Evolution of Genes, Evolution of Species: The Case of Aminoacyl-tRNA Synthetases

Y. Diaz-Lazcoz,* J.-C. Aude, †P. Nitschke,* H. Chiapello,‡ C. Landès-Devauchelle,* and J.-L. Risler*

*Université de Versailles, Génome et Informatique, Bâtiment Buffon, Versailles, France; †Institut National de Recherche en Informatique et en Automatique, Le Chesnay, France; and ‡Laboratoire de Biologie Celulaire, Institut National de la Recherche Agronomique, Versailles, France

All of the aminoacyl-tRNA synthetase (aaRS) sequences currently available in the data banks have been subjected to a systematic analysis aimed at finding gene duplications, genetic recombinations, and horizontal transfers. Evidence is provided for the occurrence (or probable occurrence) of such phenomena within this class of enzymes. In particular, it is suggested that the monomeric PheRS from the yeast mitochondrion is a chimera of the α and β chains of the standard tetrameric protein. In addition, it is proposed that the dimeric and tetrameric forms of GlyRS are the result of a double and independent acquisition of the same specificity within two different subclasses of aaRS. The phylogenetic reconstructions of the evolutionary histories of the genes encoding aaRS are shown to be extremely diverse. While large segments of the population are consistent with the broad grouping into the three Woesean domains, some phylogenetic reconstructions do not place the Archaea and the Eucarya as sister groups but, rather, show a Gram-negative bacteria/eukaryote clustering. In addition, many individual genes pose difficulties that preclude any simple evolutionary scheme. Thus, aaRS’s are clearly a paradigm of E. Jacob’s “odd jobs of evolution” but, on the whole, do not call into question the evolutionary scenario originally proposed by Woese and subsequently refined by others.

Introduction

Molecular phylogeny was born about 30 years ago (e.g., Fitch and Margoliash 1967) and, by necessity, was limited at that time to the study of chemically determined protein sequences. Since DNA sequences can now be easily and massively determined, phylogenetic studies based on molecular data (gene sequences or conceptually translated amino acid sequences, ribosomal RNA sequences) are routinely performed in many laboratories. Some analyses are even based on entire genomic sequences (Sankoff, Ferretti, and Nadeau 1997).

The most important achievement in the field, which Doolittle and Brown (1994) called “the Woesian revolution,” was probably the delineation of the three kingdoms of life (Eukaryotes, Eubacteria, and Archaeabacteria) based on the analysis of ribosomal RNA sequences (Fox et al. 1977; Woese and Fox 1977). Later, Iwabe et al. (1989) and Gogarten et al. (1989) suggested that archaeabacterial and eukaryotic nuclear genomes are sister groups, indicating that Eubacteria were the first to diverge from the universal tree. This rooting was accepted by Woese, Kandler, and Wheelis (1990), who renamed the three kingdoms as three domains (Eucarya, Bacteria, and Archaea). Although generally accepted, the scenario proposed by Woese, Iwabe, Gogarten and their co-workers has been (and is still) the subject of fierce debates (see a discussion by Doolittle and Brown 1994). A number of phylogenetic analyses, essentially based on protein sequences, have produced contradictory and apparently irreconcilable results. These studies have been summarized and criticized by Forterre et al. (1992), Doolittle and Brown (1994), Roger and Brown (1996), Tourasse (1998), and others.

Obviously, the reconstruction of the universal tree by molecular phylogenetic inference must be based on the sequences of ubiquitous and ancient macromolecules. Thus, aminoacyl-tRNA synthetases (aaRS’s) are candidates of choice, since they fulfill the above two conditions, and their key role in the fidelity of translation makes it probable that they acquired their evolutionary maturity very early (however, see Ibba et al. 1997a, 1997b). Not surprisingly, they have already been subjected to phylogenetic studies (Brown and Doolittle 1995; Nagel and Doolittle 1995; Brown et al. 1997).

What prompted us to embark on the present—and apparently redundant—evolutionary study of aaRS is the fact that a number of sequences from a great variety of organisms have recently appeared in the data banks, making it possible to perform sound analyses on each of the 20 aaRS’s. Our aim was twofold: (1) to describe the evolution of the genes of each aaRS and (2) to examine whether the evolution of the genes could be used to infer that of the species. In other words, we wanted to determine whether the reconstructed trees for the 20 proteins gave a coherent picture of evolution, and check for possible gene duplications and horizontal transfers that would complicate the task of reconstructing a universal tree with these proteins.

Hereinafter, we shall refer to all the aminoacyl-tRNA synthetases that share the same specificity for the same amino acid as a group of aaRS’s. Thus, for example, all the arginyl-tRNA synthetases belong to the same group. This is not to be confused with one of the two unrelated classes that partition the aaRS’s into the two equally populated class I and class II, each containing 10 groups (e.g., Carter 1993).
Results and Discussion

It has been known for a long time that the sequences of aaRS can be extremely different, even for proteins within the same class. For example, while the sequences of IleRS, LeuRS, ValRS, and MetRS—all belonging to class I—are rather similar (see review by Carter 1993), it proves to be much more difficult to align the sequences of MetRS and TyrRS, even though TyrRS is also a class I synthetase. Under such circumstances, the safest way to align poorly related sequences is to take into account the structures of the proteins. Indeed, the three-dimensional structures of at least 14 of the 20 synthetases have been determined (Cusack 1995, 1997). Unfortunately, most of the atomic coordinates have not been deposited with the Protein Data Bank. This is the main reason we did not attempt to perform intergroup multiple alignments, except for TrpRS and TyrRS, for
which a structural alignment has been published (Doublié et al. 1995). Thus, the phylogenetic analyses reported below are based on multiple alignments of aaRS’s belonging to the same group—having the same specificity—for which the sequences are generally better conserved. The pyramidal analyses, for their part, rely on pairwise intra- and intergroup alignments. As stated above, this can be criticized. It should be noted, however, that the Smith–Waterman algorithm is a local alignment procedure and that the threshold (Z≥14) used in the single linkage analysis is rather high. In other words, the clustering of the aaRS sequences is based on their most similar regions. While this analysis is certainly less sensitive than the analysis of global multiple alignments, the pyramidal classifications prove to be in excellent agreement with the phylogenetic reconstructions (vide infra).

Clusters and Pyramidal Classifications

In the present work, clusters of sequences were built according to their similarities based on the pairwise Z values (see Materials and Methods). It is clear that other similarity indices can be used for the same goal (e.g., Tatusov, Koonin, and Lipman 1997). Whatever the method used for grouping the sequences, it is expected that any given cluster will mainly contain orthologous proteins of similar functions, as well as paralogous sequences in the case of gene duplications. In some cases, a cluster will also contain proteins that have been acquired by horizontal transfers and mosaic or multidomain sequences that are not necessarily functionally related. Table 1 summarizes the results obtained with the 465 aaRS sequences currently available (as of July 1998) and shows that the sequences partition into 10 different clusters. All the clusters thus delineated are totally consistent with our present knowledge on aaRS. For example, the class I IleRS, LeuRS, MetRS, and ValRS have long been recognized as being closely related, as well as the class I Ila GlyRS, HisRS, ProRS, SerRS, and ThrRS (see review by Carter 1993; Cusack 1993). Conversely, it is only recently that the archaeabacterial LysRS’s from Methanococcus maripaludis, Methanobacterium thermoautotrophicum, and Methanococcus jannaschii (but not that from Sulfolobus solfataricus) and the LysRS from the spirochete Borrelia burgdorferi have been shown to be radically different from all the other LysRS’s (Ibba et al. 1997a, 1997b). These LysRS’s are grouped in cluster 10 (table 1), and their case will be discussed later.

The pyramidal classification of the phenylalanyl-tRNA synthetases is shown in figure 1a and clearly shows two groups built by the α and β chains, respectively. PheRS is normally a heterotetramer with an α2β2 quaternary structure. In any given PheRS, the short (α) and long (β) chains present very few sequence similarities. However, the X-ray structure of the tetrameric PheRS from Thermus thermophilus (Mosyak and Safro 1993) revealed an unexpected structural similarity between its α chain and the N-terminal part of its β chain. This resemblance, together with a comparison of the structure with those of other class II synthetases, enabled Mosyak and Safro (1993) to locate in both chains the three motifs characteristic of class II synthetases (Eriani et al. 1990) as depicted in figure 1b, and led to the conclusion that the large β subunit probably arose from the duplication of an ancestral catalytic domain of class II synthetases followed by subsequent insertions and deletions of polypeptides. However, as already stated, there is practically no sequence similarity between the α and β chains in any PheRS. Hence, the grouping of sequences from both chains into the same cluster came as a surprise. Examination of the cluster shows that it is the mitochondrial PheRS from yeast that makes the link between the α chains and the β chains of the other PheRS. The point here is that the yeast mitochondrial PheRS is monomeric (Sanni et al. 1991) and is so far the only PheRS of its kind. Sanni et al. (1991) noticed strong sequence similarities between the mitochondrial PheRS and the Escherichia coli α subunit, as well as 30% identities between its C-terminal part and that from the E. coli β chain. These observations are now reinforced by comparisons with more recently sequenced PheRS’s. A search against SwissProt with WU-BLAST (http://blast.wustl.edu) confirms the occurrence of two segments of high similarity between the yeast mitochondrial PheRS and the α subunit from Haemophilus influenzae, as well as an undisputable similarity between its C-terminus and that of the β chain from Synechococcus sp. (fig. 1c). It is therefore clear that the unique chain of the yeast mitochondrial PheRS is strongly linked to both the α and the β chains of other bacterial PheRS’s. This leads us to suggest that the yeast mitochondrial PheRS is a chimera of the α and β chains of a former “standard” bacterial α2β2 enzyme, resulting from genetic recombination. Nothing can tell us, however, whether the recombination occurred before or after the yeast nucleus appropriated the former mitochondrial gene. It is probably relevant to note at this point a somewhat similar situation with the GlyRS from Chlamydia trachomatis (Wagar et al. 1995) in which the α and β subunits are encoded as a single chain by a single gene.

Table 1
Clusters Obtained from the Single Linkage Analysis of Aminoacyl-tRNA Synthetases

<table>
<thead>
<tr>
<th>Cluster</th>
<th>No. of Sequences</th>
<th>Amino Acid Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>220</td>
<td>Cys, Ile, Leu, Met, Val</td>
</tr>
<tr>
<td>2</td>
<td>220</td>
<td>Gly, His, Ser, Pro, Thr</td>
</tr>
<tr>
<td>3</td>
<td>63</td>
<td>Asn, Asp, Lys</td>
</tr>
<tr>
<td>4</td>
<td>49</td>
<td>Trp, Tyr</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>Phe</td>
</tr>
<tr>
<td>6</td>
<td>32</td>
<td>Gln, Glu</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
<td>Arg</td>
</tr>
<tr>
<td>8</td>
<td>23</td>
<td>Ala</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
<td>Gly</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>Lys</td>
</tr>
</tbody>
</table>

Note.—For each cluster, the second column indicates the total number of sequences it contains, and the third column lists the cognate amino acids of the aaRS’s that have been grouped in the cluster. The GlyRS’s in cluster 2 are of the dimeric form, while those in cluster 9 are tetrameric. The last cluster (cluster 10) contains the LysRS from the Archaea Methanobacterium thermoautotrophicum, Methanococcus maripaludis, and Archaeoglobus fulgidus and that from the spirochete Borrelia burgdorferi.
Evolution of Aminoacyl-tRNA Synthetases

Fig. 1. —a, Pyramidal classification of the α and β chains of PheRS. The names of the archaeabacteria are italicized. cyto stands for cytoplasmic, and mito stands for mitochondrial. b, Schematic representation of the similarities between the α and β chains of PheRS from *T. thermophilus*. The motifs 1, 2, and 3, characteristic of class II synthetases, were located in the β chain on the basis of structural similarities. c, Schematic representation of the similarities between the yeast mitochondrial PheRS (middle) and the α and β chains of the PheRS’s from *H. influenzae* (upper) and *Synechococcus* (lower), respectively. Each segment of similarity is represented with a different shading. The figures under the segments are the percentages of identities for each segment and the corresponding Z values (in brackets).

Two more points in the pyramidal classification are worthy of note: (1) for both the α and β chains, the archaeal and eukaryotic sequences are grouped together, while the bacterial sequences make another group. This is consistent with the Woesean view of evolution. However, the sequences from *B. burgdorferi* are much closer to their archaeal/eucaryal counterparts than to the bacterial sequences. (2) two sequences, one from *M. jannaschii* (MJ1660) and the other from *M. thermoautotrophicum* (MTH1501), suggest a duplication of the PheRS α gene.

ArgRS provides quite a different example (fig. 2), in which we see that Gram-negative bacteria such as *E. coli* and *H. influenzae* are grouped with the Eucarya, while another proteobacterium (*Helicobacter pylori*) is clustered with the Gram-positive bacteria. In addition, the Archaea make up a subcluster of their own that is not directly linked to the Eucarya. Obviously, all the
aaRS did not follow identical evolutionary paths, and the case of ArgRS will be further discussed later.

Duplication of Genes Within the aaRS Family

Clusters of protein sequences such as those presented above, as well as powerful sequence retrieval systems such as SRS (Etzold, Ulyanov, and Argos 1996) or ACNUC (Gouy et al. 1985) and comparison programs like BLAST (Altschul et al. 1990) make it easy to search for gene duplications in entire genomic sequences or in data banks—provided, of course, that the sequences have not diverged too much. In table 2, we give a list—probably not exhaustive—of duplications and even triplications among genes encoding aaRS (for the moment, we do not make any distinction between a gene duplication and a probable horizontal transfer; see next section). In some cases, the sequence similarity between an authenticated aaRS and a putative duplicated gene does not span the entirety of the aaRS sequence, as is the case for the example of YOR335C (AlaRS in yeast) and YNL040W, in which the similarity is limited to the N-terminus of the AlaRS sequence. Nevertheless, table 2 shows that duplication events are far from seldom within the aaRS family, a fact that must be remembered when attempting to perform phylogenetic reconstructions. It is expected that more and more duplications within the aaRS family will be discovered as time passes and sequences accumulate.

Horizontal Transfer of aaRS Genes (LysRS, GluRS)

As shown by Médigue et al. (1991), a multivariate analysis of the codon usage of the genes from one spe-
Evolution of Aminoacyl-tRNA Synthetases

Fig. 3.—Factorial correspondence analysis of the codon usage of all the ORFs in (a) E. coli and (b) H. influenzae. Each small dot represents one gene projected onto the plane of maximum inertia. Each gene encoding an aminoacyl-tRNA synthetase is represented by a solid diamond (●) except genX (●), lysU (▲), and yadB (■).

cies can point to probable horizontal transfers (see also Guerdox-Jamet et al. 1997). We have therefore undertaken such a study on the aaRS genes from several organisms, using mainly the server accessible from http://indigo.genetique.uvsq.fr (unpublished data) that implements the factorial correspondence analysis (FCA) of the codon usage in E. coli, Bacillus subtilis, and Arabidopsis thaliana. The simple underlying idea is that a xenologous gene will likely have a codon usage different from that of similar genes in the host, and, since the establishment of the codon usage is a slow process (Diaz-Lazcoz et al. 1995), this should be detectable even if the transfer occurred a long time ago.

The FCA of the codon usage of 3,911 ORFs longer than 300 bp from E. coli is shown in figure 3a. All the genes encoding an aaRS, except three, cluster in one single group, including lysS, which encodes the constitutive LysRS, indicating that they all have a similar codon usage. One of the genes that do not cluster with the others is lysU, a nonconstitutive heat-inducible LysRS.

The situation is different for H. influenzae (fig. 3b), a bacteria close to E. coli. Here, we have two (not three) genes encoding a LysRS. One is lysU, whose gene in the FCA lies within the cluster made by the other aaRS. The second is genX, which lies far away in the FCA.

These observations can be explained in several nonexclusive ways. It is well known that the codon bias in E. coli correlates with the level of expression of the genes (Gouy and Gautier 1982). Indeed, the positions of the constituent aaRS genes in the FCA map of E. coli in the so-called class II zone reflect their high expressivity (Médiégu et al. 1991). Hence, the positions of lysU, genX, and yadB in this map and of genX in that of H. influenzae could simply mirror a lower expressivity. This does not tell us, however, whether these genes—or some of them—are paralogous or have been acquired by horizontal transfers. The latter possibility cannot be excluded since, obviously, exogenicity is not incompatible with a lower expressivity. This is particularly true for genX, the sequences of which in E. coli and H. influenzae resemble that of the LysRS from Campylobacter jejuni more closely than that of lysS or lysU in their respective hosts.

An alternative explanation could be that the genes in question are in fact pseudogenes, the differences in their codon usages being the result of genetic drift. This is certainly not the case for lysU in E. coli, which is expressed in conditions of stress (Brevet et al. 1995). As for genX, its very high sequence similarities with other LysRS’s (see above) makes it improbable, but the question remains open for yadB in E. coli.

Phylogenetic Reconstructions

Phylogenetic reconstructions have been performed on the 20 different aaRS’s as described in Materials and Methods. As judged from the bootstrap values, 13 of them gave (at least partly) reliable phylogenies, namely AspRS, ArgRS, GluRS, GlyRS, HisRS, IleRS, LeuRS, MetRS, PheRS, ProRS, ThrRS, TrpRS, and TyrRS. For all of these aaRS’s, the multiple alignments that were used in the reconstructions, as well as the resulting unrooted trees and the bootstrap values, are available by anonymous ftp from ftp://genome.genetique.uvsq.fr/pub/synthetases.

It must be stressed here that our aim was not so much to obtain the universal tree of life but, rather, to test whether the aaRS’s (or which aaRS’s) are good candidates for this purpose and whether they underwent similar or contrasting evolutionary paths. In addition, we are well aware that trying to resolve in detail the phylogeny of the different kingdoms would require more sophisticated methods (e.g., Tourasse and Gouy 1997). Indeed, the topologies of the bacterial branches are generally poorly resolved in our trees.

Trees with Woesian Topologies

Among the 13 aforementioned aaRS, 9 (AspRS, GluRS, IleRS, LeuRS, MetRS, PheRS, ProRS, TrpRS, and TyrRS) provided reconstructions with the currently accepted topology delimiting the three domains of life.
Two examples are given in figure 4, namely AspRS (a class II enzyme) and GluRS (class I). Note, however, the poor resolution of the bacterial branches and the separation of *Halobacterium salinarum* from the other Archaea in the most parsimonious tree of AspRS. In the case of ProRS (not shown), *Mycoplasma genitalium* and *Mycoplasma pneumoniae* are placed within the Archaea and Eucarya. But since mycoplasma are parasites for a wide range of hosts including animals and plants, the occurrence of horizontal gene transfers would not be surprising. Similarly, *B. burgdorferi* is classified together with the Archaea and the Eucarya in the case of MetRS, ProRS, and PheRS. These anomalies most probably point to the idiosyncratic evolution of particular genes in particular species and do not call into question the overall topologies of the trees.
Trees with Nonconventional Topologies

The trees obtained with the other four aaRS’s are, in one way or another, nonconventional. For example, the present analyses on ArgRS (fig. 5) do not link the Archaea and the Eucarya as sister groups. In addition, H. pylori, a Gram-negative proteobacterium, is classified together with the Gram-positive bacteria. Both of these observations could be made from the pyramidal classification shown in figure 2. It is probably relevant to recall here that the ArgRS gene is duplicated in A. thaliana. The two copies (see table 2) are so similar (88% identity) that they cluster together. In addition, two
genes resembling ArgRS have been found so far in Caenorhabditis elegans. Their amino acid sequences share 27% identity, and they are well separated in the reconstructed tree. In this case, it seems possible that one gene (Trembl: Q18316) would in fact encode the mitochondrial ArgRS. Finally, figure 5 shows that the yeast cytoplasmic ArgRS clusters with its mitochondrial counterpart. All of these observations, as well as the long branch between the two main groups, suggest an ancient paralogy.

The situation is similar in the case of ThrRS (fig. 5), in which the Archaea are also separated from the Eucarya and clearly clustered with Gram-positive bacteria. Such a situation has been observed with e.g., the HSP 70 chaperones (Gupta et al. 1997) and glutamine synthetase (Brown et al. 1994). Here again, the yeast cytoplasmic and mitochondrial enzymes belong to the same clade, and H. pylori is clustered with the Gram-positive bacteria. We note that there exist two ThrRS’s in B. subtilis, namely thrS and thrZ. The unusual topology of the tree can be explained parsimoniously by an early duplication of the primitive ThrRS gene, leading to what are now thrS and thrZ. During the course of evolution, the Archaea and the Gram-positive bacteria would have lost thrZ, while the Eukarya and the Gram-negative bacteria would have lost thrS.

An alternative and more general explanation for the nonconventional topologies of the ArgRS and ThrRS trees is that eukaryotes obtained a copy from a proteobacterial endosymbiont while their original genes were lost. This would explain the similarity between eukaryotic and bacterial Gram-negative ArgRS and ThrRS. Indeed, such an evolutionary scenario has been proposed for other gene phylogenies showing a Gram-negative bacteria/eukaryote clustering (see Martin 1996 for a summary).

The Case of Tryptophanyl- and Tyrosyl-tRNA Synthetases

TrpRS and TyrRS are quite controversial in the literature. A phylogenetic analysis of a multiple alignment comprising 16 different TrpRS and TyrRS sequences led Ribas de Pouplana et al. (1996) to the surprising conclusion that the eukaryotic TrpRS and TyrRS are more related to each other than to their respective prokaryotic counterparts. Their most parsimonious tree clearly classified the eukaryotic proteins within the same clade, thus suggesting an origin for the eucaryal genes separate from that of their bacterial counterparts. The high degree of structural similarity between TrpRS and TyrRS from B. stearothermophilus (Doublé et al. 1995) was taken as supporting evidence for a surprisingly recent common ancestor, in spite of the much poorer sequence similarity. Later, Brown et al. (1997) undertook a similar study, but with a larger number of sequences and a broader range of taxa, and came to a different conclusion: all of their phylogenetic reconstructions supported the separate monophyly of TrpRS and TyrRS.

We therefore undertook yet one more similar study based on 49 TrpRS and TyrRS sequences, with particular attention being paid to the structural alignment of Doublé et al. (1995) in the multiple-alignment step (see Materials and Methods). Part of our multiple alignment, from the end of the first crossover connection to the middle of the dimer interface, is shown in figure 6. Clearly, a number of residues (identical or similar) are conserved within the two groups of sequences. In addition, there are positions where a residue is highly conserved within only one group and is characteristic of this group. Although Brown et al. (1997) also took full account of the structural alignment, their published multiple-alignment of TrpRS and TyrRS sequences differs from ours in some places (but is identical in the most conserved regions). The differences probably come from the use of different alignment programs and/or from the fact that these authors aligned the TrpRS and TyrRS sequences together, while we treated them separately.

Figure 7 shows the pyramidal classification of the TrpRS and TyrRS sequences. It is clear that in each group, the archaeal and eucaryal synthetases are classified together and that the bacterial enzymes make another subgroup. One prominent feature of this classification is that the archaeal/eucaryal TyrRS’s are closer to the archaeal/eucaryal TrpRS’s than to the bacterial TyrRS’s. This is in agreement with the observations of Ribas de Pouplana et al. (1996). However, the archaeal/eucaryal TrpRS’s are (slightly but clearly) closer to their bacterial counterparts than to the archaeal/eucaryal TyrRS’s, and this is at variance with the classification of Ribas de Pouplana et al. (1996). In addition, the bacterial TrpRS’s and TyrRS’s are quite far from one another and cannot be considered as forming a group. The simplest tentative explanation for these observations is that, for one reason or another, the rate of mutations of the TyrRS genes within the Archaea and Eucarya is different from that of the TyrRS genes within the Bacteria.

Let us now turn to the phylogenetic reconstructions based on the structural multiple alignment, shown in figure 8a and b. It appears that the TrpRS sequences and the TyrRS sequences partition into two separate monophyletic groups. Within each group, the archaeal and eucaryal synthetases form one clade. The eukaryotic TrpRS’s and TyrRS’s are definitely not clustered together. In many respects, the trees in figure 8a and b are similar to those published by Brown et al. (1997) and therefore do not support the results of Ribas de Pouplana et al. (1996). Since Brown et al. (1997) suggested that these conflicting results could arise from too limited a sampling of the taxa in the work by Ribas de Pouplana et al. (1996), we removed all those sequences that were not present in this last study from our multiple alignment and reconstructed the phylogenetic tree with the same set of sequences (fig. 8c and d). Again, the eukaryotic TrpRS’s and TyrRS’s do not cluster into one monophyletic group. It is thus probable that the multiple alignment used by Ribas de Pouplana et al. (1996) was markedly different from ours and from that of Brown et al. (1997).

The Case of Lysyl-tRNA Synthetase

We have previously seen that the gene encoding LysRS is duplicated in several species and that there is
FIG. 6.—Part of the multiple alignment between TrpRS and TyrRS. The alignment between the TrpRS (W-BACST) and TyrRS (Y-BACST) from B. stearothermophilus is identical to the structural alignment published by Doublie et al. (1995). Some residues (identical or similar) that are conserved in both groups are boxed and shaded. Some residues that are conserved within only one group are boxed unshaded. The names of the TrpRS and TyrRS sequences begin with W and Y, respectively. The names of the species follow the SwissProt convention, i.e. SCHPO stands for Schizosaccharomyces pombe, METJA for M. jannaschii, etc.

some evidence of horizontal transfer in E. coli and H. influenzae. Until recently, no LysRS's had been identified among the Archaea except in S. solfataricus. Unfortunately, the reconstruction of the LysRS phylogeny results in rather low bootstrap values (not shown) that prevent a more detailed study. This synthetase, however, is absolutely unique, since it has been demonstrated recently (Ibba et al. 1997a, 1997b) that the LysRS's in four Archaea (M. jannaschii, A. fulgidus, M. maripaludis, and M. thermoautotrophicum) and one spirochete (B. burgdorferi) display sequence features that are characteristic of the class I-type synthetases, while all the LysRS's from other bacteria and eukaryotes and that from S. solfataricus belong to class II. It is not clear at the moment whether the class I-type LysRS's appeared before or after the universal ancestor gave rise to the three kingdoms of life, but the results of Ibba et al. (1997a, 1997b) demonstrate unambiguously that the tRNA\(^{Lys}\) aminoacylation function appeared independently at least twice during the course of evolution.

The Case of Glycyl-tRNA Synthetase

GlyRS has been found to exist under two different oligomeric forms, namely a \(\alpha_2\beta_2\) heterotetramer and a \(\alpha_2\) homodimer. The two forms do not share any convincing sequence similarity (Shiba et al. 1994). Until recently, the tetrameric GlyRS has been found only among the Bacteria and the dimeric enzyme only within Euca-ria, which suggested that the oligomeric structure of GlyRS distinguishes prokaryotes and eukaryotes (see Mazauric et al. 1996 and references therein). In addition, it had been suggested that the discriminator base 73 in tRNA\(^{Gly}\) could be correlated with both subunit structure and the prokaryote/eukaryote division, i.e., \(\alpha_2\beta_2/\alpha_4\) and \(\alpha_2/\alpha_2\) (Shiba et al. 1994). Finally, sequence similarities clearly placed the dimeric GlyRS within subclass IIa together with HisRS, ProRS, SerRS, and ThrRS, and the tetrameric GlyRS within subclass IIc together with the \(\beta_4\) AlaRS and the \(\alpha_2/\alpha_2\) PheRS (Cusack 1993, 1995).

The picture today is somewhat different, since a dimeric GlyRS was found in T. thermophilus (Logan et al. 1995; Mazauric et al. 1996) while genomic sequencing projects proved that the \(\alpha_2\) GlyRS is the form that exists in the Archaea, Mycoplasma sp., and other bacteria. In contrast, the tetrameric GlyRS has been observed only among the Bacteria. In addition, GlyRS from T. thermophilus is able to aminoacylate heterologous E. coli and yeast tRNA\(^{Gly}\) (Mazauric et al. 1996), which discloses any relationship between the oligomeric state of the enzyme and the nature of its host. Finally, the availability of new sequences provides compelling evidence that the closest relatives to the \(\alpha_2\) GlyRS are ThrRS and ProRS, reinforcing its assignation to subclass IIa. We are therefore led to the following conclusions:

1. Since the dimeric GlyRS is found among the three kingdoms of life, this is the glycyl-tRNA synthetase that
Fig. 7.—Pyramidal classification of all the currently available TrpRS and TyrRS sequences. The sequence labeled *Synechocystis*(2) is fragmentary. The names of the archaeabacteria are italicized.

existed originally, sharing a common subclass IIa ancestor with HisRS, ProRS, SerRS, and ThrRS. (2) The existence of another GlyRS with the same specificity, yet markedly different in both its quaternary structure and amino acid sequence, suggests that this very specificity was acquired a second time during the course of evolution among a different subclass of tetrameric aaRS's comprising PheRS and AlaRS. This may seem unlikely at first, but that the same specificity could appear independently within two subclasses of the same family is, after all, less surprising than the convergent evolution of trypsin and subtilisin (Kraut et al. 1972). Note also that the tRNA-aminoacylation function itself clearly appeared independently at least twice, once within the class I family, and a second time within the class II family. Finally, the existence of both class I– and class II–type LysRS's (vide supra) shows that this last synthetase was invented twice. Hence, the existence of two different nonorthologous GlyRS's is not that improbable. (3) Since the tetrameric GlyRS has been found only within the Bacteria, it is probable that it appeared on this branch after the Bacteria separated from the other kingdoms.

A special case should be noted for *A. thaliana*, in which a GlyRS of plastid origin (EMBL: ATAJ3069) appears to contain a single chain encoded by a single gene but containing both a α-chain at its N-terminus and a β-chain at its C-terminus. As indicated before, this is also true for the GlyRS from *C. trachomatis* (Wagar et al. 1995).
FIG. 8.—Reconstructed phylogeny based on the multiple alignment between TrpRS and TyrRS. All the TrpRS names begin with W, and all the TyrRS names begin with Y. Yc stands for yeast cytoplasm, Ym for yeast mitochondrion, and Ce for Caenorhabditis elegans. The trees in a and b are the consensus neighbor-joining and most-parsimonious trees, respectively, for the 49 currently available sequences. The trees in c and d are the consensus neighbor-joining and most-parsimonious trees, respectively, reconstructed with only those sequences that were used in the study of Ribas de Pouplana et al. (1996). Each tree was obtained after 1,000 bootstrap resamplings. The bootstrap scores are given for some key branches.
Conclusions

Aminoacyl-tRNA synthetases have long been considered suitable for case studies because of their extreme diversity in terms of amino acid sequences, quaternary structures, and modes of action, while sharing basically the same function. In the present work, we have shown that the evolutionary paths of their genes are also very diverse due to numerous events such as gene fusions, gene duplications, genetic recombinations, probable horizontal transfers, and even reacquisition of the same specificity. These proteins are a paradigm of F. Jacob's "odd jobs of evolution." If the majority of them show phylogenetic reconstructions that are consistent with the broad grouping into three domains, many individual genes themselves pose difficulties that preclude any simple evolutionary scheme. To make the picture coherent, however, we do not think it necessary to invoke, like Shiba, Motegi, and Schimmel (1997), a confused notion of "crossover between domains," even less a non-Darwinian "necessity for adaptation." As time passes, aaRS's become even more fascinating as their study brings new and unexpected insights into the processes of evolution.

Acknowledgments

We wish to thank Prof. P. Slnimski for his continuous support and interest. The all-by-all pairwise alignments, the calculations of the pairwise Z values, the delineation of the clusters of sequences, their pyramidal classifications, and the corresponding Web pages were done with the help of J.-J. Codani and E. Glemet from INRIA. We also thank the two anonymous referees for their constructive suggestions and criticisms.

LITERATURE CITED

IBBA, M., S. MORGAN, A. W. CURNOW, D. R. PRIDMORE, U. C. VOTHKNECHT, W. GARDNER, W. LIN, C. R. WOES, and


MOYSOK, L., and M. SAFRO. 1993. Phenylalanyl-tRNA synthetase from T. thermophilus has four antiparallel folds of which only two are catalytically functional. Biochimie 75:1091–1098.


MANOLO GOUY, reviewing editor

Accepted August 6, 1998