ELSEVIER

Gene 243 (2000) 105–114

GENE
AN INTERNATIONAL JOURNAL ON GENES, GENOMES AND EVOLUTION

www.elsevier.com/locate/gene

Alfalfa (*Medicago sativa*) carbamoylphosphate synthetase gene structure records the deep lineage of plants

Zongmin Zhou [a,c], Anthony E. Metcalf [a], Carol J. Lovatt [c], Bradley C. Hyman [a,b,c,*]

[a] *Department of Biology, University of California, Riverside, CA 92521-0201, USA*
[b] *Interdisciplinary Ph.D. Program in Genetics, University of California, Riverside, CA 92521-0201, USA*
[c] *Department of Botany and Plant Sciences, University of California, Riverside, CA 92521-0124, USA*

Received 8 September 1999; received in revised form 19 November 1999; accepted 3 December 1999
Received by A. Dugaiczyk

Abstract

Given the central role of carbamoylphosphate synthetases in pyrimidine and arginine metabolism in all living organisms, the absence of fundamental information regarding plant CPSase genes is a striking omission [Lawson et al., Mol. Biol. Evol. 13 (1996) 970–977; van den Hoff et al., J. Mol. Evol. 41 (1995) 813–832]. Whereas CPSase gene architecture and aa sequence have proven to be useful characters in establishing ancient and modern genetic affinities, phylogenetic analysis cannot be completed without the inclusion of plant CPSases. We describe the first isolation by molecular cloning of a plant CPSase gene (*CPAII*) derived from alfalfa (*Medicago sativa*). DNA sequence analysis reveals a proteobacterial architecture, namely closely linked *carA* and *carB* coding domains separated by a short intergenic region, and transcribed as a polycistronic mRNA. *CPAII* encodes the amino acid residues that typify a CPSase type II enzyme. In addition, an ancient internal duplication has been retained in the plant *carB* sequence. Partial nucleotide sequencing of additional clones reveals that the alfalfa genome contains multiple CPSase II gene copies which may be tissue-specific in their expression. It appears that with respect to CPSase genes, *CPAII* resembles the *carAB* gene of bacteria, and may have preserved much of this ancient gene structure in the alfalfa genome.

Keywords: Arginine metabolism; Evolution; Gene structure; Pyrimidine metabolism

1. Introduction

Carbamoylphosphate synthetase (CPSase) catalyzes the formation of carbamoylphosphate (CP), a key intermediate in the pathways leading to the biosynthesis of both pyrimidine nucleotides and arginine. *Escherichia coli* and other proteobacteria have a single CPSase enzyme [CPSase II, L-glutamine-dependent; carbon dioxide:L-glutamine amido ligase (ADP-forming, carbamate-phosphorylating); EC 6.3.5.5] that provides CP to each pathway. Independent gene duplication events within the bacterial and eukaryotic lineages have increased CPSase gene copies in eubacteria and fungal eukaryotes, which have two CPSase II enzymes that are regulated separately for pyrimidine and arginine biosynthesis (Paulus and Switzer, 1979; Lusty et al., 1983; Nyunoya and Lusty, 1984; Souciet et al., 1989).

CPSase II catalyzes the first step committed to pyrimidine biosynthesis. Enzymatic activity can be part of a multifunctional protein that also contains aspartate carbamyltransferase (ACTase) as in *Saccharomyces cerevisiae* (Souciet et al., 1989), or both ACTase and dihydroorotase (DHOase) as in *Dictyostelium discoideum* (Faure et al., 1989), *Drosophila melanogaster* (Freund and Jarry, 1987), and mammals (Mori and Tatibana, 1975; Coleman et al., 1977; Davidson and Patterson, 1979). Ureotelic mammals have an arginine pathway-specific CPSase I [carbon dioxide; ammonia ligase (ADP-forming, carbonate-phosphorylating); EC 6.3.4.16] that, as a rule, is a mitochondrial constituent, utilizes ammonia rather than glutamine, and is activated by N-acetylglutamate (NAG) (Lawson et al., 1996). A

Abbreviations: aa, amino acid; bp, base pair(s); CP, carbamoyl phosphate; CPSase, carbamoylphosphate synthetase; dNTP, deoxynucleotide triphosphate; FUE, far upstream element; kb, kilobase pair(s); M_r, relative molecular mass; NUE, near upstream element; ORF, open reading frame; polyA$^+$, polyadenylated mRNA; RPA, RNAase protection assay; RT-PCR, reverse transcriptase polymerase chain reaction; SD, Shine–Delgarno; TBE, Tris–borate–EDTA; u, unit.

* Corresponding author. Tel.: +1-909-787-5911; fax: +1-909-787-4286.
E-mail address: bhyman@citrus.ucr.edu (B.C. Hyman)

0378-1119/00/$ - see front matter
PII: S0378-1119(99)00552-1

third type, CPSase III, which is more closely related to CPSase I than CPSase II, has been identified in invertebrates and fish species (Anderson, 1980; Hong et al., 1994). CPSase III utilizes glutamine as the nitrogen-donating substrate similar to CPSase II, but like CPSase I requires NAG for catalytic activity. CPSase III appears to represent an intermediate in the evolution of the ancestral CPSase II to the CPSase I of ureotelic terrestrial vertebrates (Hong et al., 1994). Based solely on enzymatic studies, vascular plants are generally considered to have only a single CPSase II (Ong and Jackson, 1972; Jacques and Sung, 1981; Doremus and Jagendorf, 1987). This stands in contrast to other eukaryotes, even those as simple as *S. cerevisiae* and *D. discoideum*, and to some prokaryotes, including the eubacteria (e.g. *Bacillus subtilis*). Is it that investigations of enzyme activity failed to detect a second CPSase activity in plants because the plant arginine pathway-specific CPSase is also a glutamine-hydrolyzing CPSase II as in *B. subtilis* and the lower eukaryotes *S. cerevisiae* or *N. crassa*? Alternatively, did evolutionary events result in plants utilizing a strategy different from that of other eukaryotes and like that of proteobacteria?

The possibility of a second plant CPSase (mitochondrial, arginine-specific) has been proposed (Ludwig, 1991). Employing polyclonal antiserum raised against purified *N. crassa* CPSase-A (arginine pathway-specific CPSase large subunit), protein transfer analyses revealed cross-reactive polypeptides from cauliflower, tomato, rice, pea, and maize and a single mitochondrial protein from *N. crassa* (Ludwig, 1991). This report of a second plant CPSase awaits confirmation. Only one gene for each subunit of a single bacterial-like CPSase has been registered for a plant species in GenBank (*Arabidopsis thaliana*, accession nos. U40341, U73175; Williamson et al., 1996; Brandenburg et al., 1998).

There are currently over 20 complete CPSase gene sequences, including that of an Archaeon, available for phylogenetic analysis. Given the fundamental importance of pyrimidine nucleotide and arginine metabolism to all living organisms, we and others (Lawson et al., 1996; van den Hoff et al., 1995) view the lack of knowledge regarding the number of functional CPSases in plants as a startling omission. Towards resolution of this limitation, we initiated a study to determine the number and types of CPSase(s) encoded in alfalfa (*Medicago sativa*), a plant of economic importance with a complex, autotetraploid genome. Here we report the first, to our knowledge, description of plant CPSase genes.

2. Materials and methods

2.1. Plant material

Seeds of *M. sativa* L., 'Saranac' (USDA-ARS, Beltsville, MD), 'Chief' (W-L Research, Inc., Bakersfield, CA), and 'Mesilla' (New Mexico State University, Las Cruces, NM) were used in this study. Surface sterilized seeds were imbibed in distilled water overnight at room temperature, rinsed, and spread on cheese cloth saturated with a balanced salt solution in a glass baking dish. Seeds were germinated under ambient conditions in the laboratory. After 7 days, seedlings were harvested and used as the source of RNA or transferred to 1 l pots containing a standard soil mix.

2.2. Identification and isolation of alfalfa CPSase genes

A *M. sativa* (cv. Chief) genomic DNA fragment library was constructed in the bacteriophage EMBL3 vector (provided by Ann Hirsch, UCLA). Plaque hybridizations (Sambrook et al., 1989) were conducted using a 5.3 kb alfalfa cDNA fragment ^{32}P-labeled in vitro by random priming (Prime-It II, Stratagene, La Jolla, CA) known to complement an *E. coli carAB*$^{-}$ mutation (Hyman, unpublished results). Transfer hybridization procedures (Sambrook et al., 1989) were used to identify CPSase-containing EcoR1 subfragments of candidate genomic clones using the same probe. Putative CPSase subfragments were subcloned into EcoR1-digested pBluescript II SK(+) (Stratagene, La Jolla, CA) vector DNA.

2.3. Nucleotide sequencing and sequence analysis

Nucleotide sequencing was conducted using either the Sequenase ver. 2.0 DNA sequencing kit (Amersham Pharmacia Biotech, Inc., Piscataway, NJ) or by PCR using the OmniBase™ DNA sequencing system (Promega, Madison, WI). Initial sequence alignments of predicted CPS aa sequences were performed in Clustal X (Jeanmougin et al., 1998) and imported into MacClade (Maddison and Maddison, 1991). The central amidotransferase domain, corresponding to aa residues 38 to 367 of *E. coli carA* and the synthetase domain corresponding to residues 11 to 916 of *E. coli carB* (Lawson et al., 1996) were used for phylogenetic analysis. The two domains were merged and exported as a single sequence into MEGA for neighbor-joining analysis. Substitutions per site were estimated assuming a gamma distribution with parameter $(a)=2$. Reliability of internal branches was tested using 1000 bootstrap replications of the data set (Felsenstein, 1985).

2.4. Alfalfa DNA and RNA preparation

Total genomic DNA was isolated from developing leaves harvested from mature plants using the procedure of Saghai-Maroof et al. (1984). Total RNA from 7 day old seedlings or developing leaves from mature plants was isolated using RNeasy™ Plant Total RNA kit (Qiagen, Chatsworth, CA). PolyA^{+} mRNA was purified

from total RNA using a magnetic bead procedure (DYNAL, Lake Success, NY).

2.5. Transfer hybridization and RPA

Genomic DNA was digested with restriction enzymes following manufacturer's recommendations. After fractionation on 0.8% TBE-buffered agarose gels (Sambrook et al., 1989), DNA was blotted onto Hybond-N$^+$ nylon membranes (Amersham). Hybridization was performed using the ExpressHyb kit (Clontech, Palo Alto, CA). Hybridization signals were detected via autoradiography using Kodak BioMax X-ray film for 2–5 days, enhanced at -80°C with an intensifying screen. Hybridization probes were generated by the Prime-It II™ random priming kit (Stratagene, La Jolla, CA). RNA transfer hybridization procedures were performed in a similar fashion on denaturing agarose gels (Sambrook et al., 1989). Ribonuclease protection assays were conducted using a 121 bp EcoR1 alfalfa *carB* fragment (nucleotides 2914–3035) cloned into pBluescript II SK. Riboprobes were generated from this template using T3 or T7 RNA polymerases and assays performed using the RPAII kit (Ambion, Austin, TX).

2.6. PCR and RT-PCR

PCR reactions were performed in a Perkin-Elmer 480 thermal cycler (Norwalk, CO). For RT-PCR, first strand cDNA was synthesized from polyA$^+$ mRNA using the RETROscript kit (Ambion). Second strand synthesis by PCR was conducted in a 50 μl reaction volume containing 1 μM CPSase-specific primers, 1 u Taq DNA polymerase (Promega, Madison, WI), 0.2 μm dNTPs and 1× RT-PCR buffer (Ambion). Amplification was carried out for 35 cycles at 95°C for 1 min, 54°C for 1 min, and 72°C for 1.5 min.

3. Results and discussion

3.1. Identification and sequence analysis of a M. sativa CPSase gene

An alfalfa genomic DNA library was screened by plaque hybridization using a 5.3 kb alfalfa partial cDNA fragment known to complement an *E. coli carAB*$^-$ mutation. Cloned alfalfa genomic DNA fragments present in pEMBL-3 vectors were rescued from five plaques generating a positive hybridization signal. Internal *Eco*R1 restriction fragments hybridizing to the same cDNA probe were subcloned into the *E. coli* plasmid vector pBlueScript II KS for subsequent analysis.

A contiguous nucleotide sequence representing 5331 bp from one alfalfa genomic fragment (termed p7-6) encompassed a complete CPSase coding region (GenBank accession no. AF191301). The CPSase gene is composed of two ORFs (Fig. 1) separated by a 97 bp non-coding intergenic sequence (Fig. 2) that does not share similarity in nucleotide sequence with other known CPSase intergenic regions. The ORF positioned 5′ to the intergenic sequence contains 1212 nucleotides and encodes a putative 403 aa peptide with a predicted M_r of 43.4 kd (Fig. 1). This aa sequence shares significant sequence similarity to proteobacterial CPSase *carA* coding sequences known to encode glutamine amidotransferase enzymatic activity (Fig. 1). For example, the predicted *carA* protein reveals a 67% amino acid identity with the *E. coli carA* peptide. A 3216 ORF residing 3′ to the intergenic region is positioned as a +2 frameshift relative to the *carA* sequence (Fig. 2), encodes a putative 1072 aa polypeptide (predicted $M_r = 117.6$ kd), and shares 72% amino acid sequence similarity to the *E. coli carB* protein, the synthetase activity of CPSase (Fig. 1).

Inspection of the predicted alfalfa *carA* and *carB* polypeptides reveals features characteristic of a CPSase II (the glutamine-dependent, pyrimidine pathway-specific enzyme). These include the presence of highly conserved cysteine, histidine and glutamate aa residues at aa positions 293, 377 and 379, respectively, within the *carA* domain. The absence of conserved cysteines at aa residues 908 and 918 in the downstream *carB* domain, which are considered signatures of CPSase I and CPSase III enzymes (Lawson et al., 1996), further supports the conclusion that the genomic p7-6 clone encodes a CPSase II gene.

As expected, residues thought to be involved in ATP binding are present in the *carB* synthetase domain, extending from amino acids 303 to 353 and from 844 to 885. Moreover, there is an internal duplication of the first and second thirds of the *carB* domain when residues 3 to 398 are aligned with amino acids 554 to 930 (Fig. 3). This alignment reveals that 36% of the amino acid residues are identical, a value similar to that determined for *carB* internal duplications within CPSase genes from other organisms including yeast (28%; Souciet et al., 1989) and *E. coli* (39%; Nyunoya and Lusty, 1983). Although this duplication is thought to be ancient, having preceded the divergence of bacteria, the Archaea, and eukaryotes, we provide the first evidence that such a duplication is also present in the plant lineage.

The overall sequence organization of this *M. sativa* CPSase gene is highly reminiscent of proteobacterial CPSase gene architecture. For example, the *E. coli carA–carB* gene cluster encodes amidotransferase and synthetase genes separated by a 17 bp intergenic spacer (Fig. 2) which places the downstream *carB* (synthetase domain) in a translational reading frame different from that of the upstream *carA* coding sequence. The sequence organization of this *M. sativa* CPSase gene found within p7-6, coupled with the high degree of nucleotide sequence similarity when aligned with proteobacterial CPSases, raised the possibility that we had inadvertently cloned

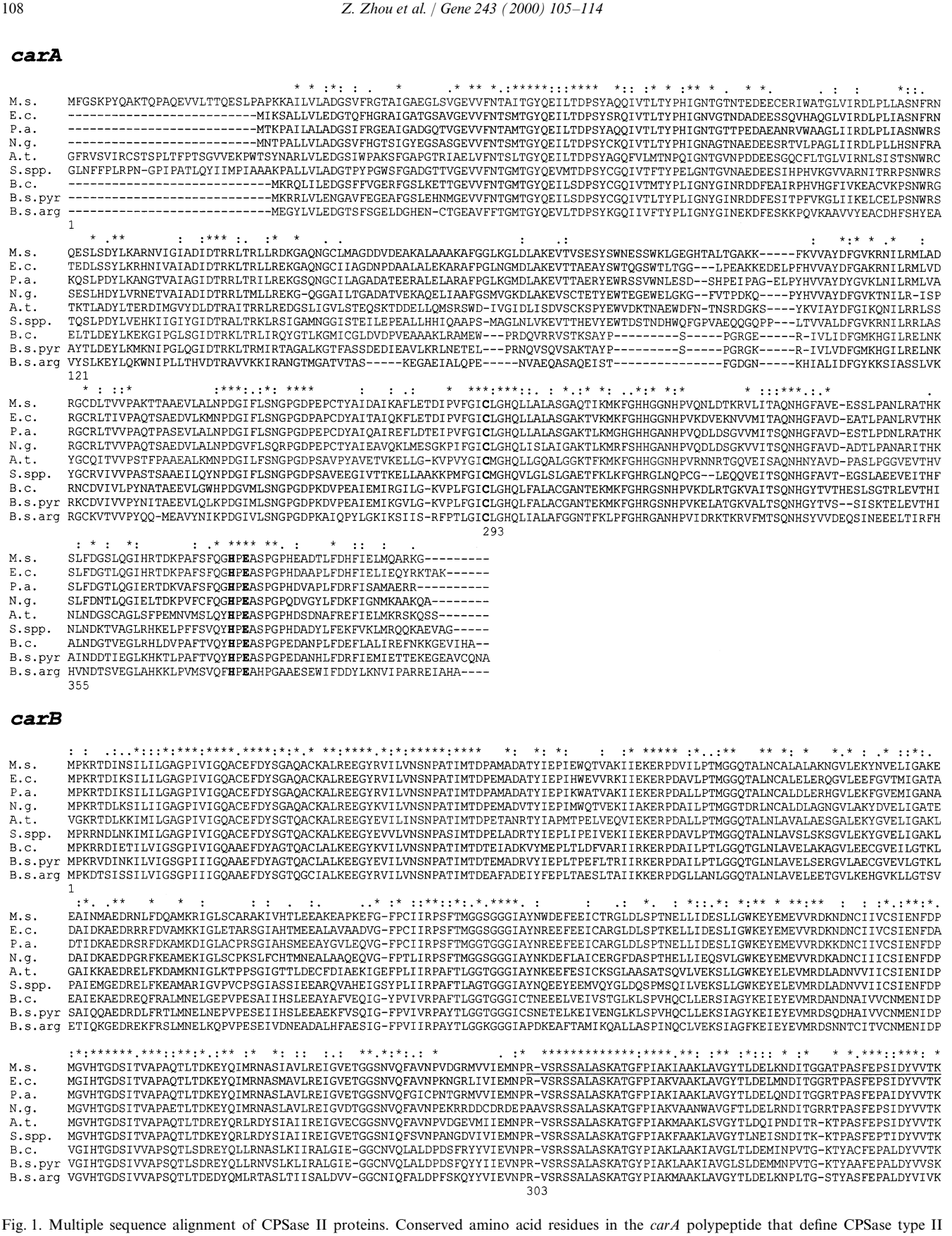

Fig. 1. Multiple sequence alignment of CPSase II proteins. Conserved amino acid residues in the *carA* polypeptide that define CPSase type II enzymes are in bold upper case letters; positions that would be occupied by cysteines in the *carB* protein that define CPSase type I and III enzymes are in italicized bold lower case letters. Long underlines indicate conserved ATP binding domains. Asterisks denote amino acid identity among all taxa; colon and period identify conservative substitutions. Numbers below alignments define critical aa residues for alfalfa *CPAII*, beginning with the translational start methionine. The aligned sequences represent members of the alfalfa *CPAII* clade presented in Fig. 3. M. s. (*M. sativa*); E. c. (*E. coli*); P. a. (*P. aeruginosa*); N. g. (*N. gonorrhoeae*); A. t. (*A. thaliana*); S. spp. (*Synechocystis* spp.); B. s. (*B. subtilis*); B. c. (*B. caldolyticus*). See legend to Fig. 5 for additional taxonomic descriptions.

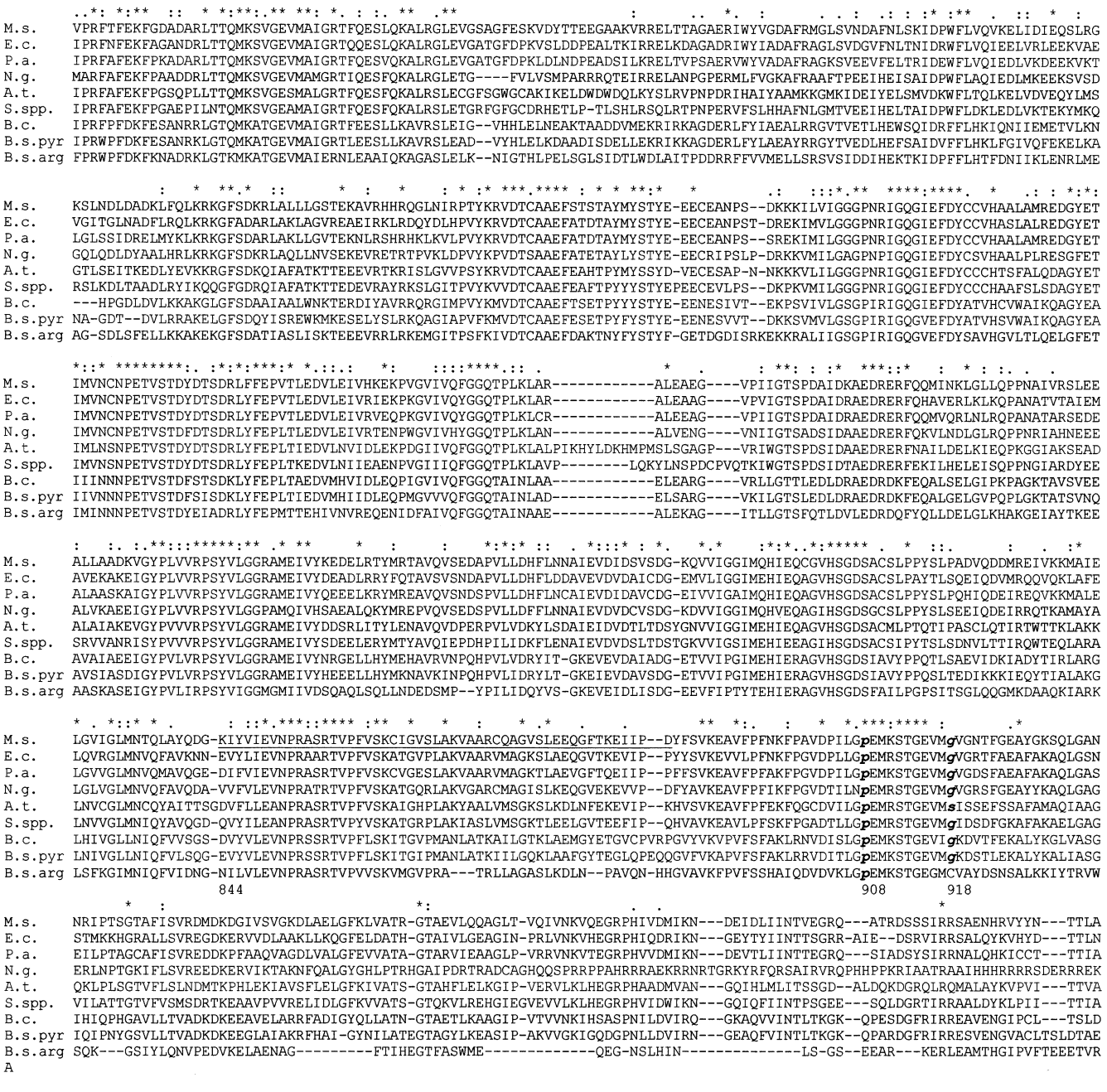

Fig. 1. (*continued*)

a CPSase gene derived from a prokaryotic contaminant. However, 731 bp of sequence information 3′ to the putative *carB* translational stop codon and 71 bp of available genomic sequence 5′ to the *carA* start codon do not share significant sequence similarity to any entry in GenBank. To address this concern further, we undertook a detailed inspection of the alfalfa sequence for appropriately placed proteobacterial translation signals. In *E. coli*, independent SD sequences are required for translation of each peptide to form a heterodimeric CPSase. Exact or near matches to the prokaryotic AGGAGG ribosome binding site were not observed in

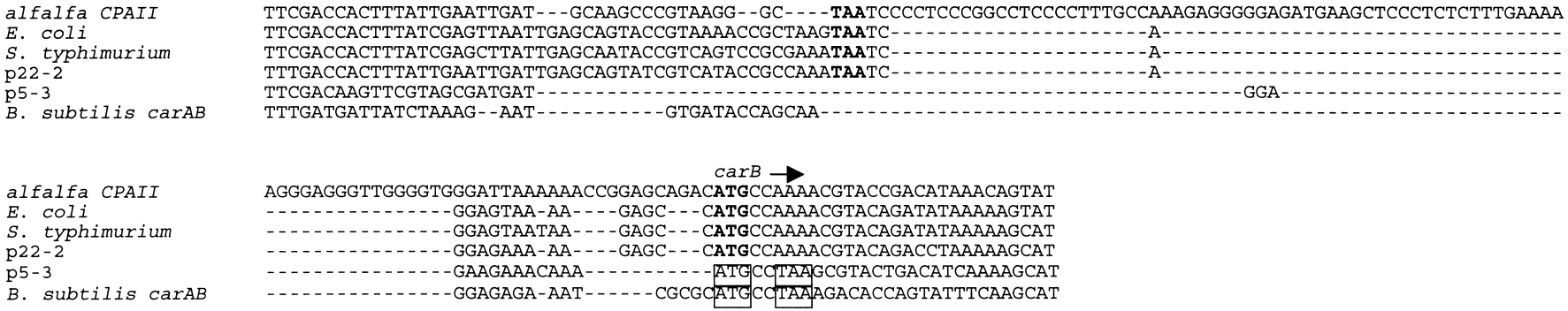

Fig. 2. Comparison of intergenic regions with CPSase II genes. Bold translational termination (TAA) and start (ATG) codons delimit frameshifted *carA* and *carB* coding domains separated by an intergenic region. Boxed ATG and TAA codons identify overlapping *carA* and *carB* genes.

```
  3-92    KRTDINSILILGAGPIVIGQACEFDYSGAQACKALREEGYRVILVNSNPATIMTDPAMADATYIEPIEWQTVAKIIEKERPDVILPTMGG
          :.:| :.||::|:||  |||: |||| .:|  |:||:|| .|:||.||.|: || . :|  ::||:. :.| .|::||:|  ::  .||
554-643   NPSDKKKILVIGGGPNRIGQGIEFDYCCVHAALAMREDGYETIMVNCNPETVSTDYDTSDRLFFEPVTLEDVLEIVHKEKPVGVIVQFGG

 93-182   QTALNCALALAKNGVLEKYNVELIGAKEEAINMAEDRNLFDQAMKRIGLSCARAKIVHTLEEAKEAPKEFGFPCIIRPSFTMGGSGGGIA
          ||:|:    ||:  .||  .| :||:. :||:.||||: |:| ::::||  :.. ||::|||| |:.. |:| ::|||:.:||.: .|.
644-727   QTPLK----LAR--ALEAEGVPIIGTSPDAIDKAEDRERFQQMINKLGLLQPPNAIVRSLEEALLAADKVGYPLVVRPSYVLGGRAMEIV

183-272   YNWDEFEEICTRGLDLSPTNELLIDESLLGWKEYEMEVVRDKNDNCIIVCSIENFDPMGVHTGDSITVAPAQTLTDKEYQIMRNASIAVL
          |: ||: .    ::::| .. :|:|: | .  | ::: |.| ::  :|   :::::. |||:|||    |: :|... : ||:    :
728-816   YKEDELRTYMRTAVQVSEDAPVLLDHFLNNAIEVDIDSVSDGKQV-VIGGIMQHIEQCGVHSGDSACSLPPYSLPADVQDDMREIVKKMA

273-362   REIGVETGGSNVQFAVNPVDGRMVVIEMNPRVSRSSALASKATGFPIAKIAAKLAVGYTLDELKNDITGGATPASFEPSIDYVVTKVPRF
           |:|| .|  |.|:| :  ||:: |||:|||.||: ::.|| .| ::||:||: ..| :|:|  :::| . .|. |       .| : |
817-894   IELGV-IGLMNTQLAYQ--DGKIYVIEVNPRASRTVPFVSKCIGVSLAKVAARCQAGVSLEE--QGFTKEIIPDYF-------SVKEAVF

363-398   TFEKFGDADARLTTQMKSVGEVMAIGRTFQESLQKA
          .|:|| ..|: |..:|||.||||::|.|| |:  |:
895-930   PFNKFPAVDPILGPEMKSTGEVMGVGNTFGEAYGKS
```

Fig. 3. Identification of the ancient duplication within the alfalfa *CPAII carB* coding region. Lipman–Pearson protein alignment was performed by aligning the 396 aa N-terminal portion of the *carB* polypeptide (aa 3 to 398) with the 377 residue C-terminal 'half' of the *carB* protein (aa 554–930). Vertical lines denote amino acid identities; colons or periods denote conservative replacement substitutions.

relevant locations upstream of either putative methionine start codons, further suggesting that our CPSase clone was not of proteobacterial origin. We have termed this *M. sativa* CPSase coding sequence *CPAII*.

3.2. Gene expression

Given the requirement for high levels of CPSase II expression in rapidly growing tissues, which include developing leaves and seedlings, we next examined *CPAII* mRNA levels in 7 day old seedlings. Using convergently-oriented primers that target nucleotide residues 720–734 and 1373–1398 present in the genomic sequence, a 668 bp product that encompasses the C-terminal 188 codons of *carA*, the 97 bp intergenic region, and the first three codons of *carB*, was successfully amplified using RT-PCR from 7 day old seedling polyA$^+$ mRNA populations. We infer from this result that *CPAII* is transcribed as a polycistronic mRNA. The identity of this cDNA as a *CPAII* transcript was confirmed by DNA sequencing. However, we were unable to detect *CPAII* mRNA in the same mRNA preparations using the less-sensitive RPA assay, suggesting that *CPAII* is not the primary supplier of CPSase to developing seedlings. We do not yet understand the reason for the low level of *CPAII* transcription. With only 71 bp of available genomic sequence upstream of the *carA* coding domain, it is not yet possible to identify elements responsible for transcriptional control. Within the 3′ untranslated sequence downstream to the putative translational stop codon for *carB* (Fig. 4), two exact matches to consensus plant poly(A) signals, termed near upstream elements NUE(s) (Rothnie, 1996) are observed. One is an AAUAAA box at position 4594, five nucleotides downstream from the *carB* termination codon. Moreover, plant poly(A) signals are typically modular (Wu et al., 1995), and are composed of multiple elements including a U/G-rich far upstream element (FUE), a NUE, and one or more (C/U)A cleavage/polyadenylation sites residing 11 to 30 nucleotides 3′ to the NUE. Such a set of elements can be found downstream of the *carB* coding sequence (Fig. 4). A U/G-rich FUE-like element occupies positions 4769–4775, followed by a NUE consensus sequence AAUGAAA beginning at nucleotide 4790, and then flanked by cleavage sites 8, 17, 24 and 29 nucleotides downstream from this NUE. The presence of an appropriately positioned, complete collection of plant mRNA 3′ end processing signals further argues against the artifactual cloning of a proteobacterial CPSase gene.

In support of our conclusion that the *CPAII* gene is not the primary supplier of CPSase to developing seedlings, it is unlikely that the low abundance *CPAII* mRNA is efficiently translated. The putative *carA* and *carB* translational start codons (positions 1 and 1310, respectively) are not flanked by nucleotides generally considered in a consensus context for dicots (Joshi et al., 1997). Interestingly, the *Arabidopsis carA* cDNA sequence registered in GenBank does not appear to be expressed in leaves (Brandenburg et al., 1998). Similarly, expression of a putative *Alnus carA* gene identified in a root nodule cDNA preparation was low to undetectable in other tissues (Lundquist et al., 1996).

We propose that plants must contain additional CPSase genes that are likely expressed in a tissue specific fashion. Preliminary evidence indicates that *CPAII* is one member of a small alfalfa CPSase gene family. We have obtained partial nucleotide sequences from two additional, independently cloned *Medicago* genomic segments (Fig. 2). These alfalfa genomic clones exhibit two different types of CPSase coding sequence architectures. One (p22-2) contains adjacent, frameshifted *carA* and *carB* genes separated by a 17 bp intergenic region reminiscent of alfalfa *CPAII*. The genomic segment encoded

```
4570                                          4594
TTG CAA GAT TTG CAC ACA CGA ATC AAA TAA AGGAAATAAAAATGAGTACAAAATTTCCGATGACCGAACAAGGTGCAGAAAAACT
leu gln asp leu his thr arg ile lys  •  <-carB

GCGCCTTGAGTTGGAAGACCTCAAAAAAGTACAGCGCCCACGTATTGTGCAAGCAATTGCTGAAGCACGTGAGCACGGAGACCTCAAAGAAAACGCTGAATAT

                    4769
TCTGCGGCGCGCGAGCAACAAAGTTTTTGCGAAGGTCGCATTAATGAAATCGAAGGTAAGCTGGGTATGCCCAGATTATCGATGTGACCAAAATCCCGCACA
```

Fig. 4. Nucleotide sequence downstream of the *carB* coding sequence in alfalfa *CPAII*. Boxed sequences represent plant mRNA 3′ processing signals (see text for details).

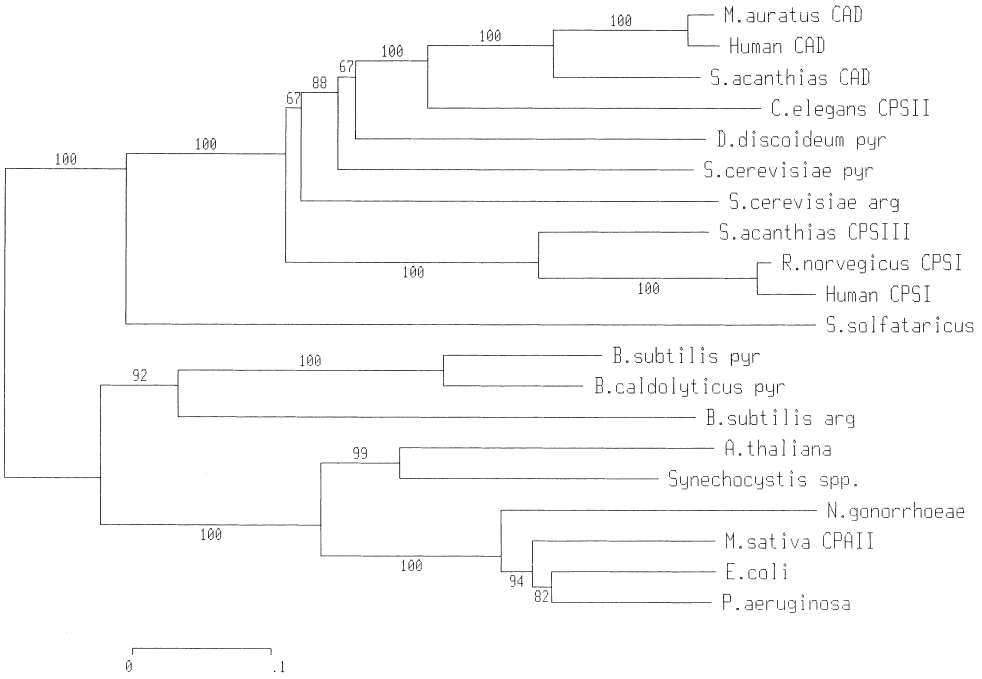

Fig. 5. Phylogenetic neighbor-joining distance tree of complete (merged *carA* and *carB*) CPS protein sequences. Confidence levels are given on each branch. Scale bar, branch length of 0.01 substitutions per position following gamma distance corrections for site variation of aa substitution rate. Sequences employed with accession nos. include: *M. auratus* (hamster, P08955); human (D78586/D90282), *S. acanthias* (shark, U18868/L31362); *C. elegans* (nematode, Z54284); *D. discoidium* (slime mold, X55433); *S. cerevisiae* (baker's yeast, M27174/K01178/K02132); *R. norvegicus* (rat, M11710); *S. solfataricus* (archaeon, U33768); *B. subtilis* (eubacteria, Z26919/M59757); *B. caldolyticus* (eubacteria, X73308); *A. thaliana* (plant, U40341/ U73175); *Synechocystis spp.* (cyanobacteria, D64002); *N. gonorrhoeae* (proteobacteria, U11295); *E. coli* (proteobacteria, J01597); *P. aeruginosa* (proteobacteria, U81259).

by p5-3 is organized as overlapping *carA* and *carB* reading frames; the *carA* TAA stop codon resides just downstream from the *carB* AUG translational start triplet. This overlapping organization is similar to CPSase gene structure found in some eubacteria (Lawson et al., 1996). These two alfalfa genes also appear to encode CPSase type II enzymes based on the presence of highly conserved cysteine, histidine and glutamate residues in the *carA* region and the absence in the *carB* domain of cysteines that typify CPSase I and III enzymes (data not presented). One clone (p22-2) is capable of generating a robust signal when used as a hybridization probe to filter-immobilized total root and leaf RNA (Potenza, unpublished results), and may represent the gene that is the primary source of CPSase within alfalfa root and leaf tissue. It is tempting to compare the organization of *CPAII* and the CPSases specified by p22-2 and p5-3 (Fig. 2) to that of bona fide bacterial operons. At present there are no known operons in the plant nuclear genome, but recent research regarding specific plant gene clusters has provided evidence reminiscent of operons (Frey et al., 1995, 1997). Moreover, the genomes of the mitochondrion and chloroplast are replete with operons (Gillham, 1994). Thus, the *carAB* genes of the alfalfa CPSase gene family reported here are consistent with linked gene organizations documented in both plant nuclear and organellar genomes.

3.4. Phylogenetic considerations

Computer-assisted translations of the *carA* and *carB* domains of *CPAII* were aligned at the amino acid level with CPSase peptides representing a wide range of taxa.

Fig. 5 depicts a neighbor-joining distance analysis from aligned aa sequences for these domains, after removing taxon-specific N- and C-terminal residues (Lawson et al., 1996). *M. sativa carA* and *carB* CPSase domains cluster with a well-supported clade composed of CPSase polypeptides representing the proteo- and eubacteria and to the exclusion of CPSases from the Archaea and all other eukaryotic sources tested. The alfalfa *CPAII* gene reveals a functional organization reminiscent of CPSase genes from proteobacteria, namely *carA* and *carB* coding domains separated by an intergenic region. Therefore, within the bacterial clade, positioning of *Medicago CPAII* within the proteobacteria was anticipated. Equivalent distance results were also achieved with heuristic searches using either minimum evolution or maximum parsimony methods (data not presented).

As with *M. sativa* CPSase, *Arabidopsis carA* and *carB* sequences also cluster within the same clade that comprises the proteo- and eubacterial CPSase genes. Neighbor-joining analysis indicates an affinity between the *Arabidopsis* and *Synechocystis* (cyanobacteria) CPSase sequences. In *Synechocystis*, the *carA* and *carB* sequences are unlinked in the genome (Kaneko et al., 1996). Using information made available from the *Arabidopsis* genome sequencing initiative, we suggest that the *carA* and *carB* coding sequences are also unlinked in the *A. thaliana* genome. BLASTN searches localize the *carA* gene to nucleotides 56122–56352 within a 84 702 contiguous sequence derived from *A. thaliana* chromosome III (accession no. AB018114). *Arabidopsis carB* does not align to this sequence, nor to any sequence yet available in the *Arabidopsis* genome database. The inability to position the *carB* coding sequences within the 28.5 kb of nucleotide sequence downstream of the *carA* gene suggests that these two coding sequences are also separated within the *Arabidopsis* genome. Within this limited set of two plant CPSase gene sequences, there is an interesting coincidence between the physical linkage of *carA* and *carB* coding regions in the plant genome and phylogenetic affinities as determined by sequence alignment. In both *Arabidopsis* and *Synechocystis* these two genes are unlinked, whereas in *Medicago* and the bacteria they are adjacent. These shared characters are also revealed in our distance analysis (Fig. 5). Clearly, CPSase gene organization has undergone considerable divergence within the plant lineage.

Reports of a single CPSase activity in plants conflict with the recognized view of multiple CPSase enzymes in other eukaryotes. Until now, no studies at the molecular level have been conducted to deduce the number and type of CPSase genes in plants. A thorough understanding of how CPSase genes evolved cannot be completed until information from plants is available. In this report, we present the results of initial studies intended to fill this void. The discovery of multiple CPSase genes in *M. sativa* complicates resolving the question at the level of the plant genome as to whether plants have only one CPSase enzyme or pyrimidine and arginine pathway CPSase-specific enzymes.

The greater similarity between plant and bacterial genes relative to affinities between plant and other eukaryotic sequences is not surprising. Reports of plant genes that are highly homologous to bacterial genes are increasing (Pear et al., 1996; Taylor et al., 1999). Specifically, the *Arabidopsis carA* gene contains putative upstream regulatory elements characteristic of the *E. coli carAB* operon promoter region (Slocum, 1999). Importantly, characterization of plant genes that specifically encode enzymes involved in pyrimidine metabolism reveal sequence affinities that are more similar to prokaryotic sequences than they are to cognate eukaryotic coding regions (Kafer and Thornburg, 1999). Whether this was the state of evolution when plants branched off from the eukaryotes (Kafer and Thornburg, 1999) or represents the results of interkingdom DNA transfer (Zambryski, 1992; Katz, 1998) is unknown at present. Resolving the questions of the numbers and types of functional CPSases in plants, and the association of CPSase genes with nuclear or organellar genomes, should contribute towards addressing these alternatives.

Acknowledgements

B.C.H. and C.J.L. acknowledge the following sources of support: USDA NRICGP Grant No. 94-37305-0798, the Southwest Consortium on Plant Genetics and Water Resources, the California Citrus Research Board, the University of California Riverside Citrus Research Center and Agricultural Experiment Station and the Academic Senate Committee on Research. We thank Dr. Julie Maley, Pamela Gosch, and Maria Avitia for participating in the initial cloning experiments.

References

Anderson, P.M., 1980. Glutamine- and N-acetylglutamate-dependent carbamoyl phosphate synthetase in elasmobranchs. Science 208, 291–293.

Brandenburg, S.A., Williamson, C.L., Slocum, R.D., 1998. Characterization of a cDNA encoding the small subunit of *Arabidopsis* carbamoyl phosphate synthetase. Plant Physiol. 117, 717.

Coleman, P.F., Suttle, D.P., Stark, G.R., 1977. Purification from hamster cells of the multifunctional protein that initiates de novo synthesis of pyrimidine nucleotides. J. Biol. Chem. 252, 6379–6385.

Davidson, J.N., Patterson, D., 1979. Alteration in structure of multifunctional protein from Chinese hamster ovary cells defective in pyrimidine biosynthesis. Proc. Natl. Acad. Sci. USA 76, 1731–1735.

Doremus, H.D., Jagendorf, A.T., 1987. Site of synthesis of the enzymes of the pyrimidine biosynthetic pathway in oat leaves. Plant Physiol. 83, 657–658.

Faure, M., Camonis, J.H., Jacquet, M., 1989. Molecular characterization of a *Dictyostelium discoideum* gene encoding a multifunctional enzyme of the pyrimidine pathway. Eur. J. Biochem. 179, 345–358.

Felsenstein, J., 1985. Confidence levels on phylogenies: an approach to using the bootstrap. Evolution 38, 16–24.

Freund, J.N., Jarry, B.P., 1987. The *rudimentary* gene of *Drosophila melanogaster* encodes four enzymic functions. J. Mol. Biol. 193, 1–13.

Frey, M., Kliem, R., Saedler, H., Gierl, A., 1995. Expression of cytochrome P450 gene family in maize. Mol. Gen. Genet. 246, 100–109.

Frey, M., Chomet, P., Glawischnig, E., Stettner, C., Grün, S., Winklmair, A., Eisenreich, W., Bacher, A., Meeley, R.B., Briggs, S.P., Simcox, K., Gierl, A., 1997. Analysis of a chemical defense mechanism in grasses. Science 277, 696–699.

Gillham, N.W., 1994. Organelle Genes and Genomes. Oxford University Press, New York. 424 pp.

Hong, J., Salo, W.L., Lusty, C.J., Anderson, P.M., 1994. Carbamyl phosphate synthetase III, an evolutionary intermediate in the transition between glutamine-dependent and ammonia-dependent carbamyl phosphate synthetases. J. Mol. Biol. 243, 131–140.

Jacques, Z., Sung, R., 1981. Regulation of pyrimidine and arginine biosynthesis investigated by the use of phaseolotoxin and 5-fluorouracil. Plant Physiol. 67, 287–291.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., Gibson, T.J., 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. 23, 383–402.

Joshi, C.P., Zhou, H., Huang, X., Chiang, V.L., 1997. Context sequences of translation codon initiation in plants. Plant Mol. Biol. 35, 993–1001.

Kafer, C., Thornburg, R., 1999. Pyrimidine metabolism in plants. Paths to Pyrimidines 7, 14–24.

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Tekeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., Tabata, S., 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* spp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res. 3, 109–136.

Katz, L., 1998. Changing perspectives on the origin of eukaryotes. Tree 13, 493–497.

Lawson, F.S., Charlebois, R.L., Dillon, J.-A.R., 1996. Phylogenetic analysis of carbamoylphosphate synthetase genes: complex evolutionary history includes an internal duplication within a gene which can root the tree of life. Mol. Biol. Evol. 13, 970–977.

Ludwig, R.A., 1991. Plants shuttle CO_2 equivalents as carbamoyl phosphate from mitochondria to chloroplasts. In: Huang, A.H.C., Taiz, L. (Eds.), Molecular Approaches to Compartmentation and Metabolic Regulation. American Society of Plant Physiologists, Rockville, MD, pp. 99–112.

Lundquist, P.-O., Pawlowski, K., Vance, C.P., 1996. Carbamoylphosphate synthetase: sequence analysis and nodule enhanced expression in nitrogen-fixing *Alnus incana* root nodules. Plant Physiol. 111, Suppl., 90.

Lusty, C.J., Widgren, E.E., Broglie, K.E., Nyunoya, H., 1983. Yeast carbamyl phosphate synthetase: structure of the yeast gene and homology to *Escherichia coli* carbamyl phosphate synthetase. J. Biol. Chem. 258, 14 466–14 472.

Maddison, M.W.P., Maddison, D.R., 1991. MacClade: Analysis of Phylogeny and Character Evolution Vers. 3.06. Sinauer Associates, Sunderland, MA.

Mori, M., Tatibana, M., 1975. Purification of homogeneous glutamine-dependent carbamyl phosphate synthetase from ascites hepatoma cells as a complex with aspartate transcarbamylase and dihydroorotase. J. Biochem. 78, 239–242.

Nyunoya, H., Lusty, C.J., 1983. The *carB* gene of *Escherichia coli*: a duplicated gene coding for the large subunit of carbamoyl–phosphate synthetase. Proc. Natl. Acad. Sci. USA 80, 4629–4633.

Nyunoya, H., Lusty, C.J., 1984. Sequence of the small subunit of yeast carbamyl phosphate synthetase and identification of its catalytic domain. J. Biol. Chem. 259, 9790–9798.

Ong, B.L., Jackson, J.F., 1972. Pyrimidine nucleotide biosynthesis in *Phaseolus aureus*. Biochem. J. 129, 583–593.

Paulus, T.J., Switzer, R.L., 1979. Characterization of pyrimidine-repressible and arginine-repressible carbamoylphosphate synthetases from *Bacillus subtilis*. J. Bacteriol. 137, 82–91.

Pear, J.E., Kawagoe, Y., Schreckengost, W.E., Delmer, D.P., Stalker, D.M., 1996. Higher plants contain homologs of bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. Proc. Natl. Acad. Sci. USA 93, 12 637–12 642.

Rothnie, H.M., 1996. Plant mRNA 3′-end formation. Plant Mol. Biol. 32, 43–61.

Saghai-Maroof, M.A., Solinam, K.M., Jorgensen, R.A., Allard, R.W., 1984. Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc. Natl. Acad. Sci. USA 81, 8014–8018.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Slocum, R.D., 1999. Regulation of pyrimidine biosynthesis in plants. Paths to Pyrimidines 7, 79–85.

Souciet, J.L., Nagy, M., Le Gouar, M., Lacroute, F., Potier, S., 1989. Organization of the yeast *URA2* gene: identification of a defective dihydroorotase-like domain in the multifunctional carbamoylphosphate synthetase–aspartate transcarbamylase complex. Gene 79, 59–70.

Taylor, N.G., Scheible, W.-R., Cutler, S., Somerville, C.R., Turner, S.R., 1999. The *irregular xylem3* locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell 11, 769–779.

van den Hoff, J.B., Jonker, A., Beintema, J.J., Lamers, W.H., 1995. Evolutionary relationships of the carbamoylphosphate synthetase genes. J. Mol. Evol. 41, 813–832.

Williamson, C.L., Lake, M.R., Slocum, R.D., 1996. A cDNA encoding carbamoyl phosphate synthetase large subunit (*carB*) from *Arabidopsis*. Plant Physiol. 111, 1354.

Wu, L., Ueda, T., Messing, J., 1995. The formation of mRNA 3′-ends in plants. Plant J. 8, 323–329.

Zambryski, P.C., 1992. Chronicles from the *Agrobacterium*-plant cell DNA transfer story. Annu. Rev. Plant Physiol. Plant Mol. Biol. 43, 465–490.