

Comprehensive Analysis of Archaeal tRNA Genes Reveals Rapid Increase of tRNA Introns in the Order Thermoproteales

Junichi Sugahara,*†¹ Kaoru Kikuta,*‡¹ Kosuke Fujishima,*† Nozomu Yachie,*†
Masaru Tomita,*†‡ and Akio Kanai*†‡

*Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan; †Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, Japan; and ‡Faculty of Environment and Information Studies, Keio University, Fujisawa, Japan

The analysis of archaeal tRNA genes is becoming more important to evaluate the origin and evolution of tRNA molecule. Even with the recent accumulation of complete genomes of numerous archaeal species, several tRNA genes are still required for a full complement of the codon table. We conducted comprehensive screening of tRNA genes from 47 archaeal genomes by using a combination of different types of tRNA prediction programs and extracted a total of 2,143 reliable tRNA gene candidates including 437 intron-containing tRNA genes, which covered more than 99.9% of the codon tables in Archaea. Previously, the content of intron-containing tRNA genes in Archaea was estimated to be approximately 15% of the whole tRNA genes, and most of the introns were known to be located at canonical positions (nucleotide position between 37 and 38) of precursor tRNA (pre-tRNA). Surprisingly, we observed marked enrichment of tRNA introns in five species of the archaeal order Thermoproteales; about 70% of tRNA gene candidates were found to be intron-containing tRNA genes, half of which contained multiple introns, and the introns were located at various noncanonical positions. Sequence similarity analysis revealed that approximately half of the tRNA introns found at Thermoproteales-specific intron locations were highly conserved among several tRNA genes. Intriguingly, identical tRNA intron sequences were found within different types of tRNA genes that completely lacked exon sequence similarity, suggesting that the tRNA introns in Thermoproteales could have been gained via intron insertion events at a later stage of tRNA evolution. Moreover, although the CCA sequence at the 3' terminal of pre-tRNA is added by a CCA-adding enzyme after gene transcription in Archaea, most of the tRNA genes containing highly conserved introns already encode the CCA sequence at their 3' terminal. Based on these results, we propose possible models explaining the rapid increase of tRNA introns as a result of intron insertion events via retrotransposition of pre-tRNAs. The sequences and secondary structures of the tRNA genes and their bulge-helix-bulge motifs were registered in SPLITSdb (<http://splits.iab.keio.ac.jp/splitsdb/>), a novel and comprehensive database for archaeal tRNA genes.

Introduction

As tRNA is a key molecule in translation, its origin and evolution are important topics widely discussed in the field of molecular evolution (Maizels and Weiner 1994; Di Giulio 1999, 2006; Fujishima et al. 2008; Randau and Söll 2008). Around 15% of all archaeal tRNA genes, which are estimated to be the most ancestral (Sun and Caetano-Anolles 2008), have been reported to contain endogenous introns (intron-containing tRNA genes). According to Marck and Grosjean (2003), in a total of 800 archaeal tRNA genes analyzed, 103 out of 136 introns of tRNA genes are located at the nucleotide position between 37 and 38 (canonical position, 37/38); the remaining 33 introns are located at other sites (noncanonical positions); 6 of these tRNA genes reportedly harbor 2 introns. Recently, we have reported that computational analysis using SPLITS, a program for detecting various types of archaeal tRNA genes (Sugahara et al. 2006), has predicted three tRNA genes that contain triple introns (Sugahara et al. 2007). The tRNA introns are known to form a structural motif consisting of a bulge-helix-bulge (BHB) consensus sequence around the intron-exon junction (Kaine et al. 1983; Diener and Moore 1998; Marck and Grosjean 2003). The canonical BHB motif is a single hairpin structure that consists of two bulges (B), which are separated by a central helix

(H) within consensus motif sequences (hBHBh'), and is recognized and cleaved by heterotetrameric ($\alpha_2\beta_2$), homodimeric (α_2), or homotetrameric (α_4) splicing endonucleases during precursor tRNA (pre-tRNA) processing (Lykke-Andersen and Garrett 1997; Li et al. 1998; Tocchini-Valentini et al. 2005a; Yoshinari et al. 2005). Relaxed BHL and HBh' motifs including the constant central helix H flanked by one helix (h or h') on either sides are also known motifs of pre-tRNA (Marck and Grosjean 2003) and coevolution between the types of the BHB motifs and their corresponding splicing endonucleases have been previously studied (Tocchini-Valentini et al. 2005b).

Unique types of introns carrying tRNA genes have recently been reported in Archaea and primitive Eukaryote. It has been reported that the hyperthermophilic archaeal parasite, *Nanoarchaeum equitans*, encodes split tRNA genes with 5' and 3' halves on separate genes (Randau, Calvin, et al. 2005; Randau, Munch, et al. 2005). We have also reported a novel type of tRNA gene, so-called permuted tRNA genes, with the 3' half of the tRNA lying upstream of the 5' half in the nuclear genome of the primitive red alga, *Cyanidioschyzon merolae* (Soma et al. 2007). Interestingly, both split and permuted tRNA genes form BHB motifs around the intron-exon junctions (Randau, Pearson, and Söll 2005; Soma et al. 2007); thus, having this common intron-processing machinery suggests an evolutionary link between disrupted (intron-containing, split, and permuted) tRNA genes in Archaea and primitive Eukaryote. Analysis of archaeal tRNA genes possessing BHB motifs is thus becoming important to elucidate the evolution of tRNA genes. Besides, it has been widely discussed whether the tRNA introns represent a primitive trait that is important in tRNA origin or whether they are

¹ Equal contribution to this work.

Key words: tRNA, Archaea, intron, bulge-helix-bulge (BHB) motif, Thermoproteales, Bioinformatics.

E-mail: akio@sfc.keio.ac.jp.

Mol. Biol. Evol. 25(12):2709–2716. 2008

doi:10.1093/molbev/msn216

Advance Access publication October 1, 2008

acquired at a later stage of tRNA evolution (Di Giulio 2006; Fujishima et al. 2008; Randau and Söll 2008; Di Giulio 2008). However, current tRNA databases are still inadequate (Sprinzl and Vassilenko 2005; Andersen et al. 2006; Putz et al. 2007; <http://lowelab.ucsc.edu/GtRNAdb/>) due to the difficulty of detecting introns located at noncanonical positions of tRNA. Besides, several tRNA genes are still unknown in certain archaeal species, suggesting the possibility of undocumented novel intron-containing or split tRNA genes.

In an attempt to identify all the tRNA genes in Archaea, we constructed an integrated tRNA gene prediction system based on a combination of SPLITS, tRNAscan-SE (Lowe and Eddy 1997), ARAGORN (Laslett and Canback 2004), Blast (Altschul et al. 1990), and tRNA-CM covariance model (Eddy and Durbin 1994). We then conducted a comprehensive prediction of tRNA gene candidates from the complete genome sequences of 47 archaeal species, which is presented here with the results of phylogenetic analysis of tRNA introns, as well as the number, location, sequence conservation, and coevolution of splicing endonucleases. All sequences, secondary structures, and splicing motifs of the tRNA genes have been registered in SPLITSdb (<http://splits.iab.keio.ac.jp/splitsdb/>), a novel and comprehensive web-based database for archaeal tRNA genes. SPLITSdb is the only database that covers all types of archaeal tRNA genes including split and intron-containing tRNA genes and will contribute to further research on the origin and evolution of tRNA and tRNA intron processing.

Materials and Methods

Genome Sequences

A total of 47 complete archaeal genomes were obtained from National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). The complete list of species and accession numbers are shown in supplementary table S2 (Supplementary Material online).

Comprehensive Prediction of tRNA Genes

An overview of the comprehensive prediction of tRNA genes is shown in supplementary figure S1 (Supplementary Material online). The tRNA gene candidates are predicted from each genome sequence based on four different programs. Split tRNA gene candidates were predicted by SPLITS with “-t” search mode. Intron-containing tRNA and other tRNA gene candidates, including the selenocysteiny tRNA gene, were predicted by SPLITS, tRNAscan-SE, and ARAGORN. Pyrrolysyl tRNA gene candidates were predicted based on the BlastN search using *Methanosarcina barkeri* DNA sequences encoding the pyrrolysyl tRNA genes as a query. The results of ARAGORN and BlastN were further evaluated using tRNA-CM covariance model (Eddy and Durbin 1994) for more detail about the cloverleaf structural features and consensus sequences of the candidates. Next, all the output data were merged and overlapping tRNA gene candidates were eliminated. In addition, initial predictions of tRNA introns and the secondary structure of splicing motifs were made by SPLITS.

Then, the location of the introns and the type of splicing motifs were manually annotated according to (Marck and Grosjean 2003); and all the predicted tRNA genes were confirmed manually to maintain a highly reliable tRNA gene data set by satisfying the following conditions: 1) tRNA gene candidates that fulfill the criteria of the cloverleaf structure of archaeal tRNAs (Marck and Grosjean 2002) and 2) tRNA gene candidates that are able to decode all sense codons in each species.

Phylogenetic Analysis Based on rRNA Sequences

The 16S rRNA sequences from 47 archaeal species were aligned (as rDNA sequence) using ClustalW 2.0 (Larkin et al. 2007) with the following parameters: multiple alignment parameters, gap opening = 22.5; gap extension penalty = 0.83, and delay divergent sequence = 25%; and pairwise alignment parameter, gap opening = 22.5. These parameters gave the highest performance in RNA sequence alignment (Wilm et al. 2006). Aligned rRNA sequences were manually improved by trimming both ends to maintain the consensus sequence. Bayesian trees were constructed using MrBayes software version 3.1.2 with *C. merolae* 18S rRNA as an out-group (Huelsenbeck and Ronquist 2001). The model of sequence evolution was determined by MrModeltest software version 2.2. The phylogenetic tree files were visualized using Interactive Tree Of Life (Letunic and Bork 2007).

Sequence Similarity Analysis of Archaeal tRNA Introns

In order to analyze the conserved tRNA introns that seem to be created via intron insertion events, we calculated the sequence similarity score (%) for every archaeal tRNA intron. For this purpose, we aligned all possible pairs of tRNA introns having similar length (allowing ± 1 nt length differentiation), and the sequence having the highest sequence similarity was selected as the matched sequence. If the length of the sequences was different, a single gap was allowed to optimize the overall sequence similarity. We excluded pairs of tRNA introns derived from tRNA genes corresponding to the same amino acids (e.g., tRNA^{Gly} in *Pyrobaculum calidifontis* and *Pyrobaculum arsenaticum*) to eliminate the introns conserved among the orthologous tRNA genes.

Results and Discussion

Construction of SPLITSdb System and Prediction of Archaeal tRNA Genes

To extract complete tRNA gene sets in archaeal species, we conducted a comprehensive prediction of tRNA genes from 47 archaeal genomes using the SPLITSdb system, which implemented multiple types of tRNA prediction programs and manual confirmation (supplementary fig. S1, Supplementary Material online). As a result, 2,143 tRNA gene candidates including 437 intron-containing tRNA genes were extracted (supplementary table S1, Supplementary Material online). The candidates covered an average of

99.9% of all 61 sense codons and the initiator codon in the 47 species using the SPLITSdb system, versus an average of 91.8% and 91.2% using tRNAscan-SE and ARAGORN, respectively (supplementary table S2, Supplementary Material online). The same numbers of tRNA gene candidates were often predicted by tRNAscan-SE, ARAGORN, and SPLITSdb system in a single species; however, the codon coverage of the SPLITSdb system exceeded that of other prediction programs. This is because tRNAscan-SE and ARAGORN mispredict the anticodon of tRNA genes due to tRNA intron insertion at noncanonical positions in certain cases (Sugahara et al. 2006) and also misidentifies AUA decoding tRNA^{Leu} genes possessing the posttranscriptionally modified CAT anticodon. The SPLITSdb system has accomplished reliable extraction of tRNA gene candidates with reasonable number and little redundancy due to the combination of the prediction of tRNA genes containing noncanonical introns, correct identification of tRNA anticodons, and elimination of irregular tRNA genes.

Only 26 archaeal tRNA genes containing multiple introns were previously reported (Marck and Grosjean 2003; Sugahara et al. 2007). To determine novel tRNA gene candidates containing multiple introns, the predicted 437 intron-containing tRNA genes were classified into three types of tRNA: 364 with a single intron, 63 with double introns, and 10 with triple introns. As a result, we found 47 novel tRNA gene candidates containing multiple introns. Although tRNA gene candidates containing triple introns were only reported previously in tRNA^{Pro} genes of *Thermofilum pendens* (Sugahara et al. 2007), we further documented 7 novel candidates containing triple introns in other two species; tRNA^{Glu} (UUC), tRNA^{Glu} (CUC), tRNA^{Gln} (UUG), tRNA^{Gln} (CUG), tRNA^{Thr} (UGU), and tRNA^{Pro} (UGG) genes in *P. calidifontis* (fig. 1a and supplementary fig. S2, Supplementary Material online) and tRNA^{Pro} (UGG) genes in *Pyrobaculum islandicum* (fig. 1b). Although most of those tRNA genes contained at least one intron at a canonical position (37/38), we found two tRNA gene candidates with all three introns located at noncanonical positions, tRNA^{Glu} (CUC) and tRNA^{Glu} (UUC) in *P. calidifontis* (fig. 1a).

Phylogenetic Analysis of tRNA Introns and Its Splicing Endonucleases

The proportion of split and intron-containing tRNA genes and the coevolution of tRNA introns and their splicing endonucleases were analyzed based on the phylogeny of small subunit rRNA among the 47 species (fig. 2). We initially observed that the numbers of tRNA introns in archaeal species possessing heterotetrameric endonucleases (all Crenarchaeal and Nanoarchaeal species, and *Methanopyrus kandleri* in Euryarchaea) (Tocchini-Valentini et al. 2005a; Yoshinari et al. 2005) were significantly high. Moreover, we also found marked enrichment of tRNA introns in the archaeal order Thermoproteales (*T. pendens*, *P. calidifontis*, *Pyrobaculum aerophilum*, *P. islandicum*, and *P. arsenaticum*), a member of the Crenarchaeota encoding heterotetrameric endonucleases. A significant 68.7% of all Thermoproteales tRNA genes were found to be intron

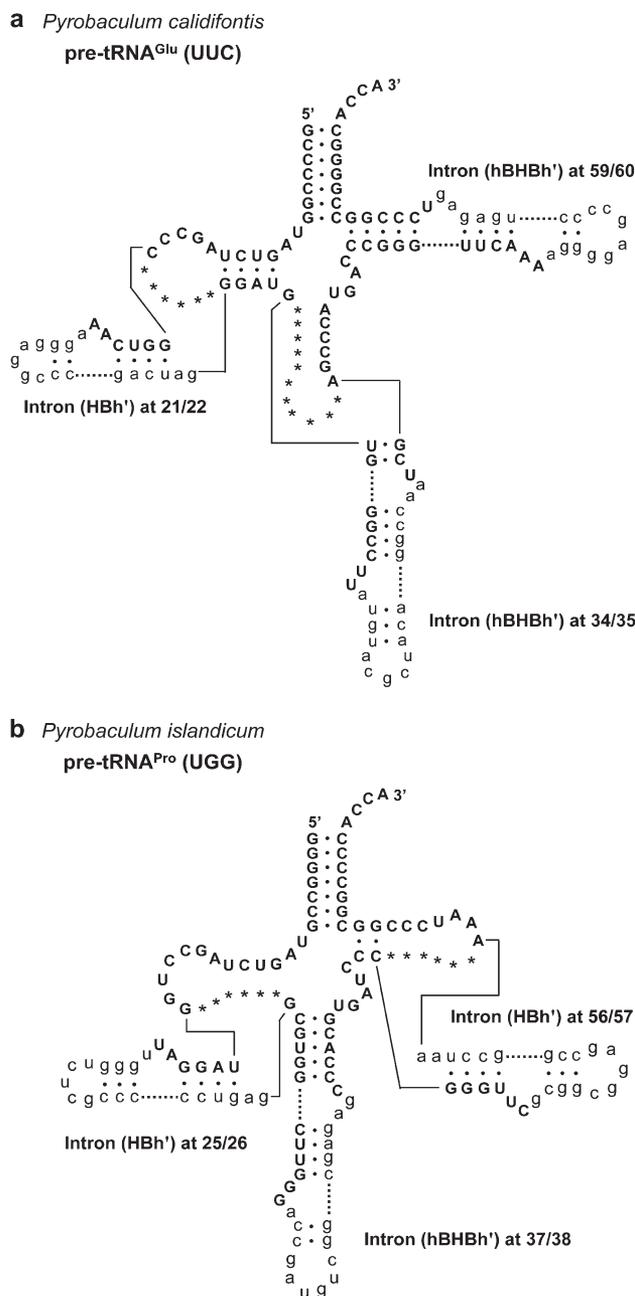


FIG. 1.—Examples of archaeal pre-tRNAs containing three introns. (a) Pre-tRNA^{Glu} (UUC) from *P. calidifontis*. (b) Pre-tRNA^{Pro} (UGG) from *Pyrobaculum islandicum*. The locations and topological types of BHB motifs are presented for each intron. Predicted intron sequences are shown in lower case. BHB motif abbreviations are according to Marck and Grosjean (2003).

containing, whereas 36.4%, 9.2%, and 5.8% of other archaeal species tRNA genes were found to be intron containing with either heterotetrameric, homodimeric, or homotetrameric endonucleases, respectively (fig. 2). Furthermore, 85.7% (54 out of 63) of the tRNA genes containing double introns and 100% (10 out of 10) of the tRNA genes containing triple introns found in all archaeal species were concentrated in Thermoproteales. In particular, 87% of both *T. pendens* and *P. calidifontis* tRNA genes were found

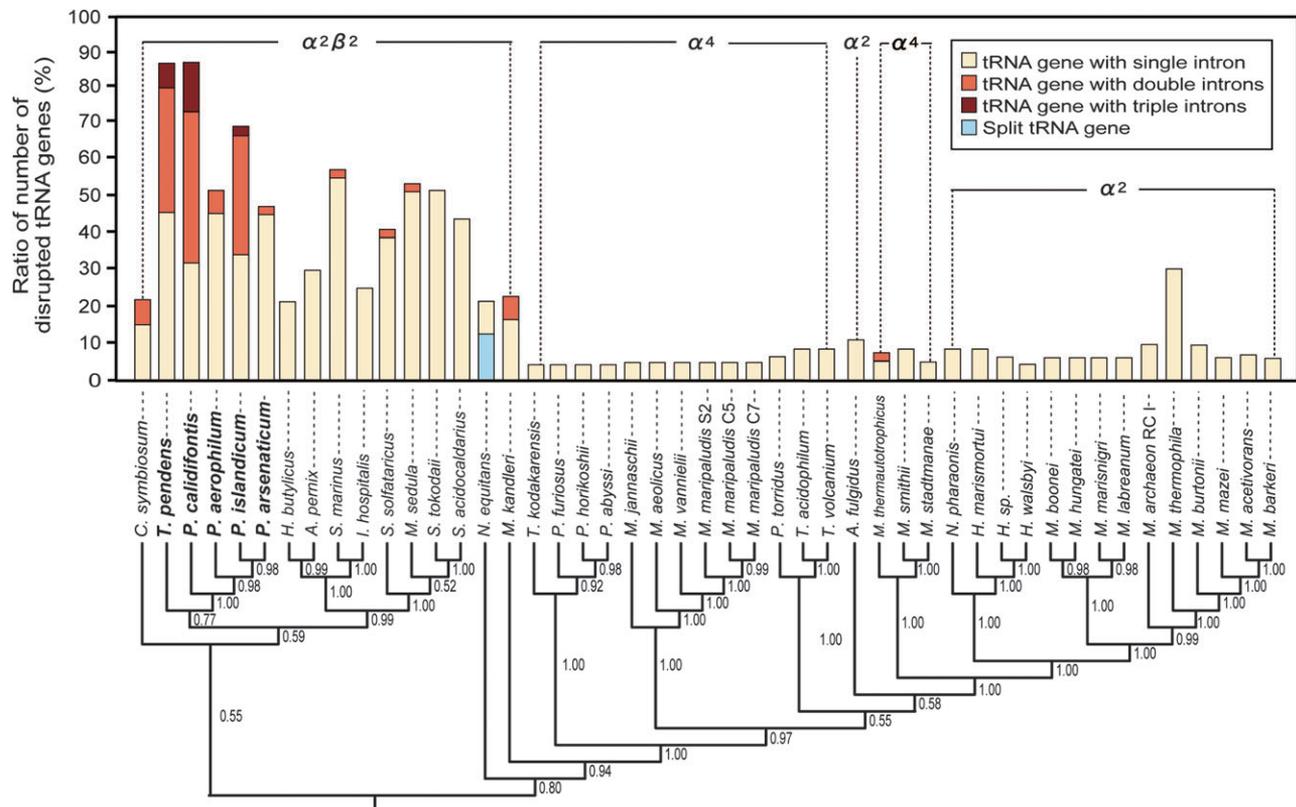


FIG. 2.—Ratio of number of split and intron-containing tRNA genes in Archaea. The ratio of the number of split or intron-containing tRNA genes per total number of tRNA genes in each archaeal genomes are shown for a total of 47 archaeal species (32 Euryarchaeota, 14 Crenarchaeota, and 1 Nanoarchaeota). Endonuclease architectures of pre-tRNA in each species are also denoted ($\alpha_2\beta_2$: heterotetramer; α_4 : homotetramer; and α_2 : homodimer). Thermoproteales, which are found to possess the largest amount of intron-containing tRNA genes, are indicated in bold characters. Posterior probability values are denoted at each branch.

to be intron-containing tRNA genes and half of these genes contained more than two introns.

To analyze the relative effect of the number of tRNA introns and their divergence of location in the tRNA genes, we compared the distribution of tRNA introns based on the nucleotide position of tRNA gene in Thermoproteales and other archaeal species (fig. 3a). In total, 32 nt positions of tRNA genes in the 47 archaeal species were found to contain introns; 20 out of the 32 positions were found to be Thermoproteales-specific intron location, 2 were found to be specific to other archaeal species, and the remaining 10 positions were found to be common between Thermoproteales and other archaeal species. Although 83.4% of tRNA introns in archaeal species except Thermoproteales were concentrated at a canonical position (nucleotide position at 37/38) of tRNA genes, only 24.6% of tRNA introns in Thermoproteales were found at a canonical position, indicating wide diversity of tRNA intron locations in this specific order. Moreover, with the exception of Thermoproteales, all archaeal species tended to contain introns in the 5' half of tRNA gene, whereas tRNA introns in Thermoproteales tended to be equally located in both 5' and 3' halves of tRNA gene. Especially, archaeal species that encoded homotetrameric or homodimeric endonucleases completely lacked introns at the 3' half of the tRNA gene (supplementary fig. S3, Supplementary Material online).

Rapid Increase of tRNA Introns in the Order Thermoproteales

In this paper, comprehensive prediction and analysis of the tRNA genes from 47 archaeal species have discovered an extensive number of tRNA introns in five species of the order Thermoproteales. Further, the locations of tRNA introns in Thermoproteales were found to be widely spread across the nucleotide positions of tRNA genes compared with those in other archaeal species, and surprisingly, 20 intron positions were specific to the order. It is important to determine whether the tRNA introns located at Thermoproteales-specific positions represent an ancient form of tRNA genes or whether these introns have been gained during a later stage of archaeal evolution. Therefore, we further conducted sequence similarity analysis of archaeal tRNA introns and compared the conservation rate among the tRNA intron groups shown in figure 3a: tRNA introns in archaeal species except Thermoproteales, tRNA introns in Thermoproteales located at common positions with other archaeal species, and tRNA introns located at Thermoproteales-specific positions. As a result, the tRNA introns in Thermoproteales were found to be highly conserved within different tRNA species compared with those in other archaeal species (fig. 3b). Especially, tRNA introns located at Thermoproteales-specific nucleotide positions showed the most significant sequence conservation. For example,

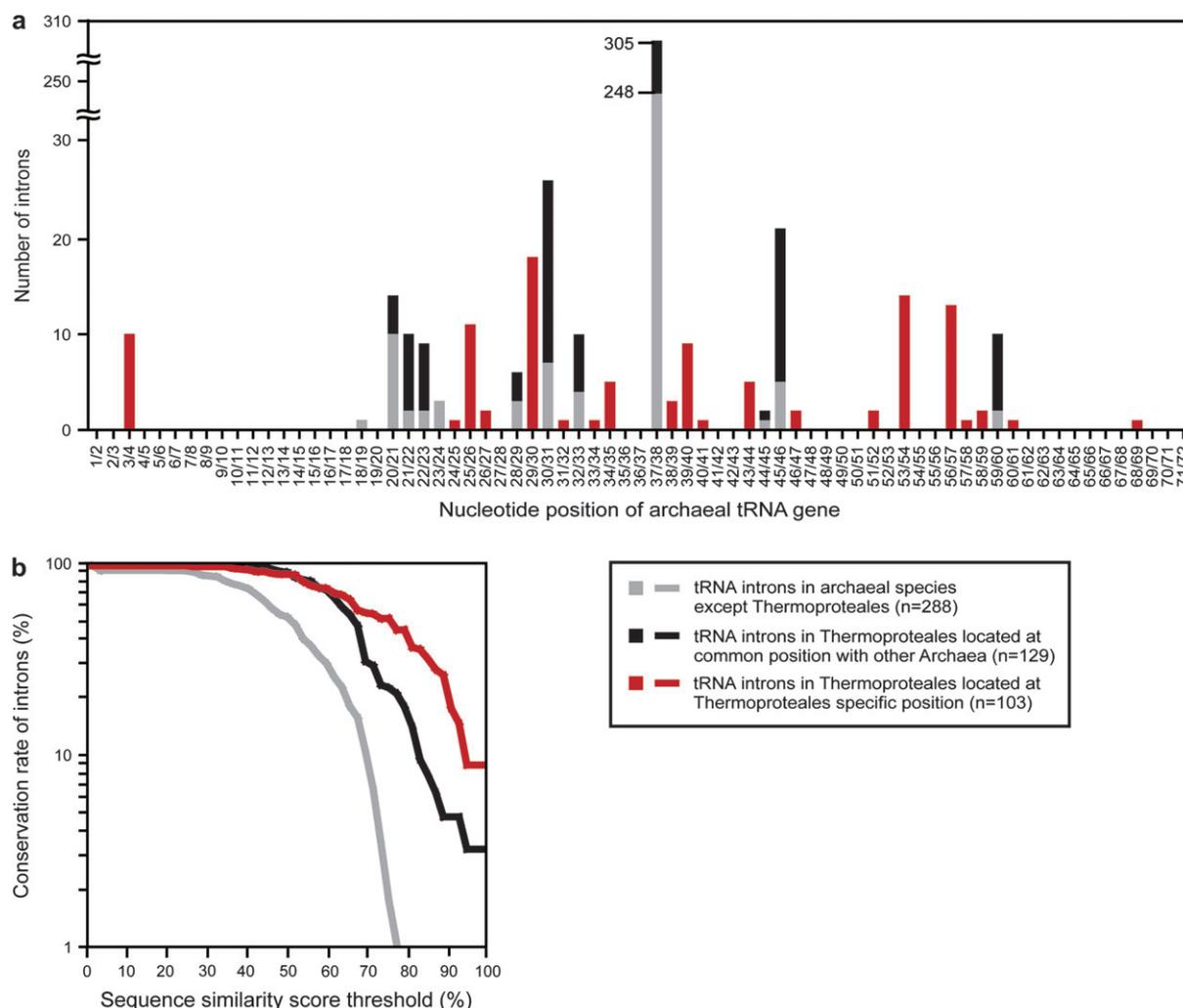


FIG. 3.—Summary of location and sequence conservation of tRNA introns in Archaea. (a) Distribution of tRNA introns in archaeal tRNA genes. Numbering of the nucleotide position of tRNA genes is according to Sprinzl and Vassilenko (2005). tRNA introns were classified into three groups: tRNA introns in all archaeal species except Thermoproteales (gray), tRNA introns in Thermoproteales located at common positions with other archaeal tRNA introns (black), and tRNA introns in Thermoproteales located at Thermoproteales-specific positions (red). (b) Conservation rate of tRNA introns for each group. Conservation rate is defined as the number of tRNA introns above the sequence similarity score threshold per total number of tRNA introns in each group. Colors are the same as those in (a). See Materials and Methods for details of sequence similarity score calculation.

at sequence similarity thresholds above 80%, 42.7% (44 out of 103) of tRNA introns located at Thermoproteales-specific positions were conserved in different tRNA species versus 16.2% (21 out of 129) of tRNA introns of Thermoproteales located at common positions and 0% (0 out of 288) of other archaeal tRNA introns.

Examples of identical introns conserved in different tRNA genes in Thermoproteales are presented in figure 4. The introns located at Thermoproteales-specific positions 22/23 and 43/44 of the tRNA^{Asn} (GTT) and tRNA^{Ile} (GAT) genes in *T. pendens* have 100% sequence similarity, even though the exon sequences of the tRNA genes themselves lack overall similarity (83.1%), suggesting different evolutionary backgrounds (fig. 4a). From the genetic point of view, it is difficult to consider that these introns were individually created in two tRNA genes and maintained their sequences through evolution because the mutation rate of the intron region could be higher than that of the exon

region and would be rapidly modified. Therefore, we consider that intron insertion events could explain why similar introns are found in divergent tRNA species in Thermoproteales, and the high conservation rate of tRNA introns located at Thermoproteales-specific positions strongly suggests that these introns have been newly gained at a later stage of tRNA evolution.

Moreover, we also found a positive correlation between the ratio of the number of tRNA genes possessing a 3'-terminal CCA sequence and the conservation of tRNA introns (supplementary fig. S4, Supplementary Material online). Although the ratio of the number of CCA sequence-possessing tRNA genes remained 30.3% (651 out of 2,143) of all archaeal tRNA genes, approximately 70% (49 out of 71) of the tRNA genes containing highly conserved introns (sequence similarity ranging from 80% to 100%) were found to be encoding the CCA sequence at their 3' terminal. In Archaea, the CCA sequence is generally added by

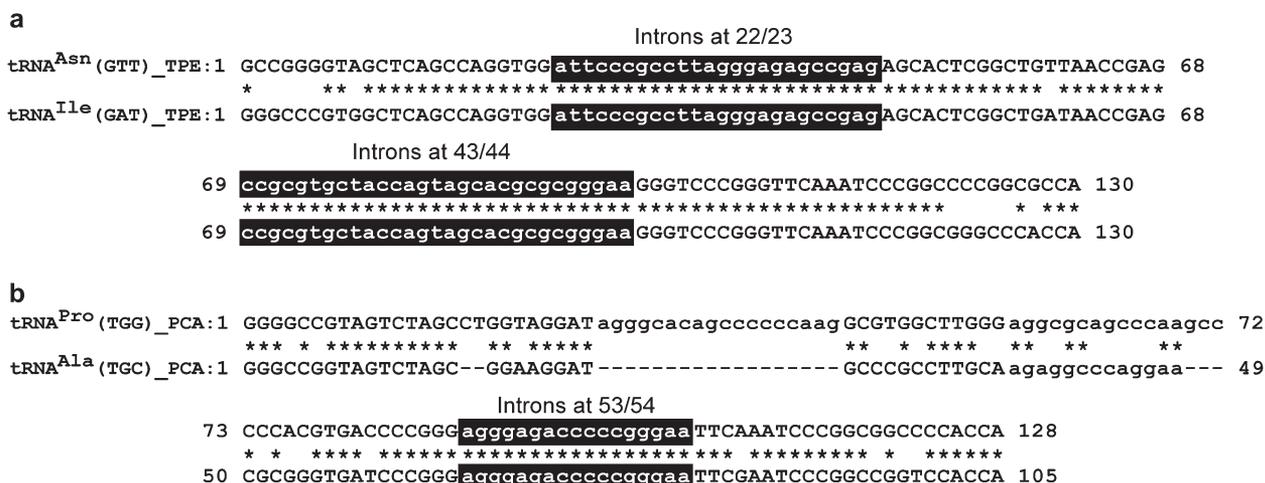


FIG. 4.—Sequence alignment of tRNA genes containing identical intron sequences. (a) Sequences of tRNA^{Asn} (GTT) and tRNA^{Ile} (GAT) genes in *Thermofilum pendens*. (b) Sequences of tRNA^{Pro} (TGG) and tRNA^{Ala} (TGC) genes in *Pyrobaculum calidifontis*. Asterisks denote matching nucleotides in the aligned tRNA gene sequences. All intron sequences are shown in lower case. Identical intron sequences are shown as outline characters on a colored background. The sequences were aligned using ClustalW 2.0 (Larkin et al. 2007) with default parameters; then sequences were manually confirmed to conduct correct structural alignment of tRNA gene.

a CCA-adding enzyme after gene transcription (Yue et al. 1996); therefore, the high rate of CCA sequences found at the 3' terminal of tRNA genes which contain highly conserved introns suggests that intron insertion could have occurred via a retrotransposition event (supplementary fig. S5a, Supplementary Material online). Further, we speculate that it is possible that the *trans*-splicing event found in Nanoarchaea which generates two chimeric tRNAs from two alternative 5'-tRNA halves corresponding to a single 3'-tRNA half (Randau, Calvin, et al. 2005) may also create identical introns in different tRNA genes (supplementary fig. S5b, Supplementary Material online). The sequences of the 3' halves of tRNA^{Pro} (TGG) and tRNA^{Ala} (TGC) genes possessing identical introns at 53/54 found in *P. calidifontis* (fig. 4b) had significantly higher similarity (87.7%) than those of the 5' halves (46.5%), suggesting the possibility that two different tRNA genes containing identical introns are created by retrotransposition via *trans*-splicing of intron-containing pre-tRNA fragments.

Why are tRNA introns of these species increased? Considering the splicing endonuclease evolution, the heterotetrameric endonuclease is estimated to have emerged after a gene duplication event of homotetrameric endonuclease that strictly recognizes only the hBHBh' motif at a specific region of tRNA genes (supplementary fig. S3, Supplementary Material online) (Tocchini-Valentini et al. 2005b). Because all species of the order Thermoproteales possess heterotetrameric endonuclease which is capable of splicing out tRNA introns forming a relaxed HBh' motif located at various noncanonical positions, we speculate that the expansion of substrate specificity of splicing endonucleases may have contributed to the increase of tRNA introns. As another explanation, it is suggested that the presence of an intron in archaeal tRNA genes is a protective mechanism against integration of mobile genetic elements such as conjugative plasmids and viruses because nonintrinsic tRNA genes are ideal targets of site-specific recombination (Randau and Söll 2008). Further investigation

of archaeal tRNA genes including their splicing endonucleases and mobile genetic elements are necessary to substantiate this view.

Implementation and Web Interface of SPLITSdb

We have successfully completed encoding of all tRNA genes required for reading 61 sense codons and an initiator codon in 45 of the 47 species. All sequence and secondary structure data of the tRNA genes and the introns are registered in SPLITSdb (PostgreSQL version 8.2.4 running on a Mac OS X server), a user-friendly web interface database developed for easy viewing and retrieval of tRNA gene information by CGI scripts as outlined in figure 5. The options, "tRNA search" or "BHB search," enable the user to search and screen for the required tRNA genes or BHB motifs. This system allows users to search the tRNA genes or BHB motifs by selecting from one or several search conditions including "Organism," "Anticodon," "tRNA type," "Intron motif," or "Intron location" (fig. 5a and b). The search result is initially displayed in a table format (fig. 5c), and more detailed information such as sequences and secondary structures is displayed (fig. 5d). Zipped files containing all the information of tRNA genes and their BHB motifs are also downloadable (fig. 5c). The home page gives an overview of SPLITSdb that allows for easy navigation throughout the database interface (fig. 5e). As around 90 archaeal genomes are being sequenced and more are continually being sequenced (Lioliou et al. 2008), SPLITSdb is regularly being updated with the latest archaeal tRNA sequences.

Comparative studies of disrupted tRNA genes possessing a common BHB splicing motif that is found in various primitive species will be important to evaluate the origin and evolution of tRNA genes during the early stages of life. Our comprehensive data set containing novel tRNA gene candidates having multiple introns, together with SPLITSdb, which is the only database that includes split

- Andersen ES, Rosenblad MA, Larsen N, Westergaard JC, Burks J, Wower IK, Wower J, Gorodkin J, Samuelsson T, Zwieb C. 2006. The tmRDB and SRPDB resources. *Nucleic Acids Res.* 34:D163–D168.
- Diener JL, Moore PB. 1998. Solution structure of a substrate for the archaeal pre-tRNA splicing endonucleases: the bulge-helix-bulge motif. *Mol Cell.* 1:883–894.
- Di Giulio M. 1999. The non-monophyletic origin of the tRNA molecule. *J Theor Biol.* 197:403–414.
- Di Giulio M. 2006. The non-monophyletic origin of the tRNA molecule and the origin of genes only after the evolutionary stage of the last universal common ancestor (LUCA). *J Theor Biol.* 240:343–352.
- Di Giulio M. 2008. Transfer RNA genes in pieces are an ancestral character. *EMBO Rep.* 9:820.
- Eddy SR, Durbin R. 1994. RNA sequence analysis using covariance models. *Nucleic Acids Res.* 22:2079–2088.
- Fujishima K, Sugahara J, Tomita M, Kanai A. 2008. Sequence evidence in the archaeal genomes that tRNAs emerged through the combination of ancestral genes as 5' and 3' tRNA halves. *PLoS ONE.* 3:e1622.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* 17:754–755.
- Kaine BP, Gupta R, Woese CR. 1983. Putative introns in tRNA genes of prokaryotes. *Proc Natl Acad Sci USA.* 80:3309–3312.
- Larkin MA, Blackshields G, Brown NP, et al. (13 co-authors). 2007. Clustal W and Clustal X version 2.0. *Bioinformatics.* 23:2947–2948.
- Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32:11–16.
- Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics.* 23:127–128.
- Li H, Trotta CR, Abelson J. 1998. Crystal structure and evolution of a transfer RNA splicing enzyme. *Science.* 280:279–284.
- Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC. 2008. The genomes on line database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.* 36:D475–D479.
- 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.
- Lykke-Andersen J, Garrett RA. 1997. RNA-protein interactions of an archaeal homotetrameric splicing endoribonuclease with an exceptional evolutionary history. *EMBO J.* 16: 6290–6300.
- Maizels N, Weiner AM. 1994. Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proc Natl Acad Sci USA.* 91:6729–6734.
- 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.* 8:1189–11232.
- Marck C, Grosjean H. 2003. Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications. *RNA.* 9:1516–1531.
- Putz J, Dupuis B, Sissler M, Florentz C. 2007. Mamit-tRNA, a database of mammalian mitochondrial tRNA primary and secondary structures. *RNA.* 13:1184–1190.
- Randau L, Calvin K, Hall M, Yuan J, Podar M, Li H, Söll D. 2005. The heteromeric Nanoarchaeum equitans splicing endonuclease cleaves non-canonical bulge-helix-bulge motifs of joined tRNA halves. *Proc Natl Acad Sci USA.* 102:17934–17939.
- Randau L, Munch R, Hohn MJ, Jahn D, Söll D. 2005. Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5'- and 3'-halves. *Nature.* 433:537–541.
- Randau L, Pearson M, Söll D. 2005. The complete set of tRNA species in Nanoarchaeum equitans. *FEBS Lett.* 579:2945–2947.
- Randau L, Söll D. 2008. Transfer RNA genes in pieces. *EMBO Rep.* 9:623–628.
- Soma A, Onodera A, Sugahara J, Kanai A, Yachie N, Tomita M, Kawamura F, Sekine Y. 2007. Permuted tRNA genes expressed via a circular RNA intermediate in Cyanidioschyzon merolae. *Science.* 318:405–407.
- Sprinzi M, Vassilenko KS. 2005. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* 33:D139–D140.
- 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.
- Sugahara J, Yachie N, Sekine Y, Soma A, Matsui M, Tomita M, Kanai A. 2006. SPLITS: a new program for predicting split and intron-containing tRNA genes at the genome level. *In Silico Biol.* 6:411–418.
- Sun FJ, Caetano-Anolles G. 2008. Evolutionary patterns in the sequence and structure of transfer RNA: early origins of archaea and Viruses. *PLoS Comput Biol.* 4(3):e1000018.
- Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP. 2005a. Coevolution of tRNA intron motifs and tRNA endonuclease architecture in Archaea. *Proc Natl Acad Sci USA.* 102:15418–15422.
- Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP. 2005b. Structure, function, and evolution of the tRNA endonucleases of Archaea: an example of subfunctionalization. *Proc Natl Acad Sci USA.* 102:8933–8938.
- Wilm A, Mainz I, Steger G. 2006. An enhanced RNA alignment benchmark for sequence alignment programs. *Algorithms Mol Biol.* 1:19.
- Yoshinari S, Fujita S, Masui R, Kuramitsu S, Yokobori S, Kita K, Watanabe Y. 2005. Functional reconstitution of a crenarchaeal splicing endonuclease in vitro. *Biochem Biophys Res Commun.* 334:1254–1259.
- Yue D, Maizels N, Weiner M. 1996. CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the CCA-adding enzyme from the archaeal hyperthermophile *Sulfolobus shibatae*. *RNA.* 2:895–908.

Takashi Gojobori, Associate Editor

Accepted September 24, 2008