Intraphylum Diversity and Complex Evolution of Cyanobacterial Aminoacyl-tRNA Synthetases

Ignacio Luque,[1] María Loreto Riera-Alberola,* Alfonso Andújar,* and Jesús A. G. Ochoa de Alda†[1]*

*Instituto de Bioquímica Vegetal y Fotosíntesis, Consejo Superior de Investigaciones Científicas and Universidad de Sevilla, Avda Américo Vespucio, Seville, Spain; and †Departamento de Biología Molecular y Celular, IE Universidad, Campus de Santa Cruz la Real, C/Cardenal Zúñiga, Segovia, Spain

A comparative genomic analysis of 35 cyanobacterial strains has revealed that the gene complement of aminoacyl-tRNA synthetases (AARSs) and routes for aminoacyl-tRNA synthesis may differ among the species of this phylum. Several genes encoding AARS paralogues were identified in some genomes. In-depth phylogenetic analysis was done for each of these proteins to gain insight into their evolutionary history. GluRS, HisRS, ArgRS, ThrRS, CysRS, and Glu-Q-RS showed evidence of a complex evolutionary course as indicated by a number of inconsistencies with our reference tree for cyanobacterial phylogeny. In addition to sequence data, support for evolutionary hypotheses involving horizontal gene transfer or gene duplication events was obtained from other observations including biased sequence conservation, the presence of indels (insertions or deletions), or vestigial traces of ancestral redundant genes. We present evidences for a novel protein domain with two putative transmembrane helices recruited independently by distinct AARS in particular cyanobacteria.

Introduction

Aminoacyl-tRNA synthetases (AARSs) are the enzymes that load tRNAs with their cognate amino acids, thus having a central role in translation. The capacity of an AARS to decipher the genetic code relies on their double specificity for a particular amino acid and the corresponding isoacceptor transfer RNAs (tRNAs; Ibba and Soll 2000, 2004). This group of enzymes has been partitioned into two different classes (of 11 and 10 enzymes each), proteins of the same class being related in sequence and structure and having a common phylogenetic origin (Eriani et al. 1990). The catalytic domain of class I AARS is characterized by a Rossman fold structure and conserved HIGH and KSMKS motifs in the active site, whereas the class II catalytic domain is formed by a seven-stranded β structure with three α helices and contains three degenerate conserved motifs (Ibba and Soll 2000, 2004; O'Donoghue and Luthey-Schulten 2003). All steps involved in the expression of genetic information, including the aminoacylation step, must have a high degree of accuracy. Thus, some AARS possess proofreading activities located in editing domains that prevent the release of incorrectly acylated tRNAs (Jakubowski 2004).

Proteins in living organisms are generally composed of 20 different amino acids, and it has long been assumed that every cell should be equipped with a set of 20 AARS, one for each amino acid. However, this scheme is far from universal as mitochondria, chloroplasts, and most prokaryotic cells contain a smaller number of AARS. For instance, GlnRS and AsnRS are missing from eukaryotic organelles and most prokaryotes where the corresponding aminoacyl-tRNAs are synthesized by indirect routes (Ibba et al. 1997, 2000; Feng et al. 2004) that involve the misacylation of $tRNA^{Gln}$ or $Asn\text{-}tRNA^{Asn}$ with glutamate or aspartate by a nondiscriminating GluRS (ND-GluRS) or a nondiscriminating AspRS (ND-AspRS), respectively, followed by transamidation of the charged glutamate or aspartate in a reaction catalyzed by a tRNA-dependent amido transferase (AdT) (Wilcox and Nirenberg 1968; Curnow et al. 1996, 1997; Gagnon et al. 1996; Raczniak et al. 2001). ND-GluRS and ND-AspRS are thus enzymes of relaxed specificity that function on their cognate $tRNA^{Glu}$ or $tRNA^{Asp}$ and on noncognate $tRNA^{Gln}$ or $tRNA^{Asn}$, respectively. This is in contrast to the discriminating enzymes, D-GluRS and D-AspRS, that act solely on their cognate tRNAs. In bacteria, AdT is a trimeric enzyme encoded by the *gatA*, *gatB*, and *gatC* genes (Curnow et al. 1997; Nakamura et al. 2006). Particular bacterial species can be predicted to utilize the indirect pathway for the synthesis of $Gln\text{-}tRNA^{Gln}$ or $Asn\text{-}tRNA^{Asn}$ by the absence of GlnRS- or AsnRS-encoding genes from their genomes and the simultaneous presence of *gat* genes.

The universality of both the genetic code and the amino acid composition of living organisms suggest that AARS arose very early in the history of life. Many evolutionary studies have focused on AARS due to their central role in cell physiology and to their close relation to the origin and evolution of the genetic code (Diaz-Lazcoz et al. 1998; Woese et al. 2000; Ribas de Pouplana and Schimmel 2001). AARS genes frequently show evidence of horizontal gene transfer (HGT) (Brown and Doolittle 1999; Woese et al. 2000; Brown et al. 2003; Dohm et al. 2006; Zhaxybayeva et al. 2006), which is in contrast to genes of other components of the translation apparatus that tend to be inherited vertically and are therefore more suitable for the inference of the phylogenetic relation of living organisms (Doolittle and Handy 1998; Beiko et al. 2005; Ciccarelli et al. 2006). Although HGT is limited by many physical and biological barriers (Thomas and Nielsen 2005; Sorek et al. 2007), it is nowadays acknowledged as a leading force in evolution especially in prokaryotes but also in eukaryotic organisms (Beiko et al. 2005; Gogarten and Townsend 2005; Lerat et al. 2005). Evolutionary divergence of redundant genes originated by HGT or gene duplication events often leads to gene loss or the generation of pseudogenes; however in some cases, it may contribute to increase the protein repertoire of an organism by the acquisition of new functions (Gogarten and Townsend 2005). Thus,

[1] These two authors contributed equally.

Key words: aminoacyl-tRNA synthetase, cyanobacteria, horizontal gene transfer, gene duplication, domain shuffling, phylogeny.

E-mail: ignacio.luque@ibvf.csic.es.

Mol. Biol. Evol. 25(11):2369–2389. 2008
doi:10.1093/molbev/msn197
Advance Access publication September 4, 2008

AARS homologs involved in a variety of functions including tRNA modification and amino acid synthesis have been described (Sissler et al. 1999; Schimmel and Ribas De Pouplana 2000; Roy et al. 2003; Dubois et al. 2004; Salazar et al. 2004).

Computer-assisted analysis of protein or DNA sequences is a valuable tool for the study of the phylogenetic relationship of proteins or living organisms and for tracing their evolutionary history. Current likelihood-based techniques allow a wide variety of phylogenetic inferences from sequence data and robust statistical assessment of all results including the potential monophyly of specific groups and the identification of potential gene transfer events (Huelsenbeck and Crandall 1997; Whelan et al. 2001; Shimodaira 2002; Poptsova and Gogarten 2007).

Cyanobacteria are a monophyletic group of Gram-negative bacteria (Honda et al. 1999; Zhaxybayeva et al. 2006) considered to have invented the process of oxygenic photosynthesis, thus having a determinant role in the transition of the Earth's atmosphere to an oxygen-rich state (Rye and Holland 1998; Brocks et al. 1999; Xiong and Bauer 2002; Mulkidjanian et al. 2006; Tomitani et al. 2006; Xiong 2006; Knoll 2008). The genomic sequences of more than 30 cyanobacterial species have become available in the last decade, providing a valuable tool for the study of these organisms. Genome sizes vary from 1.66 to 9 Mbp illustrating the genetic diversity of this phylum. Despite some efforts (Beauchemin et al. 1973; Luque et al. 2002, 2006), little attention has been paid to AARS and gene translation in cyanobacteria, and many elements and molecular mechanisms remain unexplored. In this work, we have undertaken a global survey of 35 cyanobacterial genomes to characterize the gene set encoding AARS and homologous proteins. Phylogenetic analyses have been carried out for each AARS or AARS-like protein to gain insight into their evolutionary history. Functional consequences derived from the evolutionary course of different AARS have been further explored and are discussed.

Materials and Methods

The pipeline of bioinformatic methods is shown in figure 1.

Data Retrieval

The genome sequences used in this work are shown in table 1. Small ribosomal RNA (rRNA) and protein sequences used for analysis are indicated in supplementary file 1 (Supplementary Material online).

Small rRNA sequences of 35 cyanobacteria and 4 other bacteria to be used as outgroups (supplementary file 1, Supplementary Material online) were obtained from the National Center for *Biotechnology Information* (NCBI) microbial genomes database (Wheeler et al. 2007) at http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi after a BlastN search using the *rrn16Sa* gene of *Synechocystis* PCC 6803 (GI: 16329170) as query. Homologous sequences longer than 1,300 nt were retrieved for phylogenetic analysis (supplementary file 1, Supplementary Material online).

AARS protein sequences were obtained from the same database after a BlastP search using the corresponding protein sequence from *Synechocystis* PCC 6803, *Gloeobacter violaceus* PCC 7421, *Nostoc* sp. PCC 7120, or *Prochlorococcus marinus* MIT9313 as query. The existence of Class I LysRS in cyanobacteria was explored as described above but using the sequence of *Methanococcus jannaschii* LysRS (GI: 3183557) as query. This enzyme was found to be absent from the 35 cyanobacterial genomes examined. Cyanobacterial AARS sequences were retrieved together with representative sequences of other phyla including those showing a Blosum62 alignment score within the score range and/or close to the minimum score for cyanobacterial sequences.

Protein sequences for the construction of the cyanobacterial species tree were selected according to Ciccarelli et al. (2006), who have recently described 29 bona fide clusters of orthologous groups suitable for the inference of the tree of life based on their universal distribution and their reluctance to HGT. These cyanobacterial sequences, together with those of *Bacillus subtilis* and *Escherichia coli* used to root the tree, were retrieved from the NCBI microbial genomes database after a BlastP search using the sequence of *Synechocystis* PCC 6803 as query. We selected 25 clusters of orthologous groups (COGs), out of the 29 reported by Ciccarelli, to carry out our analysis: COG0012, COG0016, COG0048, COG0049, COG0052, COG0087, COG0091, COG0092, COG0093, COG0094, COG0096, COG0097, COG0098, COG0100, COG0102, COG0172, COG0184, COG0186, COG0197, COG0200, COG0201, COG0202, COG0256, COG0495, and COG0522. Some COGs described by Ciccarelli (COG0080, COG0081, COG0103, and COG0533) were not retrieved because they were incomplete, fragmented, or absent from the shotgun genome sequence of *Nostoc punctiforme* and/or *Crocosphaera watsonii*, a problem frequently encountered in genome sequencing associated with genes reluctant to HGT (Sorek et al. 2007).

In cases where multiple alignments revealed sequences lacking N-terminal amino acids presumably due to misannotation of the start codon, TBlastN searches were used to inspect the DNA sequence and reassign, if required, the translation start point.

Sequence Alignment

Small rRNA sequences were aligned using the Ribosomal Database Project II pipeline, release 9.46 (Cole et al. 2007) at http://rdp.cme.msu.edu/, which takes into account the secondary structure of rRNA sequences.

We chose MAFFT version 5 (Katoh et al. 2005) for protein sequence alignments because we observed that, in agreement with a recent report (Talavera and Castresana 2007), it allows a better identification of homologous positions than MUSCLE (Edgar 2004) or ClustalW (Thompson et al. 1994).

Selection of Conserved Positions in Multiple Alignments

Conserved blocks in multiple alignments were selected with the Gblocks 0.91b program (Castresana

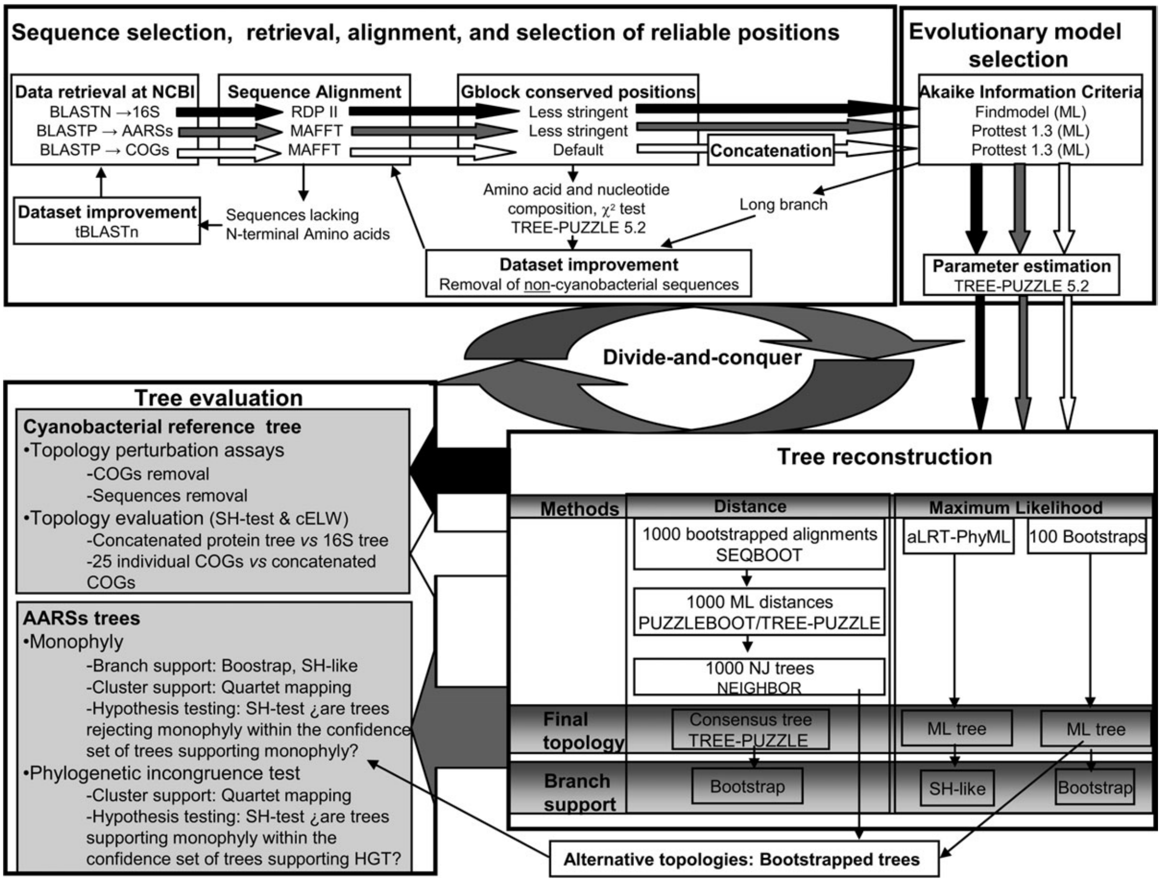

Sequence selection, retrieval, alignment, and selection of reliable positions
Data retrieval at NCBI
BLASTN →16S
BLASTP → AARSs
BLASTP → COGs
Sequence Alignment
RDP II
MAFFT
MAFFT
Gblock conserved positions
Less stringent
Less stringent
Default
Concatenation
Dataset improvement
tBLASTn
Sequences lacking
N-terminal Amino acids
Amino acid and nucleotide
composition, χ² test
TREE-PUZZLE 5.2
Long branch
Dataset improvement
Removal of non-cyanobacterial sequences
Evolutionary model selection
Akaike Information Criteria
Findmodel (ML)
Prottest 1.3 (ML)
Prottest 1.3 (ML)
Parameter estimation
TREE-PUZZLE 5.2
Divide-and-conquer
Tree evaluation
Cyanobacterial reference tree
•Topology perturbation assays
-COGs removal
-Sequences removal
•Topology evaluation (SH-test & cELW)
-Concatenated protein tree vs 16S tree
-25 individual COGs vs concatenated COGs
AARSs trees
•Monophyly
-Branch support: Boostrap, SH-like
-Cluster support: Quartet mapping
-Hypothesis testing: SH-test ¿are trees rejecting monophyly within the confidence set of trees supporting monophyly?
•Phylogenetic incongruence test
-Cluster support: Quartet mapping
-Hypothesis testing: SH-test ¿are trees supporting monophyly within the confidence set of trees supporting HGT?
Tree reconstruction
Methods
Distance
Maximum Likelihood
1000 bootstrapped alignments
SEQBOOT
1000 ML distances
PUZZLEBOOT/TREE-PUZZLE
1000 NJ trees
NEIGHBOR
aLRT-PhyML
100 Bootstraps
Final topology
Consensus tree
TREE-PUZZLE
ML tree
ML tree
Branch support
Bootstrap
SH-like
Bootstrap
Alternative topologies: Bootstrapped trees

FIG. 1.—Diagram of the phylogenetic analysis workflow. White boxes represent the four main steps of the procedure for the construction and analysis of different data sets: small rRNA sequences (16S, dark arrows), AARS protein sequences (gray arrows), and COGs (white arrows). Sequences were identified and retrieved after Blast search from the NCBI and then homologous positions were aligned using MAFFT (for proteins) or the Ribosome Database pipeline (for rRNA). Reliable positions were obtained using Gblock and the optimal evolutionary models for the resulting alignment were selected using the AIC implemented in Prottest (for proteins) and Findmodel (for rRNA). The results obtained at different steps of the workflow were used to improve the data sets by removing, if possible, long-branch attraction artifacts, sequences with heterogeneous amino acid composition, and N-terminal misannotations. The optimal evolutionary model was used for tree reconstruction by a distance and ML methods. The resolution of AARSs was improved following a divide-and-conquer strategy. Finally, different phylogenetic hypotheses were evaluated statistically using confidence sets of optimal and alternative topologies as well as different branch supports. For a detailed description and references, see Materials and Methods.

2000) at http://molevol.ibmb.csic.es/Gblocks_server.html that extracts alignment positions that can be used reliably in phylogenetic analysis (Talavera and Castresana 2007). Gblocks settings were selected depending on the data sets because a relaxed selection of blocks is better for short alignment, whereas a stringent selection is more adequate for longer ones (Talavera and Castresana 2007).

In rRNA alignments, the "Minimum length of a block" parameter was set to five in order to select many short conserved blocks. Gaps were removed, and the minimum number of sequences for conserved and flanking positions was set to half the number of sequences plus one. The maximum number of contiguous nonconserved positions was set to eight.

To select blocks in single protein alignments (i.e., AARS alignments), we removed gap positions within the final blocks and used less stringent block selection than under the default conditions, that is, setting the minimum length of blocks to two, relaxing the number of contiguous nonconserved positions, and fixing the minimum number of sequences for conserved positions to half the number of sequences plus one.

To select blocks in protein alignments to be concatenated, we used the default settings of the Gblock program, which are more stringent with respect to the aforementioned because the minimum length of blocks is set to ten and the contiguous nonconserved positions is set to eight. Independent concatenation of blocks obtained from the 25 multiple alignments of COGs resulted in a supermatrix of 5,012 positions.

Data Set Improvement

Alignments of conserved positions were analyzed by the χ^2 test implemented in Tree-Puzzle 5.2 (Schmidt et al. 2002) to detect sequences with a significant heterogeneity in amino acid or nucleotide composition and those causing long branching after maximum likelihood (ML) tree reconstruction. Such sequences were removed (or in some cases replaced by a sequence not showing a significant compositional bias from a closely related species) unless they were

Table 1
Cyanobacterial Genome sequences Used in This Work

Organism	GenBank Accession Number	Reference or Sequencing Institution
Gloeobacter violaceus PCC 7421	BA000045	Nakamura et al. (2003)
Synechococcus OS-A (JA-3-3Ab)	CP000239	The Institute for Genomic Research
Synechococcus OS-B'(JA-2-3B'(2-13))	CP000240	The Institute for Genomic Research
Synechococcus PCC 6301	AP008231.1	Nagoya University, Japan
Synechococcus PCC 7942	CP000101	Department Of Energy Joint Genome Institute
Synechococcus WH 5701	AANO00000000	J. Craig Venter Institute
Synechococcus CC9311	CP000435.1	The Institute for Genomic Research
Synechococcus RS9917	AANP00000000	J. Craig Venter Institute
Synechococcus RS9916	AAUA00000000	J. Craig Venter Institute
Synechococcus WH 8102	BX548020.1	Palenik et al. (2003)
Synechococcus CC9605	CP000110.1	Department Of Energy Joint Genome Institute
Synechococcus BL107	AATZ00000000	J. Craig Venter Institute
Synechococcus CC9902	CP000097.1	Department Of Energy Joint Genome Institute
Synechococcus WH 7805	NZ_AAOK00000000	J. Craig Venter Institute
Prochlorococcus marinus MIT9312	CP000111.1	Department Of Energy Joint Genome Institute
Prochlorococcus marinus AS9601	CP000551.1	J. Craig Venter Institute
Prochlorococcus marinus MIT 9301	CP000576.1	Gordan and Betty Moore Foundation Marine Microbiology Initiative
Prochlorococcus marinus CCMP1986 (MED4)	BX548174.1	Rocap et al. (2003)
Prochlorococcus marinus MIT 9515	CP000552.1	J. Craig Venter Institute
Prochlorococcus marinus NATL1A	CP000553.1	J. Craig Venter Institute
Prochlorococcus marinus NATL2A	CP000095.2	Department Of Energy Joint Genome Institute
Prochlorococcus marinus MIT 9211	AALP00000000	J. Craig Venter Institute
Prochlorococcus marinus CCMP1375 (SS120)	AE017126.1	Dufresne et al. (2003)
Prochlorococcus marinus MIT 9303	CP000554.1	J. Craig Venter Institute
Prochlorococcus marinus MIT 9313	BX548175.1	Rocap et al. (2003)
Thermosynechoccus elongatus BP-1	BA000039.2	Nakamura et al. (2002)
Synechocystis sp. PCC6803	BA000022.2	Kaneko et al. (1995, 1996) and Kaneko and Tabata 1997
Cyanothece sp. CCY0110	NZ_AAXW00000000	J. Craig Venter Institute
Crocosphaera watsonii	AADV00000000	Department Of Energy Joint Genome Institute
Anabaena variabilis ATCC 29413	CP000117.1	CP000117.1
Nostoc (Anabaena) sp. PCC 7120	BA000019.2	Kaneko et al. (2001)
Nodularia spumigena CCY9414	AAVW00000000	Netherlands Institute of Ecology and J Craig Venter Institute
Nostoc punctiforme PCC 73102 (ATCC 29133)	AAAY00000000	Meeks et al. (2001)
Lyngbya sp. PCC 8106 (CCY9616)	AAVU01000001	J Craig Venter Institute
Trichodesmium erythraeum IMS101	CP000393.1	Department Of Energy Joint Genome Institue

cyanobacterial sequences. In the later case, such sequences were removed, and trees with and without them were compared to analyze how their presence affected tree topology. Although all sequences passed the χ^2 test in most alignments of individual COGs, more than half of the concatenated COG sequences (20 out of 36) did not pass the amino acid composition test. To cope with a putative amino acid composition bias, we reduced the number of concatenated sequences progressively to obtain two sets of data: one of 22 concatenated COGs obtained after removing 4 sequences with atypical amino acid frequencies and another of 13 concatenated COG sequences in which all except one outgroup passed the χ^2 test.

Another source of tree topology instability could be the overrepresentation of some strains in the guide tree (i.e., marine *Synechococcus* and *Prochlorococcus*). We used Jalview 2.3 (Clamp et al. 2004) to reduce the number of these sequences in the data set (25 concatenated COGs) from 36 to 29 and to 17 by selecting those with less than 95% and 90% identities, respectively.

Model Selection

We used the Akaike information criterion (AIC) for evolutionary model selection (Sullivan and Joyce 2005). The AIC for a particular tree (t) is calculated as follows:

$$\mathrm{AIC}_t = -2\ln L_t + 2k_t,$$

where $\ln L_t$ is the maximum log likelihood of the data, given a particular tree and model, and k_t is the number of parameters in the model. For two competing models, a difference of 2 in AIC indicates substantial support for the model having the lowest AIC.

For each alignment of conserved positions, we first tested different substitution models with a proportion of invariable sites (I) and a Γ distribution to approximate among-site rate variations. Then, we selected the appropriate number of Γ rates, from 4 to 16.

The evolutionary model for small rRNA sequences was selected using Findmodel at http://hcv.lanl.gov/content/hcv-db/findmodel/findmodel.html, a Web implementation of Modeltest (Posada 2006). Findmodel results indicated that the best-fit models for the small rRNA were the Tamura-Nei (TN) (Tamura and Nei 1993) and the general time-reversible (GTR) (Tavaré 1986) substitution models with a proportion of I and a Γ distribution, denoted TN $+ \mathrm{I} + \Gamma$ and GTR $+ \mathrm{I} + \Gamma$, respectively. Because these models are equivalent (i.e., they showed an AIC difference of less than two), we chose the least parameterized TN model with five discrete categories (TN $+ \mathrm{I} + 5\Gamma$).

Evolutionary models for proteins were selected with Prottest 1.3 (Abascal et al. 2005) among empirical substitution matrices such as WAG, Dayhoff, Jones–Taylor–Thornton, VT, and specific improvements such as $+\mathrm{I}$, $+\Gamma$, and/or $+\mathrm{F}$ (observed amino acid frequencies) to account for the evolutionary constraints imposed by conservation of protein structure and function. In addition to AIC, Prottest uses other statistics (AIC and Bayesian information criterion [BIC]) to find the candidate model that best fits the data at hand. Although the empirical WAG matrix (Whelan and Goldman 2001) was the most suitable for the proteins under study, the specific improvements varied from one to another (supplementary file 2, Supplementary Material online), indicating different evolutionary constraints for each AARS.

Parameter Estimation

Once the model was established, we recalculated the parameters (transition/transversion, Y/R transition, and Γ alpha parameters as well as the fraction of invariable sites) with Tree-Puzzle 5.2 (supplementary file 2, Supplementary Material online).

Tree Reconstruction

We used two approaches to build phylogenetic trees from alignments, a distance method (Neighbor-Joining [NJ]) and ML method.

In order to obtain the distance tree (NJ tree), 1,000 bootstrap samples of the alignments of conserved residues were generated using program SEQBOOT from the PHYLIP package version 3.6 (Felsenstein 1989), and a pairwise ML distance matrix was calculated for each bootstrap sample using PUZZLEBOOT 1.03 (Roger A and Holmer M, unpublished data) allowing the analysis of multiple data sets with Tree-Puzzle 5.2 for each selected substitution model and the parameters set as estimated for the original data set. Phylogenetic trees from the 1,000 ML distance matrices were calculated with program NEIGHBOR from the PHYLIP package version 3.6. Finally, a consensus tree (NJ) was obtained using Tree-Puzzle 5.2 that provides the log likelihood of the topology obtained under the selected substitution model and the nonparametric bootstrap support for each branch and collapses branches with a bootstrap support lower than 50%. As a rule, the consensus tree improved the likelihood of the topology obtained by ML with Tree-Puzzle 5.2 during parameter estimation.

ML trees were estimated using the program aLRT-PhyML (Anisimova and Gascuel 2006) that performs an approximated likelihood ratio test (aLRT) for all branches and allows several branch support options. We obtained branch support by two methods: 100 bootstrapped samples and an aLRT nonparametric branch support based on a Shimodaira–Hasegawa–like procedure (SH-like) with the above-mentioned substitution model and parameters. In a few cases, the evolutionary model required the use as equilibrium amino acid frequencies those observed in the alignment (indicated as +F), instead of the original amino acid frequencies obtained from the data set used to generate the model. In these cases, because this option is not available in aLRT-PHYML, we calculated the WAG + I + Γ + F ML tree with program PhyML (Guindon and Gascuel 2003) modified for Prottest 1.3.

Tree Topology Evaluation and Hypothesis Testing

To evaluate the uncertainty of choosing the optimal tree among different tree topologies, we used tests available in the program Tree-Puzzle 5.2, the SH test and the expected likelihood weight (cELW) (Kishino and Hasegawa 1989; Shimodaira and Hasegawa 1999; Goldman et al. 2000; Strimmer and Rambaut 2002). These tests provide confidence set for a given tree topology that allows the comparison of two different phylogenetic hypotheses and the evaluation of different tree topologies under the selected model and parameters. All tests were done with $P < 0.05$ set as statistically significant and using the resampling of the estimated log-likelihood methods with 1,000 replications. The SH test compares the likelihoods of multiple tree topologies and, for each topology, yields a *P* value (Shimodaira and Hasegawa 1999) that represents the possibility that the tree is the true tree given the data and the model. A significant *P* value indicates that a topology has a likelihood that is significantly different from that of the ML topology and can be rejected. The candidate trees that are not significantly worse than the optimal tree are used to construct the confidence set of trees. The SH test is particularly safe to use and is a good option when the number of candidate trees is small (Shimodaira 2002). Using this conservative method to infer a confidence set of trees was complemented with the cELW method, which considers the possibility of potential misspecification of the investigated trees (Strimmer and Rambaut 2002), that is, the true tree is not included in the set of candidate trees.

Data Set Partition and Tree Assembly

Trees from ML and NJ analyses were congruent. The latter was chosen for most figures (except for the concatenated tree which corresponds to a PhyML tree) because they are generally better resolved than ML–Tree-Puzzle trees and more conservative than the PhyML trees (Tree-Puzzle collapses those branches showing less than 50% support). Most phylogenetic trees of AARS sequences allowed broad phylogenetic relationships to be distinguished but frequently did not resolve closely related species. In order to increase the resolution of some tree branches, we used a “divide-and-conquer” approach (Delsuc et al. 2005), in which we selected sequences in particular branches plus an external sequence to be used as an outgroup. Gblock selection of the aligned subset of sequences resulted in an increase of conserved positions and, hence of the phylogenetic information, allowing the construction of a subtree with improved resolution. Subtrees were assembled into the global tree at the insertion point of the corresponding cluster.

Phylogenetic Incongruence Test

Whenever the topology of a tree suggested a putative horizontal transfer to a cyanobacterium, we applied a phylogenetic incongruence test based on a method published by Poptsova and Gogarten (2007). However, instead of the AU test we used the SH test to compare two different phylogenetic hypothesis (i.e., the occurrence or not of an HGT event). In this case, the optimal tree suggesting an HGT event was compared with an alternative tree in which all cyanobacterial sequences were forced to be monophyletic by placing the branch of the sequence putatively acquired by HGT at the position observed in the species tree (a topology unavailable among the bootstrapped trees). The null hypothesis (both trees are equally good explanations of the data) was compared with the alternative hypothesis (both trees are not equally good explanation of the data).

Monophyly Test

Monophyly of cyanobacterial AARSs was inferred from the branch support (NJ bootstrap, ML bootstrap, and SH-like) for a clade grouping all, and only, cyanobacterial AARS sequences. In addition to this, a quartet mapping for the group of interest was carried out (Daubin and Ochman 2004). Quartet mapping rapidly assesses the support for a group of sequences or the possibility of HGT events (supplementary file 4, Supplementary Material online). Clusters of cyanobacterial sequences having a branch support of 100 for the four methods were considered monophyletic (supplementary file 4, Supplementary Material online). Bootstraps showing lower branch support were analyzed by searching among the bootstrapped trees those rejecting monophyly and evaluating with the SH test if they were within the confidence set of the tree supporting monophyly. Monophyly is rejected if the clade is not found in any of the nonrejected trees (Shimodaira 2002).

Other Bioinformatic Methods

Jalview 2.3 (Clamp et al. 2004), TreeView 1.6.6 (Page 1996), and TreeDyn (Chevenet et al. 2006) were used for sequence or tree editing, respectively.

Prediction of transmembrane helices was performed with the following programs, TMHMM version 2.0 at http://www.cbs.dtu.dk/services/TMHMM-2.0/, DAS at http://www.sbc.su.se/~miklos/DAS/, and HMMTOP at http://www.enzim.hu/hmmtop/.

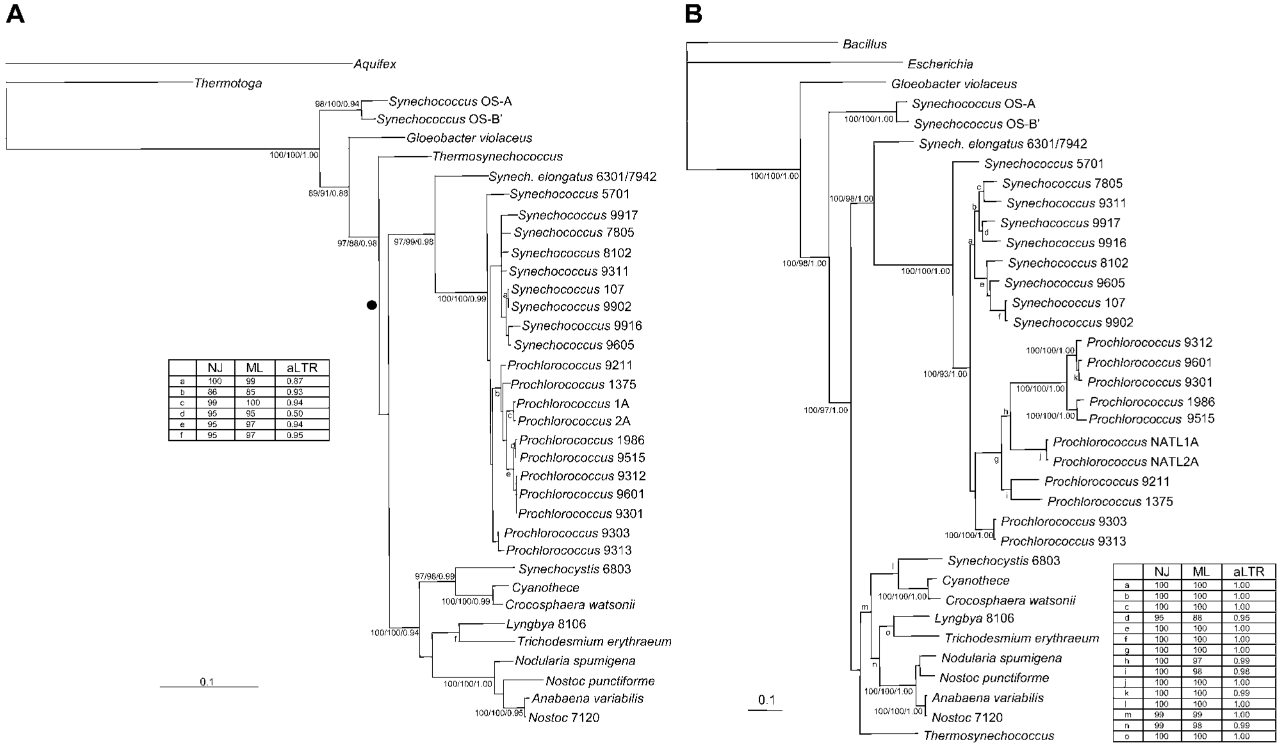
Disruption of the AsnRS Gene in *Synechococcus elongatus* PCC 7942 and Characterization of the Mutants

A 2,200-bp fragment of the *Synechococcus elongatus* PCC 7942 genome containing the *asnS* gene encoding AsnRS was amplified by polymerase chain reaction (PCR) using primers Asn-TRNA-AD-1F (5′-TGCGAGCTCAGCAGCGAT-3′) and Asn-TRNA-AD-1R (5′-ACCTGATGCAGCAGGTCAAA-3′) and cloned in the pTZ57R vector (Fermentas, Burlington, Ontario, Canada) generating plasmid pCA12. The chloramphenicol-resistance cassette C.C1 from pRL178 (Elhai and Wolk 1988) was inserted in pCA12 in the internal *Xmn*I site located 92 bp downstream of the ATG start codon of *asnS* generating plasmids pCA13 or between the two internal *Xmn*I sites of *asnS* generating plasmid pCA14 (with the concomitant deletion of the 287 *Xmn*I fragment). Both plasmids were transformed into *S. elongatus* as described (Golden and Sherman 1984). Integration of the disrupted *asnS* gene and complete segregation of the mutants were addressed by Southern hybridization and PCR (supplementary file 5, Supplementary Material online). Growth curves were carried out under constant illumination (75 μmol photons m^{-2} s^{-1}) in cultures in BG11 medium (Rippka 1988) bubbled with a stream of air/1% CO_2. Growth was monitored by quantification of the chlorophyll content in 1 ml culture volume as described (MacKinney G 1941).

Results and Discussion

The Phylogenetic Tree of Cyanobacteria

In order to have a reference phylogenetic tree for the 35 species used in this study (table 1), we undertook two different approaches, one derived from the alignment of the 16S rRNA sequences and a second from an end-to-end stacking alignments of orthologous proteins (Ciccarelli et al. 2006) currently available from genomic data. Our 16S rRNA–based tree using *Aquifex* and *Thermotoga* as outgroups (fig. 2*A*) was in good agreement with previously reported 16S trees constructed with a wide range of sequences (Honda et al. 1999; Turner et al. 1999; Ochoa de Alda et al. 2005; Hess 2008). An identical topology was observed using *Bacillus* and *Escherichia* as outgroups, except for the position of the root (fig. 2*A*), a phenomenon previously observed in analyses of other rRNA sequences (Cao et al. 1994). Thus, although the topology of most branches was consistent with other work, further analysis was required to increase the accuracy of the tree by reducing stochastic and sampling errors. We undertook an alternative approach based on the concatenation of 25 bona fide orthologous proteins that, after removing gaps and unreliable sites (i.e., those that have undergone many changes over time), generated an alignment of 5,012 residues. Model selection by Prottest indicated that, whatever the criterion used (AIC or BIC), the best evolutionary model to fit these data was the parameter-rich WAG + I + 16Γ + F model that accounts for substitution rate heterogeneity among sites and stationary amino acid frequencies estimated from the data set. Whereas AIC indicated that the model was not overparameterized, the averaged likelihood per site (Seo et al. 2005) of the concatenated sequences was close to the corresponding value obtained from independent COGs, suggesting that the model was not underparameterized (supplementary file 3, Supplementary Material online). Robustness against other systematic errors, such as site saturation or long-branch attraction, which are frequent in concatenated sequence phylogenies (Delsuc et al. 2005), was ascertained by using a Bayesian Markov chain Monte Carlo sampler under the site heterogeneous mixture model CAT + 16Γ + F (Lartillot et al. 2007) and by perturbation analysis (detailed in supplementary file 3, Supplementary Material online). In the latter analyses, trees were constructed from new data sets generated by reducing the number of COGs in the concatenated sequences to remove those showing a compositional bias, or the number of species to remove fast-evolving strains and to account for overrepresentation of closely related sequences. Moreover, in order to account for site saturation, we increased the



	NJ	ML	aLTR
a	100	99	0.87
b	86	85	0.93
c	99	100	0.94
d	95	95	0.50
e	95	97	0.94
f	95	97	0.95

	NJ	ML	aLTR
a	100	100	1.00
b	100	100	1.00
c	100	100	1.00
d	95	88	0.95
e	100	100	1.00
f	100	100	1.00
g	100	100	1.00
h	100	97	0.99
i	100	98	0.98
j	100	100	1.00
k	100	100	0.99
l	100	100	1.00
m	99	99	1.00
n	99	98	0.99
o	100	100	1.00

FIG. 2.—Cyanobacterial species tree. Sequences used for the analysis are indicated in supplementary file 1 (Supplementary Material online). Numbers indicate branch support from NJ/ML/aLRT analyses. Only those with two values above 95% are shown. (*A*) 16S-based tree. The position of the root when *Escherichia* and *Bacillus* were used as outgroups is indicated by a black dot. (*B*) Phylogenetic tree based on the stacking alignments end to end of 25 COGs.

proportion of reliable sites by applying more stringent block selection. Finally, trees were subjected to statistical evaluation (supplementary file 3, Supplementary Material online). Our results indicated that the tree shown in figure 2*B*, constructed with the 25 concatenated COGs and the WAG + I + 16Γ + F model, was robust and was therefore taken as a reference for the phylogenetic relation of the cyanobacterial species used in this study. However, it is important to point out that our analysis does not statistically reject the possibility that *Thermosynechococcus elongatus* diverge immediately after *Synechococcus* OS-A and *Synechococcus* OS-B′.

The tree shown in figure 2*B* is congruent with a tree of 11 cyanobacteria constructed by the analysis of quartets (Zhaxybayeva et al. 2006) and with recent trees based on the concatenated alignment of hundreds of protein families from 24 and 13 genomes, respectively (Shi and Falkowski 2008; Swingley et al. 2008). The position of *T. elongatus* is uncertain in our analyses and is also conflicting among published trees, which may be due to a biased evolution associated with its thermophilic lifestyle. Despite this uncertainty, the topology of the tree shown in figure 2*B* is robust enough to be used as a reference in our study, and it is partially corroborated by several rare genomic changes including putative HGT and gene duplication events that are described below.

Gene Complement for AARS and Homologous Proteins in Cyanobacteria

A genomic survey of genes encoding AARS and homologous proteins was carried out in 35 cyanobacterial genomes (table 1 and table 2). The sequences of AARS and homologous proteins were identified by their annotation combined with Blast analysis (Altschul et al. 1997) and MAFFT annealing (Katoh et al. 2005). All cyanobacterial genomes were found to contain full-length genes in single copy for ArgRS, IleRS, LeuRS, MetRS, ValRS, GluRS, TrpRS, and TyrRS; the α and β subunits of GlyRS, HisRS, ProRS, SerRS, AlaRS, AspRS, and LysRS; and the α and β subunits of PheRS (table 2). All genomes were found to lack a gene for GlnRS, whereas a subset including *Gloeobacter*, *Synechococcus* OS-A, *Synechococcus* OS-B′, *Prochlorococcus*, and marine *Synechococcus* species, also lacked a gene for AsnRS (table 2, see below). Duplicated full-length genes were found for ThrRS in three genomes and for CysRS in one genome (table 2). Truncated ORFs were detected for genes encoding ArgRS, CysRS, ThrRS, AspRS, AlaRS, TyrRS, and TrpRS and the α subunit of GlyRS. Genes for several AARS paralogs were also detected. For instance, every genome contained a *hisZ* gene encoding a protein homologous to HisRS that functions as one of the subunits of the adenosine triphosphate–phosphorybosyl transferase complex catalyzing the first step in the histidine biosynthetic pathway (Sissler et al. 1999). *Synechococcus elongatus* (PCC 7942 and PCC 6301), marine *Synechococcus* species, and some *Prochlorococcus* were found to contain a *yadB* gene (table 2) encoding Glu-Q-RS, a paralogue of GluRS involved in $tRNA^{Asp}$ modification (Campanacci et al. 2004; Dubois et al. 2004; Salazar et al. 2004). ORFs encoding YbaK, a homologue of a ProRS domain that functions in trans hydrolyzing misacylated $Ala\text{-}tRNA^{Pro}$ and $Cys\text{-}tRNA^{Pro}$ (Wong et al. 2003;

Table 2
Gene Complement for AARS and Homologous Proteins in Cyanobacterial Genomes

	R	C	I	L	M	V	Q	E	W	Y	G	H	P	T	S	A	N	D	K	F	HisZ	Glu-Q-RS	YbaK
Gloeobacter violaceus	▨						■	▨				▨		▨			■					■	■
Synechococcus OS-A	▨						■	▨				▨		▨			■					■	■
Synechococcus OS-B’	▨						■	▨				▨		▨			■					■	■
Synech. elongatus 6301/7942	▨						■	▨				▨		▨								▨	■
Synechococcus 5701	▨						■	▨				▨		▨			■					▨	■
Synechococcus 9311	▨						■	▨				▨		▨			■					▨	■
Synechococcus 9917	▨						■	▨				▨		▨			■					▨	■
Synechococcus 9916	▨						■	▨				▨		▨			■					▨	■
Synechococcus 8102	▨						■	▨				▨		▨			■					▨	■
Synechococcus 9605	▨						■	▨				▨		▨			■					▨	■
Synechococcus 107	▨						■	▨				▨		▨			■					▨	■
Synechococcus 9902	▨						■	▨				▨		▨			■					▨	■
Prochlorococcus 9312	▨						■	▨				▨		▨			■					■	■
Prochlorococcus 9601	▨						■	▨				▨		▨			■					■	■
Prochlorococcus 9301	▨						■	▨				▨		▨			■					■	■
Prochlorococcus 1986	▨						■	▨				▨		▨			■					■	■
Prochlorococcus 9515	▨						■	▨				▨		▨			■					■	■
Prochlorococcus NATL1A	▨						■	▨				▨		▨			■					■	■
Prochlorococcus NATL2A	▨						■	▨				▨		▨			■					■	■
Prochlorococcus 9211	▨						■	▨				▨		▨			■					▨	■
Prochlorococcus 1375	▨						■	▨				▨		▨			■					▨	■
Prochlorococcus 9303	▨						■	▨				▨		▨			■					■	■
Prochlorococcus 9313	▨						■	▨				▨		▨			■					■	■
Thermosynechococcus	▨						■	▨				▨		▨								■	■
Synechocystis 6803	▨						■	▨				▨		▨								■	■
Cyanothece	▨	2					■	▨				▨		▨ 2								■	■
Crocosphaera watsonii	▨						■	▨				▨		▨								■	■
Anabaena variabilis	▨						■	▨				▨		▨ 2								■	
Nostoc 7120	▨						■	▨				▨		▨ 2								■	
Nodularia spumigena	▨						■	▨				▨		▨								■	■
Nostoc punctiforme	▨						■	▨				▨		▨								■	
Lyngbya 8106	▨						■	▨				▨		▨								■	■
Trichodesmium erythraeum	▨						■	▨				▨		▨								■	■

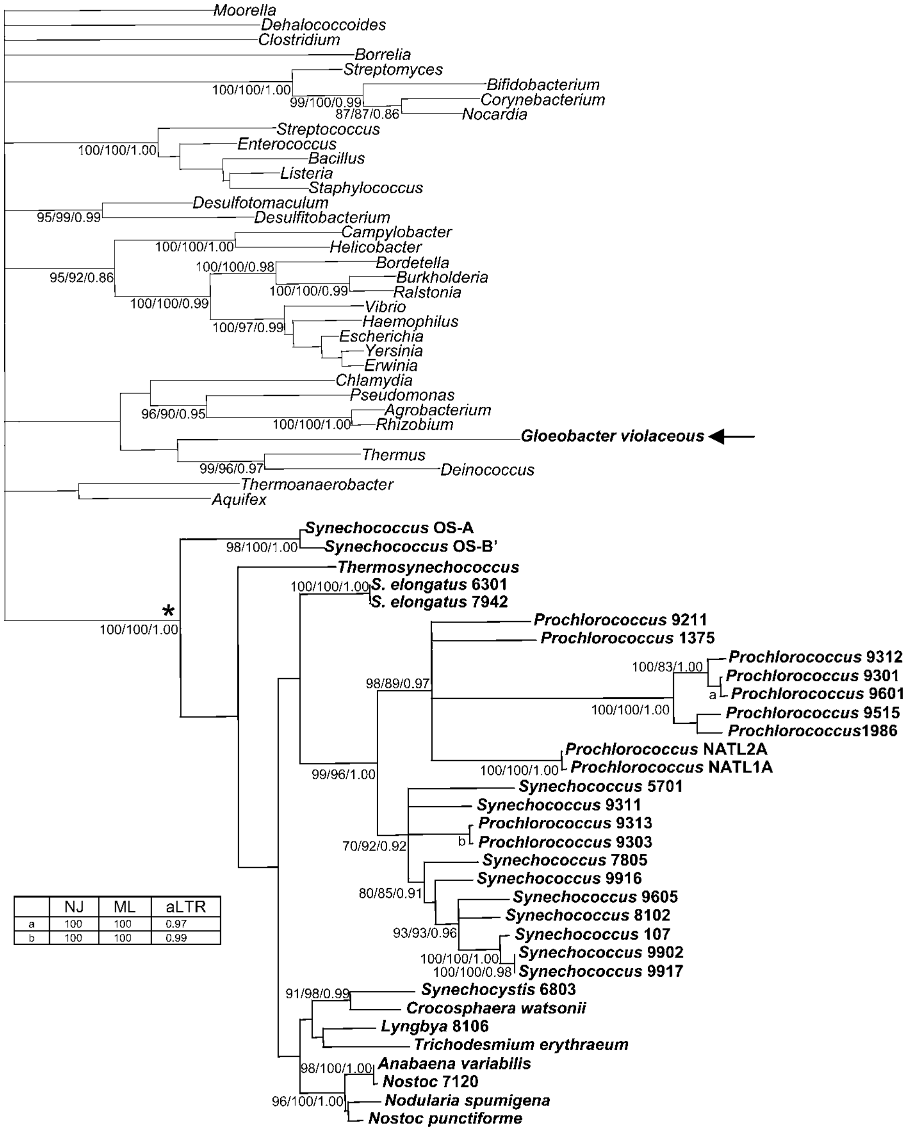
NOTE.—Cyanobacterial species are indicated at the left. AARS are indicated at the top using the one-letter code for their cognate amino acid. G refers to the α and β subunit of GlyRS. F refers to the α and β subunit of PheRS. AARS paralogs are indicated by their complete name. Black cells indicate the absence of a gene for the corresponding AARS in the genome. All other colors indicate that the corresponding gene is present. Gray indicates AARS associated with probable HGT events and that may have a paraphyletic origin in cyanobacteria. Cells with the numeral 2 indicate the existence of two full-length copies of the gene.

An and Musier-Forsyth 2004; Ruan and Soll 2005), were found in the genomes of *Nostoc* sp. PCC 7120, *Anabaena variabilis*, and *N. punctiforme* (table 2).

Sequences of cyanobacterial AARS and homologous proteins were subjected to in-depth phylogenetic analysis. Sequences for most cyanobacterial AARS including CysRS, IleRS, LeuRS, MetRS, ValRS, TrpRS, and TyrRS; the α and β subunits of GlyRS, ProRS, SerRS, AlaRS, AsnRS, AspRS, and LysRS; and the α and β subunits of PheRS clustered together in phylogenetic analyses suggesting a common ancestor for these proteins. Monophyly testing (see Materials and Methods) supported a monophyletic origin for most of these synthetases. However, an alternative topology (nonmonophyletic) could not be rejected for ProRS, TyrRS, and TrpRS or the α subunit of PheRS (supplementary file 4, Supplementary Material online). In contrast, the topology of trees for ArgRS, GluRS, HisRS, and ThrRS suggested a paraphyletic origin for these synthetases in this bacterial group. These observations are supported by statistical tests in which the monophyly of these groups is rejected (see Materials and Methods and supplementary file 4 [Supplementary Material online]). Other anomalies detected included the presence of insertions within GluRS, ValRS, LeuRS, and IleRS in particular cyanobacteria. These cases and some other features of cyanobacterial AARS and homologous proteins are described in detail below.

Absence of GlnRS and/or AsnRS. The Use of the Indirect tRNA^{Gln} and tRNA^{Asn} Aminoacylation Pathways in Cyanobacteria

The absence of GlnRS activity from a cyanobacterium was first reported two decades ago (Schön et al. 1988). This situation seems to be general in the cyanobacterial phylum as every genome sequenced lacks GlnRS-encoding genes (table 2) and contains *gatA*, *gatB*, and *gatC*, indicating that these organisms utilize the indirect pathway for the synthesis of Gln-tRNA^{Gln}. Therefore, the *gltX* gene encoding GluRS, which is found in single copy in every genome examined, must encode a ND-GluRS able to glutamylate both tRNA^{Glu} and



	NJ	ML	aLTR
a	100	100	0.97
b	100	100	0.99

FIG. 3.—Phylogenetic tree based on the sequence of GluRS. Cyanobacterial species are shown in bold. The position of *Gloeobacter violaceus* GluRS is indicated by an arrow. An asterisk indicates the insertion point of a nested subtree. Accession numbers for sequences used in the analysis are indicated in supplementary file 1 (Supplementary Material online). Numbers indicate branch support from NJ/ML/aLRT analyses. Only those with two values above 95% are shown.

tRNA^{Gln}. In the phylogenetic tree shown in figure 3, most cyanobacterial GluRS form a coherent cluster with the exception of *G. violaceus* GluRS which clusters with GluRS from phylogenetically divergent bacteria distantly related to cyanobacteria, suggesting the horizontal acquisition by *Gloeobacter* of a foreign *gltX* gene or the acquisition, after the early diversification of *Gloeobacter*, of a foreign *gltX* by a common ancestor for all other cyanobacteria. These hypotheses were evaluated by a phylogenetic incongruence test (see Materials and Methods) that rejected the monophyly of cyanobacterial GluRSs (*P* values 0.0040 and 0.0000, respectively), supporting the occurrence of a putative HGT event.

Remarkably, some but not all cyanobacterial genomes also lacked a gene encoding an AsnRS (table 2). In these cyanobacteria, the indirect pathway for the synthesis of Asn-tRNA^{Asn} is required to be functional, so their AspRS must thus be nondiscriminating enzymes. The apparent monophyly of cyanobacterial AspRS (supplementary file 4, Supplementary Material online) raised the question of whether AspRS is discriminatory or nondiscriminatory in species also containing AsnRS. To address this issue, we have used insertional mutagenesis to disrupt the gene encoding AsnRS in *S. elongatus* PCC 7942 (see Materials and Methods and supplementary file 5 [Supplementary Material online]). Knock out mutants were viable and grew at the same rate as the wild type under standard growth conditions (supplementary file 5, Supplementary Material online), indicating that AspRS is a nondiscriminating enzyme and both the direct and the indirect pathways for the synthesis of Asn-tRNA^{Asn} are operational in *S. elongatus* and probably in other cyanobacteria. This functional

Table 3
Cyanobacterial Groups according to the Presence of AsnRS and AsnB

Group I—AspRS + AsnRS + AsnB	*Thermosynechococcus* *Crocosphaera watsonii* *Nostoc punctiforme* *Nodularia spumigena*
Group II—AspRS + AsnRS	*Synechocystis elongatus* 6301 *S. elongatus* 7942 *Synechocystis* 6803 *Lyngbya* 8106 *T. erythraeum* *Nostoc* 7120 *Anabaena variabilis*
Group III—AspRS + AsnB	*Prochlorococcus* 1375 *Prochlorococcus* NATL1A *Prochlorococcus* 9312 *Synechococcus* 9605 *Synechococcus* 9902 *Synechococcus* 9311 *Synechococcus* 8102 *Synechococcus* 7805 *Synechococcus* 9917
Group IV—AspRS	*Synechococcus* OS-A *Synechococcus* OS-B′ *Gloeobacter violaceus* *Synechococcus* 5701 *Prochlorococcus* 9211 *Prochlorococcus* NATL2A *Prochlorococcus* 1986 *Prochlorococcus* 9515 *Prochlorococcus* 9601 *Prochlorococcus* 9301 *Prochlorococcus* 9303 *Prochlorococcus* 9313 *Synechococcus* 107 *Synechococcus* 9916

redundancy may have allowed the loss of AsnRS genes in the species where it is currently missing.

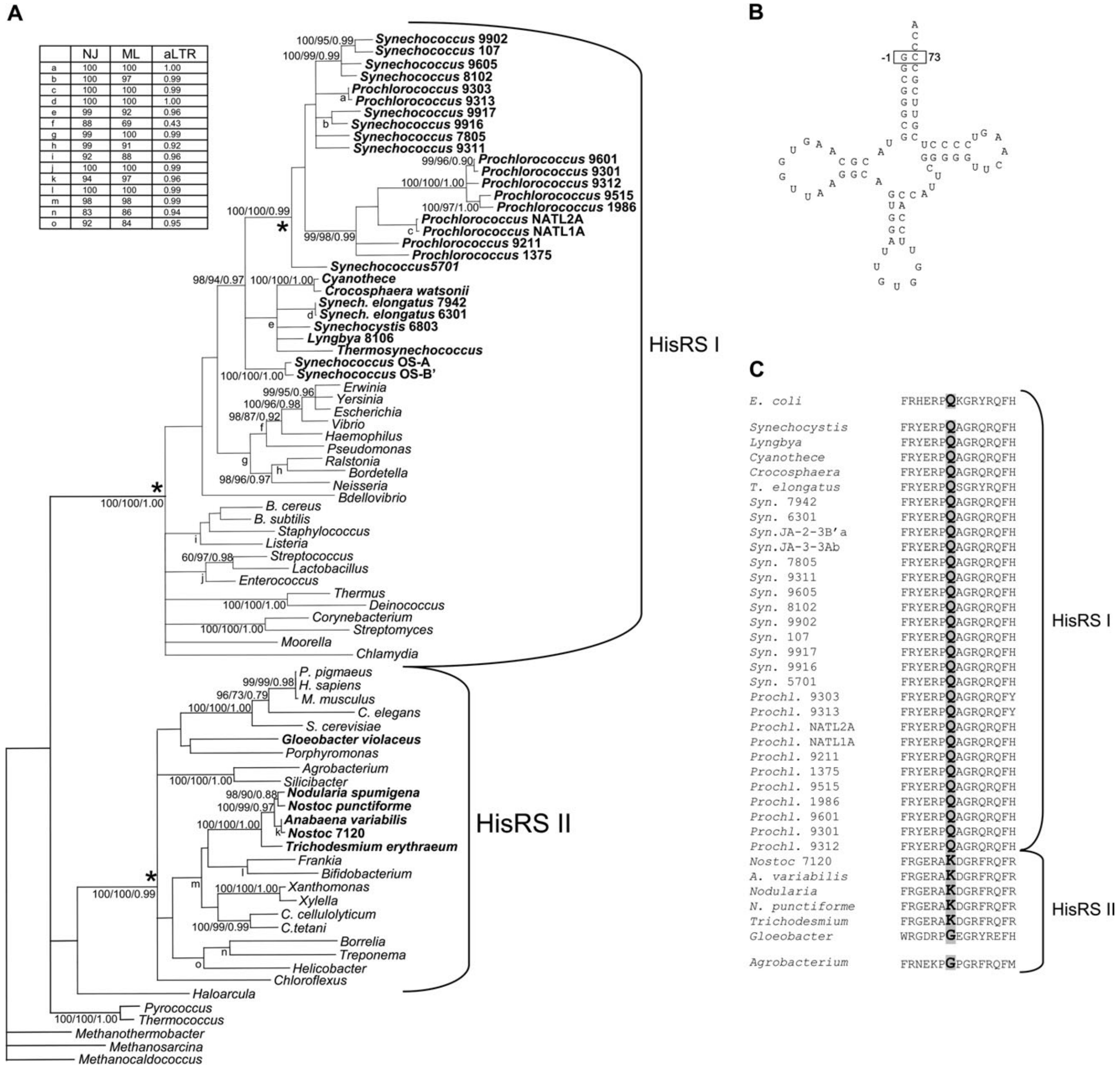
The *asnA* and *asnB* genes encoding the two asparagine synthetases described (Nakamura et al. 1981; Scofield et al. 1990) are missing from *S. elongatus*, which may have accounted for the selective conservation of the indirect aminoacylation pathway as the only route for the synthesis of the amino acid asparagine, a situation that has been described in other prokaryotes (Becker and Kern 1998; Curnow et al. 1998; Min et al. 2002). The *asnA* is absent from cyanobacteria, whereas *asnB* was found in some genomes (table 3). Based on the presence of AsnRS-encoding genes and *asnB*, we were able to classify cyanobacteria into four groups predicted to utilize distinct metabolic pathways for the synthesis of asparagine and of Asn-tRNA$^{\mathrm{Asn}}$ (table 3). In groups II and IV, which lack *asnB*, the indirect aminoacylation route may be the only pathway for Asn biosynthesis. In contrast, Asn can be synthesized by asparagine synthetase in groups I and III. Cyanobacteria of group III are most striking because they are predicted to synthesize Asn as a free amino acid, despite they cannot charge it onto tRNA$^{\mathrm{Asn}}$ due to the lack of AsnRS (tables 2 and 3). Therefore, the Asn synthesized by AsnB in these organisms would be utilized for purposes other than tRNA loading and ribosomal protein synthesis.

Paraphyletic Origin of HisRS and ArgRS

In the phylogenetic tree shown in figure 4*A*, it can be seen that most cyanobacterial HisRS form a cluster with other bacterial HisRS, whereas those from *G. violaceus*, *Trichodesmium erythraeum*, and heterocyst-forming cyanobacteria, a group of nitrogen fixers able to undergo cell differentiation including *Nostoc* sp. PCC 7120, *A. variabilis*, *Nodularia spumigena*, and *N. punctiforme*, are located within a separate cluster with HisRS sequences from eukaryotes and some bacteria. The presence of a eukaryotic-like HisRS in particular eubacteria has been reported and has been attributed to acquisition by HGT (Bond and Francklyn 2000; Ardell and Andersson 2006). Following the nomenclature used by Bond and Francklyn, the bacterial-type HisRS are called HisRS I and the eukaryotic or eukaryotic-like HisRS present in bacteria are called HisRS II. Clustering of *Nostoc* sp. PCC 7120, *A. variabilis*, *N. spumigena*, *N. punctiforme*, and *T. erythraeum* HisRS in the phylogenetic tree indicates a single event of HGT from a eukaryote or from a prokaryote bearing a HisRS II to a common ancestor of these species. *Gloeobacter violaceus* HisRS localizes to a separate branch within the HisRS II cluster suggesting an independent HGT event. Note that the occurrence of HisRS I in *Lyngbya*, which in our reference tree clusters with *Trichodesmium*, indicates that the HGT event is ancient in cyanobacterial evolution. None of the cyanobacterial genomes checked contains coexisting HisRS I and HisRS II, so the acquisition of a foreign HisRS by HGT must have been followed by the loss of the original one. The presence of HisRS I or HisRS II in different cyanobacteria indicates that HisRS has had a paraphyletic origin in this phylum.

In bacteria, the G-1C73 pair of tRNA$^{\mathrm{His}}$ is a major identity element (Yan and Francklyn 1994; Hawko and Francklyn 2001; Connolly et al. 2004). C73 is likely contacted by Gln118 in *E. coli* HisRS (Hawko and Francklyn 2001). Some α-proteobacteria bearing an eukaryotic-like HisRS II, which contains Gly instead of Gln at position 118, have been found to carry a tRNA$^{\mathrm{His}}$ with a eukaryotic identity determinant (i.e., A instead of C at position 73) which has been interpreted as the result of the adaptive evolution of a resident bacterial-type tRNA$^{\mathrm{His}}$ to favor recognition by a foreign eukaryotic HisRS (Ardell and Andersson 2006; Wang et al. 2007). We found that all cyanobacterial tRNA$^{\mathrm{His}}$, including those of strains encoding HisRS II, have a G-1C73 bacterial identity determinant (fig. 4*B*). In all cyanobacterial HisRS I, Gln is also the residue homologous to Gln118 of *E. coli* (fig. 4*C*). Among the cyanobacterial HisRS II, *Nostoc* sp. PCC 7120, *A. variabilis*, *N. spumigena*, *N. punctiforme*, and *T. erythraeum* have Lys at position 118, whereas *G. violaceus* has Gly (fig. 4*C*). The presence of Lys118 suggests that the horizontally acquired HisRS II may have evolved to better recognize a resident tRNA$^{\mathrm{His}}$ with a bacterial identity determinant.

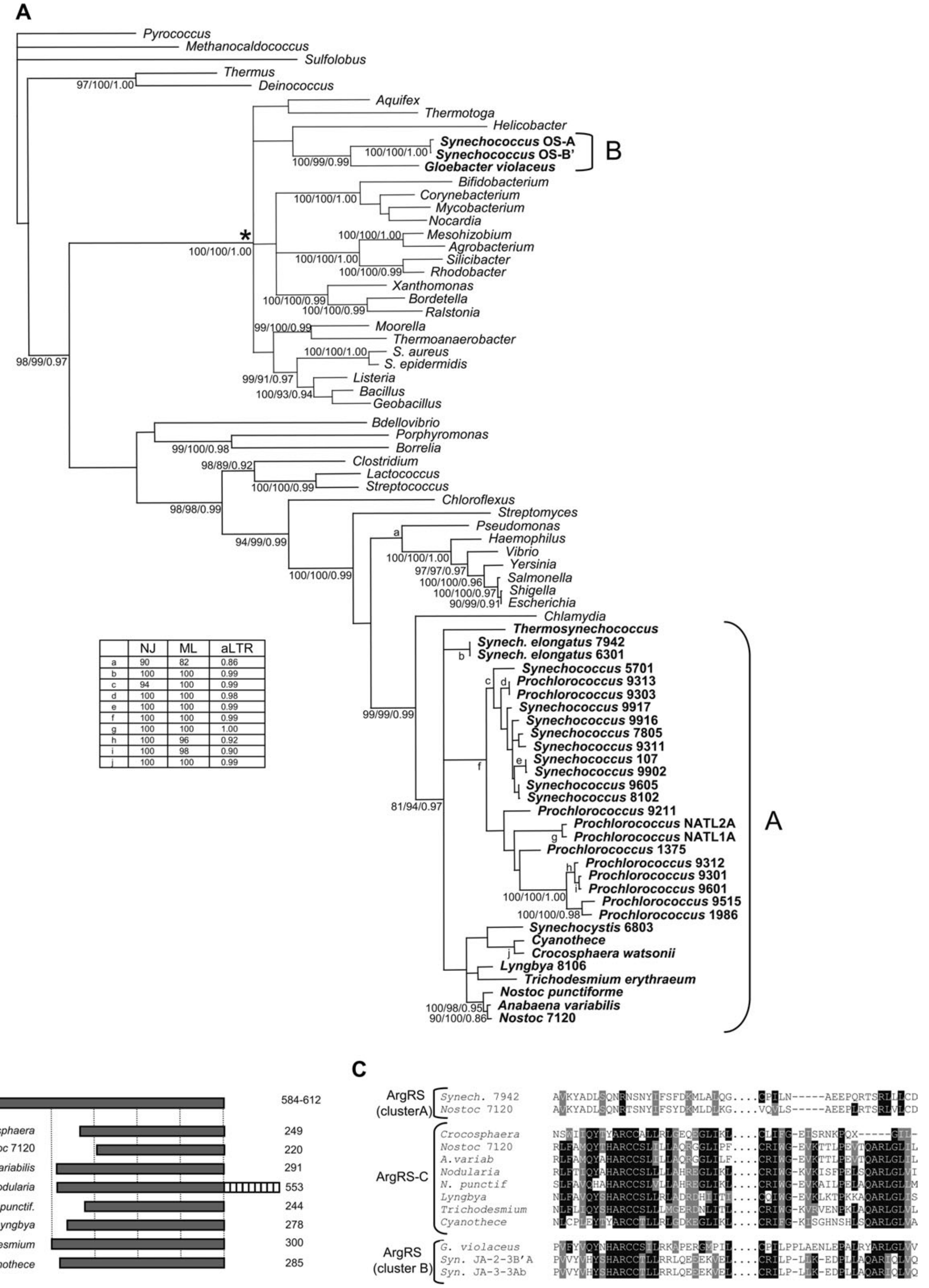
In the phylogenetic tree shown in figure 5*A*, it can be observed that cyanobacterial ArgRS form two distinct groups. Most cyanobacterial ArgRS sequences group in a cluster (cluster A) with ArgRS from proteobacteria and *Chlamydia trachomatis*, whereas ArgRS from *G. violaceus*, *Synechococcus* OS-A, and *Synechococcus* OS-B′ form


A
B
C
100/95/0.99 Synechococcus 9902
100/99/0.99 Synechococcus 107
Synechococcus 9605
Synechococcus 8102
a Prochlorococcus 9303
Prochlorococcus 9313
Synechococcus 9917
b Synechococcus 9916
Synechococcus 7805
Synechococcus 9311
99/96/0.90 Prochlorococcus 9601
Prochlorococcus 9301
100/100/1.00 Prochlorococcus 9312
Prochlorococcus 9515
100/97/1.00 Prochlorococcus 1986
100/100/0.99
Prochlorococcus NATL2A
c Prochlorococcus NATL1A
99/98/0.99 Prochlorococcus 9211
Prochlorococcus 1375
Synechococcus5701
98/94/0.97 100/100/1.00 Cyanothece
Crocosphaera watsonii
Synech. elongatus 7942
d Synech. elongatus 6301
e Synechocystis 6803
Lyngbya 8106
Thermosynechococcus
Synechococcus OS-A
100/100/1.00 Synechococcus OS-B'
99/95/0.96 Erwinia
100/96/0.98 Yersinia
Escherichia
98/87/0.92 Vibrio
f Haemophilus
Pseudomonas
g Ralstonia
h Bordetella
98/96/0.97 Neisseria
Bdellovibrio
*
100/100/1.00
B. cereus
B. subtilis
Staphylococcus
i Listeria
60/97/0.98 Streptococcus
Lactobacillus
j Enterococcus
Thermus
100/100/1.00 Deinococcus
Corynebacterium
100/100/1.00 Streptomyces
Moorella
Chlamydia
HisRS I
P. pigmaeus
99/99/0.98 H. sapiens
96/73/0.79 M. musculus
100/100/1.00 C. elegans
S. cerevisiae
Gloeobacter violaceus
Porphyromonas
Agrobacterium
100/100/1.00 Silicibacter
98/90/0.88 Nodularia spumigena
100/99/0.97 Nostoc punctiforme
100/100/1.00 Anabaena variabilis
k Nostoc 7120
Trichodesmium erythraeum
Frankia
l Bifidobacterium
m 100/100/1.00 Xanthomonas
Xylella
*
100/100/0.99
C. cellulolyticum
100/99/0.99 C.tetani
Borrelia
n Treponema
o Helicobacter
Chloroflexus
HisRS II
Haloarcula
100/100/1.00 Pyrococcus
Thermococcus
Methanothermobacter
Methanosarcina
Methanocaldococcus
-1 G C 73
HisRS I
HisRS II
E. coli FRHERPQKGRYRQFH
Synechocystis FRYERPQAGRQRQFH
Lyngbya FRYERPQAGRQRQFH
Cyanothece FRYERPQAGRQRQFH
Crocosphaera FRYERPQAGRQRQFH
T. elongatus FRYERPQSGRYRQFH
Syn. 7942 FRYERPQAGRQRQFH
Syn. 6301 FRYERPQAGRQRQFH
Syn.JA-2-3B'a FRYERPQAGRQRQFH
Syn.JA-3-3Ab FRYERPQAGRQRQFH
Syn. 7805 FRYERPQAGRQRQFH
Syn. 9311 FRYERPQAGRQRQFH
Syn. 9605 FRYERPQAGRQRQFH
Syn. 8102 FRYERPQAGRQRQFH
Syn. 9902 FRYERPQAGRQRQFH
Syn. 107 FRYERPQAGRQRQFH
Syn. 9917 FRYERPQAGRQRQFH
Syn. 9916 FRYERPQAGRQRQFH
Syn. 5701 FRYERPQAGRQRQFH
Prochl. 9303 FRYERPQAGRQRQFY
Prochl. 9313 FRYERPQAGRQRQFY
Prochl. NATL2A FRYERPQAGRQRQFH
Prochl. NATL1A FRYERPQAGRQRQFH
Prochl. 9211 FRYERPQAGRQRQFH
Prochl. 1375 FRYERPQAGRQRQFH
Prochl. 9515 FRYERPQAGRQRQFH
Prochl. 1986 FRYERPQAGRQRQFH
Prochl. 9601 FRYERPQAGRQRQFH
Prochl. 9301 FRYERPQAGRQRQFH
Prochl. 9312 FRYERPQAGRQRQFH
Nostoc 7120 FRGERAKDGRFRQFR
A. variabilis FRGERAKDGRFRQFR
Nodularia FRGERAKDGRFRQFR
N. punctiforme FRGERAKDGRFRQFR
Trichodesmium FRGERAKDGRFRQFR
Gloeobacter WRGDRPGEGRYREFH
Agrobacterium FRNEKPGPGRFRQFM

	NJ	ML	aLTR
a	100	100	1.00
b	100	97	0.99
c	100	100	0.99
d	100	100	1.00
e	99	92	0.96
f	88	69	0.43
g	99	100	0.99
h	99	91	0.92
i	92	88	0.96
j	100	100	0.99
k	94	97	0.96
l	100	100	0.99
m	98	98	0.99
n	83	86	0.94
o	92	84	0.95

FIG. 4.—(*A*) Phylogenetic tree based on the sequence of HisRS. Cyanobacterial species are shown in bold. Asterisks indicate the insertion point of a nested subtree. Accession numbers for sequences used in the analysis are indicated in supplementary file 1 (Supplementary Material online). Numbers indicate branch support from NJ/ML/aLRT analyses. Only those with two values above 95% are shown. (*B*) Cloverleaf structure of the *Nostoc* 7120 tRNA^{His}. The G-1C73 prokaryotic identity determinant is boxed. (*C*) Alignment of the HisRS region containing amino acid 118. Amino acids at this position are shown in bold on a gray background.

a separate cluster (cluster B) which includes ArgRS from bacteria of diverse phylogenetic affiliation. It is important to note that these three species were the earliest to diverge in cyanobacterial evolution (fig. 2*B*) (Swingley et al. 2008). In the genomes of nine cyanobacteria of cluster A, *Nostoc* sp. PCC 7120, *A. variabilis*, *N. punctiforme*, *T. erythraeum*, *C. watsonii*, *N. spumigena*, *Lyngbya* sp. PCC 8106, and *Cyanothece* sp. CCY0110, we found short ORFs, which we have named *argRS-C*, encoding putative proteins of 220–300 amino acids (except in *Nodularia*, where it is longer due to an unrelated C-terminal extension) with a limited degree of similarity to the C-terminal part of ArgRS (fig. 5*B*). The best hit for these sequences in BlastP analyses are cyanobacterial ArgRS sequences from cluster B. The alignment in figure 5*C* shows that the putative products of these ORFs are more similar to the C-terminus of ArgRS from cluster B than to those of cluster A. A plausible interpretation of these observations is that a foreign ArgRS horizontally acquired after the early divergence of *G. violaceus*, *Synechococcus* OS-A, and *Synechococcus* OS-B′ (see fig. 2) would have taken over displacing the ArgRS originally present in cyanobacteria, *argRS-C* being thus vestigial sequences (pseudogenes) of the original ArgRS gene in cyanobacteria of cluster A. This hypothesis is



	NJ	ML	aLTR
a	90	82	0.86
b	100	100	0.99
c	94	100	0.99
d	100	100	0.98
e	100	100	0.99
f	100	100	0.99
g	100	100	1.00
h	100	96	0.92
i	100	98	0.90
j	100	100	0.99

FIG. 5.—(*A*) Phylogenetic tree based on the sequence of ArgRS. Cyanobacterial species are shown in bold. An asterisk indicates the insertion point of a nested subtree. Accession numbers for sequences used in the analysis are indicated in supplementary file 1 (Supplementary Material online). Numbers indicate branch support from NJ/ML/aLRT analyses. Only those with two values above 95% are shown. (*B*) Schematic representation of ArgRS and ArgRS-C. The bars representing proteins are not shown to scale. Vertical dotted lines connect regions showing sequence similarity. (*C*) Sequence alignment of ArgRS-C and ArgRS sequences from clusters A and B. Dots represent a tract of approximately 60–80 amino acids not depicted in the alignment. Conserved residues are highlighted.

supported by our phylogenetic incongruence test in which ArgRS monophyly is rejected (*P* value 0.0000). Alternatively, the two ArgRS genes might have arisen by an ancient gene duplication event thereafter being one or the other selectively lost in different species throughout evolution.

Horizontal Acquisition and Duplication of ThrRS Genes

Threonyl-tRNA synthetases often contain an editing domain that varies according to their origin, and this has allowed their classification into different groups (Dock-Bregeon et al. 2000; Korencic et al. 2004). All cyanobacterial ThrRS are of the eubacterial type and contain an N1–N2 editing domain. In the phylogenetic tree shown in figure 6, it can be seen that whereas most cyanobacterial ThrRS cluster together in a coherent group, ThrRS from the *Prochlorococcus* species cluster with ThrRS from γ-proteobacteria, which are among the most abundant bacteria in the oceans, indicating a possible lateral acquisition of a ThrRS-encoding gene after diversification of the *Prochlorococcus* genus. This hypothesis has been proposed by other researchers (Zhaxybayeva et al. 2006) and is supported by our phylogenetic incongruence test (*P* value 0.0000) in which monophyly of cyanobacterial ThrRS sequences was rejected. Further support to this was found by comparing the length of the N1 editing subdomain which is approximately 60 amino acids long in the *Prochlorococcus* enzyme, similar to the length in γ-proteobacteria, whereas in other cyanobacterial ThrRS, it is 14–28 amino acids.

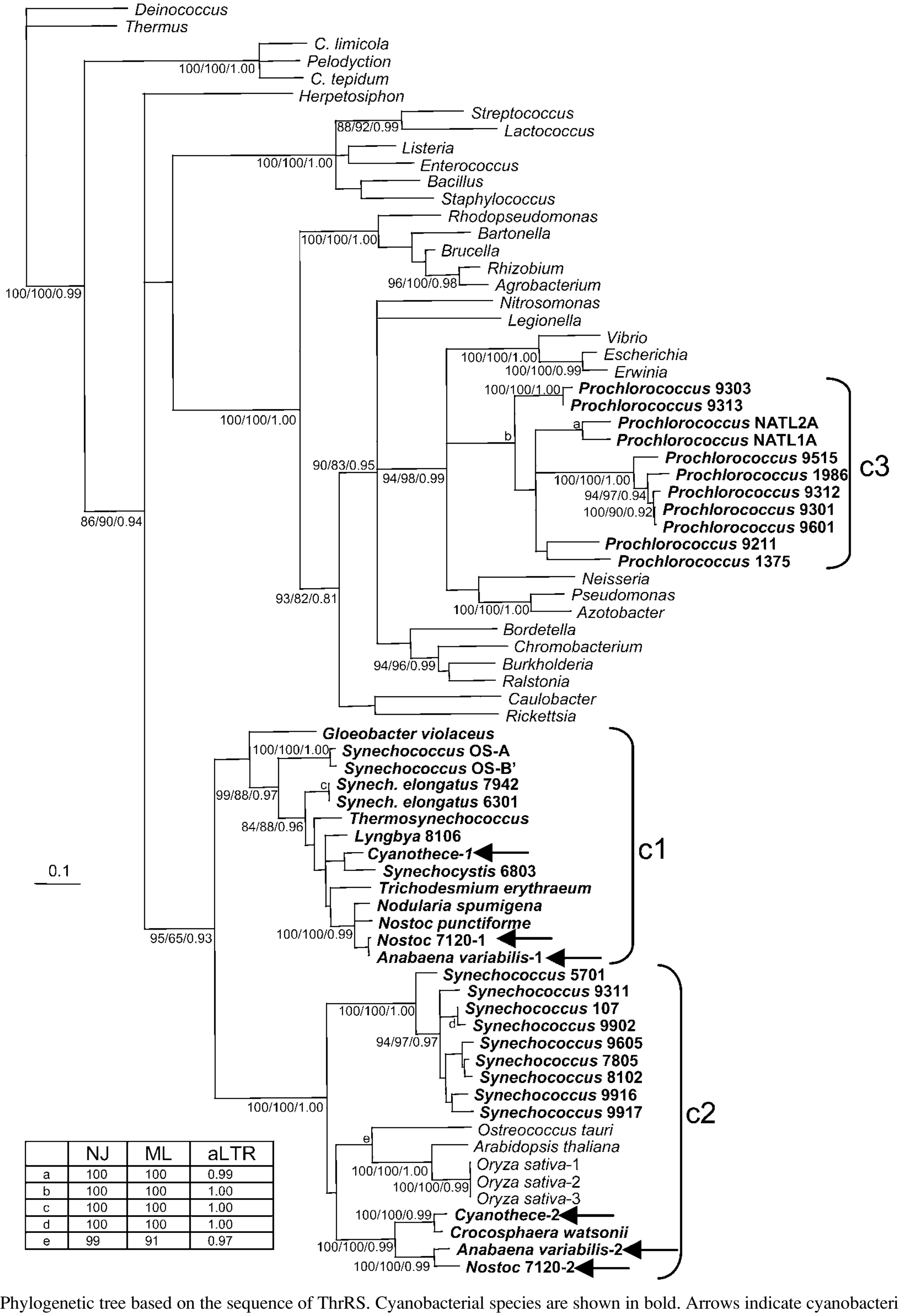
Cyanobacterial ThrRS, excluding *Prochlorococcus* ThrRS, clustered together in our phylogenetic tree; however, the topology of this cluster is very different from that of our reference tree (figs. 2*B* and 6). Early diversification of ThrRS in two major branches leads to two groups that we have named ThrRS-c1 and ThrRS-c2. Each cyanobacterium (excluding *Prochlorococcus*) contains one ThrRS gene that may encode either a ThrRS-c1 or a ThrRS-c2, except for *Nostoc* sp. PCC 7120, *A. variabilis*, and *Cyanothece* which contain one of each type. ThrRS-c1 and ThrRS-c2 diverge mainly in the C-terminal part between amino acids 234 and 449 (numbering is as in the ThrRS-c1 from *Nostoc* sp. PCC 7120) which includes the catalytic domain. Conservation of many columns specific for each class in sequence alignments and the presence of two specific indels (data not shown) reinforce the existence of two distinct classes of ThrRS and support the topology of the tree shown in figure 6. ThrRS from both classes have a high level of sequence similarity (∼50% identities), and they are more similar to each other than to other ThrRS. This suggests that an ancient gene duplication event within the cyanobacterial phylum originated ThrRS-c1 and ThrRS-c2, most species having lost one of the two redundant genes throughout evolution. Gene duplication and linage specific gene loss could give rise to hidden paralogy. In the case of *Prochlorococcus*, the acquisition of a foreign ThrRS by HGT may have allowed the loss of both ThrRS-c1 and ThrRS-c2. Algae and plant chloroplasts would have inherited ThrRS-c2 (fig. 6). The conservation of duplicated redundant genes over evolutionary time is unlikely unless they confer an advantage to their host, either by the increased gene dosage or by the evolutionary acquisition of asymmetry; that is, different efficacy, different expression profiles, different susceptibility to inhibitors etc. (Krakauer and Nowak 1999). The products of the two ThrRS genes coexisting in *Nostoc* sp. PCC 7120, *A. variabilis*, and *Cyanothece* are apparently functional as deduced from the conservation of motifs important for activity. It is probable that they have acquired slightly different functionality. In *E. coli*, ThrRS is a homodimer (Sankaranarayanan et al. 1999). A reason that may have accounted for the conservation of two ThrRS genes could be the possibility of forming heterodimers that were functionally more efficient than homodimers. Experimental tests of these hypotheses are being done in an attempt to understand the reasons for the selective conservation of duplicated ThrRS genes.

Duplication of a CysRS Gene

Cyanobacterial CysRS cluster together in our phylogenetic tree (fig. 7*A*) and appear to have had a monophyletic origin according to our analysis (supplementary file 4, Supplementary Material online). Besides, the topology of the tree based on the sequence of CysRS is in general coherent with our reference tree (figs. 2*B* and 7*A*). However, the presence of two full-length genes encoding CysRS in the *Cyanothece* genome is remarkable. Their putative products CysRS-1 and CysRS-2 are highly similar (63% identities). *Crocosphaera watsonii* that is closely related to *Cyanothece* contains a single full-length CysRS highly similar to *Cyanothece* CysRS-2 (92% identities) and a short ORF, that we named *orf61* (GI|67920768), putatively encoding a protein of 61 residues, which seems to be vestigial for the presence of an ancient CysRS-1 encoding gene that had undergone an internal deletion of approximately 400 amino acids (fig. 7*B* and *C*). Another important observation is that CysRS from genera phylogenetically related to *Cyanothece* and *C. watsonii* tend to form two branches, grouping either with CysRS-1 or with CysRS-2 (clusters 1 and 2 in fig. 7, respectively). These observations are consistent with a gene duplication event that took place in a common ancestor of these cyanobacteria yielding genes encoding CysRS-1 and CysRS-2. However, it is important to note that there are not many sequence features that differentiate CysRS of clusters 1 and 2. On the contrary, all these sequences are highly similar (63–69% identities), which may indicate that gene duplication occurred shortly before speciation, so that the two CysRS genes may have diverged mostly within the distinct lineages. Besides, the high level of sequence similarity may have permitted homologous recombination between both genes, which could have contributed to blurring the differences between CysRS-1 and CysRS-2. Thus, the position of *Trichodesmium* CysRS outside clusters 1 and 2 might result from extensive recombination between the two copies.

A Putative Novel Domain in Some Cyanobacterial AARS

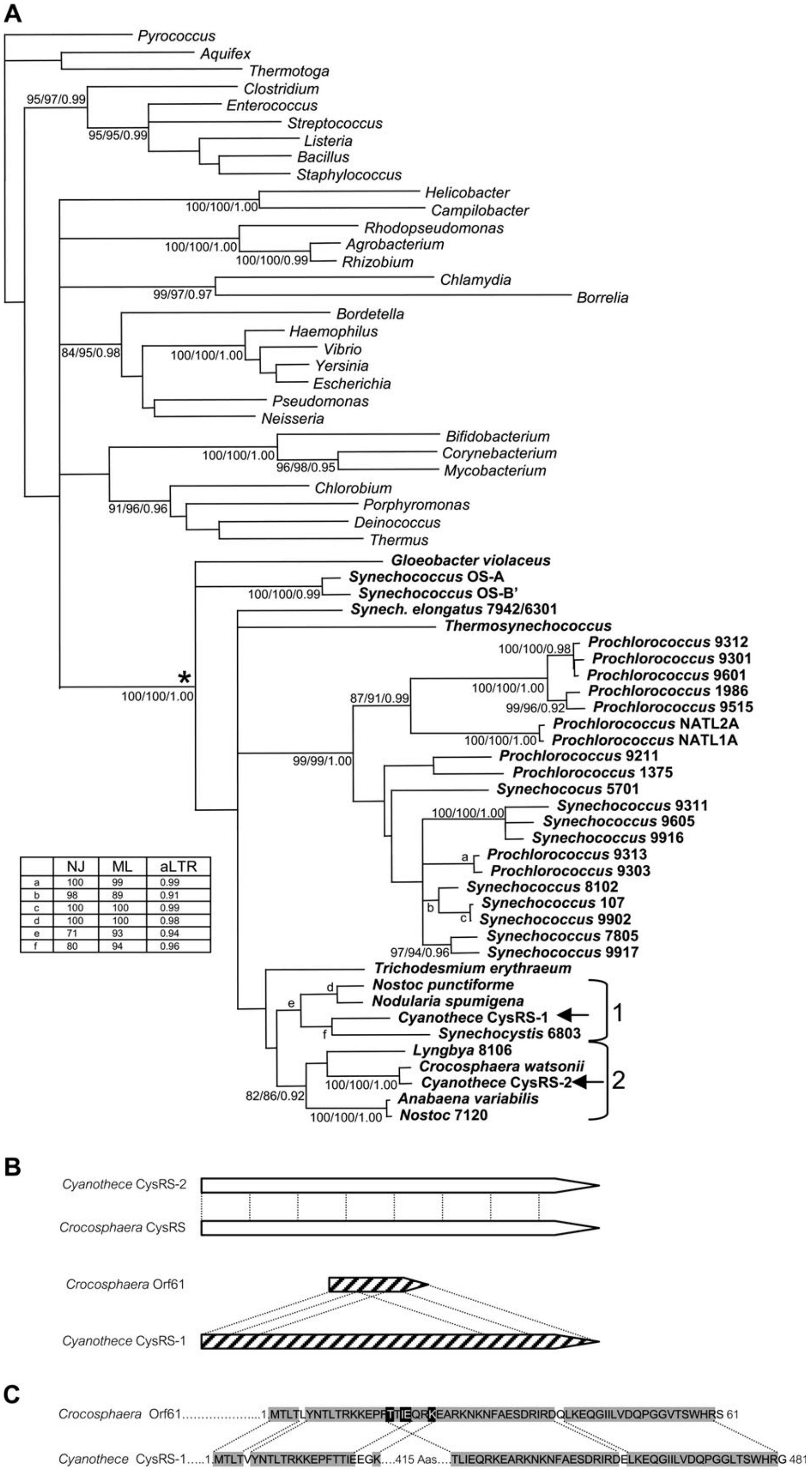
The GluRS encoded in the *T. erythraeum* genome is anomalously long (881 amino acids vs. 476–530 amino



	NJ	ML	aLTR
a	100	100	0.99
b	100	100	1.00
c	100	100	1.00
d	100	100	1.00
e	99	91	0.97

FIG. 6.—Phylogenetic tree based on the sequence of ThrRS. Cyanobacterial species are shown in bold. Arrows indicate cyanobacteria that contain two ThrRS. Accession numbers for sequences used in the analysis are indicated in supplementary file 1 (Supplementary Material online). Numbers indicate branch support from NJ/ML/aLRT analyses. Only those with two values above 95% are shown.

acids of most GluRS) due to a long C-terminal extension that includes two subregions (fig. 8). The first one contains 7 repeats of a 28 amino acid–long sequence, plus an additional truncated repeat, whereas the second C-terminal subregion contains 137 amino acids and bears two putative transmembrane helices. Strikingly, insertions exhibiting a limited but recognizable sequence similarity with the C-terminal subregion of *T. erythraeum* GluRS and bearing two putative transmembrane helices were found in ValRS from heterocyst-forming cyanobacteria (*Nostoc* sp. PCC



	NJ	ML	aLTR
a	100	99	0.99
b	98	89	0.91
c	100	100	0.99
d	100	100	0.98
e	71	93	0.94
f	80	94	0.96

FIG. 7.—(*A*) Phylogenetic tree based on the sequence of CysRS. Cyanobacterial species are shown in bold. Arrows indicate cyanobacteria that contain two CysRS. Accession numbers for sequences used in the analysis are indicated in supplementary file 1 (Supplementary Material online). Numbers indicate branch support from NJ/ML/aLRT analyses. Only those with two of the tree values above 95% are shown. (*B*) Comparison of CysRS and CysRS-like genes in *Cyanothece* and *Crocosphaera watsonii*. Dotted lines connect regions showing sequence similarity. (*C*) Sequence comparison of *Cyanothece* CysRS-1 and *Crocosphaera* Orf61. Dotted lines connect similar sequences. Conserved residues are highlighted in gray. Residues of Orf61 similar to both N-terminal and C-terminal sequences of CysRS-1 are highlighted in black. The position of 415 amino acids of *Cyanothece* CysRS-1 not depicted here is shown.

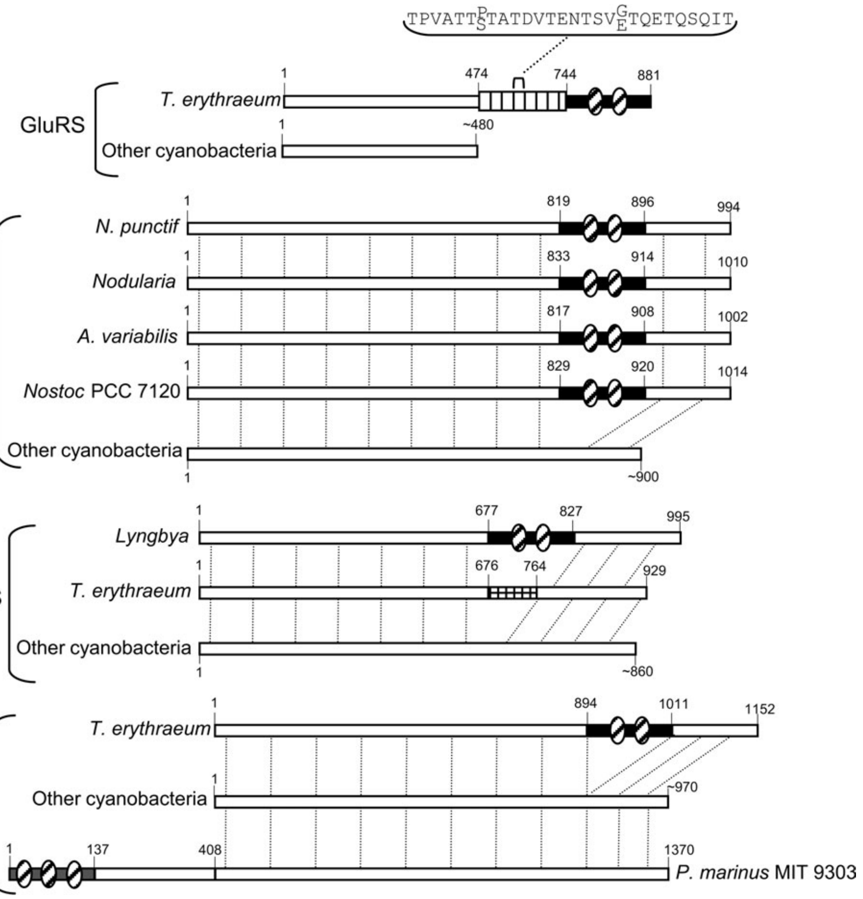


FIG. 8.—Cyanobacterial AARS containing a nonrelated sequence insertion. AARS are represented as bars. CAAD is shown in black. Putative transmembrane helices are depicted as broken-line ovals. Repetitive sequences in *Trichodesmium erythraeum* GluRS are shown as contiguous boxes. An internal insertion in *T. erythraeum* LeuRS is indicated as a gridded segment. An N-terminal extension to *Prochlorococcus marinus* MIT 9303 IleRS is shown, a subregion homologous to an ORF found upstream in other Prochlorococcus strains is depicted in gray. Dotted lines connect regions with sequence similarity.

7120, *A. variabilis*, *N. punctiforme* and *N. spumigena*), LeuRS from *Lyngbya*, and IleRS from *T. erythraeum* (fig. 8). The presence of an unrelated insertion in LeuRS from *T. erythraeum* at the same position as in *Lyngbya* indicates that extra amino acids in such region are probably not deleterious for this synthetase (fig. 8). ORFs encoding proteins 100–200 amino acids long homologous to this region and all containing two putative transmembrane domains were found in cyanobacteria and plants. One of these is the photosystem I peripheral protein TMP14 (thylacoid-associated protein 14 also known as PSI-P) from *Arabidopsis thaliana* chloroplasts (Hansson and Vener 2003; Khrouchtchova et al. 2005). No other protein was found to bear a similar domain. This region appended to some cyanobacterial class I AARS may constitute a novel domain that we have named CAAD (for cyanobacterial AARS appended domain). The origin and evolution of modular proteins, including AARS, is often associated with the gain or loss of particular domains, a phenomenon known as domain shuffling (Doolittle 1995; Ostermeier and Benkovic 2001; Schmidt and Davies 2007). The presence of CAAD in four distinct AARS indicates four independent events of domain recruitment. Most domains appended to the catalytic and anticodon-binding domains of AARS throughout evolution are involved in tRNA binding or in editing functions (Frugier et al. 2000; Kaminska et al. 2000; Francin et al. 2002; Deniziak et al. 2007). We do not know whether CAAD has a role in AARS function, but the presence of transmembrane helices suggests that it could mediate the association of AARS with the thylacoid or the plasma membrane. To our knowledge, there are no published data showing association of AARS with membranes, an issue that we are currently testing by experimental approaches. It would be worthwhile to check if recruitment of CAAD would have had any effect on the activity of the synthetases that carry it. It is worth noting that IleRS from *P. marinus* MIT9303 bears an N-terminal extension that, although sequence unrelated to CAAD, contains two or three putative transmembrane helices. This region has probably been recruited by fusion with the ORF located just upstream in all other *Prochlorococcus* species (fig. 8).

Presence of Glu-Q-RS of Diverse Classes in Cyanobacterial Species

Some cyanobacterial genomes contain a gene encoding a GluRS-like protein named Glu-Q-RS. These polypeptides are homologous to GluRS but are considerably shorter (293–316 amino acids vs. 476–530 of GluRS) lacking the C-terminal anticodon-binding region (Salazar et al. 2001, 2004; Campanacci et al. 2004; Dubois et al. 2004). Recent investigations have shown that Glu-Q-RS (also termed YadB in *E. coli*) is involved in tRNA^{Asp} modification by glutamylation (Dubois et al. 2004; Salazar et al. 2004; Blaise et al. 2005).

Genes encoding Glu-Q-RS are present in some genera with a scattered distribution in the cyanobacterial radiation, like marine *Synechococcus* species, some but not all *Prochlorococcus* and two closely related freshwater strains, *S. elongatus* PCC 7942 and *S. elongatus* PCC 6301 (table 2). Dubois et al. have observed the existence of three subgroups of Glu-Q-RS characterized by conserved specific sequences at the position of the HIGH motif of class I AARS. Glu-Q-RS from *Prochlorococcus* and marine *Synechococcus* species bear the HxGN sequence and belongs to subgroup 2, whereas Glu-Q-RS from *S. elongatus* has the HxGS sequence and belongs to subgroup 1, suggesting a paraphyletic origin for Glu-Q-RS in this phylum (monophyly *P* value 0.0000).

Glu-Q-RS does not seem to be an essential enzyme in cyanobacteria as a group, but it may confer a selective advantage to the genera that have it. We have been able to observe that *yadB* is transcribed in *S. elongatus* PCC 7942 (data not shown). The presence of a Glu-Q-RS-encoding gene in *P. marinus* MIT 9211 and in *P. marinus* CCMP1375 (SS120) is particularly striking because their genomes have been subjected to an evolutionary process of compaction that has led to the loss of many nonessential genes (Rocap et al. 2003; García-Fernandez et al. 2004; Dufresne et al. 2005).

Conclusion

We have carried out an extensive genomic survey of the complement of genes encoding AARS and homologous proteins in cyanobacteria. We have found that the set of AARS-encoding genes vary from one cyanobacteria to another due to the lack of some genes or the presence of duplicated genes. Moreover, genes that are present in all cyanobacteria do not always show a harmonious phylogeny. GluRS, HisRS, ArgRS, and ThrRS show evidences of a paraphyletic origin likely due to interphylum horizontal transfers, which adds further support to the idea of AARS being somehow prone to HGT (Brown and Doolittle 1999; Woese et al. 2000; Brown et al. 2003; Beiko et al. 2005; Dohm et al. 2006; Zhaxybayeva et al. 2006).

Our analyses have uncovered two putative cases of early gene duplication in cyanobacteria involving ThrRS and CysRS genes. Duplicated ThrRS- and CysRS-encoding genes are estimated to have been present in some cyanobacterial species for billion years. Gene duplication events are considered to be key turning points in evolution generating a functional redundancy that allows one of the duplicated genes to “freely” evolve by genetic drift, accumulating mutations that either inactivate it or, more rarely, lead to the acquisition of new functions (Nowak et al. 1997; Prince and Pickett 2002). Although evolutionary conservation of functionally relevant motifs suggests that duplicated ThrRS and CysRS genes are functional, experimental approaches would be required to demonstrate it and to assess whether duplicated genes have acquired functional differences. A putative novel protein domain, termed CAAD, appended to some class I AARS in particular cyanobacterial genera has been identified by our analyses. Although its function is unknown, its independent acquisition by distinct synthetases strongly opposes the possibility of representing neutral insertions. A most interesting feature in this domain is the presence of two putative transmembrane helices, which may suggest a possible association of AARS with membranes, a phenomenon not described so far for any prokaryotic or eukaryotic AARS.

Supplementary Material

Supplementary files 1–5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by grants BMC2003-06686 and BFU2007-66589/BMC from Ministerio de Educación y Ciencia (Spain), SEK7/04 from Junta de Castilla y León, a program in support of research of IE Universidad and a Proyecto Intramural Especial from Consejo Superior de Investigaciones Científicas. We are indebted to Beatriz Rosón for COGs retrieval, alignment, and block selection; to Federico Abascal for his help using Prottest 1,3; and to Antonia Herrero, Agustín Vioque, Javier Castresana, Naiara Rodríguez-Ezpeleta, Hernán Dopazo, and Jean Houmard for helpful comments on the manuscript. We are also indebted to Antonia Herrero and Enrique Flores for constant encouragement during the course of this work.

Literature Cited

Abascal F, Zardoya R, Posada D. 2005. Prottest: selection of best-fit models of protein evolution. Bioinformatics. 21: 2104–2105.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

An S, Musier-Forsyth K. 2004. Trans-editing of Cys-tRNA^{Pro} by *Haemophilus influenzae* YbaK protein. J Biol Chem. 279: 42359–42362.

Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol. 55:539–552.

Ardell DH, Andersson SG. 2006. TFAM detects co-evolution of tRNA identity rules with lateral transfer of histidyl-tRNA synthetase. Nucleic Acids Res. 34:893–904.

Beauchemin N, Larue B, Cedergren RJ. 1973. The characterization of the RNAs and aminoacyl-tRNA synthetases of the blue-green alga *Anacystis nidulans*. Arch Biochem Biophys. 156:17–25.

Becker HD, Kern D. 1998. *Thermus thermophilus*: a link in evolution of the tRNA-dependent amino acid amidation pathways. Proc Natl Acad Sci USA. 95:12832–12837.

Beiko RG, Harlow TJ, Ragan MA. 2005. Highways of gene sharing in prokaryotes. Proc Natl Acad Sci USA. 102: 14332–14337.

Blaise M, Becker HD, Lapointe J, Cambillau C, Giege R, Kern D. 2005. Glu-Q-tRNA(Asp) synthetase coded by the *yadB* gene, a new paralog of aminoacyl-tRNA synthetase that glutamylates tRNA^{Asp} anticodon. Biochimie. 87:847–861.

Bond JP, Francklyn C. 2000. Proteobacterial histidine-biosynthetic pathways are paraphyletic. J Mol Evol. 50:339–347.

Brocks JJ, Logan GA, Buick R, Summons RE. 1999. Archean molecular fossils and the early rise of eukaryotes. Science. 285:1033–1036.

Brown JR, Doolittle WF. 1999. Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. J Mol Evol. 49:485–495.

Brown JR, Gentry D, Becker JA, Ingraham K, Holmes DJ, Stanhope MJ. 2003. Horizontal transfer of drug-resistant aminoacyl-transfer-RNA synthetases of anthrax and Gram-positive pathogens. EMBO Rep. 4:692–698.

Campanacci V, Dubois DY, Becker HD, et al. (14 co-authors). 2004. The *Escherichia coli yadB* gene product reveals a novel aminoacyl-tRNA synthetase like activity. J Mol Biol. 337: 273–283.

Cao Y, Adachi J, Hasegawa M. 1994. Eutherian phylogeny as inferred from mitochondrial DNA sequence data. Jpn J Genet. 69:455–472.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17:540–552.

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. 2006. Toward automatic reconstruction of a highly resolved tree of life. Science. 311:1283–1287.

Clamp M, Cuff J, Searle SM, Barton GJ. 2004. The Jalview Java alignment editor. Bioinformatics. 20:426–427.

Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM, Tiedje JM. 2007. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. Nucleic Acids Res. 35:D169–D172.

Connolly SA, Rosen AE, Musier-Forsyth K, Francklyn CS. 2004. G-1:C73 recognition by an arginine cluster in the active site of *Escherichia coli* histidyl-tRNA synthetase. Biochemistry. 43:962–969.

Curnow AW, Hong K, Yuan R, Kim S, Martins O, Winkler W, Henkin TM, Soll D. 1997. Glu-tRNA^{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. Proc Natl Acad Sci USA. 94:11819–11826.

Curnow AW, Ibba M, Soll D. 1996. tRNA-dependent asparagine formation. Nature. 382:589–590.

Curnow AW, Tumbula DL, Pelaschier JT, Min B, Soll D. 1998. Glutamyl-tRNA^{Gln} amidotransferase in *Deinococcus radiodurans* may be confined to asparagine biosynthesis. Proc Natl Acad Sci USA. 95:12838–12843.

Chevenet F, Brun C, Banuls AL, Jacq B, Christen R. 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics. 7:439.

Daubin V, Ochman H. 2004. Quartet mapping and the extent of lateral transfer in bacterial genomes. Mol Biol Evol. 21: 86–89.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet. 6:361–375.

Deniziak M, Sauter C, Becker HD, Paulus CA, Giege R, Kern D. 2007. *Deinococcus* glutaminyl-tRNA synthetase is a chimer between proteins from an ancient and the modern pathways of aminoacyl-tRNA formation. Nucleic Acids Res. 35:1421–1431.

Diaz-Lazcoz Y, Aude JC, Nitschke P, Chiapello H, Landes-Devauchelle C, Risler JL. 1998. Evolution of genes, evolution of species: the case of aminoacyl-tRNA synthetases. Mol Biol Evol. 15:1548–1561.

Dock-Bregeon A, Sankaranarayanan R, Romby P, Caillet J, Springer M, Rees B, Francklyn CS, Ehresmann C, Moras D. 2000. Transfer RNA-mediated editing in threonyl-tRNA synthetase. The class II solution to the double discrimination problem. Cell. 103:877–884.

Dohm JC, Vingron M, Staub E. 2006. Horizontal gene transfer in aminoacyl-tRNA synthetases including leucine-specific subtypes. J Mol Evol. 63:437–447.

Doolittle RF. 1995. The multiplicity of domains in proteins. Annu Rev Biochem. 64:287–314.

Doolittle RF, Handy J. 1998. Evolutionary anomalies among the aminoacyl-tRNA synthetases. Curr Opin Genet Dev. 8: 630–636.

Dubois DY, Blaise M, Becker HD, Campanacci V, Keith G, Giege R, Cambillau C, Lapointe J, Kern D. 2004. An aminoacyl-tRNA synthetase-like protein encoded by the *Escherichia coli yadB* gene glutamylates specifically tRNA^{Asp}. Proc Natl Acad Sci USA. 101:7530–7535.

Dufresne A, Garczarek L, Partensky F. 2005. Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biol. 6:R14.

Dufresne A, Salanoubat M, Partensky F, et al. (18 co-authors). 2003. Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. Proc Natl Acad Sci U S A. 100:10020–10025.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797.

Elhai J, Wolk CP. 1988. A versatile class of positive-selection vectors based on the nonviability of palindrome-containing plasmids that allows cloning into long polylinkers. Gene. 68: 119–138.

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

Felsenstein J. 1989. Phylogeny Inference Package (Version 3.2). Cladistics. 5:164–166.

Feng L, Sheppard K, Namgoong S, Ambrogelly A, Polycarpo C, Randau L, Tumbula-Hansen D, Soll D. 2004. Aminoacyl-tRNA synthesis by pre-translational amino acid modification. RNA Biol. 1:16–20.

Francin M, Kaminska M, Kerjan P, Mirande M. 2002. The N-terminal domain of mammalian Lysyl-tRNA synthetase is a functional tRNA-binding domain. J Biol Chem. 277:1762–1769.

Frugier M, Moulinier L, Giege R. 2000. A domain in the N-terminal extension of class IIb eukaryotic aminoacyl-tRNA synthetases is important for tRNA binding. EMBO J. 19:2371–2380.

Gagnon Y, Lacoste L, Champagne N, Lapointe J. 1996. Widespread use of the Glu-tRNA^{Gln} transamidation pathway among bacteria. A member of the alpha purple bacteria lacks glutaminyl-tRNA synthetase. J Biol Chem. 271:14856–14863.

García-Fernandez JM, de Marsac NT, Díez J. 2004. Streamlined regulation and gene loss as adaptive mechanisms in *Prochlorococcus* for optimized nitrogen utilization in oligotrophic environments. Microbiol Mol Biol Rev. 68:630–638.

Gogarten JP, Townsend JP. 2005. Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol. 3: 679–687.

Golden SS, Sherman LA. 1984. Optimal conditions for genetic transformation of the cyanobacterium *Anacystis nidulans* R2. J Bacteriol. 158:36–42.

Goldman N, Anderson JP, Rodrigo AG. 2000. Likelihood-based tests of topologies in phylogenetics. Syst Biol. 49:652–670.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52:696–704.

Hansson M, Vener AV. 2003. Identification of three previously unknown in vivo protein phosphorylation sites in thylakoid membranes of *Arabidopsis thaliana*. Mol Cell Proteomics. 2:550–559.

Hawko SA, Francklyn CS. 2001. Covariation of a specificity-determining structural motif in an aminoacyl-tRNA synthetase and a tRNA identity element. Biochemistry. 40: 1930–1936.

Hess WR. 2008. Comparative genomics of marine cyanobacteria and their phages. In: Herrero A, Flores E, editors. The cyanobacteria. Molecular biology, genomics and evolution. Norfolk (UK): Caister Academic Press. p. 89–116.

Honda D, Yokota A, Sugiyama J. 1999. Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine Synechococcus strains. J Mol Evol. 48:723–739.

Huelsenbeck JP, Crandall KA. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annu Rev Ecol Syst. 28:437–466.

Ibba M, Becker HD, Stathopoulos C, Tumbula DL, Soll D. 2000. The adaptor hypothesis revisited. Trends Biochem Sci. 25:311–316.

Ibba M, Curnow AW, Soll D. 1997. Aminoacyl-tRNA synthesis: divergent routes to a common goal. Trends Biochem Sci. 22:39–42.

Ibba M, Soll D. 2000. Aminoacyl-tRNA synthesis. Annu Rev Biochem. 69:617–650.

Ibba M, Soll D. 2004. Aminoacyl-tRNAs: setting the limits of the genetic code. Genes Dev. 18:731–738.

Jakubowski S. 2004. Accuracy of aminoacyl-tRNA synthetases: proofreading of amino acids. In: Ibba M, Francklyn C, Cusack S, editors. Aminoacyl-tRNA synthetases. Austin (TX): Landes Biosciences. p. 384–396.

Kaminska M, Deniziak M, Kerjan P, Barciszewski J, Mirande M. 2000. A recurrent general RNA binding domain appended to plant methionyl-tRNA synthetase acts as a cis-acting cofactor for aminoacylation. EMBO J. 19:6908–6917.

Kaneko T, Nakamura Y, Wolk CP, et al. (19 co-authors). 2001. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. DNA Res. 8:205–213; 227–253.

Kaneko T, Sato S, Kotani H, et al. (21 co-authors). 1996. Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res. 3:109–136.

Kaneko T, Tabata S. 1997. Complete genome structure of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. Plant Cell Physiol. 38:1171–1176.

Kaneko T, Tanaka A, Sato S, Kotani H, Sazuka T, Miyajima N, Sugiura M, Tabata S. 1995. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. DNA Res. 2:153–166 191–158.

Katoh K, Kuma K-i, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33:511–518.

Khrouchtchova A, Hansson M, Paakkarinen V, Vainonen JP, Zhang S, Jensen PE, Scheller HV, Vener AV, Aro EM, Haldrup A. 2005. A previously found thylakoid membrane protein of 14kDa (TMP14) is a novel subunit of plant photosystem I and is designated PSI-P. FEBS Lett. 579: 4808–4812.

Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol. 29:170–179.

Knoll AH. 2008. Cyanobacteria and earth history. In: Herrero A, Flores E, editors. The cyanobacteria: molecular biology, genomics and evolution. Norfolk (UK): Caister Academic Press. p. 1–19.

Korencic D, Ahel I, Schelert J, Sacher M, Ruan B, Stathopoulos C, Blum P, Ibba M, Soll D. 2004. A freestanding proofreading domain is required for protein synthesis quality control in Archaea. Proc Natl Acad Sci USA. 101:10260–10265.

Krakauer DC, Nowak MA. 1999. Evolutionary preservation of redundant duplicated genes. Semin Cell Dev Biol. 10: 555–559.

Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol. 7(Suppl 1):S4.

Lerat E, Daubin V, Ochman H, Moran NA. 2005. Evolutionary origins of genomic repertoires in bacteria. PLoS Biol. 3:e130.

Luque I, Andujar A, Jia L, Zabulon G, de Marsac NT, Flores E, Houmard J. 2006. Regulated expression of glutamyl-tRNA synthetase is directed by a mobile genetic element in the cyanobacterium *Tolypothrix* sp. PCC 7601. Mol Microbiol. 60:1276–1288.

Luque I, Contreras A, Zabulon G, Herrero A, Houmard J. 2002. Expression of the glutamyl-tRNA synthetase gene from the cyanobacterium *Synechococcus* sp PCC 7942 depends on nitrogen availability and the global regulator NtcA. Mol Microbiol. 46:1157–1167.

MacKinney. 1941. Absorption of light by chlorophyll solutions. J Biol Chem. 140:315–322.

Meeks JC, Elhai J, Thiel T, Potts M, Larimer F, Lamerdin J, Predki P, Atlas R. 2001. An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium. Photosynth Res. 70:85–106.

Min B, Pelaschier JT, Graham DE, Tumbula-Hansen D, Soll D. 2002. Transfer RNA-dependent amino acid biosynthesis: an essential route to asparagine formation. Proc Natl Acad Sci USA. 99:2678–2683.

Mulkidjanian AY, Koonin EV, Makarova KS, et al. (12 co-authors). 2006. The cyanobacterial genome core and the origin of photosynthesis. Proc Natl Acad Sci USA. 103: 13126–13131.

Nakamura A, Yao M, Chimnaronk S, Sakai N, Tanaka I. 2006. Ammonia channel couples glutaminase with transamidase reactions in GatCAB. Science. 312:1954–1958.

Nakamura M, Yamada M, Hirota Y, Sugimoto K, Oka A, Takanami M. 1981. Nucleotide sequence of the *asnA* gene coding for asparagine synthetase of *E. coli* K-12. Nucleic Acids Res. 9:4669–4676.

Nakamura Y, Kaneko T, Sato S, et al. (18 co-authors). 2002. Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. DNA Res. 9:123–130.

Nakamura Y, Kaneko T, Sato S, et al. (16 co-authors). 2003. Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. DNA Res. 10:137–145.

Nowak MA, Boerlijst MC, Cooke J, Smith JM. 1997. Evolution of genetic redundancy. Nature. 388:167–171.

Ochoa de Alda JAG, Del Pico A, Pedraza A, Houmard J. 2005. Caracterización, clasificación y filogenia de adenilil y guanilil ciclasas de cianobacterias. Oppidum. 1:311–355.

O'Donoghue P, Luthey-Schulten Z. 2003. On the evolution of structure in aminoacyl-tRNA synthetases. Microbiol Mol Biol Rev. 67:550–573.

Ostermeier M, Benkovic S. 2001. Evolution of protein function by domain swapping. In: Arnold F, editor. Evolutionary protein design. San Diego (CA): Academic Press. p. 29–77.

Page RD. 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci. 12:357–358.

Palenik B, Brahamsha B, Larimer FW, et al. (12 co-authors). 2003. The genome of a motile marine *Synechococcus*. Nature. 424:1037–1042.

Poptsova MS, Gogarten JP. 2007. The power of phylogenetic approaches to detect horizontally transferred genes. BMC Evol Biol. 7:45.

Posada D. 2006. ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online. Nucleic Acids Res. 34:W700–W703.

Prince VE, Pickett FB. 2002. Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet. 3:827–837.

Raczniak G, Becker HD, Min B, Soll D. 2001. A single amidotransferase forms asparaginyl-tRNA and glutaminyl-tRNA in *Chlamydia trachomatis*. J Biol Chem. 276: 45862–45867.

Ribas de Pouplana L, Schimmel P. 2001. Aminoacyl-tRNA synthetases: potential markers of genetic code development. Trends Biochem Sci. 26:591–596.

Rippka R. 1988. Isolation and purification of cyanobacteria. Methods Enzymol. 167:3–27.

Rocap G, Larimer FW, Lamerdin J, et al. (24 co-authors). 2003. Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. Nature. 424:1042–1047.

Roy H, Becker HD, Reinbolt J, Kern D. 2003. When contemporary aminoacyl-tRNA synthetases invent their cognate amino acid metabolism. Proc Natl Acad Sci USA. 100:9837–9842.

Ruan B, Soll D. 2005. The bacterial YbaK protein is a Cys-tRNA^{Pro} and Cys-tRNA^{Cys} deacylase. J Biol Chem. 280:25887–25891.

Rye R, Holland HD. 1998. Paleosols and the evolution of atmospheric oxygen: a critical review. Am J Sci. 298:621–672.

Salazar JC, Ambrogelly A, Crain PF, McCloskey JA, Soll D. 2004. A truncated aminoacyl-tRNA synthetase modifies RNA. Proc Natl Acad Sci USA. 101:7536–7541.

Salazar JC, Zuniga R, Raczniak G, Becker H, Soll D, Orellana O. 2001. A dual-specific Glu-tRNA^{Gln} and Asp-tRNA^{Asn} amidotransferase is involved in decoding glutamine and asparagine codons in *Acidithiobacillus ferrooxidans*. FEBS Lett. 500:129–131.

Sankaranarayanan R, Dock-Bregeon AC, Romby P, Caillet J, Springer M, Rees B, Ehresmann C, Ehresmann B, Moras D. 1999. The structure of threonyl-tRNA synthetase-tRNA^{Thr} complex enlightens its repressor activity and reveals an essential zinc ion in the active site. Cell. 97:371–381.

Scofield MA, Lewis WS, Schuster SM. 1990. Nucleotide sequence of *Escherichia coli asnB* and deduced amino acid sequence of asparagine synthetase B. J Biol Chem. 265: 12895–12902.

Schimmel P, Ribas De Pouplana L. 2000. Footprints of aminoacyl-tRNA synthetases are everywhere. Trends Biochem Sci. 25:207–209.

Schmidt EE, Davies CJ. 2007. The origins of polypeptide domains. Bioessays. 29:262–270.

Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. Tree-Puzzle: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 18: 502–504.

Schön A, Kannangara CG, Gough S, Söll D. 1988. Protein biosynthesis in organelles requires misaminoacylation of tRNA. Nature. 331:187–190.

Seo TK, Kishino H, Thorne JL. 2005. Incorporating gene-specific variation when inferring and evaluating optimal evolutionary tree topologies from multilocus sequence data. Proc Natl Acad Sci USA. 102:4436–4441.

Shi T, Falkowski PG. 2008. Genome evolution in cyanobacteria: the stable core and the variable shell. Proc Natl Acad Sci USA. 105:2510–2515.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol. 51:492–508.

Shimodaira H, Hasegawa M. 1999. Multiple comparisons of loglikelihoods with applications to phylogenetic inference. Mol Biol Evol. 16:1114–1116.

Sissler M, Delorme C, Bond J, Ehrlich SD, Renault P, Francklyn C. 1999. An aminoacyl-tRNA synthetase paralog with a catalytic role in histidine biosynthesis. Proc Natl Acad Sci USA. 96:8985–8990.

Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. 2007. Genome-wide experimental determination of barriers to horizontal gene transfer. Science. 318:1449–1452.

Strimmer K, Rambaut A. 2002. Inferring confidence sets of possibly misspecified gene trees. Proc Biol Sci. 269:137–142.

Sullivan J, Joyce P. 2005. Model selection in phylogenetics. Annu Rev Ecol Evol Syst. 36:445–466.

Swingley WD, Blankenship RE, Raymond J. 2008. Integrating Markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families. Mol Biol Evol. 25:643–654.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 56:564–577.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 10:512–526.

Tavaré S. 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. Lect Math Life Sci. 17:57.

Thomas CM, Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol. 3:711–721.

Thompson JD, Higgins DG, Gibson TJ. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

Tomitani A, Knoll AH, Cavanaugh CM, Ohno T. 2006. The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. Proc Natl Acad Sci USA. 103:5442–5447.

Turner S, Pryer KM, Miao VP, Palmer JD. 1999. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. J Eukaryot Microbiol. 46:327–338.

Wang C, Sobral BW, Williams KP. 2007. Loss of a universal tRNA feature. J Bacteriol. 189:1954–1962.

Wheeler DL, Barrett T, Benson DA, et al. (30 co-authors). 2007. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 35:D5–D12.

Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol. 18:691–699.

Whelan S, Lio P, Goldman N. 2001. Molecular phylogenetics: state-of-the-art methods for looking into the past. Trends Genet. 17:262–272.

Wilcox M, Nirenberg M. 1968. Transfer RNA as a cofactor coupling amino acid synthesis with that of protein. Proc Natl Acad Sci USA. 61:229–236.

Woese CR, Olsen GJ, Ibba M, Soll D. 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol Mol Biol Rev. 64:202–236.

Wong FC, Beuning PJ, Silvers C, Musier-Forsyth K. 2003. An isolated class II aminoacyl-tRNA synthetase insertion domain is functional in amino acid editing. J Biol Chem. 278: 52857–52864.

Xiong J. 2006. Photosynthesis: what color was its origin? Genome Biol. 7:245.

Xiong J, Bauer CE. 2002. Complex evolution of photosynthesis. Annu Rev Plant Biol. 53:503–521.

Yan W, Francklyn C. 1994. Cytosine 73 is a discriminator nucleotide in vivo for histidyl-tRNA in *Escherichia coli*. J Biol Chem. 269:10022–10027.

Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF, Papke RT. 2006. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. Genome Res. 16:1099–1108.

Laura Katz, Associate Editor

Accepted August 6, 2008