Aminoacyl-tRNA Synthetases and the Evolution of the Genetic Code

Matt Gordon

October 31, 2001

Much recent work on the origins of the genetic code have centered on the role played by the class of enzymes known as aminoacyl-tRNA synthetases. Because of their central role in the genetic code, understanding their place in evolution is believed to be key in understanding genetic evolution itself.

Aminoacyl-tRNA synthetases catalyze the attachment of amino acids to their corresponding transfer RNAs. In the context of translation, transfer RNAs act as the mediator between amino acids and nucleic acids. Each transfer RNA consists of 76 nucleotides that fold into an L-shaped structure. One leg of the L contains the amino acid attachment site, and the other end contains the anticodon, the three-nucleotide sequence that pairs with the codon in the messenger RNA for the coded amino acid. Any given organism has at least one synthetase for each amino acid, and the synthetase for an amino acid recognizes its corresponding tRNA and "charges" that tRNA with the correct amino acid, which then gets incorporated into the polypeptide by the ribosome. Hence, the tRNA synthetases give the amino acid-codon specificity to the genetic code, and are a good place to study the origins of the genetic code itself.

One current area of study on aminoacyl-tRNA synthetases is in classifying the various synthetases within an organism by structure[1]. The most striking fact about the structural and sequence data on the synthetases is that they divide very clearly into two groups, known as class I and class II, with ten synthetases in each class. The classes differ significantly in several ways. Class I synthetases have an active site domain composed of alternating β strands and α helices, whereas the class II synthetase active sites have a seven stranded β sheet with flanking α helices. These differences can be seen fairly easily by doing a protein sequence alignment of the various tRNA synthetases in a given organism[2, 3]. An example of such an alignment is given in Figure 1. Note that the synthetases for valine, leucine, isoleucine, and methionine, all of which are class I synthetases, reside on a separate section of the tree from the synthetases for serine, threonine, proline, and histidine, all of which are class II synthetases.

The obvious evolutionary implications of this subclass grouping is the existence of two separate ancestor aminoacyl-tRNA synthetases, from which the current synthetases all evolved. But this still does not tell us much about the origin of the genetic code itself.

Another intriguing aspect of the synthetase classes is the existence of complementary synthetase pairs across classes. Class I and class II synthetases bind to opposite sides of the tRNA acceptor stem. Information from synthetase crystal structures has shown that synthetases from class I and class II can be grouped into pairs that are not sterically forbidden from binding simultaneously to a tRNA acceptor stem. In other words, for each class I synthetase bound to an acceptor stem, there is a class II synthetase that can bind to the same acceptor stem simultaneously without steric hindrance. Furthermore, there is a high degree of correlation between assignments of similar codons and steric pairings. These pairings, along with the class separations noted above, suggest an earlier time when the genetic code was less varied, and less accurate. Ancient class I and class II synthetases may have started as amino-acylated ribozymes (RNA enzymes) that facilitated amino-acylation of the precursor tRNAs with whatever amino acids happened to be in a locally high concentration. Selective pressures then may have driven them to separate and diversify their aminoacylation activities, in order to produce greater chemical diversity in their products, while at the same time making them more specific to particular substrate amino acids. This sort of evolutionary picture would help account for the class separation of the tRNA synthetases, as well as the steric compatibility of the two classes.

Another aspect of aminoacyl-tRNA synthetases that has been the subject of research has been their phylogenetic evolution. Woese, *et al*[4] performed an extensive phylogenetic analysis of all twenty tRNA synthetases, using phylogenetic trees derived from sequence analysis, as well as other tools such as amino acid similarity matrices. They used sequence data derived from 72 organisms from *Bacteria*, *Archae*, and eukaryotes. Typical examples are reproduced in figures 2 and 3. The phylogeny of the leucine tRNA synthetase follows what Woese, *et al* refer to as the "canonical evolutionary pattern," which essentially says that examples of a protein from Bacteria and those from Archaea show notable, qualitative differences, as well as quantitative differences, that are conserved across organisms in the same category. Similarly, the same protein from a eukaryote will show qualitative and quantitative differences with those from Archaea and Bacteria, but will show more similarity to those from Archaea than from Bacteria. As seen in figure 2, this structure is clearly present in the phylogeny of LeuRS.

The phylogeny of the methionine tRNA synthetase, however, shows a marked departure from this structure. While there is clearly a Bacterial branch of the tree, there is quite a bit crossover between the Bacterial and Archaeal branches, and the eukaryote branch appears to be more closely related to a local extension of the Bacterial branch than to the rest of the Archaeal branch. Many other synthetases show non-canonical phylogenies of this sort. These sorts of phylogenies strongly imply extensive and promiscuous lateral gene transfer in the evolution of the synthetases, which has very profound implications for the development of the universality of the genetic code. Woese, *et al* argue that the tRNA synthetases are natural candidates for widespread horizontal gene transfer, because they are "modular" cell components, interacting only with free amino acids and tRNAs, and because of the universality of the genetic code. Indeed, the existence of multiple synthetases for a given amino acid in some organisms is yet more evidence for the horizontal transfer of the corresponding genes.

As the generators of the universal genetic code, the aminoacyl-tRNA synthetases were clearly essential to the evolution of the genetic code itself, and further study in both of these areas, structural as well as sequential, has potential to illuminate the origins of the genetic code and shed light on the origin of life itself.

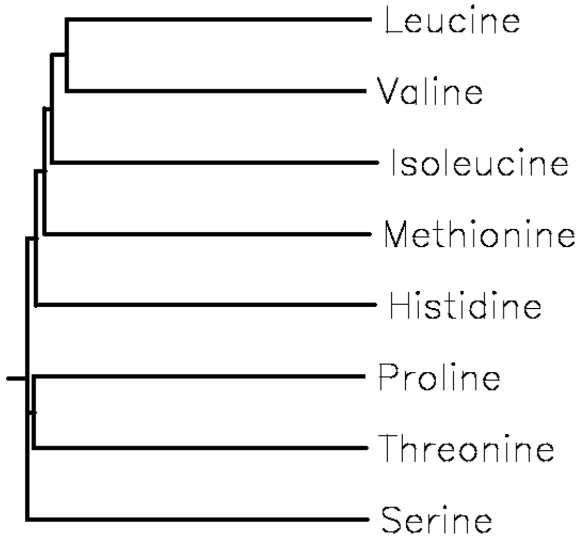

Figure 1: Tree based on the alignment of several *E. Coli* tRNA synthetases.

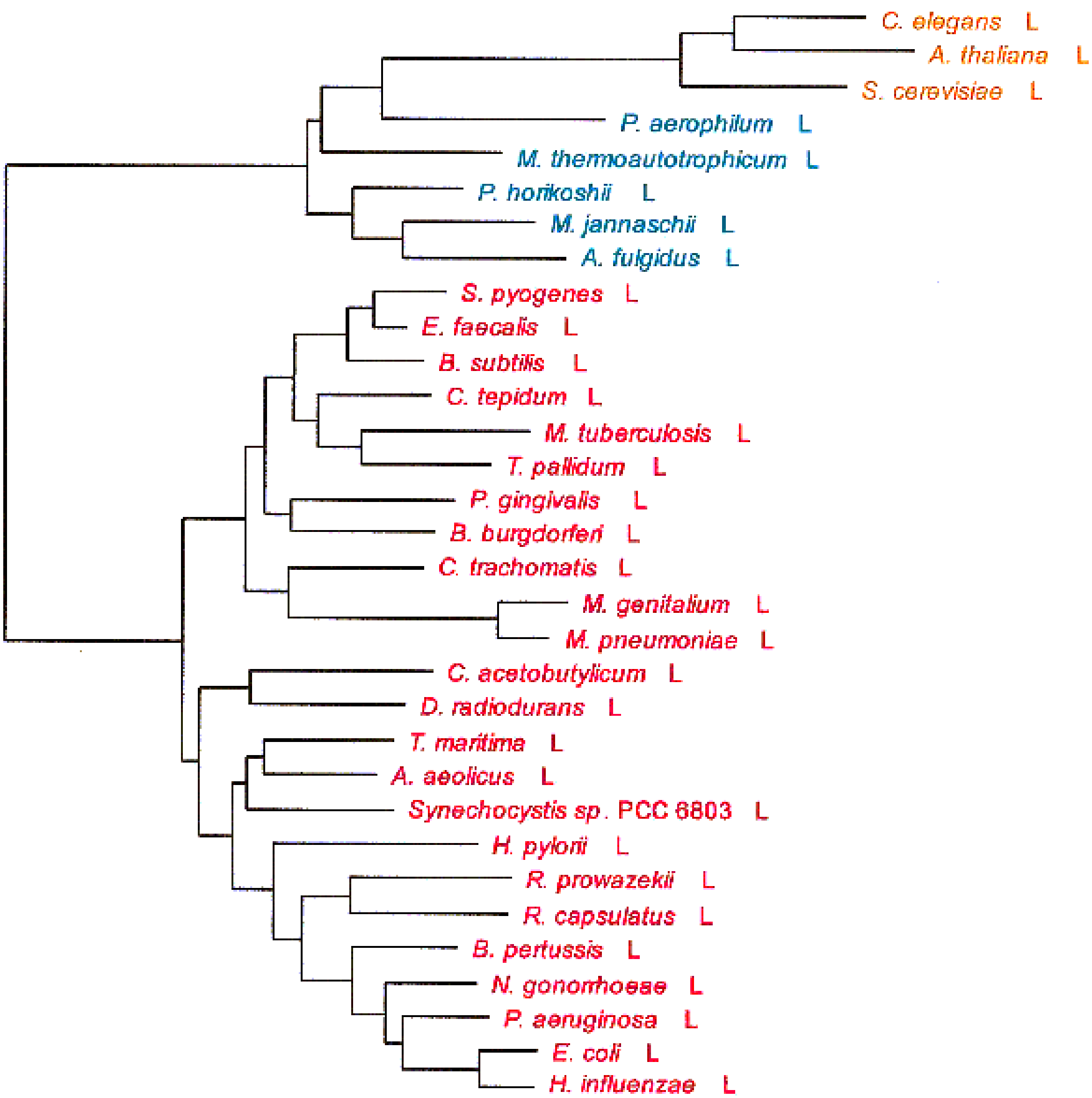

Figure 2: Phylogenetic tree for leucyl tRNA synthetase. Bacteria appear in red, Archaea in blue, and eukaryotes in orange.

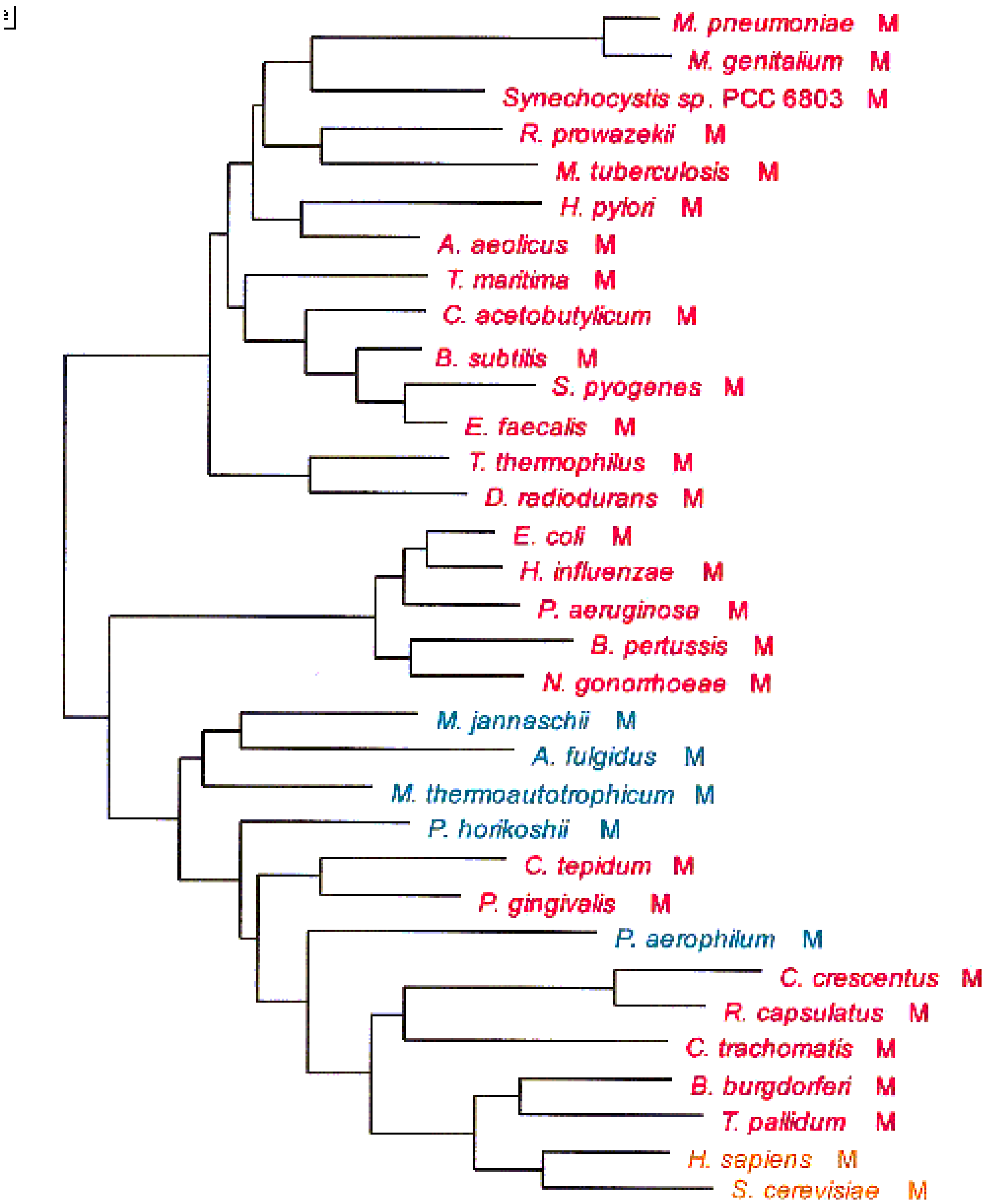

Figure 3: Phylogenetic tree for methioninyl tRNA synthetase.

Bibliography

[1] Lluís Ribas de Pouplana and Paul Schimmel. Aminoacyl-tRNA synthetases: potential markers of genetic code development. *TRENDS in Biochemical Sciences*, 26(10):591 – 596, Oct 2001.

[2] Julie D. Thompson, Desmond G. Higgins, and Toby J. Gibson. CLUSTAL W. Computer program.

[3] J. Felsenstein. PHYLIP (Phylogeny Inference Package) version 3.5c, 1993.

[4] Carl R. Woese, Gary J. Olsen, Michael Ibba, and Dieter Söll. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiology and Molecular Biology Review*, 64(1):202–236, Mar 2000.