Tri-split tRNA is a transfer RNA made from 3 transcripts that provides insight into the evolution of fragmented tRNAs in archaea

Kosuke Fujishima[a,b], Junichi Sugahara[a,b], Kaoru Kikuta[a,c], Reiko Hirano[a], Asako Sato[a], Masaru Tomita[a,b,c], and Akio Kanai[a,b,c,1]

[a]Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0017, Japan; [b]Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa 252-8520, Japan; and [c]Faculty of Environment and Information Studies, Keio University, Fujisawa 252-8520, Japan

Edited by Dieter Söll, Yale University, New Haven, CT, and approved December 16, 2008 (received for review August 20, 2008)

Transfer RNA (tRNA) is essential for decoding the genome sequence into proteins. In Archaea, previous studies have revealed unique multiple intron-containing tRNAs and tRNAs that are encoded on 2 separate genes, so-called split tRNAs. Here, we discovered 10 fragmented tRNA genes in the complete genome of the hyperthermoacidophilic Archaeon *Caldivirga maquilingensis* that are individually transcribed and further *trans*-spliced to generate all of the missing tRNAs encoding glycine, alanine, and glutamate. Notably, the 3 mature tRNA^{Gly}'s with synonymous codons are created from 1 constitutive 3′ half transcript and 4 alternatively switching transcripts, representing tRNA made from a total of 3 transcripts named a ''tri-split tRNA.'' Expression and nucleotide sequences of 10 split tRNA genes and their joined tRNA products were experimentally verified. The intervening sequences of split tRNA have high identity to tRNA intron sequences located at the same positions in intron-containing tRNAs in related Thermoproteales species. This suggests that an evolutionary relationship between intron-containing and split tRNAs exists. Our findings demonstrate the first example of split tRNA genes in a free-living organism and a unique tri-split tRNA gene that provides further insight into the evolution of fragmented tRNAs.

tRNA intron | RNA processing | molecular evolution | *trans*-splicing | *Caldivirga maquilingensis*

The origin and evolution of tRNA is one of the most important subjects being discussed in the field of molecular evolution, with varying hypotheses being proposed (1–6). Three types of tRNA genes have previously been identified in archaeal genomes: nonintronic tRNA, which is encoded on a single gene with no intron sequence; intron-containing tRNA, which is encoded on a single gene with a maximum of 3 introns punctuating the mature tRNA sequence at various locations (7–9); and *trans*-spliced tRNA—so-called split tRNA—which has 5′ and 3′ halves encoded on 2 separate genes found only in the hyperthermophilic archaeal parasite, *Nanoarchaeum equitans* (10). Interestingly, split tRNA and intron-containing tRNA share a common bulge–helix–bulge (BHB) consensus motif around the intron/leader–exon boundaries that can be cleaved by the same tRNA splicing endonucleases (11, 12). BHB motifs are further classified by their structure into the canonical form (hBHBh′) and relaxed forms (HBh′ and BHL). The discovery of these tRNA genes raised the question of whether ancestral tRNA was encoded on a single gene or on separate genes. Our previous study has shown clear phylogenetic relationships among these 3 types of archaeal tRNAs (13). Because *N. equitans* is a parasite with indications of a massive genome reduction (14), whether its tRNAs represent the ancient form of tRNA or a later product of genome reduction is still unclear. Therefore, we have been conducting comprehensive prediction and analysis of tRNA sequences in various species on the basis of our original software, SPLITS (8, 9, 13, 15, 16). We have especially focused on the hyperthermoacidophilic archaeon *Caldivirga maquilingensis*, which belongs to the deep branching archaeal order Thermoproteales and was isolated from an acidic hot spring in the Philippines (17), because it is the only species that lacks 6 tRNA genes for a full complement of the codon table. In this study, we have successfully found all of the missing tRNAs (corresponding to Gly, Ala, Glu) as novel split tRNAs in *C. maquilingensis* on the basis of computational analysis and experimental verification. Our study presents the first example of split tRNA genes in a free-living organism and a novel type of tRNA produced from 3 individual pretRNA transcripts. We also found sequence evidence for an evolutionary relationship between fragmented tRNA genes and intron-containing tRNA genes in related species. From an evolutionary point of view, the discovery of fragmented tRNA molecules found in the deep-branched archaea provides a unique opportunity to test various hypotheses to explain the origin of tRNA.

Results and Discussion

Computational Prediction of Novel Split tRNAs in *C. maquilingensis*. A comprehensive prediction of archaeal tRNA genes remains a challenging task in the field of bioinformatics, because various types of disrupted tRNA genes are found in archaea. To provide a complete picture of tRNA diversity and evolution in the archaeal domain, we recently developed a novel tRNA database, SPLITSdb (18), by retrieving a set of reliable archaeal tRNA sequences using 3 tRNA prediction software packages: tRNAscan-SE (19), ARAGORN (20), and SPLITS (9). Application of this tRNA prediction procedure to the recently sequenced genome of the hyperthermoacidophilic archaeon *C. maquilingensis* predicted 4 of the 6 missing tRNAs: tRNA^{Gly} (CCC), tRNA^{Ala} (CGC/TGC), and tRNA^{Glu} (TTC) as the typical 2-fragmented type of split tRNAs. Each of the split tRNA pairs possessed a flanking leader sequence at the 5′ or the 3′ end to meet each other in the cell by forming long complementary RNA helices. Also, a BHB motif for recognition by tRNA endonuclease was predicted at the leader–exon boundaries

Author contributions: K.F. and A.K. designed research; K.F., R.H., A.S., and A.K. performed research; J.S. and K.K. contributed new reagents/analytic tools; K.F., J.S., K.K., M.T., and A.K. analyzed data; and K.F. and A.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. NC_009954 (*Caldivirga maquilingensis* IC-167), NC_003364 (*Pyrobaculum aerophilum* str. IM2), NC_009376 (*P. arsenaticum* DSM 13514), NC_009073 (*P. calidifontis* JCM 11548), NC_008701 (*P. islandicum* DSM 4184), NC_008698 (*Thermofilum pendens* Hrk 5), and NC_010525 (*T. neutrophilus* V24Sta)] and the DDBJ database (split tRNA genes) [AB470216 (tRNA^{Gly}CCC), AB470217 (tRNA^{Gly}TCC), AB470218 (tRNA^{Gly}GCC), AB470219 (tRNA^{Ala}CGC), AB470220 (tRNA^{Ala} TGC), and AB470221 (tRNA^{Glu}TTC)].

[1]To whom correspondence should be addressed. E-mail: akio@sfc.keio.ac.jp.

This article contains supporting information online at www.pnas.org/cgi/content/full/0808246106/DCSupplemental.

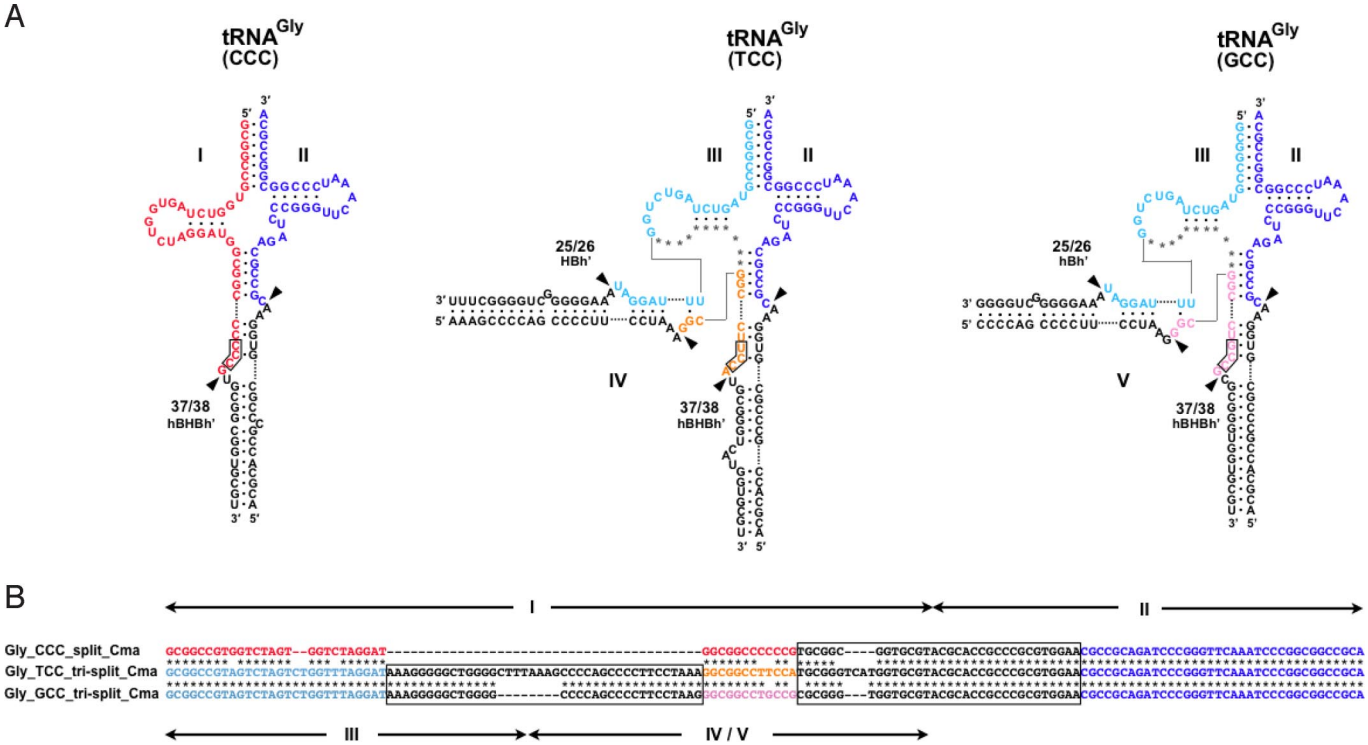

Fig. 1. Alternative splicing of 5 split tRNA^{Gly} transcripts in *C. maquilingensis*. (*A*) Schematic diagrams of the 3 novel tRNA^{Gly} assembly models. Three potential mature joined tRNAs formed by alternative assembly of 5 split tRNA^{Gly} transcripts (I–V) are shown: I + II, tRNA^{Gly} (CCC); III + IV + II, tRNA^{Gly} (TCC); and III + V + II, tRNA^{Gly} (GCC). Anticodons are boxed. The locations of bulge–helix–bulge (BHB) splicing motifs represent each leader–exon junction (arrowhead). (*B*) Nucleotide sequence alignment of the 3 pretRNA^{Gly} candidates with the inclusion of gaps (−). Identical nucleotide sequences are indicated by asterisks. Leader sequences are boxed. Exon sequences are colored by transcript (I, red; II, blue; III, light blue; IV, orange; and V, pink).

[supporting information (*SI*) Fig. S1]. Although 4 tRNA species were found as 2-fragmented split tRNAs in the genome, the remaining 2 synonymous tRNA^{Gly}'s (TCC/GCC) were not identified by any of the tRNA prediction software packages.

To our surprise, 3 additional fragmented tRNA^{Gly} genes were found because of further investigation of the missing tRNA sequences by a Blat search (21) and identification of the possible leader sequences. Particularly, 2 of these genes possessed a 12-bp exon sequence with 2 leader sequences on both ends that were complementary to that of other split tRNA^{Gly}'s. In an attempt to obtain the overall picture of tRNA^{Gly}'s, we performed computational assembly of putative tRNA^{Gly} fragments using SPLITS. Each pair was manually evaluated on the basis of the complementality of leader sequences and the BHB structure at the leader–exon junction. The assembly model in Fig. 1*A* clearly represents the reproduction of 3 synonymous tRNA^{Gly}'s from 5 tRNA^{Gly} fragments (I–V) representing tRNA^{Gly} (CCC) produced from 2 fragments (I + II) and tRNA^{Gly}'s (TCC/GCC) produced from 3 fragments (III + IV/V + II). We named these identified 3-fragmented tRNAs "tri-split tRNA," since it is a unique example found in a living organism. Interestingly, the 3′ half tRNA^{Gly} fragment (II) was constitutively used in all 3 synonymous tRNA^{Gly}'s, while the other 4 transcripts (I, III, IV, V) were alternatively joined to produce specific tRNA^{Gly}'s. Furthermore, sequence alignments of split and tri-split tRNA^{Gly} sequences have shown that both the exon and the leader sequences are highly conserved (Fig. 1*B*), suggesting that two different types of split tRNAs may be derived from the same ancestral origin.

The analysis of leader–exon junctions of the newly found split tRNA pairs has revealed that either canonical (hBHBh′) or relaxed (HBh′) BHB motifs exist at proposed positions. However, an unusual single-nucleotide mismatch was found within the central H helix of tRNA^{Gly} (CCC/TCC) and tRNA^{Glu} (TTC). In Archaea, three types of splicing endonucleases have been reported: heterotetrameric ($\alpha_2\beta_2$), homodimeric (α_2), and homotetrameric (α_4) (22). A heterotetrameric enzyme is able to process both canonical and relaxed BHB motifs at various tRNA positions (12) and is even capable of cleaving BHB motifs with variable bulge lengths, as in the bulge–helix–bulge of 2-4-3 and 4-4-4 nucleotides (nt) in size (11). To examine the consistency of the BHB motifs and the 2 putative tRNA endonucleases (Cmaq_0348 and Cmaq_1747) found in *C. maquilingensis*, we performed phylogenetic analysis of archaeal tRNA splicing endonucleases to gain a better understanding of the 2 proteins. The 2 proteins are found to be located within the branch of α- and β-subunits, respectively, which clearly show that the splicing endonuclease in *C. maquilingensis* belongs to the heterotetrameric splicing endonuclease family (Fig. S2 and *SI Materials and Methods*). Hence, the single-nucleotide mismatch at the central 4-bp helix of the BHB motif may represent a unique example of the broad substrate specificity of heterotetrameric splicing endonucleases. In summary, we have predicted 10 unique split tRNA genes in the genome of *C. maquilingensis* that are assumed to be transcribed and processed by the heterotetrameric splicing endonuclease to produce the missing 6 tRNAs (Fig. 2*A*).

Next, to examine whether these split tRNA genes are actually expressed in the cell, we first conducted computational analysis of the promoter sequence. On the basis of the preference of the promoter sequences, archaeal species can be classified into group A (TA-rich) or group B (G-rich) (23). The tRNA promoter sequences of group A genomes have been carefully analyzed in *Pyrobaculum aerophilum* (AAANNTTAAAAA) and in *N. equitans* (AAAATTTTTAAAT), which are positioned at conserved distances from the transcription initiation site to produce leaderless tRNA transcripts (24). We found that

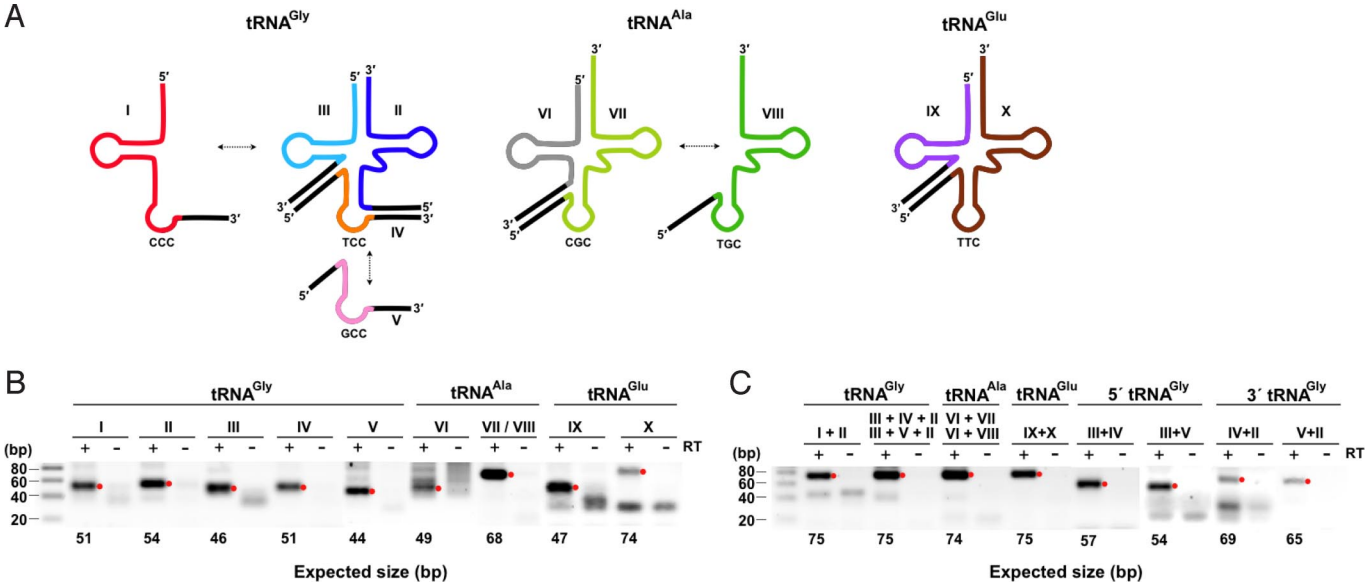

Fig. 2. A combination of 10 novel split tRNAs and their gene expression. (*A*) Assembly models of 6 different mature joined tRNAs in *C. maquilingensis*. The combination of 5 transcripts forms 3 mature $tRNA^{Gly}$ (CCC/TCC/GCC), 3 transcripts form 2 mature $tRNA^{Ala}$ (CGC/TGC), and 2 transcripts form 1 mature $tRNA^{Glu}$ (TTC). Exon sequences are colored by transcript (I, red; II, blue; III, light blue; IV, orange; V, pink; VI, gray; VII, light green; VIII, green; IX, purple; and X, brown). Expression of (*B*) each split tRNA gene and (*C*) mature joined tRNAs and intermediate $tRNA^{Gly}$'s was confirmed by RT-PCR. Samples without reverse transcriptase were used as negative controls (RT−) to ensure that only RNA transcripts were amplified (RT+). EtBr-stained DNA bands were detected as negative images by Molecular Imager FX Pro (Bio-Rad Laboratories). Red marks indicate each position of target transcripts.

C. maquilingensis also belongs to group A, and, as expected, it possesses similar TA-rich consensus sequences (GAAANGCT-TATAAA) upstream of each split tRNA gene (Fig. S3). The strict conservation of guanosines at the first and sixth positions of the consensus sequence is unique but could be evolutionarily related to the G-rich sequence adjacent to the *P. aerophilum* tRNA promoter. These features suggest that the 10 split genes are individually expressed.

Expression, Processing, and Functionality of Split tRNAs. To experimentally verify the expression and processing of the 10 split tRNA genes, reverse transcriptase–PCR (RT-PCR) was performed against *C. maquilingensis* total RNA, using specific primers for each transcript (Table S1). Amplified PCR products of all 10 split tRNAs (I–X) were detected at the predicted size in a RT-dependent manner (Fig. 2*B*), indicating that all split tRNAs are actually expressed in the cell. We further examined whether these split tRNAs produce mature joined tRNAs in the fashion of our predicted assembly models by using combinations of primers for different split tRNAs [e.g., sense primer for I and antisense primer for II amplify mature $tRNA^{Gly}$ (CCC)]. As a result, all 6 mature joined tRNAs, $tRNA^{Gly}$ (CCC/TCC/GCC), $tRNA^{Ala}$ (CGC/TGC), and $tRNA^{Glu}$ (TTC), were successfully amplified (Fig. 2*C*) and sequencing analysis proved that all mature tRNA sequences are cleaved and ligated at the predicted leader–exon junctions (Fig. S4*A*). In addition, we performed a BLAST search and reconfirmed that all 6 mature joined tRNA sequences are not encoded in any region of the *C. maquilingensis* genome. According to the maturation model of tri-split tRNAs, 3 processing pathways can take place: (*i*) initiation from the assembly of 5′ halves, (*ii*) initiation from the assembly of 3′ halves, and (*iii*) initiation from either half, thereby indicating that 2 different intermediate forms (5′ half and 3′ half) of the tri-split $tRNA^{Gly}$'s (TCC/GCC) may exist. To reveal the precise processing step of tri-split tRNAs, we examined the presence of 5′ tRNA intermediates (III + IV/V) and 3′ tRNA intermediates (IV/V + II) by RT-PCR, using sequence-specific primers. We detected all 4 predicted tri-split tRNA intermediates with their expected sizes (Fig. 2*C*), and nucleotide sequence analysis confirmed that the PCR products are tRNA intermediates derived from split $tRNA^{Gly}$'s (Fig. S4*B*), indicating that maturation of the tri-split tRNAs can initiate from either half.

Finally, we conducted 2 additional analyses to verify the functionality of mature joined tRNAs in *C. maquilingensis*. First, the mature joined tRNA sequences were aligned with orthologus tRNA sequences from related Thermoproteales species (*Thermophilum pendens* and *P. aerophilum*), which showed perfect conservation of all tRNA identity elements necessary for the recognition of cognate amino acids (Fig. S5). Second, we conducted an in vitro tRNA aminoacylation assay to access the functionality of the mature joined tRNAs of 2-fragmented $tRNA^{Gly}$ (CCC) and 3-fragmented $tRNA^{Gly}$ (TCC), using recombinant *C. maquilingensis* glycyl-tRNA synthetase (GlyRS). We successfully obtained clear separation of glycylated $tRNA^{Gly}$ and uncharged $tRNA^{Gly}$ on acid/urea gels in a glycine- and enzyme-dependent manner (Fig. S6 and *SI Materials and Methods*). All these results indicate that mature joined tRNAs produced from split and tri-split tRNAs are actually functional tRNAs involved in protein biosynthesis.

Evolutionary Relationship Between Split tRNA and Intron-Containing tRNA. Previously, split tRNAs have been found only in the fast-evolving archaeal parasite *N. equitans*, which possesses a highly reduced genome (25). In this study, we report a unique split tRNA in the free-living archaeon *C. maquilingensis*, which has no sign of genome minimization, indicating that the split tRNA gene is not a derivative of DNA rearrangement accompanied by genome reduction. In addition, most of the split tRNAs found in the 2 species correspond to different cognate amino acids and their intervened positions are not compatible (Fig. S7). Despite the evolutionary distance of the 2 species belonging to distinct phyla Nanoarchaeota and Crenarchaeota, split tRNAs found in both species share many common features such as a long complementary leader sequence for assembly, individual expression by a conserved promoter sequence, and alternatively used tRNA pieces to produce tRNAs with synon-

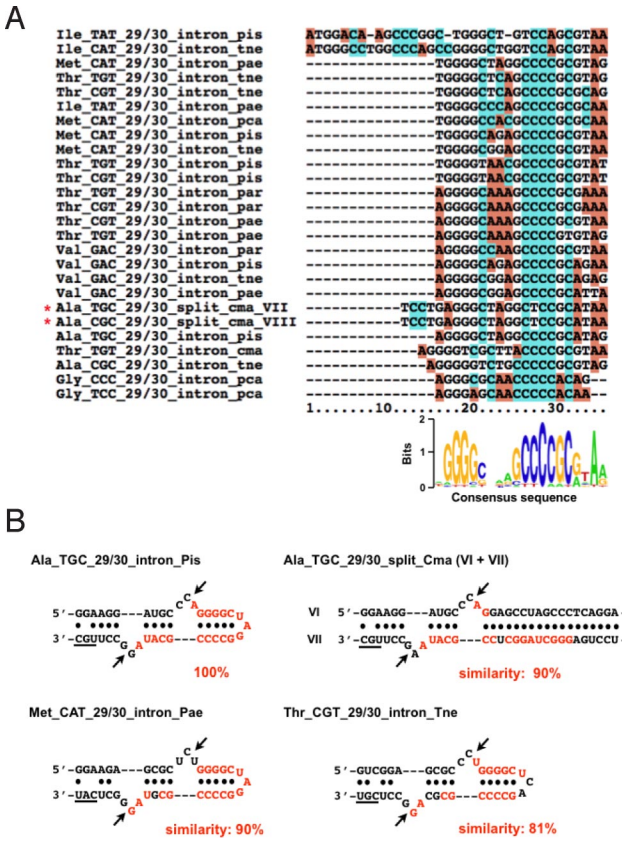

Fig. 3. Consensus between split tRNAs and intron-containing tRNAs in Archaea. (*A*) Nucleotide sequence alignment of 26 archaeal tRNA intron and leader sequences intervening at location 29/30 of 6 species in the Thermoproteales order. The visualization of consensus sequences was created by WebLogo (31). Abbreviations of the species names are as follows: Cma, *C. maquilingensis*; Pae, *Pyrobaculum aerophilum*; Par, *P. arsenaticum*; Pca, *P. calidifontis*; Pis, *P. islandicum*; and Tne, *Thermoproteus neutrophilus*. (*B*) RNA secondary structure of the intron/leader–exon junction at position 29/30 of 3 intron-containing tRNAs and split tRNA^{Ala} (TGC). Sequence identity (percentage) is denoted for each intron/leader sequence compared with the intron sequence of intron-containing tRNA^{Ala} (TGC) from *P. islandicum* (100%).

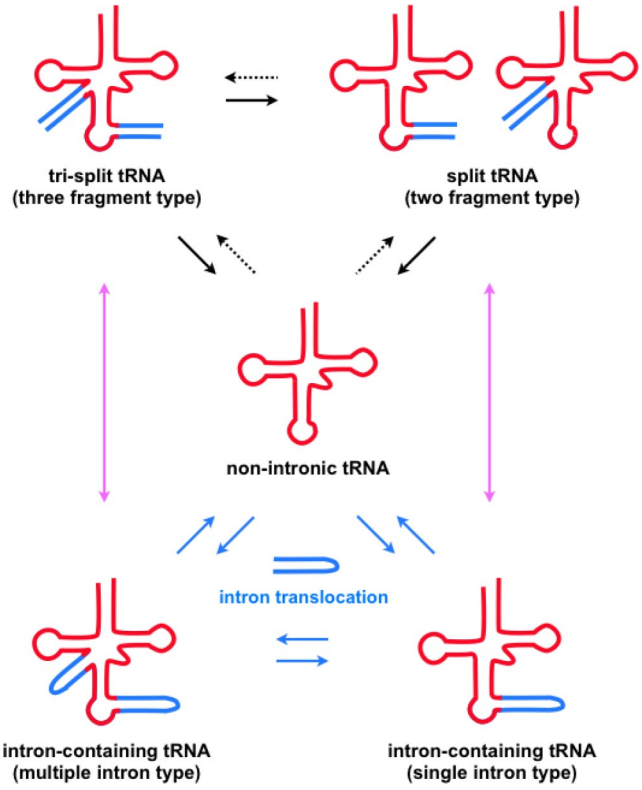

Fig. 4. Schematic of archaeal tRNA evolution. The relation of 5 unique types of tRNAs is represented with exon (red) and intron/leader (blue) sequences indicated. Possible evolutionary direction via tRNA splicing (black arrow), reverse direction of tRNA splicing (dotted black arrow), intron translocation (blue arrow), and sequence evidence indicating the relationship between split tRNA and intron-containing tRNA (see Fig. 3 for details) (pink arrows) are shown.

ymous anticodons. Therefore, we next investigated the evolutionary relationship between the split tRNAs and other type of tRNAs.

In our previous research, we found a leader sequence of *N. equitans* split tRNA^{Lys} and an intron from *P. aerophilum* tRNA^{Arg}'s possessing similar sequences with 50% identity that were both intervened at an identical nucleotide position, representing a unique example connecting the 2 types of tRNA genes (13). In the current study, a total 3 nucleotide positions were intervened by the leader sequences of *C. maquilingensis* split tRNAs, such as positions 25/26 and 37/38 in tRNA^{Gly}, 29/30 in tRNA^{Ala}, and 25/26 in tRNA^{Glu}. A comparison of the split positions and intron positions from 400 archaeal intron-containing tRNAs revealed two intervening positions (25/26, 29/30) as Thermoproteales-specific intron locations. Significant sequence similarity was observed among the 5′ leader sequences of split tRNA^{Ala} (CGC/TGC) and tRNA intron sequences located at position 29/30 (Fig. 3*A*). Notably, the tRNA intron of tRNA^{Ala} (TGC) from *P. islandicum* shared 90% sequence identity with the 2 leader sequences (VII and VIII). This new sequence evidence strongly suggests that at least some intron-containing tRNA genes are derived from split tRNA genes or vice versa. Intriguingly, similar intron sequences were also observed in various tRNA species located at 29/30 with a conserved length of 20 nt corresponding to different amino acids (Fig. 3*B*). These results further support the recently proposed intron translocation scenario in the order Thermoproteales (18). However, we could not find intron sequences that are similar to the leader sequences of tRNA^{Gly} (CCC/TCC/GCC) and tRNA^{Glu} (CGC/TGC).

On the basis of our current research and the recent knowledge of tRNAs in archaea, we have summarized a model representing the relationships among different types of tRNAs (Fig. 4). Sequence evidence supports the connection between tri-split and split (Fig. 1*B*) and the connection between split and intron-containing tRNAs (Fig. 3*B*). We also show that an intron translocation event may play an important role in the gain and loss of intron sequences among diverse tRNA species, leading to the transition of tRNA types between intron-containing tRNAs and nonintronic tRNAs. The evolutionary order of the fragmented and unfragmented types of tRNA genes is an important topic that has recently been debated (26, 27). Because tRNA is a typical target of the genetic mobile element, one explanation for the late acquisition of tRNA gene disruption is to prevent the attack of conjugative plasmids and viruses (27). That the intervening position of the split tRNAs in *C. maquiligensis* is found at the Thermoproteales-specific intron position further supports the late acquisition of split tRNA genes. On the contrary, if intron-containing tRNA predates split tRNA, multiple copies of identical tRNA fragments would likely be produced. However, neither redundant copies nor pseudo-tRNAs were found for the alternatively used split tRNA genes (II, III, and VI) in the genome of *C. maquilingesis*. Moreover, producing a split tRNA

gene from an intron-containing tRNA gene by placing the promoter at the exact upstream position and further acquiring a long complementary leader sequence at the flanking region seems unlikely to occur (28), because these modifications have to take place while maintaining the function of tRNA.

Besides, we assume that evolutionary transition of tRNA types could not be explained solely by sequence modification at the DNA level. Rather, since tRNA and tRNA-like sequences serve as a potential target of reverse transcriptase and replicase (29, 30), modification in the tRNA sequences (various sequence products generated by tRNA processing) could later become stabilized in the genome via retrotransposition (i.e., reverse transcription and integration of RNA into the genome). The similar intron sequences found within diverse tRNA species further support this hypothesis. In conclusion, split and tri-split tRNA genes found in the free-living archaeon *C. maquilingensis* represent a unique example of novel tRNA genes and provide further insight into the evolutionary traits of fragmented and unfragmented tRNA genes. So far we have obtained the complete genome sequences of only 50 archaeal species, which are too few to answer the underlying question of their evolutionary relationships. Thus, finding and characterizing numerous novel tRNA genes within metagenomic sequences will be our next goal to gain a clear overview of the evolutionary transition of tRNA types in archaea.

Materials and Methods

Prediction of Split tRNA Genes of *C. maquilingensis*. tRNA with 2 fragments was predicted by scanning the whole *C. maquilingensis* IC-167 genome (NC_009954) by the transmode (specialized for finding split tRNAs) of SPLITS with the following parameters: $-t$, (switch to *trans*-splicing mode), and $-p$, 0.6 (change cutoff value of position weight matrix). Archaea-specific training data sets [prepared in the SPLITS (ref. 9) software package] were configured. tRNA with 3 fragments was predicted by an online Blat search [UCSC Archaeal genome browser (ref. 21)], using the exon tRNA sequences of related Thermoproteales as a query. The existence of a leader sequence was manually checked.

RT-PCR Analysis and Nucleotide Sequencing. The active culture of *C. maquilingensis* (DSM 13496) was obtained from DSMZ, the German Resource Centre for Biological Material. The culture was initially treated with the RNAprotect Bacteria Reagent (Qiagen) for stabilizing RNA and then total RNA was isolated using the RNeasy Midi Kit (Qiagen), with the exception of performing purification of RNA by phenol–chloroform extraction. The RNA fraction was further treated with RNase-free DNase I (Takara). RT-PCR of candidate tRNA genes was carried out using a ReverTra Dash Kit and KOD FX (TOYOBO Biochemicals). PCR was performed for 28 cycles at 98 °C (10 s), 59 °C (2 s), and 74 °C (1 s). The PCR products were separated using 3% NuSieve 3:1 agarose gel electrophoresis (Cambrex Bio Science) and stained with ethidium bromide. Gels were visualized with a Molecular Imager FX Pro (Bio-Rad Laboratories). The primers used in the RT-PCR analysis are summarized in Table S1. RT-PCR products were further purified by using an illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) and subcloned into the *Srf*I-digested pPCR-Script vector (Stratagene). The nucleotide sequences of the DNA inserts were determined by an ABI3100 DNA Sequencer (Applied Biosystems).

ACKNOWLEDGMENTS. We thank R. Martin, H. Kochiwa, Y. Watanabe, M. Matsui, A. Shinhara, and K. Takane from our laboratory for their insightful comments and technical support. We also thank A. Soma (Rikkyo University, Japan) and Y. Kikuchi (Toyohashi University of Technology, Japan) for their beneficial advice. This work is supported by the Japan Society for the Promotion of Science, by a grant from the Ministry of Education, Culture, Sports, Science and Technology of Japan (The 21st Century Center of Excellence Program, entitled Understanding and Control of Life's Function via Systems Biology), and by research funds from the Yamagata prefectural government and Tsuruoka City, Japan.

1. Di Giulio M (1992) On the origin of the transfer RNA molecule. *J Theor Biol* 159(2):199–214.
2. Dick TP, Schamel WA (1995) Molecular evolution of transfer RNA from two precursor hairpins: implications for the origin of protein synthesis. *J Mol Evol* 41(1):1–9.
3. Maizels N, Weiner AM, Yue D, Shi PY (1999) New evidence for the genomic tag hypothesis: archaeal CCA-adding enzymes and tDNA substrates. *Biol Bull* 196(3):331–333; discussion 333–334.
4. Nagaswamy U, Fox GE (2003) RNA ligation and the origin of tRNA. *Orig Life Evol Biosph* 33(2):199–209.
5. Schimmel P, Giege R, Moras D, Yokoyama S (1993) An operational RNA code for amino acids and possible relationship to genetic code. *Proc Natl Acad Sci USA* 90(19):8763–8768.
6. Tanaka T, Kikuchi Y (2001) Origin of the cloverleaf shape of transfer RNA -the double-hairpin model: implication for the role of tRNA intron and the long extra loop. *Viva Orig* 29:134–142.
7. 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(10):1189–1232.
8. 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(5):671–681.
9. Sugahara J, *et al.* (2006) SPLITS: a new program for predicting split and intron-containing tRNA genes at the genome level. *In Silico Biol* 6(5):411–418.
10. Randau L, Munch R, Hohn MJ, Jahn D, Soll D (2005) Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5′- and 3′-halves. *Nature* 433(7025):537–541.
11. Randau L, *et al.* (2005) The heteromeric Nanoarchaeum equitans splicing endonuclease cleaves noncanonical bulge-helix-bulge motifs of joined tRNA halves. *Proc Natl Acad Sci USA* 102(50):17934–17939.
12. Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP (2005) Coevolution of tRNA intron motifs and tRNA endonuclease architecture in Archaea. *Proc Natl Acad Sci USA* 102(43):15418–15422.
13. 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(2):e1622.
14. Makarova KS, Koonin EV (2005) Evolutionary and functional genomics of the Archaea. *Curr Opin Microbiol* 8(5):586–594.
15. Hallam SJ, *et al.* (2006) Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. *Proc Natl Acad Sci USA* 103(48):18296–18301.
16. Soma A, *et al.* (2007) Permuted tRNA genes expressed via a circular RNA intermediate in Cyanidioschyzon merolae. *Science* 318(5849):450–453.
17. Itoh T, Suzuki K, Sanchez PC, Nakase T (1999) Caldivirga maquilingensis gen. nov., sp. nov., a new genus of rod-shaped crenarchaeote isolated from a hot spring in the Philippines. *Int J Syst Bacteriol* 49(Pt 3):1157–1163.
18. Sugahara J, *et al.* (2008) Comprehensive analysis of archaeal tRNA genes reveals rapid increase of tRNA introns in the order Thermoproteales. *Mol Biol Evol* 25(12):2709–2716.
19. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25(5):955–964.
20. Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32(1):11–16.
21. Schneider KL, Pollard KS, Baertsch R, Pohl A, Lowe TM (2006) The UCSC Archaeal Genome Browser. *Nucleic Acids Res* 34(Database issue):D407–D410.
22. Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP (2005) Structure, function, and evolution of the tRNA endonucleases of Archaea: an example of subfunctionalization. *Proc Natl Acad Sci USA* 102(25):8933–8938.
23. Torarinsson E, Klenk HP, Garrett RA (2005) Divergent transcriptional and translational signals in Archaea. *Environ Microbiol* 7(1):47–54.
24. Randau L, Schroder I, Soll D (2008) Life without RNase P. *Nature* 453(7191):120–123.
25. Waters E, *et al.* (2003) The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci USA* 100(22):12984–12988.
26. Di Giulio M (2008) Transfer RNA genes in pieces are an ancestral character. *EMBO Rep* 9(9):820; author reply 820–821.
27. Randau L, Soll D (2008) Transfer RNA genes in pieces. *EMBO Rep* 9(7):623–628.
28. 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(3):343–352 .
29. 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(15):6729–6734.
30. Marquet R, Isel C, Ehresmann C, Ehresmann B (1995) tRNAs as primer of reverse transcriptases. *Biochimie* 77(1–2):113–124.
31. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188–1190.