Mohanta, N., Mohanta, Y. K., and Bae, H. (2015b). Genome-wide identification of calcium dependent protein kinase gene family in plant lineage shows presence of novel D-x-D and D-E-L motifs in EF-hand domain. *Front. Plant Sci.* 6:1146. doi: 10.3389/fpls.2015.01146
- Mohanta, T. K., Mohanta, N., Mohanta, Y. K., Parida, P., and Bae, H. (2015c). Genome-wide identification of calcineurin B-like (CBL) gene family of plants reveals novel conserved motifs and evolutionary aspects in calcium signaling events. *BMC Plant Biol.* 15:189. doi: 10.1186/s12870-015-0543-0
- Mohanta, T. K., Mohanta, N., Parida, P., Panda, S. K., Ponpandian, L. N., and Bae, H. (2016). Genome-wide identification of mitogen-activated protein kinase gene family across fungal lineage shows presence of novel and diverse activation loop motifs. *PLoS ONE* 11:e0149861. doi: 10.1371/journal.pone.0149861
- Mohanta, T. K., Pudake, R. N., and Bae, H. (2017). Genome-wide identification of major protein families of cyanobacteria and genomic insight into the circadian rhythm. *Eur. J. Phycol.* 52, 1–17. doi: 10.1080/09670262.2016.1251619
- Navarre, W. W., and Schneewind, O. (1999). Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol. Mol. Biol. Rev.* 63, 174–229.
- Niimura, Y., and Nei, M. (2007). Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE* 2:e708. doi: 10.1371/journal.pone.0000708
- Nirenberg, M., and Leder, P. (1964). RNA codewords and protein synthesis. *Science* (80-). 145, 1399–1407.
- Ohno, S. (1970). *Evolution by Gene Duplication*. London: George Allen and Unwin.
- Palumbi, S. (1994). Genetic divergence, reproductive isolation, speciation. *Annu. Rev. Ecol. Syst.* 25, 547–572. doi: 10.1146/annurev.es.25.110194.002555
- Paquin, B., Kathe, S. D., Nierzwicki-bauer, S. A., and Shub, D. A., (1997). Origin and evolution of group I introns in cyanobacterial tRNA genes. Origin and Evolution of Group I Introns in Cyanobacterial tRNA Genes. *J. Bacteriol.* 179, 6798–6806. doi: 10.1128/jb.179.21.6798-6806.1997
- Randau, L., Münch, R., Hohn, M. J., Jahn, D., and Söll, D. (2005). Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5'- and 3'-halves. *Nature* 433, 537–541. doi: 10.1038/nature03233
- Rasmussen, M. D., and Kellis, M. (2012). Unified modeling of gene duplication, loss, and coalescence using a locus tree. *Genome Res.* 22, 755–765. doi: 10.1101/gr.123901.111
- Reinhold-Hurek, B., and Shub, D. (1992). Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* 357, 173–176. doi: 10.1038/357173a0
- Rogers, H. H., Bergman, C. M., and Griffiths-Jones, S. (2010). The evolution of tRNA genes in *Drosophila*. *Genome Biol. Evol.* 2, 467–477. doi: 10.1093/gbe/evq034
- Roy, H., and Ibba, M. (2008). RNA-dependent lipid remodeling by bacterial multiple peptide resistance factors. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4667–4672. doi: 10.1073/pnas.0800006105
- Rudi, K., and Jakobsen, K. S. (1997). Cyanobacterial tRNA^{Leu}(UAA) group I introns have polyphyletic origin. *FEMS Microbiol. Lett.* 156, 293–298. doi: 10.1111/j.1574-6968.1997.tb12743.x
- Rudi, K., and Jakobsen, K. S. (1999). Complex evolutionary patterns of tRNA^{Leu}(UAA) group I introns in the cyanobacterial radiation. *J. Bacteriol.* 181, 3445–3451.
- Schirmermeister, B. E., Anisimova, M., Antonelli, A., and Bagheri, H. C. (2011). Evolution of cyanobacterial morphotypes: taxa required for

- improved phylogenomic approaches. *Commun. Integr. Biol.* 4, 424–427. doi: 10.4161/cib.16183
- Schopf, J. W., and Packer, B. M. (1987). Early archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. *Science* (80-). 237, 70–73.
- Shi, J. P., Martinis, S. A., and Schimmel, P. (1992). RNA tetraloops as minimalist substrates for aminoacylation. *Biochemistry* 31, 4931–4936. doi: 10.1021/bi00136a002
- Shi, P. Y., Weiner, A. M., and Maizels, N. (1998). A top-half tDNA minihelix is a good substrate for the eubacterial CCA-adding enzyme. *RNA* 4, 276–284.
- Shimizu, N., Okamoto, T., Moriyama, E. N., Takeuchi, Y., Gojobori, T., and Hoshino, H. (1989). Patterns of nucleotide substitutions and implications for the immunological diversity of human immunodeficiency virus. *FEBS Lett.* 250, 591–595. doi: 10.1016/0014-5793(89)80802-6
- Smith, D., and Yarus, M. (1989). Transfer RNA structure and coding specificity. *J. Mol. Biol.* 206, 489–501. doi: 10.1016/0022-2836(89)90496-8
- Soma, A. (2014). Circularly permuted tRNA genes: their expression and implications for their physiological relevance and development. *Front. Genet.* 5:63. doi: 10.3389/fgene.2014.00063
- Sugahara, J., Fujishima, K., Morita, K., Tomita, M., and Kanai, A. (2009). Disrupted tRNA gene diversity and possible evolutionary scenarios. *J. Mol. Evol.* 69, 497. doi: 10.1007/s00239-009-9294-6
- Sun, F. J., and Caetano-Anollés, G. (2008). The origin and evolution of tRNA inferred from phylogenetic analysis of structure. *J. Mol. Evol.* 66, 21–35. doi: 10.1007/s00239-007-9050-8
- Tamura, K., Stecher, G., Peterson, D., Filipiński, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tang, D. T., Glazov, E., A., McWilliam, S. M., Barris, W. C., and Dalrymple, B. P. (2009). Analysis of the complement and molecular evolution of tRNA genes in cow. *BMC Genomics* 10:188. doi: 10.1186/1471-2164-10-188
- Teufel, A. I., Liu, L., and Liberles, D. A. (2016). Models for gene duplication when dosage balance works as a transition state to subsequent neo- or sub-functionalization. *BMC Evol. Biol.* 16:45. doi: 10.1186/s12862-016-break0616-1
- Ting, C. T., Tsauro, S. C., Sun, S., Browne, W. E., Chen, Y. C., Patel, N. H., et al. (2004). Gene duplication and speciation in *Drosophila*: evidence from the *Odysseus* locus. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12232–12235. doi: 10.1073/pnas.0401975101
- Tocchini-Valentini, G. D., Fruscoloni, P., and Tocchini-Valentini, G. P. (2009). Processing of multiple-intron-containing pre-tRNA. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20246–20251. doi: 10.1073/pnas.0911658106
- Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Stat. Sin.* 21, 5–42.
- Vernot, B., Stolzer, M., Goldman, A., and Durand, D. (2008). Reconciliation with non-binary species trees. *J. Comput. Biol.* 15, 981–1006. doi: 10.1089/cmb.2008.0092
- Weiner, A. M., and Maizels, N. (1987). tRNA-like structures tag the 3' ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 84, 7383–7387. doi: 10.1073/pnas.84.21.7383
- Weiner, A. M., and Maizels, N. (1999). The genomic tag hypothesis: modern viruses as molecular fossils of ancient strategies for genomic replication, and clues regarding the origin of protein synthesis. *Biol. Bull.* 196, 327–330. doi: 10.2307/1542962
- Wilusz, J. E. (2015). Controlling translation via modulation of tRNA levels. *Wiley Interdiscip. Rev. RNA* 6, 453–470. doi: 10.1002/wrna.1287
- Withers, M., Wernisch, L., and dos Reis, M. (2006). Archaeology and evolution of transfer RNA genes in the *Escherichia coli* genome. *RNA* 12, 933–942. doi: 10.1261/rna.2272306
- Xu, X., and Reid, N. (2011). On the robustness of maximum composite likelihood estimate. *J. Stat. Plan. Inference* 141, 3047–3054. doi: 10.1016/j.jspi.2011.03.026
- Yang, Z., and Yoder, A. D. (1999). Estimation of the transition/transversion rate bias and species sampling. *J. Mol. Evol.* 48, 274–283. doi: 10.1007/PL00006470
- Zhang, Z., and Gerstein, M. (2003). Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. *Nucleic Acids Res.* 31, 5338–5348. doi: 10.1093/nar/gkg745
- Zhao, J., Teufel, A. I., Liberles, D. A., and Liu, L. (2015). A generalized birth and death process for modeling the fates of gene duplication. *BMC Evol. Biol.* 15:275. doi: 10.1186/s12862-015-0539-2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Mohanta, Syed, Ameen and Bae. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.