tRNA pieces in archaeal genomes, what do they tell us?

Kosuke Fujishima
Institute for Advanced Biosciences, Keio University

It is hypothesized that the evolutionary flow of life can be divided into two stages: a prebiotic world based on nucleic acids (presumably RNA) as functional polymers and the subsequent protein-based world. According to the Central Dogma of molecular biology, genetic information in modern living systems flows from DNA to RNA and next to proteins. In this scheme, gene is a basic module encoding information for non-coding RNA or proteins. To explain the origin and early evolution of genes, Walter Gilbert proposed the Exon Theory of Genes (also known as Exon Shuffling Theory), which states that the first genes were made of small pieces of RNA, from which evolved longer, more complex genes through shuffling and fusing small exons (*1*). However, this argument was based on the orientation of exon and introns present in the protein-encoding genes in eukaryotic genomes and therefore limited to explain the gene evolution in the earlier stages of life.

Gilbert's argument was primarily based on the analysis of eukaryotic genomes. Recently, the Exon Theory of Genes has been related to microbial organisms through a discovery that a hyperthermophilic archaeal parasite *Nanoarchaeum equitans* has *trans*-spliced tRNA, also called split tRNA, in which the 5' and 3' tRNA halves are encoded by two separate minigenes (*2*). To create a mature tRNA, two short RNAs transcribed from these genes are assembled via long complimentary leader sequences, followed by an enzymatic cleavage and ligation reaction that takes place at a specific RNA structure known as Bulge-Helix-Bulge (BHB) motif (Fig. 1).

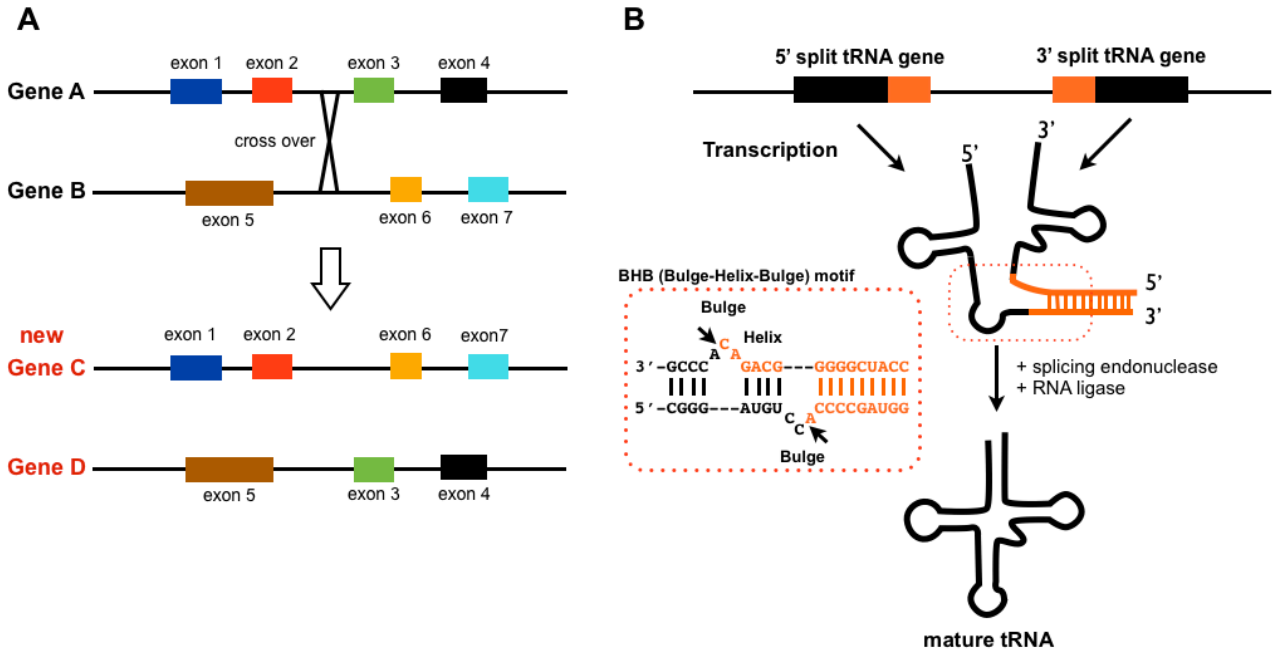

Figure 1. Illustration of the 'Exon Theory of Genes' and tRNA *trans*-splicing

(A) 'Exon Theory of Genes' is a hypothesis introduced by Walter Gilbert, according to which exon shuffling in the primitive species gave rise to a variety of genes that produced different proteins. For example, recombination between gene A and gene B would create new mosaic genes C and gene D that partially

possesses the previous exons. This process allows for creating various multi-domain proteins. (B) Two tRNA fragments are individually transcribed, but become connected due to complementarity in a long GC-rich leader sequence (orange). The exon-leader junction forms a Bulge-Helix-Bulge (BHB) motif, which is recognized and excised (arrow) by a tRNA splicing endonuclease and joined by RNA ligase.

Since tRNA is known as a universally conserved RNA molecule that decodes the genetic information for translation to proteins, the discovery of small genes that encode tRNA halves has brought forward an idea that it might be a ancestral trait of genes. This is in line with a hypothesis that ancient tRNA was once encoded as a simple hairpin RNA (minihelix) that served as an amino acid carrier in primordial translation or a donor of replication, and later became the top half of the present tRNA molecule (Fig. 2). Indeed, the top half sequence is highly conserved among different tRNA species in all three domains of life and recognized by various ribozymes and enzymes (*3*).

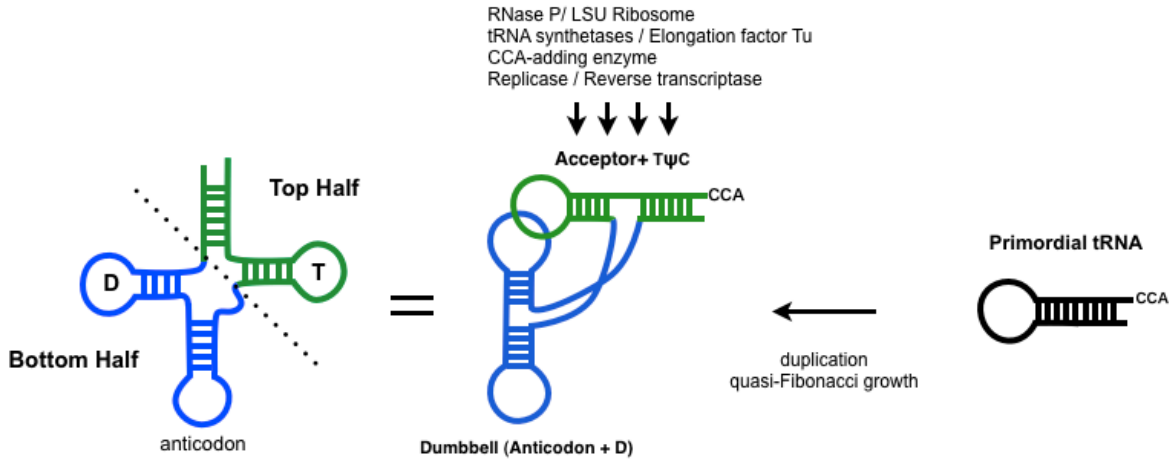

Figure 2. Representation of top/bottom half of tRNA in a cloverleaf and L-shaped structure. A schematic view of the two-domain structure of present tRNA, which emerged from a primordial hairpin RNA with a CCA sequence at the 3' end. Broadly, duplication model and quasi-Fibonacci growth model (gradual addition of structural components) have been proposed. The acceptor + pseudouridine (TψC) arm (green) comprise an independent structural domain apart from the anticodon + dihydrouridine (D) arm dumbbell (blue) in the L-shaped tertiary structure and serves as a platform for various RNA/protein enzymes.

In contrast, the bottom half (including the D-arm), which has more diverse length and sequence, was presumably added later in evolution to cope with the increasing amino acid repertoire.

Since *N. equitans* is a parasite and possesses a highly reduced genome, the origin of split tRNA gene can be also explained through genome reduction. This explanation, however, has been challenged in 2008 by our discovery of split tRNA genes in a free-living hyperthermoacidophilic archaeon, *Caldivirga maquilingensis,* isolated from acidic mud spring. This organism shows no sign

of genome reduction. Surprisingly, the tRNAs in *C. maquilingensis* are split into a maximum of three pieces, and some RNA fragments are even used in an alternative fashion to originate mature tRNAs with different anticodons (*4*) (Fig. 3). Positions of the splicing junctions varied among split tRNA genes and hardly overlap with that of *N. equitans*, which is inconsistent with the idea that these genes represent the ancient form of tRNA.

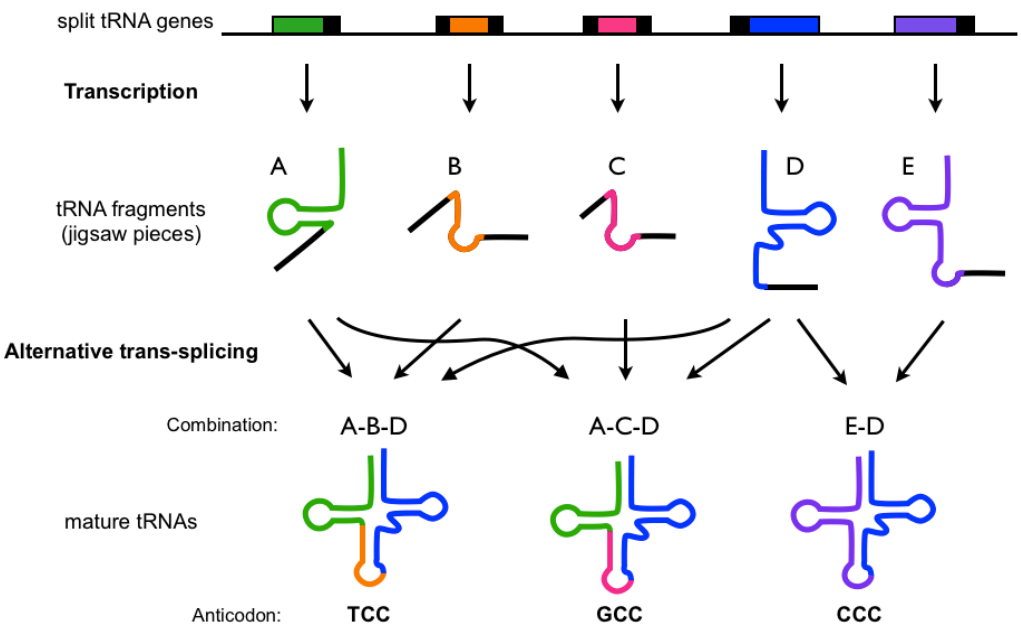

Figure 3. Complexed alternative *trans*-splicing of split tRNAs in *C. maquilingensis*.
Split tRNA genes are encoded at different loci of *C. maquilingensis* genome. tRNA fragments (A-E) are individually transcribed and assembled via *trans*-splicing based on the leader sequences (black) located at one or both ends. Sequences A and D are used multiple times in a different combination, as in a jigsaw puzzle.

In addition, a number of split tRNA genes have been recently found in the Desulfurococcales branch of archaea, expanding the population of split genes to diverse archaeal species (*5*). Examination of their gene arrangement combined with phylogenetic analysis has indicated that split tRNAs was a late acquisition, most likely created through local genome rearrangement. This means that split tRNAs in the archaeal genome might not be direct homologs but rather analogs of ancestral tRNAs (*6*).

Phylogenetic analysis of splicing endonuclease and RNA ligase suggests that the last common ancestor of archaea and eukaryotes possessed both enzymes, thus supporting the early origin of RNA *trans*-splicing mechanism. Indeed, BHB motif is known as a major landmark of RNA splicing for tRNA/mRNA intron removal and pre-rRNA processing in archaea. In line with the RNA-based Exon Theory of Genes, it would be expected that similar, BHB-based RNA *trans*-splicing was also used to produce ancestral, functional RNA molecules, but so far no such splicing was found outside of split tRNA. Why this is so remains unclear, as our computational prediction of BHB motif in all possible combinations of *C. maquilingensis* transcripts (~two million combinations) estimate that at least 1% of such RNA pairs should form a BHB motif with similar free energy to that in split tRNAs. It should be noted, however, that free energy is not the only relevant criterion, as joined RNA products must

undergo retroposition (reverse transcription and integration into the genome) to become stabilized in the genome as a novel gene. Hence, evolutionary studies of genome, transcriptome complexity and RNA-related enzymes in the deep-branching organisms will become increasingly important to evaluate further the RNA-based Exon Theory of Genes in the early stages of life.

References

1. W. Gilbert, S. J. de Souza, M. Long, *Proc Natl Acad Sci U S A* **94**, 7698 (Jul 22, 1997).
2. L. Randau, R. Munch, M. J. Hohn, D. Jahn, D. Soll, *Nature* **433**, 537 (Feb 3, 2005).
3. N. Maizels, A. M. Weiner, *The RNA World, Second Edition* R.F. Gesteland and J.F. Atkins, eds. Cold Spring Harbor Press, 79 (1999).
4. K. Fujishima *et al*., *Proc Natl Acad Sci U S A* **106**, 2683 (Feb 24, 2009).
5. P. P. Chan, A. E. Cozen, T. M. Lowe, *Genome Biol* **12**, R38 (Apr 13, 2011).
6. A. S. Rodin, E. Szathmary, S. N. Rodin, *Biol Direct* **6**, 14 (2011).