J Mol Evol (2000) 50:413–423
DOI: 10.1007/s002390010044

JOURNAL OF MOLECULAR EVOLUTION

Duplication and Quadruplication of *Arabidopsis thaliana* Cysteinyl- and Asparaginyl-tRNA Synthetase Genes of Organellar Origin

Nemo M. Peeters, Anne Chapron, Anatoli Giritch,* Olivier Grandjean, Dominique Lancelin, Tatiana Lhomme, Arnaud Vivrel, Ian Small

Station de Génétique et Amélioration des Plantes, INRA, Route de St Cyr, 78026 Versailles Cedex, France

Received: 8 October 1999 / Accepted: 23 January 2000

Abstract. Two cysteinyl-tRNA synthetases (CysRS) and four asparaginyl-tRNA synthetases (AsnRS) from *Arabidopsis thaliana* were characterized from genome sequence data, EST sequences, and RACE sequences. For one CysRS and one AsnRS, sequence alignments and prediction programs suggested the presence of an N-terminal organellar targeting peptide. Transient expression of these putative targeting sequences joined to jellyfish green fluorescent protein (GFP) demonstrated that both presequences can efficiently dual-target GFP to mitochondria and plastids. The other CysRS and AsnRSs lack targeting sequences and presumably aminoacylate cytosolic tRNAs. Phylogenetic analysis suggests that the four AsnRSs evolved by repeated duplication of a gene transferred from an ancestral plastid and that the CysRSs also arose by duplication of a transferred organelle gene (possibly mitochondrial). These case histories are the best examples to date of capture of organellar aminoacyl-tRNA synthetases by the cytosolic protein synthesis machinery.

Key words: Aminoacyl-tRNA synthetase — Mitochondria — Plastids — Gene duplication — Dual-targeting

Introduction

Aminoacyl-tRNA synthetases (aaRSs) are ubiquitous and essential components in translation needed to produce aminoacyl-tRNA, the link between the codon information and the peptide sequence. Because of their ancient history, well-understood role, and high degree of conservation, they have been natural choices for many phylogenetic studies (Brown and Doolittle 1995; Diaz-Lazcoz et al. 1998; Hashimoto et al. 1998; Kim et al. 1998; Nagel and Doolittle 1995; Shiba et al. 1998; Taupin and Leberman 1999; Wolf et al. 1999). Unfortunately (in some respects), it appears that the evolutionary history of aaRS genes has not always been straightforward, with several proposed examples of gene duplications, fusions, and gain by horizontal transfer (Diaz-Lazcoz et al. 1998; Doolittle and Handy 1998; Lamour et al. 1994; Wolf et al. 1999; Handy and Doolittle 1999). Prokaryotes contain a basic set of 18–20 aaRs, but eukaryotes have more, as they contain a second compartment (mitochondrial) capable of translation. The extra aaRS genes in eukaryotes are presumed to be derived from those contained in the original mitochondrial endosymbiont (for a review of mitochondrial origins, see Gray et al. 1999) and subsequently transferred to the nucleus, as no known extant mitochondrial genomes encode an aaRS. Plants received a third and more recent influx of aaRS genes with the acquisition of plastids (for reviews of plastid origins, see Douglas 1998; Martin et al. 1998; Turner et al. 1999). Like mitochondria, plastids have lost most of their original genes, some of which have been transferred to the nucleus. However, as these

* On leave from Institute of Cell Biology and Genetic Engineering, National Academy of Sciences of Ukraine, Zabolotnogo str. 148, Kiev-022, Ukraine
Correspondence to: Ian Small; *e-mail:* small@versailles.inra.fr

transfers are more recent [*Porphyra purpurea* plastids still contain a couple of aaRS genes (Reith and Munholland 1995)], there is more hope of being able to retrace the evolutionary events involved.

We are in the process of surveying all the tRNA and aaRS sequences of the model plant *Arabidopsis thaliana* (our "taaRSAt" database is accessible on the world wide web: http://www.inra.fr/Internet/Produits/TAARSAT/) One might have expected up to 60 different aaRSs in plants, i.e., one set of 20 enzymes for each translation compartment. However, as in other eukaryotes, the situation is more complicated, as there are several cases where one gene has replaced the function of another because its gene product is dual-targeted to two compartments. Dual-targeting to the cytosol and mitochondria was shown for *Arabidopsis* alanyl-tRNA synthetase (Mireau et al. 1996) and is probably also the case for several other plant aaRSs (Small et al. 1999). Dual-targeting is also possible to mitochondria and plastids, as demonstrated for *Arabidopsis* histidyl-tRNA (Akashi et al. 1998) and methionyl-tRNA (Menand et al. 1998) synthetases. In the case of HisRS and MetRS, the cytosolic enzymes in plants are encoded by entirely different genes of typical eukaryotic origin. In the present work, we report two new aaRSs dual-targeted to both organelles: an asparaginyl-tRNA synthetase (AsnRS) and a cysteinyl-tRNA synthetase (CysRS). However, in these cases, the cytosolic isoforms are very similar to their organellar counterparts, resulting in closely related enzymes in all three compartments. Moreover, phylogenetic analysis suggests that the four AsnRS genes and two CysRS genes all have organelle origins. These two groups of enzymes provide the clearest examples to date of organellar aaRSs that have usurped the role of their (now vanished) cytosolic counterparts.

Materials and Methods

PCR and Sequencing

Standard molecular biology techniques were performed using the protocols described by Ausubel et al. (1994). Oligonucleotides were purchased from Genosys (Cambridge, UK). Sequencing was performed by GenomeExpress (Roscoff, France). Oligonucleotide sequences and full details of cloning procedures are available on request. All sequence names, sources, and accession numbers used in the present work are listed in Table 1. Poly(A^+) RNA was prepared from *Arabidopsis thaliana*, ecotype Columbia, using an mRNA purification kit from Amersham Pharmacia Biotech (Uppsala, Sweden). 5′ RACE was performed with the 5′ RACE System for Rapid Amplification of cDNA Ends, Version 2.0, from Gibco BRL (Bethesda, MD, USA). The specific gene primers used are available on request.

Expression of the different AsnRS and CysRS genes of *Arabidopsis* was tested by performing PCR on different cDNA sources using specific primers for each gene and comparing the level of expression with a control. All oligonucleotides were designed to encompass a known intron (except those for SYNC2_ARATH, for which only the cDNA sequence is available, and those for ROC1, which has no intron in its gene). The exon/intron structures of the different genes are given in the taaRSAt database. The PCR conditions were 4 min at 94°C, followed by 25 cycles of 1 min at 94°C, 1 min at 65°C, and 1 min at 72°C, and ending with 5 min at 72°C. The PCR conditions were chosen to allow rough comparisons of expression levels. The template was either cDNA made from RNA extracted from roots, rosette leaves, stems, inflorescences, or siliques or genomic DNA from 5-week-old plants of *Arabidopsis thaliana* ecotype Wassilewskii. Total RNA was treated with DNase I (RNase-free; Amersham Pharmacia Biotech) and the cDNA

Table 1. Sequences used in the present analysis[a]

	Source	Accession No.
Asparaginyl-tRNA synthetases		
Bacteria		
SYN_BACSU	*Bacillus subtilis*	L47709
SYN_ECOLI	*Escherichia coli*	M33145
SYN_HAEIN	*Haemophilus influenzae*	U32810
SYN_LACDE	*Lactobacillus delbrueckii*	X89438
SYN_SYNY3	*Synechocystis* sp.	D64006
Eukarya		
SYN_CAEEL	*Caenorhabditis elegans*	Z71262
SYN_HUMAN	*Homo sapiens*	AJ000334
SYNC1_ARATH	*Arabidopsis thaliana*	AF170909
SYNC2_ARATH	*Arabidopsis thaliana*	AF170910
SYNC3_ARATH	*Arabidopsis thaliana*	AC008148[b]
SYNO_ARATH	*Arabidopsis thaliana*	AJ222644
SYNC_YEAST	*Saccharomyces cerevisiae*	P38707
SYNM_YEAST	*Saccharomyces cerevisiae*	P25345
Aspartyl-tRNA synthetases		
Eukarya		
SYDC_ARATH	*Arabidopsis thaliana*	AL035440[b]
SYDM_ARATH	*Arabidopsis thaliana*	AL031394[b]
SYDC_YEAST	*Saccharomyces cerevisiae*	P04802
Cysteinyl-tRNA synthetases		
Bacteria		
SYC_AZOBR	*Azospirillum brasilense*	X99587
SYC_AQUAE	*Aquifex aeolicus*	AAC07125
SYC_BACSU	*Bacillus subtilis*	D26185
SYC_ECOLI	*Escherichia coli*	X56234
SYC_HAEIN	*Haemophilus influenzae*	U32693
SYC_HELPY	*Helicobacter pylori*	U05676
SYC_RHOCA	*Rhodobacter capsulatus*	RRC03443[c]
SYC_RICPR	*Rickettsia prowazekii*	CAA14555
SYC_SYNY3	*Synechocystis* sp.	D90914
SYC_THEMA	*Thermotoga maritima*	AAD35801
Arachaea		
SYC_ARCFU	*Arachaeoglobus fulgidus*	AE001076
Eukarya		
SYC_DROME	*Drosophila melanogaster*	AAD34748
SYC_HUMAN	*Homo sapiens*	L06845
SYC_SCHPO	*Schizosaccharomyces pombe*	Q09860
SYCC_ARATH	*Arabidopsis thaliana*	AB009048[b]
SYCO_ARATH	*Arabidopsis thaliana*	AC005311[b]

[a] Sequence names conform to the usual SWISS-PROT style for aaRSs, but most are not yet available from SWISS-PROT.

[b] Accession numbers of genome sequences, from which we predicted coding sequences. AB009048 was sequenced by the Institute for Genomic Research, Rockville, MD, USA; AC005311, by the Kazusa DNA Research Institute, Chiba, Japan; and AC008148, by the DNA Sequencing and Technology Center, Stanford University, Palo Alto, CA, USA.

[c] SYC_RHOCA is accessible at the *Rhodobacter capsulatus* genome sequencing web site: http://rhodol.uchicago.edu/capsulapedia/capsulapedia/capsulapedia.shtml.

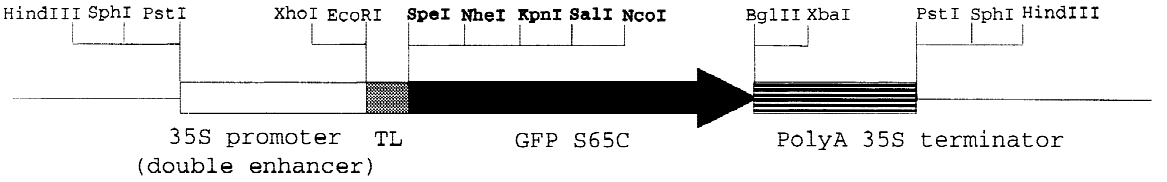

Fig. 1. GFP expression cassette of pOL GFPS65C. Useful restriction sites are depicted. The unique sites of the multiple cloning site (mcs) are in *boldface*. If no presequence is cloned in the mcs, the first ATG used is in the *Nco*I site and corresponds to the first methionine of the GFP protein. This vector contains a dual-enhancer 35S promoter from cauliflower mosaic virus (CaMV), the translation leader sequence from tobacco etch virus (TL), and the 35S polyadenylation signal from CaMV.

was synthesized using the SuperScript Preamplification System for First Strand cDNA Synthesis (Gibco BRL).

GFP Expression Vectors and Constructs

RecA–GFP and CoxIV–GFP fusion proteins were expressed from the same constructs used by Akashi et al. (1998) based on the plasmid pCK GFP-S65C (Reichel et al. 1996). The GFP expression cassette (promoter, GFP coding sequence, terminator) of pCK GFP-S65C was cut out by *Hin*dIII and religated in the opposite orientation, to prevent any LacZ promoter-driven GFP expression. This vector was called pFF GFP-S65C. Subsequently a small multicloning site was introduced in the *Nco*I cloning site surrounding the GFP intitiation codon. For this purpose, two complementary 5′-phosphorylated oligonucleotides (Table 2) were hybridized together and ligated into the *Nco*I site to form pOL GFP-S65C. Putative presequences can now be cloned in an orientated way in the sites of this new vector (Fig. 1).

The targeting presequence corresponding to the first 71 amino acids of SYNO_ARATH was PCR-amplified (from *A. thaliana* Columbia genomic DNA; PCR conditions were 30 s at 94°C, 45 s at 65°C, and 45 s at 72°C, 30 cycles) and cloned into the *Nco*I site of pFF GFP-S65C. In a similar way, the targeting presequence corresponding to the first 64 amino acids of SYCO_ARATH was PCR-amplified (from *A. thaliana* Columbia genomic DNA; PCR conditions were 30 s at 94°C, 45 s at 55°C, and 45 s at 72°C, 30 cycles) and cloned into the *Spe*I/*Sal*I sites of pOL GFP-S65C. Putative presequence–GFP clones were sequenced in both directions with an oligonucleotide hybridizing in the 35S promoter (5′-3′)GGACCTCGAGAATTCTCAA and in the GFP sequence (5′-3′)TTTGTGCCCATTAACATCAC.

Tobacco protoplast transformation and MitoTracker Red CMXRos (Molecular Probes, Eugene, OR, USA) staining were carried out in the same way as by Akashi et al. (1998). Protoplasts were examined with a Leica TCS-NT confocal laser scanning microscope (Leica Microsystems, Heidelberg, Germany) with an argon/krypton laser (Omnichrome, Chino, USA) equipped with an acoustooptical tunable filter (AOTF) excitation system. For GFP (GFP-S65C; abs, 479 nm; em, 507 nm) monitoring, excitation was at 488 nm. The emission signal was separated through a short-pass filter (RSP580), and the short-wavelength signal corresponding to GFP fluorescence was band-pass filtered (BP530/30) and collected in the green channel. The long-wavelength signal was long-pass filtered (LP590) to collect the chlorophyll autofluorescence (red channel). When the protoplasts were stained with MitoTracker Red (abs, 578 nm; em, 599 nm), excitation was at 488 and 568 nm. Both chlorophyll autofluorescence and MitoTracker Red emission were collected through a long-pass filter (LP590). The AOTF was adjusted such that the 488-nm excitation gave no discernible MitoTracker Red emission, and the 568-nm excitation gave no discernible GFP emission.

Alignment and Phylogenetic Analysis

All sequence alignments were performed using the program CLUSTAL X (version 1.8) (Thompson et al. 1997). The trees were built using maximum-parsimony algorithms with the PAUP (Swofford 1993) program contained in the GCG package (Wisconsin Package Version 10.0; Genetics Computer Group, Madison, WI, USA). The parameters for the tree reconstruction were heuristic search, stepwise addition that increases the tree size least, and no branch swapping. The number of bootstrap trials used was 100. Alignments were also analyzed by a distance matrix method (neighbor-joining method, CLUSTAL X) or by maximum likelihood using PUZZLE 4.0.2 (Strimmer and von Haeseler 1996). The trees were displayed with the TreeView program (version 1.5) (Page 1996).

Results

AsnRS and CysRS cDNAs

Two EST clones, G6D2T7 (accession No. N96875) and G8H10T7 (accession No. N96875), encoding putative AsnRSs were retrieved from the ABRC (*Arabidopsis* Biological Resource Center, Ohio State University, Columbus, OH, USA), subcloned, and sequenced. These two ESTs were identical. The full-length cDNA (accession No. AF170909) encodes a putative cytosolic AsnRS that we refer to as SYNC1_ARATH. A partial cDNA clone, sequenced by Aubourg et al. (1998) (accession No. AJ222645), was extended by 5′ RACE-PCR to obtain a full-length cDNA (accession No. AF170910) encoding a second putative cytosolic AsnRS that we call SYNC2_RATH. A BAC (accession No. AC008148) containing a gene encoding a third *Arabidopsis* AsnRS, which we refer to as SYNC3_ARATH, was recently sequenced by the DNA Sequencing and Technology Center, Stanford University, Palo Alto, CA, USA. There are no EST accessions corresponding to this gene at the moment. A fourth full-length AsnRS cDNA (accession No. AJ222644) was described by Aubourg et al. (1998) and encodes an AsnRS with a putative organellar targeting sequence. We refer to this protein as SYNO_ARATH.

Two *Arabidopsis* genes encoding probable CysRSs have been deposited in GenBank during the systematic sequencing of the genome by the Institute for Genomic Research, Rockville, MD, USA, and the Kazusa DNA Research Institute, Chiba, Japan. For both of these genes, ESTs were available and permitted easy prediction of the probable protein sequence. One of these CysRSs (SYCC_ARATH; BAC accession No. AB009048, EST accession Nos. AB015096 and T04427)

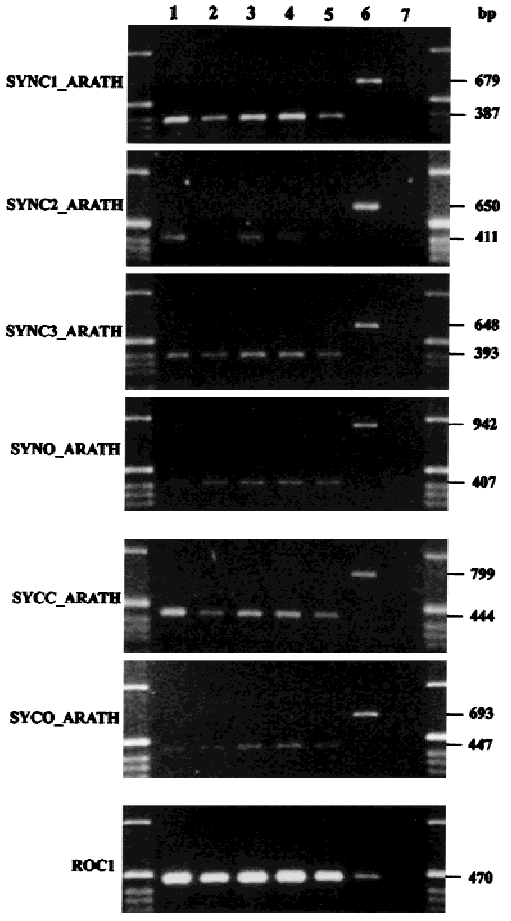

Fig. 2. Expression of *Arabidopsis* AsnRS and CysRS genes. PCR was performed on different DNA sources: cDNA from (1) roots, (2) leaves, (3) stems, (4) inflorescences, (5) siliques, and (6) genomic DNA. Lane 7 is a control with no added template. A 1-kb ladder from Gibco BRL is present on *both sides* of each gel image. The expected size of the genomic and cDNA amplification products is given to the *right* of each image. Fortunately, the oligonucleotides chosen without knowledge of the gene sequence to amplify the gene encoding SYNC2_ARATH gave a higher molecular weight amplification product for the genomic DNA, estimated to be 650 bp.

lacks an N-terminal extension and presumably encodes the cytosolic enzyme; the other (SYCO_ARATH; BAC accession No. AC005311, EST accession No. N96709) carries a putative N-terminal targeting sequence. Our predictions of the exon/intron structure of these various genes and the protein sequences are available at http://www.inra.fr/Internet/Produits/TAARSAT/.

AsnRSs and CysRSs Expression

ESTs exist for all the genes studied here except for SYNC3_ARATH, but to verify expression, PCR experiments were carried out on various cDNA sources (Fig. 2). Spliced mRNAs were detected for all six aaRS genes in all tissues tested. The amplification was carried out under semiquantitative conditions and so the results suggest that SYNC1_ARATH and SYNC3_ARATH are the most highly expressed AsnRSs (in terms of steady-state mRNA levels) and that SYCC_ARATH is more highly expressed than SYCO_ARATH.

Sequence Alignments

AsnRSs and aspartyl-tRNA synthetases (AspRSs) are related (Shiba et al. 1998) and can occasionally be confused. The four *Arabidopsis* AsnRSs were aligned with known AsnRSs from other organisms and with aspartyl-tRNA synthetases, including the putative cytosolic and mitochondrial AspRSs from *Arabidopsis* (Fig. 3). The results confirm that SYNC1_ARATH, SYNC2_ARATH, SYNC3_ARATH, and SYNO_ARATH are very probably AsnRSs and not AspRSs. The alignments also reveal heterogeneous N termini for these four sequences and indicate that SYNC2_ARATH has important sequence differences in the highly conserved motifs 2 and 3, characteristic for all class II aaRSs (Eriani et al. 1990; Moras 1992). SYNC1_ARATH, SYNC2_ARATH, and SYNC3_ARATH all contain a long insertion between motif 1 and motif 2, and SYNC3_ARATH contains a second long insertion nearer the N terminus.

The two *Arabidopsis* CysRS sequences were aligned with other known CysRSs; part of the alignment is given in Fig. 4, while the complete alignment can be seen at http://www.inra.fr/Internet/Produits/TAARSAT/CysRS/syc.msf.html.

Organellar Targeting

The N-terminal part of the CysRS alignment (Fig. 4) clearly shows a 60-amino acid extension for the SYCO_ARATH sequence. Prediction programs suggested that this could be an organellar targeting sequence. MitoProt (Claros and Vincens 1996), searching for putative mitochondrial targeting sequences, strongly predicted (score of 0.97) a 53-amino acid N-terminal targeting peptide. ChloroP (Emanuelsson et al. 1999), searching for chloroplast targeting sequences, predicted (score of 0.56) a 34-amino acid N-terminal targeting peptide. The first 64 amino acids of this sequence were fused to the GFP reporter protein and tested by transient expression in tobacco mesophyll protoplasts. GFP fluorescence was found in both mitochondria and chloroplasts in transformed protoplasts (Fig. 5D). Such double-targeting is rare (Small et al. 1998); typical mitochondrial or plastid targeting sequences direct GFP highly specifically to one or the other organelle but never both (Figs. 5B and C) (Köhler et al. 1997a, b).

Eukaryotic AsnRSs contain N-terminal extensions (Shiba et al. 1998), which complicate the identification of targeting sequences. After testing with MitoProt and ChloroP, only the *A. thaliana* SYNO_ARATH sequence was predicted to contain organellar targeting determinants: a 51-amino acid mitochondrial targeting sequence (score of 0.99) and/or a 63-amino acid chloroplast tar-

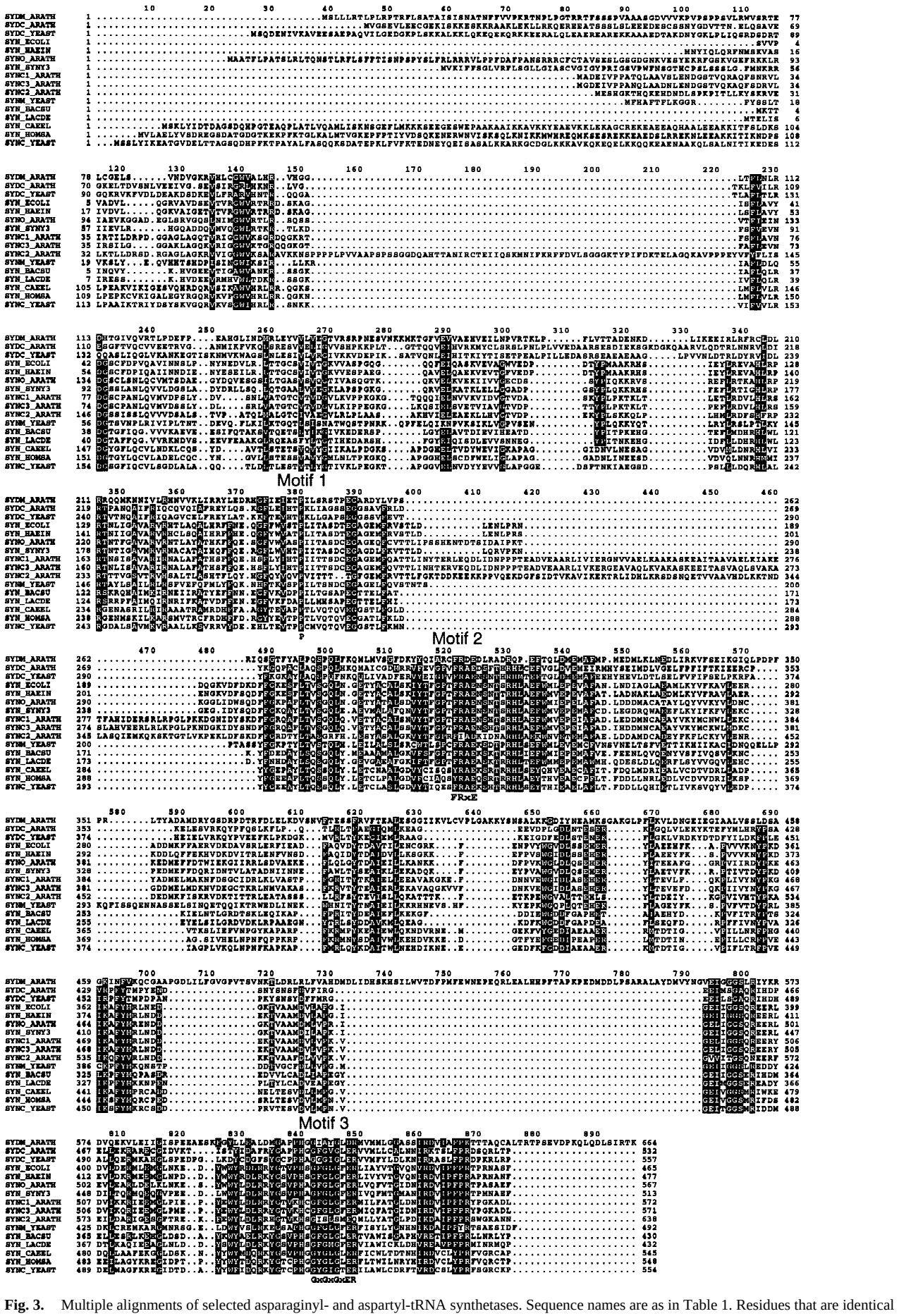

Fig. 3. Multiple alignments of selected asparaginyl- and aspartyl-tRNA synthetases. Sequence names are as in Table 1. Residues that are identical in at least 10 sequences of 16 are *shaded.* The consensus sequences of the three characteristic motifs of class II aminoacyl-tRNA synthetases are indicated *under* the alignment.

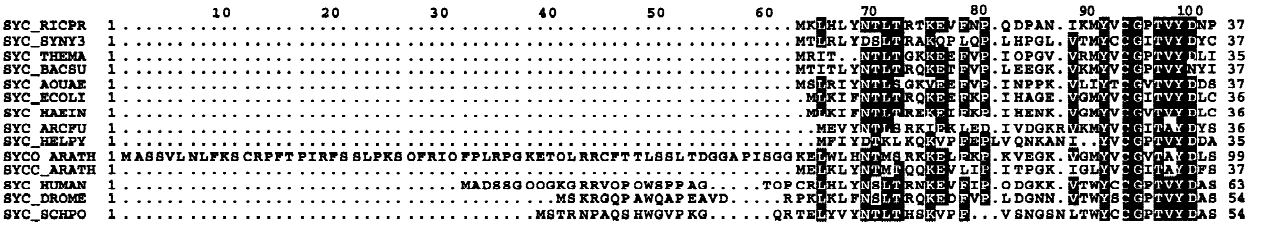

Fig. 4. The N-terminal part of a multiple alignment of cysteinyl-tRNA synthetases. Sequence names are as in Table 1. Residues that are identical in at least eight sequences of 11 are *shaded.*

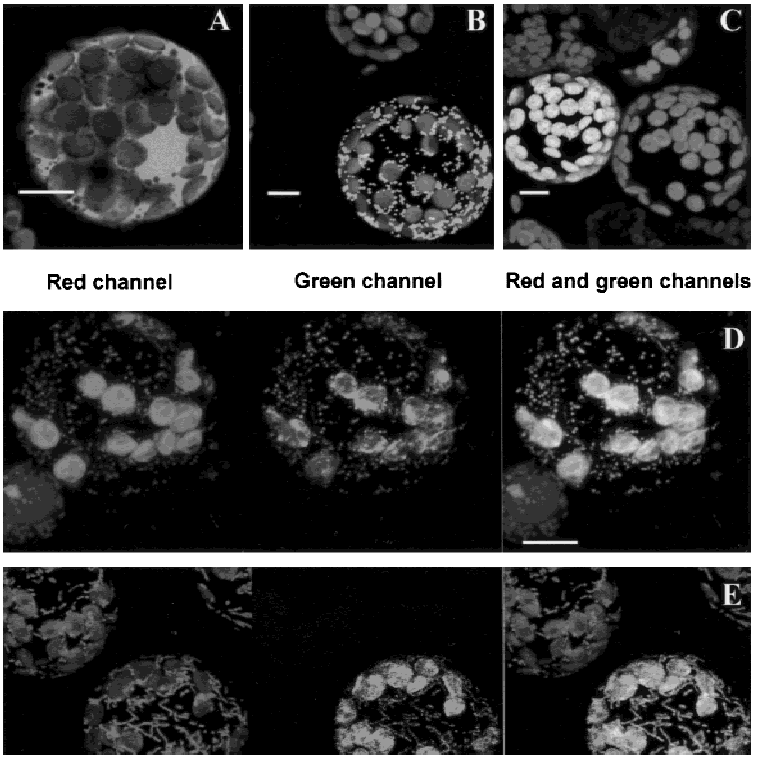

Fig. 5. Transient expression of GFP fusions in tobacco protoplasts. Cells expressing **(A)** GFP, expressed from the pOL GFPS65C vector; **(B)** CoxIV–GFP fusion protein; **(C)** RecA–GFP fusion protein; **(D)** SYCO–GFP fusion protein; and **(E)** SYNO–GFP fusion protein. B, C, and E show a mixture of transformed and untransformed protoplasts with the red and green channels superimposed. In D and E, protoplasts were stained with MitoTracker Red. Separate red channel and green channels and the two superimposed are shown. The scale bar is 10 μm in all images.

geting sequence (score of 0.58). When the first 71 amino acids were fused to GFP and tested in the protoplast system, this presequence directed the fusion protein to both organelles, like the CysRS presequence (Fig. 5E).

Phylogenetic Analysis

The AsnRS alignment in Fig. 3 and a complete CysRS alignment were used for phylogenetic analysis using distance matrices, maximum-parsimony (Fig. 6), and maximum-likelihood methods (not shown). All three methods clearly show a close relationship among SYNC1_ARATH, SYNC2_ARATH, and SYNC3_ARATH and generally group SYNO_ARATH and SYN_SYNY3 (from *Synechocystis*) together with them (Fig. 6A). Some alternative trees with less support (not shown) place the *E. coli* and *H. influenzae* sequences as the closest relatives of SYN_SYNY3, SYNO_ARATH, or both. However, as other proteobacteria (e.g., *Rickettsia*) lack an AsnRS, it seems possible that an ancestor of *E. coli* and *H. influenzae* acquired the enzyme by horizontal transfer. All four *Arabidopsis* AsnRSs are clearly distinct from the known eukaryotic cytosolic enzymes, which form a separate well-defined group. Removal of the long insertions from the *Arabidopsis* cytosolic sequences had no significant effect on the results. Using related AspRS sequences as an outgroup, the root of this tree can be placed with considerable confidence and confirms the clear distinction between the animal and fungal cytosolic AsnRSs, on one hand, and the bacterial and plant enzymes, on the other. The yeast mitochondrial AsnRS (SYNM_YEAST) groups within the bacterial sequences but not close to the plant sequences. However, the phylogenetic position of this sequence must be interpreted with prudence, as the amino acid composition deviates significantly from the other AsnRSs, implying an evolutionary bias that may invalidate the algorithms used to make the trees.

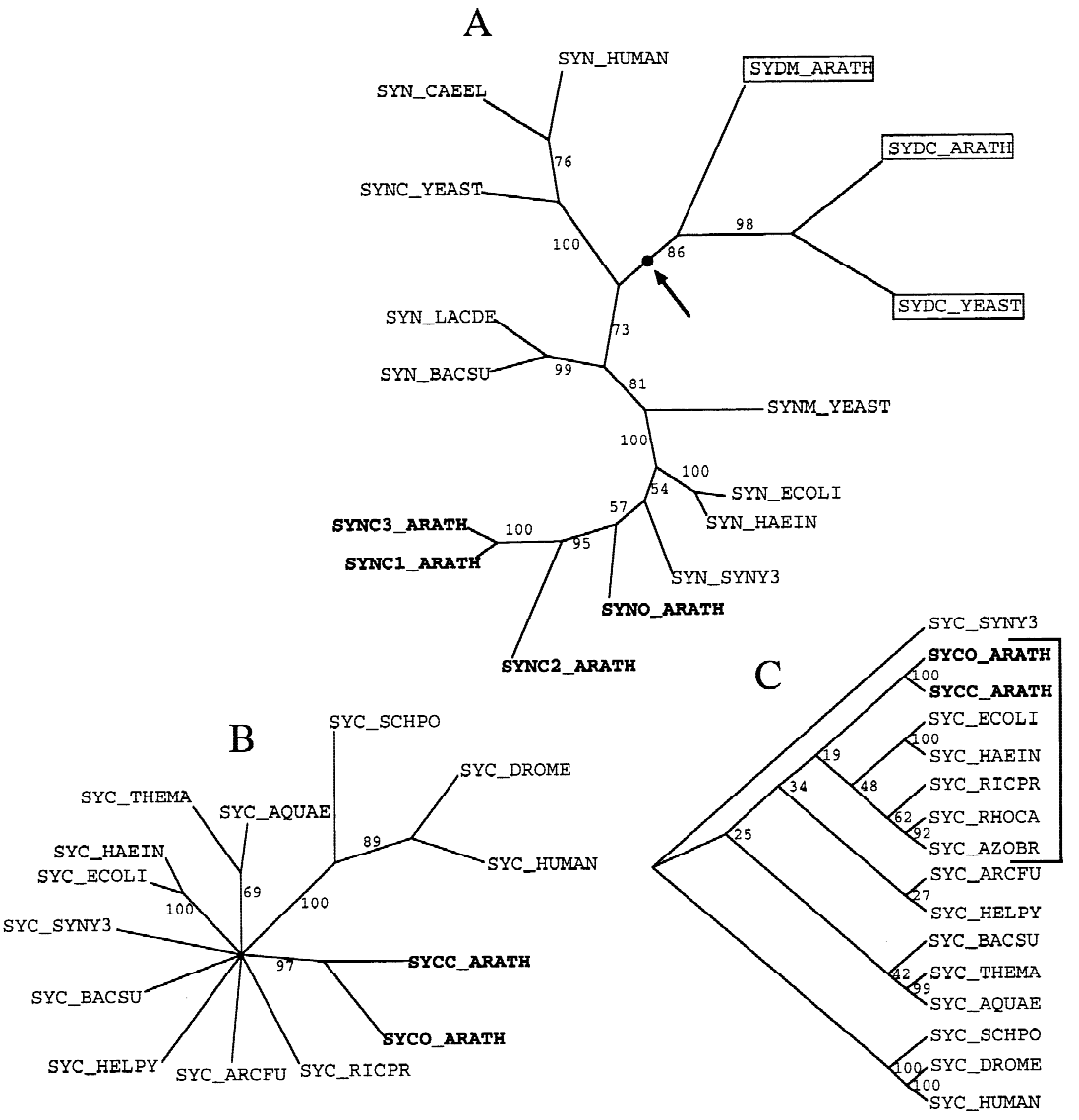

Fig. 6. Phylogenetic trees of asparaginyl-tRNA synthetases **(A)** and cysteinyl-tRNA synthetases **(B).** The trees are based on amino acid sequence alignments generated by CLUSTAL X and generated by maximum-parsimony analysis with PAUP. Branches with less than 50% bootstrap support have been collapsed. In A, the probable root of the tree (as deduced by the inclusion of several AspRS sequences as an outgroup) is *arrowed.* Sequences with biased amino acid composition unsuitable for phylogenetic reconstructions (tested by PUZZLE) or strongly suspected to be derived from relatively recent horizontal transfers were not included in the analysis shown here (with the exception of SYNM_YEAST and SYNC2_ARATH, both of which have a biased amino acid composition). Analyses with the full set of available AsnRS and CysRS sequences did not significantly alter the positions of the plant sequences. **C** A distance tree of cysteinyl-tRNA synthetases using CLUSTAL X. The bootstrap value from 100 replicates is indicated *along branches.* Nomenclature is the same as in Table 1. The *Arabidopsis* sequences are in *boldface.*

Similarly, all three methods show a close relationship between SYCC_ARATH and SYCO_ARATH and collect other known eukaryotic cytosolic enzymes in a distinct and well-supported group (Figs. 6B and C). The closest relatives of the *Arabidopsis* enzymes are not evident, but a neighbor-joining tree suggests that sequences from proteobacteria (such as *Azospirillum brasilense, Escherichia coli, Haemophilus influenzae, Rhodobacter capsulatus,* and *Rickettsia prowazekii*) might be closest (Fig. 6C). This suggests that the common ancestor of SYCC_ARATH and SYCO_ARATH possibly derived from the ancestral mitochondrial enzyme.

Discussion

Subcellular Localization

Predicting the subcellular location of proteins from sequence data is not yet an exact science, therefore some

form of experimental verification is necessary. We chose to use GFP fusions to verify the targeting properties of the predicted targeting peptides. This method has the advantage of testing for mitochondrial and plastid targeting simultaneously, which is essential if dual-targeting is a possibility. The fact that a presequence–GFP fusion is targeted to a particular organelle does not prove beyond all doubt that the protein from which the presequence was taken is targeted to the same organelle, but it is strong evidence in favor. We therefore consider it likely that SYNO_ARATH and SYCO_ARATH are indeed imported into both mitochondria and chloroplasts. The designation of SYNC1_ARATH, SYNC2_ARATH, SYNC3_ARATH, and SYCC_ARATH as cytosolic proteins relies essentially on negative observations. These proteins either lack N-terminal extensions or possess one which lacks the expected characteristics of organelle targeting sequences. Judging from the RT-PCR results (Fig. 2) and EST frequency, these four proteins are more highly expressed than the putative organellar proteins, as one would expect for cytosolic proteins. We therefore think that our attribution of these five aaRSs to different compartments is probably correct, and would provide for AsnRS and CysRS activity in each of the three translation systems where they are required. We have found no evidence in genome data or ESTs that any other types of AsnRS or CysRS genes exist in *Arabidopsis* or other plants. However, with 40% of the *Arabidopsis* genome still to be sequenced, the presence of other genes cannot be entirely ruled out.

Duplicated Genes

It is unusual to find closely related duplicated genes encoding aaRSs. In general, when two aaRSs with the same amino acid specificity are present, the sequences are divergent, implying that one of the two results from horizontal transfer (reviewed by Diaz-Lazcoz et al. 1998; Wolf et al. 1999). In such cases, the assumption is that the two enzymes have distinct functions or expression patterns or are targeted to different compartments. The presence of three very similar putative cytosolic AsnRSs in *Arabidopsis* (SYNC1_ARATH, SYNC2_ARATH, and SYNC3_ARATH) in addition to the organellar AsnRS is therefore rather puzzling. We are unaware of any other examples of four related aaRSs being present in one organism. The branch leading to SYNC2_ARATH is long in all trees generated by methods that measure substitution rates, suggesting that it is accumulating mutations more rapidly than SYNC1_ARATH and SYNC3_ARATH. The latter two are very similar in sequence (Fig. 3) and expression (Fig. 2), suggesting that they originate from a recent duplication. Moreover, the normally perfectly conserved arginine and glutamate residues of motif 2 and the glycine and arginine residues of motif 3 are not conserved in SYNC2_ARATH. This is surprising since they have been shown to interact with the ATP substrate in the *Thermus thermophilus* AsnRS–ATP crystal complex (Berthet-Colominas et al. 1998). It is therefore quite possible that SYNC2_ARATH is incapable of carrying out aminoacylation, since it seems likely that it cannot bind ATP. Nevertheless, the gene is expressed, as shown by RT-PCR (Fig. 2), the existence of ESTs (Table 2), and our ability to recover a full-length spliced cDNA by 5′ RACE. The absence of frameshifts or premature stop codons despite considerable sequence drift strongly implies that the protein is produced and has a function. Whether or not it functions as an AsnRS will need to be confirmed.

Dual-Targeting

It is common to find two distinct aaRSs with the same amino acid specificity in eukaryotes, one mitochondrial and the other cytosolic (Kumazawa et al. 1989; Nabholz et al. 1997). In most of these cases, it is probable that one of the two genes encoding these proteins is derived from the original organellar gene. The mitochondrial translation system in many eukaryotes is highly divergent, with unusual tRNAs (Ueda et al. 1992) and often variations in the genetic code (Gray et al. 1999; Kurland 1992). These differences may prevent the same aaRSs from being used in both compartments. In plants, the situation is a little different for two major reasons: the acquisition of plastids has provided an influx of new tRNA and aaRS genes, and the mitochondrial translation system has diverged very little from the eubacterial standard. This has allowed considerable sharing of aaRSs and tRNAs between compartments (Small et al. 1999). In particular, plant mitochondria contain a number of tRNAs expressed from insertions of plastid DNA in the mitochondrial genome. The presence of "chloroplast-like" $tRNA^{His}$ and $tRNA^{Met}$ in *Arabidopsis* mitochondria was put forward as an explanation for the dual-targeting of *Arabidopsis* HisRS and MetRS to mitochondria and plastids (Akashi et al. 1998; Menand et al. 1998). The same argument applies to SYNO_ARATH; all higher plant mitochondria examined to date [including *Arabidopsis* (Unseld et al. 1997)] contain a $tRNA^{Asn}$ almost-identical to the corresponding plastid $tRNA^{Asn}$. The phylogenetic analyses (Fig. 6A) strongly suggest that SYNO_ARATH derives from a plastid gene. Unlike cyanobacteria, bacteria close to the presumed ancestor of mitochondria [e.g., *Rickettsia* (Andersson et al. 1998; Gray 1998)] lack an AsnRS and thus mitochondria probably also lacked an AsnRS to start with. This is not entirely certain, however, as the yeast mitochondrial AsnRS shows some similarity to bacterial enzymes and may be of mitochondrial origin. In organisms lacking an AsnRS, asparaginyl-$tRNA^{Asn}$ is generated by transamidation of aspartyl-$tRNA^{Asn}$ (Curnow et al. 1996). In principle, the aminoacylation reaction is more efficient than transamidation, which is a

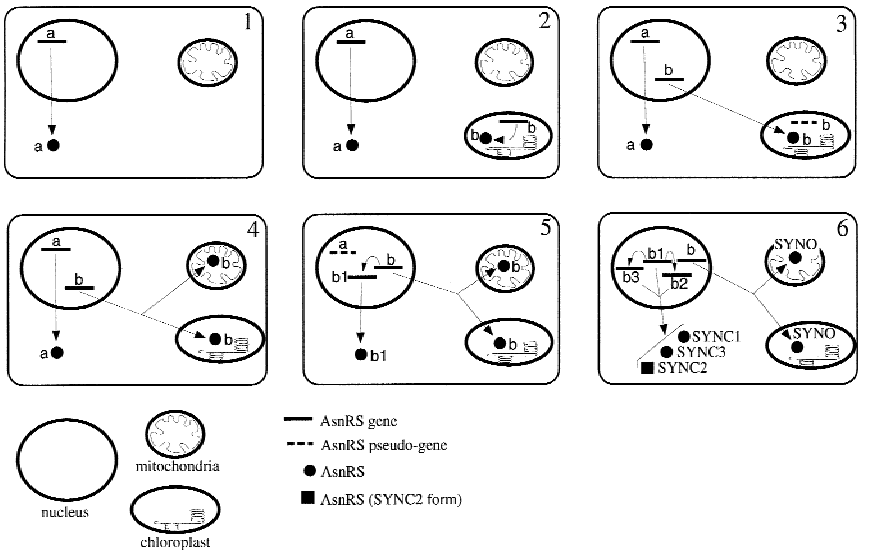

Fig. 7. Possible evolutionary history of asparaginyl-tRNA synthetases (AsnRSs) in *Arabidopsis thaliana.* "a" is the typical eukaryotic AsnRS, encoded by the nucleus; "b" is the ancestral plastid form. (1) An endosymbiotic event giving rise to the mitochondria was at the origin of eukaryotic cells. At this stage, mitochondria probably had no AsnRS; $tRNA^{Asn}$ was produced by transamidation of Asp-$tRNA^{Asn}$. (2) At the origin of plants was a second endosymbiotic event involving a cyanobacterium. The newly formed plastid encoded its own AsnRS. (3) The plastid AsnRS gene was transferred to the nucleus, and the protein retargeted to the plastid. Subsequently the plastid gene was lost. (4) Acquisition of mitochondrial targeting ability by the formerly plastid-specific AsnRS. (5) Duplication of the plastid-derived gene and replacement of the former cytosolic AsnRS by the product of the new gene (the so-called "cytosolic capture"). (6) Second and third duplications giving rise to the presence of three putative cytosolic AsnRSs: SYNC1, SYNC2, and SYNC3. The SYNC2 form is indicated differently to suggest that it may not be a functional AsnRS.

significantly more complex route to aminoacylate tRNA and generally costs two ATP molecules versus one (Curnow et al. 1996; Ibba et al. 1997; Shiba et al. 1998). This may have provided a selective advantage for double-targeting of AsnRS to both organelles (if mitochondria did lack an AsnRS at this point).

Generally it is not difficult to show a close relationship between plant plastid aaRSs and their homologues from *Synechocystis* (Fig. 6A) (Menand et al. 1998; unpublished data available in the previously cited taaRSAt database) or between cytosolic aaRSs and their homologues from animals or fungi (again, see the taaRSAt database), but it is much more difficult to detect specific relationships between mitochondrial aaRSs and their homologues from α-proteobacteria (close to the presumed ancestors of mitochondria). This is probably primarily because the divergence between mitochondria and α-proteobacteria is much more ancient than the divergence between plastids and cyanobacteria or the divergence between plants and other eukaryotes. SYCO_ARATH has no close relationship to CysRSs from either *Synechocystis* or eukaryotes and thus, by elimination, may be of mitochondrial origin. There is weak support for this hypothesis from the distance tree (Fig. 6C). Judging from the GFP fusion, SYCO_ARATH is dual-targeted to both organelles. *Arabidopsis* mitochondrial $tRNA^{Cys}$ is a typical mitochondrial, not "plastid-like," tRNA and thus the mitochondrial and plastid CysRS isoforms have different substrates. The advantage for double-targeting is thus less evident than for SYNO_ARATH.

Cytosolic Capture

The most original feature of the AsnRSs and CysRSs described here is the close relationship between the organellar and the cytosolic isoforms. It has been suggested that the cytosolic AlaRS, ValRS, and ThrRS enzymes are encoded by genes derived from mitochondria (Doolittle and Handy 1998), and indeed the presence of this "eubacterial-like" ValRS in the amitochondrial protists *Giardia* and *Trichomonas* has been used as evidence that these organisms must be derived from eukaryotes that did have mitochondria (Hashimoto et al. 1998). However, the phylogenies published for these enzymes to date do not show a clear specific relationship to bacteria close to organelle ancestors but, rather, a general similarity to eubacterial enzymes. It is possible that these genes originated elsewhere than in mitochondria, as other plausible scenarios for the acquisition of eubacterial genes by eukaryotes have been put forward (Doolittle 1998; Lopez-Garcia and Moreira 1999). The phylogenies presented here are equally inconclusive concerning the origin of the two *Arabidopsis* CysRSs but, on the contrary, are quite explicit concerning the origin of the AsnRSs; all four *Arabidopsis* genes are very likely to have derived from an ancestral plastid gene. This is the best evidence to date for replacement of a cytosolic aaRS by an organellar counterpart. A schema giving a possible evolutionary scenario is shown in Fig. 7. There is a faint possibility that there was no cytosolic AsnRS to replace;

the patchy distribution of AsnRSs suggests that this enzyme appeared relatively late in evolution, but exactly when and where it arose is not clear (Shiba et al. 1998). Until more eukaryotes have been sampled, particularly protozoa, we cannot be certain whether or not the ancestor of plants possessed an AsnRS before it received one with the acquisition of plastids.

References

Akashi K, Grandjean O, Small I (1998) Potential dual targeting of an *Arabidopsis* archaebacterial-like histidyl-tRNA synthetase to mitochondria and chloroplasts. FEBS Lett 431:39–44

Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396:133–140

Aubourg S, Cheron A, Kreis M, Lecharny A (1998) Structure and expression of an asparaginyl-tRNA synthetase gene located on chromosome IV of *Arabidopsis thaliana* and adjacent to a novel gene of 15 exons. Biochim Biophys Acta 1398:225–231

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1994) Current protocols in molecular biology. John Wiley and Sons, New York

Berthet-Colominas C, Seignovert L, Hartlein M, Grotli M, Cusack S, Leberman R (1998) The crystal structure of asparaginyl-tRNA synthetase from *Thermus thermophilus* and its complexes with ATP and asparaginyl-adenylate: The mechanism of discrimination between asparagine and aspartic acid. EMBO J 17:2947–2960

Brown JR, Doolittle WF (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc Natl Acad Sci USA 92:2441–2445

Claros MG, Vincens P (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur J Biochem 241:779–786

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

Diaz-Lazcoz Y, Aude J-C, Nitschké P, Chiapello H, Landès-Devauchelle C, Risler J-L (1998) Evolution of genes, evolution of species: The case of aminoacyl-tRNA synthetases. Mol Biol Evol 15:1548–1561

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

Doolittle WF (1998) You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet 14:307–311

Douglas SE (1998) Plastid evolution: Origins, diversity, trends. Curr Opin Genet Dev 8:655–661

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci 8:978–984

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

Gray MW (1998) *Rickettsia*, typhus and the mitochondrial connection. Nature 396:109–110

Gray MW, Burger G, Lang BF (1999) Mitochondrial evolution. Science 283:1476–1481

Handy J, Doolittle RF (1999) An attempt to pinpoint the phylogenetic introduction of glutaminyl-tRNA synthetase among bacteria. J Mol Evol 49:709–715

Hashimoto T, Sánchez LB, Shirakura T, Müller M, Hasegawa M (1998) Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny. Proc Natl Acad Sci USA 95:6860–6865

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

Kim HS, Vothknecht UC, Hedderich R, Celic I, Soll D (1998) Sequence divergence of seryl-tRNA synthetases in archaea. J Bacteriol 180:6446–6449

Köhler RH, Cao J, Zipfel WR, Webb WW, Hanson MR (1997a) Exchange of protein molecules through connections between higher plant plastids. Science 276:2039–2042

Köhler RH, Zipfel WR, Webb WW, Hanson MR (1997b) The green fluorescent protein as a marker to visualize plant mitochondria *in vivo*. Plant J 11:613–621

Kumazawa Y, Yokogawa T, Hasegawa E, Miura K, Watanabe K (1989) The aminoacylation of structurally variant phenylalanine tRNAs from mitochondria and various nonmitochondrial sources by bovine mitochondrial phenylalanyl-tRNA synthetase. J Biol Chem 264:13005–13011

Kurland CG (1992) Evolution of mitochondrial genomes and the genetic code. Bioessays 14:709–714

Lamour V, Quevillon S, Diriong S, N'Guyen VC, Lipinski M, Mirande M (1994) Evolution of the Glx-tRNA synthetase family: The glutaminyl enzyme as a case of horizontal gene transfer. Proc Natl Acad Sci USA 91:8670–8674

Lopez-Garcia P, Moreira D (1999) Metabolic symbiosis at the origin of eukaryotes. Trends Biochem Sci 24:88–93

Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik KV (1998) Gene transfer to the nucleus and the evolution of chloroplasts. Nature 393:162–165

Menand B, Marechal-Drouard L, Sakamoto W, Dietrich A, Wintz H (1998) A single gene of chloroplast origin codes for mitochondrial and chloroplastic methionyl-tRNA synthetase in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 95:11014–11019

Mireau H, Lancelin D, Small ID (1996) The same *Arabidopsis* gene encodes both cytosolic and mitochondrial alanyl-tRNA synthetases. Plant Cell 8:1027–1039

Moras D (1992) Structural and functional relationships between aminoacyl-tRNA synthetases. Trends Biochem Sci 17:159–164

Nabholz CE, Hauser R, Schneider A (1997) *Leishmania tarentolae* contains distinct cytosolic and mitochondrial glutaminyl-tRNA synthetase activities. Proc Natl Acad Sci USA 94:7903–7908

Nagel GM, Doolittle RF (1995) Phylogenetic analysis of the aminoacyl-tRNA synthetases. J Mol Evol 40:487–498

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

Reichel C, Mathur J, Eckes P, Langenkemper K, Koncz C, Schell J, Reiss B, Maas C (1996) Enhanced green fluorescence by the expression of an *Aequorea victoria* green fluorescent protein mutant in mono- and dicotyledonous plant cells. Proc Natl Acad Sci USA 93:5888–5893

Reith M, Munholland J (1995) Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. Plant Mol Biol Rep 13: 333–335

Shiba K, Motegi H, Yoshida M, Noda T (1998) Human asparaginyl-tRNA synthetase: Molecular cloning and the inference of the evolutionary history of Asx-tRNA synthetase family. Nucleic Acids Res 26:5045–5051

Small I, Wintz H, Akashi K, Mireau H (1998) Two birds with one stone: Genes that encode products targeted to two or more compartments. Plant Mol Biol 38:265–277

Small I, Akashi K, Chapron A, Dietrich A, Duchêne AM, Lancelin D, Maréchal-Drouard L, Menand B, Mireau H, Moudden Y, Ovesna J, Peeters N, Sakamoto W, Souciet G, Wintz H (1999) The strange evolutionary history of plant mitochondrial tRNAs and their aminoacyl-tRNA synthetases. J Hered 90:333–337

Strimmer K, von Haeseler A (1996) Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964–969

Swofford DL (1993) PAUP: Phylogenetic analysis using parsimony, version 3.1.1. University of Washington, Seattle

Taupin CM, Leberman R (1999) Archaeabacterial seryl-tRNA synthetases: Adaptation to extreme environments and evolutionary analysis. J Mol Evol 48:408–420

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

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

Ueda T, Yotsumoto Y, Ikeda K, Watanabe K (1992) The T-loop region of animal mitochondrial $tRNA^{Ser}$(AGY) is a main recognition site for homologous seryl-tRNA synthetase. Nucleic Acids Res 20: 2217–2222

Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nature Genet 15:57–61

Wolf YI, Aravind L, Grishin NV, Koonin EV (1999) Evolution of aminoacyl-tRNA synthetases-analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. Genome Res 9:689–710???