# Reading Evolutionary History of Aminoacyl-tRNA Synthetases from Genome Sequences

**Kiyotaka Shiba** kshiba@jfcr.or.jp
**Hiromi Motegi** hmotegi@jfcr.or.jp
**Tetsuo Noda** tnoda@jfcr.or.jp
Cancer Institute
Japanese Foundation for Cancer Research
Kami-Ikebukuro, Toshima, Tokyo 170, Japan

**Abstract**

*Aminoacyl-tRNA synthetases (ARSs) are believed to have arisen early in the evolution of life as the essential components that establish the link between triplet codons and amino acids. We have cloned and sequenced eight cDNAs for human cytoplasmic ARSs. Along with twelve sequences that have been reported from other laboratories, a set of 20 human cytoplasmic ARS genes is now available. We compared these human ARSs with ∼400 sequences of ARS currently available from various organisms and deduced the possible evolutionary history of these enzymes. The availability of complete sets of ARSs from thirteen organisms (*H. sapiens, S. cerevisiae, E. coli, H. influenzae, H. pylori, N. gonorrhoeae, S. pyogenes, M genitalium, M. pneumoniae, Synechocystis sp., M jannaschii, M. thermoautotrophicum*, and* A. fulgidus*) made systematic analyses of the evolution of this gene family possible. In this paper, we will focus on two topics; (1) the acquisition of new structural domains to the core enzyme domains in higher eukaryotes and their possible role in the formation of multi-synthetase supra-molecular complexes, and (2) the existence of eukaryotic-like ARSs in some bacterial genomes, and the relationship of this occurrence to tRNA recognition.*

## 1 Introduction

The aminoacyl-tRNA synthetases (ARSs) catalyze the specific attachment of amino acids to the 3' end of cognate tRNA substrates. This reaction is essential for the accurate translation of the genetic code, and thus, the family of enzymes is believed to be amongst the oldest proteins. Each organism has 18-20 ARSs for each amino acid, and these ARSs are divided into two classes of ten each based on shared structural motifs in their catalytic domains [1]. The goal of our researches is to reveal the evolutional paths of ARSs, and in his paper, we will introduce our current studies on this topic.

## 2 Results and Discussion

### 2.1 Cloning of cDNAs for human tRNA synthetases

Information about primary structures of ARSs has been rapidly accumulated in these years. It should be pointed out, however, that these genes have some bias toward prokaryotic and lower eukaryotic origins. To perform systematic analyses of the evolution of ARSs across taxa, we started the cloning of genes for human ARSs as a representative of higher eukaryotes. For this purpose, we established a general strategy, an alignment-guided cross-species PCR, by which synthetase genes from any organism can be isolated using available sequence data of prokaryote and lower eukaryote ARS sequences [19].

In this method, first, degenerated primers were designed from conserved regions that were identified from multiple sequence alignments of available sequences. And, these primers were used in RT-PCRs of human mRNA. Amplified fragments were used to isolate full length cDNAs from a cDNA library. From these efforts, we have isolated and determined the sequences for human cytoplasmic IleRS [18], AlaRS [15], GlyRS [16], LysRS [17], LeuRS (unpublished), MetRS (unpublished), AsnRS (unpublished) and PheRS (unpublished). Along with sequences that were published from other laboratories, now a complete set of 20 ARSs sequences for human cytoplasmic ARSs is available. This alignment-guided cross-species approach has been also applied to clone genes for ARSs from other organisms than human such as a *M. tuberculosis* [13].

### 2.2 Evolution of ARSs in higher eukaryotes

Higher eukaryotic aminoacyl-tRNA synthetases are distinguished from enzymes form other organisms in their abilities to form supracomplexes through self assembly or association with protein synthesis machinery [6,8]. The enzymes including IleRS, LeuRS, MetRS, AspRS, Glu-ProRS (fusion), GlnRS, LysRS and ArgRS have been consistently identified as the components of the multi-tRNA synthetases complex in higher eukaryotes [9]. When we compared primary structures of human ARSs with ones from lower eukaryotes, bacteria and archaebacteria, appendix domains are frequently found at N-terminal or C-terminal end of human ARSs. For instance, human cytoplasmic IleRS has a C-terminal domain of two repeats of an 90 aa sequence which is dispensable for aminoacylation activity [18]. Similarly, the N-terminal 65 aa extra sequence of human LysRS that is absent from prokaryotic LysRSs has been shown to be dispensable for *in vitro* and *in vivo* activity [17]. The acquisition of these additional domains in higher eukaryotic ARSs could be related to their abilities to form multi-synthetase complexes. To prove the possibilities, Rho *et al.* performed yeast two-hybrid experiment using the C-terminal appendix of human IleRS as a bait. By screening a cDNA library made from a human HeLa mRNA, the internal 317 aa of human Glu-ProRS bifunctional enzyme was selected [11]. The internal region of human Glu-ProRS has three repeats of 57 aa and this motif is also found in the N-terminal appendix of human GlyRS [16], HisRS [20], and TrpRS [2].

Thus, the acquisition of appendix domains and the formation of multi-synthetase complex characterize the evolution of ARSs in higher eukaryotes.

### 2.3 Rapidly evolving ARSs and slowly evolving ARSs

Comparisons of the primary structures of each of human ARSs with the corresponding enzymes from *S. cerevisiae, E. coli, H. influenzae* and *M. genitalium* have revealed various modes of structural diversification. For instance, LysRS, AlaRS and ThrRS showed very high sequence similarities among all 5 species. Enzymes for Cys, Asn, Asp and His also shared high similarities but they also have many species-specific insertions/deletions. Enzymes for Gly, Pro Trp, Leu, Arg and Tyr shared very low sequence identities among the 5 species examined. Some ARSs showed species barriers in tRNA recognition and others did not. For instance, human and *E. coli* GlyRSs showed species-specific tRNA

recognition [4, 16] whereas AlaRS didn't [12, 15]. It seems that there is some relationships between species barriers in tRNA recognition and enzyme evolution. This differences in ARS structures would be a good target for a development of species-specific drugs [3].

Interestingly, reported genome structures of *M jannaschii, M. thermoautotrophicum*, and *A. fulgidus* showed that these archaeal cells do not have the ORF that resembles to the known LysRSs (The CysRS-like ORF is also absent in *M. jannaschii* and *M. thermoautotrophicum*). We have carefully re-examined the genome sequences and have found two ORFs from each genome that have weak sequence similarities to ARSs [10]. One is the ORF:Mj1660 whose product shows weak similarities to *M. jannaschii* PheRS in limited regions, and the other is the ORF:MJ0539 whose product shows similarity to CysRSs from various organisms. In the course of the analyses of these two ORFs in our laboratory, Ibba *et al.* reported that the ORF:MJ0539 from *Methanococcus maripaludis* has LysRS activity [5]. The LysRSs from Eucarya, Bacteria and the Crenarchaeota subdomain of Archaea (or Eocytes) show a high degree of sequence similarity and belong to class II enzyme [17]. The newly discovered LysRS in some of Archaea is the first example of the class I-like LysRS.

### 2.4 Eukaryotic type ARSs found in some bacteria

In some of 20 synthetases, sequences have diverged so that they are characteristic of the organisms in which a given synthetase is located. Alignments of the available sequences for a given synthetase showed that sorting of these sequences into domain-specific groups (Bacteria, Eucarya, and Archaea) [14]. The groupings of sequences into discrete domains were done for AspRS, GluRS, LeuRS, TrpRS and PheRS from different organisms. The result agreed with the expected arrangement into three domains. In contrast, we found that GlyRS, IleRS and ProRS crossover from one domain to another.

IleRS of the most of bacterial and eukaryotic mitochondrial enzymes is characterized by the presence near the C-terminus of a $Cys_4$ cluster that is coordinated to zinc [7]. In contrast, this motif is absent from the cytoplasmic eukaryotic enzymes and from the archaebacterial enzymes [14, 18]. But two bacterial enzymes (*T. thermophilus* and *M. tuberculosis*) lack the C-terminal zinc containing motif and in this respect, and in alignments of the whole sequences, these two bacterial enzymes are eukaryote-like (Figure 1) [13, 14].

GlyRS showed a similar cross-over of from one domain to another. The enzymes essentially have no sequence similarities between human and *E. coli* types [16]. The eukaryotic and archaebacterial enzymes are of the $\alpha_2$ type (human type), while most of the bacterial enzymes have the $\alpha_2\beta_2$ quaternary structure (*E. coli* type). In contrast, five bacterial enzymes have the eukaryotic form of GlyRS (Figure 1) [14].

Similarly, ProRS showed the crossover, where a large insertion among a catalytic domain is characteristic of the enzymes from eukaryotes and archaebacteria . This and other differences divide the ProRSs into two groups and several of the bacterial ProRS are eukaryote-like (Figure 1) [14].

Analyses of MetRS turned to be equally striking in a different way. The sequence comparison of MetRS show clearly the grouping of eukaryotic and archaebacterial enzymes as distinct from those

in bacteria. Four examples of eukaryote-like methionyl-tRNA synthetases in Bacteria, however, were identified, including three instances in Proteobacteria (*N. gonorrhoeae, H. influenzae* and *E. coli*) and one in Spirochaetes (*B. burgdorferi*) (Figure 1) [14].

It is worth for pointing out that, even a synthetase crosses from one domain to another domain, their sequences for substrate tRNAs are not the case for crossover. The synthetase must adapt to a tRNA sequence not found in the domain with which the synthetase is normally associated [14].

The exact mechanism of transfer of a synthetase from one taxonomic domain to another could be explained after much information on ARSs from many organisms are available. At a moment, horizontal gene transfer could be a possible candidate to explain it, but further studies on genomes would be needed to conclude.

## References


[1] Eriani, G., Delarue, M., Poch, O., Gangloff, J. and Moras, D.,"Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs," *Nature*, 347:203–206, 1990.

[2] Fleckner, J., Rasmussen, H. H. and Justesen, J., "Human interferon $\gamma$ potently induces the synthesis of a 55-kDa protein ($\gamma$ 2) highly homologous to rabbit peptide chain release factor and bovine tryptophanyl tRNA synthetase," *Proc. Natl. Acad. Sci. USA*, 88:11520–11524, 1991.

[3] Heacock, D., Forsyth, C. J., Shiba, K. and Musier-Forsyth, K.,"Synthesis and aminoacyl-tRNA synthetase inhibitory activity of prolyl adenylate analogs," *Bioorganic Chemistry*, 24:273–289, 1996.

[4] Hipps, D., Shiba, K., Henderson, B. and Schimmel, P., "Operational RNA code for amino acids: specie-specific aminoacylation of minihelices switched by a single nucleotide," *Proc. Natl. Acad. Sci. USA*, 92:5550–5552, 1995.

[5] Ibba, M., Morgan, S., Curnow, A. W., Pridmoret, D. R., Vothknecht U. C., Gardner W., Lin, W., Woese, C. R. and Söll, D., "A euryarchaeal lysyl-tRNA synthetase: resemblance to class I synthetases," *Science*, in press, 1997.

[6] Kisselev, L. L. and Wolfson, A. D., "Aminoacyl-tRNA synthetases from higher eukaryotes," *Prog. Nucleic Acid Res. Mol. Biol.*, 48:83–142, 1994.

[7] Landro, J. A. and Schimmel, P., "Zinc-dependent cell growth conferred by mutant tRNA synthetase," *J. Biol. Chem.*, 269:20217–20220, 1994.

[8] Mirande, M., "Aminoacyl-tRNA synthetase family from prokaryotes and eukaryotes: structural domains and their implications," *Prog. Nucleic Acid Res. Mol. Biol.*, 40:95–142, 1994.

[9] Mirande, M., Le Corre, D. and Waller, J.-P., "A complex from cultured chinese hamster ovary cells containing nine aminoacyl-tRNA synthetases: thermolabile leucyl-tRNA synthetase from the

tsH1 mutant cell line is an integral component if this complex," *Eur. J. Biochem.*, 147:281–289, 1985.

[10] Mirande, M., Le Corre, D. and Waller, J.-P., "A complex from cultured chinese hamster ovary cells containing nine aminoacyl-tRNA synthetases: thermolabile leucyl-tRNA synthetase from the tsH1 mutant cell line is an integral component if this complex," *Eur. J. Biochem.*, 147:281–289, 1985.

[11] Motegi, H., Noda, T. and Shiba, K., "Does the ORF-Mj1660 in *M. jannaschii* code for a new aminoacyl-tRNA synthetase?," *17th International tRNA Workshop (Kazusa)*, abstract:pp7–6, 1997.

[12] Rho, S. B., Lee, K. H., Kim, J. W., Shiba, K., Jo,Y. J. and Kim. S., "Interaction between human tRNA synthetases involves repeated sequence elements," *Proc. Natl. Acad. Sci. USA*, 93:10128–10133, 1996.

[13] Tracy L. Ripmaster, T. L., Shiba, K. and Schimmel, P.,"Wide cross-species aminoacyl-tRNA synthetase replacement *in vivo*: Yeast cytoplasmic alanine enzyme replaced by human polymyositis serum antigen," *Proc. Natl. Acad. Sci. USA*, 92:4932–4936, 1995.

[14] Sassanfar, M., Kranz, J. E., Gallant, P., Schimmel, P. and Shiba., K., "A eubacterial *Mycobacterium tuberculosis* tRNA synthetase is eukaryote-like resistant to a eubacterial-specifc antisynthetase drug," *Biochemistry*, 35:9995–10003, 1996.

[15] Shiba, K., Motegi, H. and Schimmel, P., "Maintaining genetic code through adaptations of tRNA synthetases to taxonomic domains," *Trends Biochem. Sci.*, 22(12): in press, 1997.

[16] Shiba, K., Ripmaster, T., Suzuki, N., Nichols, R., Plotz, P., Noda, T. and Schimmel, P., "Human alanyl-tRNA synthetase: conservation in evolution of catalytic core and microhelix recognition," *Biochemistry*, 34:10340–10349, 1995.

[17] Shiba, K., Schimmel, P., Motegi, H. and Noda. T., "Human glycyl-tRNA synthetase: wide divergence of primary structure from bacterial counterpart and species-specific aminoacylation," *J. Biol. Chem.*, 269:30049–30055, 1994.

[18] Shiba, K., Stello, T., Motegi, H., Noda, T., Musier-Forsyth, K. and Schimmel. P., "Human lysyl-tRNA synthetase accepts nucleotide 73 variants and rescues *Escherichia coli* double-defective mutan," *J. Biol. Chem.*, 272:22809–22816, 1997.

[19] Shiba, K., Suzuki, N., Shigesada, K., Namba, Y., Schimmel, P. and Noda, T., "Human cytoplasmic isoleucyl-tRNA synthetase: selective divergence of the anticodon-binding domain and acquisition of a new structural unit," *Proc. Natl. Acad. Sci. USA*, 91:7435–7439, 1994.

[20] Shiba, K., Suzuki,N., Shigesada, K., Schimme, P. and Noda, T., "Isolation of higher eukaryote aminoacyl-tRNA synthetase genes by an alignment-guided cross-species PCR: application to human isoleucine tRNA synthetase," *15th International tRNA Workshop (Cap d'Agde)*, abstract:pp364, 1993.

[21] Tsui, F. W. and Siminovitch, L., "Isolation, structure and expression of mammalian genes for histidyl-tRNA synthetase," *Nucleic Acids Res.*, 15:3349–3367, 1987.

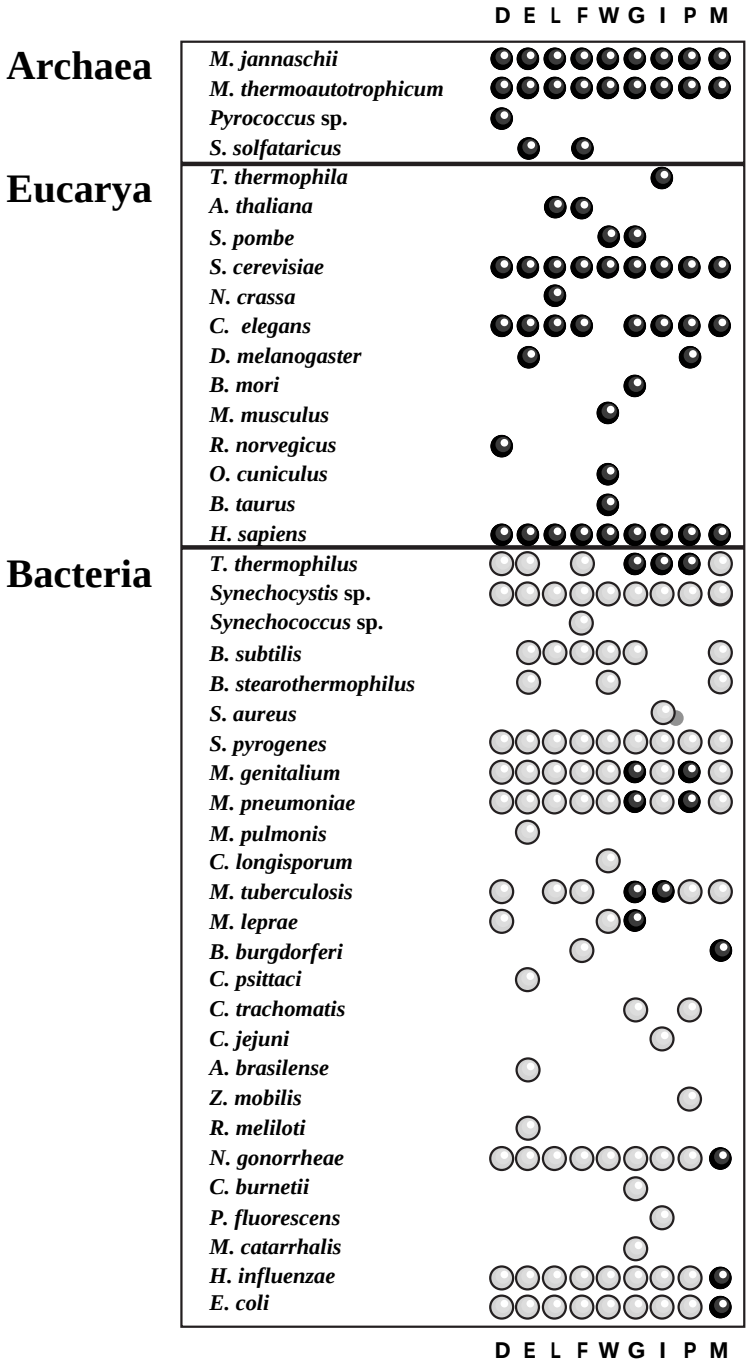


Figure 1: Organisms and types of nine different aminoacyl-tRNA synthetases (D, aspartyl-; E, glutamyl-; L, leucyl-; F, phenylalanyl-; W, tryptophanyl-; G, glycyl-; I, isoleucyl-;P, prolyl-;M, methionyl-tRNA synthetases) can be assigned. Black circles represent Archaea- or Eucarya-like synthetases and white circles represent Bacteria-like synthetases. A clinical strain of Staphylococcus aureus has an extra-chromosomally encoded Archaea- or Eucarya-like isoleucyl-tRNA synthetase, which is shown as a small circle [14].