Evolution of Higher-Plant Glutamine Synthetase Genes: Tissue Specificity as a Criterion for Predicting Orthology[1]

Jeff J. Doyle
L. H. Bailey Hortorium, Cornell University

Evolutionary relationships among members of the nuclear multigene family encoding isozymes of glutamine synthetase (GS) in flowering plants were studied by using parsimony and distance methods. Analyses using subsets of the data indicated that the consistency index is not a sufficient measure of phylogenetic data quality and that assumptions about the information content of third-codon positions and transitions may be misleading. Tissue-specific expression patterns of GS isozymes were not uniformly useful as predictors of orthology. Chloroplast and cytosolic genes were shown to form two groups of orthologous sequences. Within the cytosolic group, however, genes expressed in the nodule did not form a single orthologous group, suggesting that a regulatory shift has occurred in the pea-alfalfa lineage. Organismal relationships inferred from the limited sampling are generally in concert with current phylogenetic hypotheses.

Introduction

Because of their abundance and/or biological importance, many of the best characterized protein-encoding genes of the plant nucleus belong to small- or medium-sized multigene families. Phylogenetic analyses have been conducted on some of these gene families, including those encoding the small subunit of ribulose bisphosphate carboxylase/oxygenase (*rbc*S; Meagher et al. 1989), chlorophyll a/b binding protein (*cab;* Demmin et al. 1989), and seed storage proteins (Gibbs et al. 1989).

Those wishing to reconstruct organismal phylogenies by using duplicated genes are faced with the problem of determining whether particular genes from the species being studied are orthologous and thus permit legitimate phylogenetic comparison (Fitch 1970). Expression differences among multigene family members could potentially represent clues for predicting orthology, but expression patterns can also be misleading. For example, mistaken assessments of orthology caused by shifts in tissue-specific lysozyme expression among bird taxa led to phylogenetic conclusions contradicted both by conventional taxonomy and by other molecular markers (Arnheim et al. 1973; Hindenberg et al. 1974; Wilson et al. 1977). Examples of regulatory changes within a multigene family are also known from vertebrate globins (e.g., see Tagle et al. 1988).

Glutamine synthetase (GS; E.C.6.3.1.2), which catalyzes the assimilation of ammonia into glutamine, is encoded by a small multigene family (three to perhaps as many as five genes) in flowering plants, with a single nuclear gene encoding a chloroplast isozyme and with several genes encoding cytosolic enzymes (Gebhardt et al. 1986;

1. Key words: glutamine synthetase, orthology, multigene family, parsimony, phylogeny, homoplasy.

Address for correspondence and reprints: Jeff J. Doyle, L. H. Bailey Hortorium, Cornell University, Ithaca, New York 14853.

Mol. Biol. Evol. 8(3):366–377. 1991.

0737-4038/91/0803-0008$02.00

Tingey et al. 1987, 1988; Bennett et al. 1989; Sakamoto et al. 1989). Leguminous plants possess GS isozymes that are predominantly expressed in their root nodules, the site of symbiotic nitrogen fixation. In the present study, phylogenetic analyses of GS genes suggest that, while patterns of tissue-specific expression may in some cases be useful criteria for inferring orthology, this is not always the case, even within a single gene family.

Material and Methods

Sequences and Alignments

Sixteen GS nucleotide sequences were analyzed, representing coding regions for chloroplastic, cytosolic nonnodular, and cytosolic nodular isozymes of both dicotyledonous and monocotyledonous flowering plants (fig. 1). A rat GS cDNA sequence (van de Zande et al. 1988) served as out-group. Plant sequences are all ~1 kb in length and possess 70%–90% nucleotide identity throughout coding regions; 3′ and 5′ extensions in sequences encoding chloroplast isozymes (Tingey et al. 1987; Lightfoot et al. 1988; Snustad et al. 1988; Sakamoto et al. 1989; Freeman et al. 1990) were excluded from all analyses conducted. For the present study, published alignments were used when available; no additional gaps were required when all plant sequences were aligned. Alignment of the rat sequence with plant sequences required the addition of (*a*) five gaps of one codon each in the rat sequence and (*b*) two gaps, one of one codon and another of four codons, to all plant sequences. These gaps were inserted by inspection.

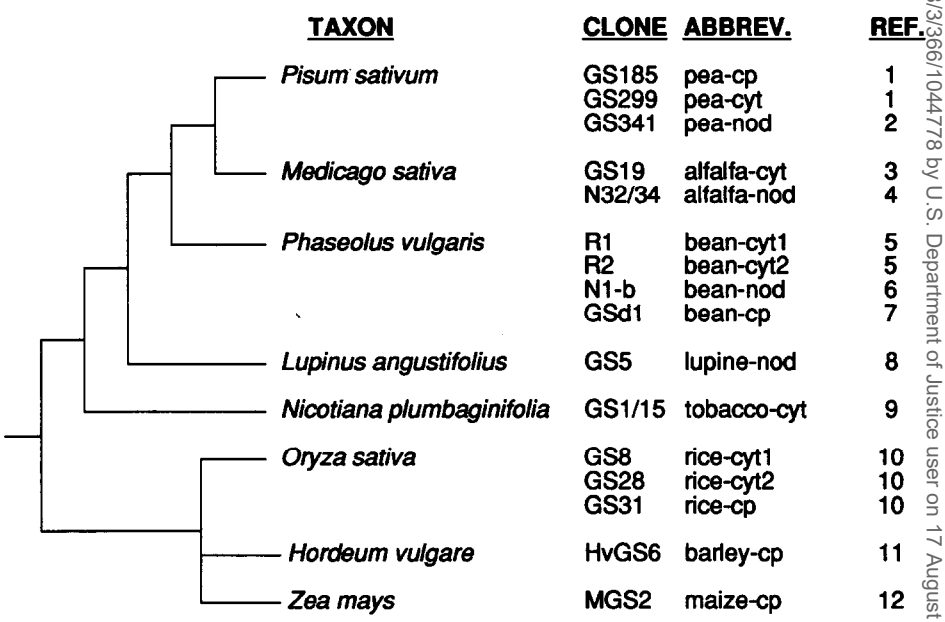

FIG. 1.—Plant GS gene sequences. References are as follows: 1, Tingey et al. 1988; 2, Tingey et al. 1987; 3, Tischer et al. 1986; 4, Dunn et al. 1988; 5, Gebhardt et al. 1986; 6, Bennett et al. 1989; 7, Lightfoot et al. 1988; 8, Grant et al. 1989; 9, Tingey and Coruzzi 1987; 10, Sakamoto et al. 1989; 11, Freeman et al. 1990; and 12, Snustad et al. 1988. All sequences except that of reference 4 are cDNA clones; clones contain complete coding regions, except for those of references 4 (258 bp) and 8 (945 bp). Phylogenetic relationships among taxa follow the data of Polhill (1981) and Cronquist (1988).

Cladistic Analyses

Characters for cladistic analyses were coded as unordered and were given equal weights in all runs. Gaps were not counted as characters; nucleotide positions within gapped regions were coded as missing data.

Phylogenetic analyses of the full data set and of subsets thereof were performed using various options available in the Hennig86 (version 1.5; Farris 1988) and/or PAUP (version 3.0; Swofford 1989) software packages. Exhaustive searches for shortest trees were performed when possible; heuristic options were used, however, on the full data set. Where multiple equally parsimonious trees were obtained, strict-consensus algorithms were utilized to find those clades that appeared in all trees.

The bootstrap option of PAUP 3.0 was employed to estimate levels of support for various clades (Felsenstein 1985). Strength of support for cladistic relationships was also investigated by a consideration of slightly longer (suboptimal) trees, followed by strict-consensus analysis.

Various subsets of the data were analyzed, in an effort to study the effects, on the most parsimonious topologies, of (1) outgroup choice, (2) transversions versus transitions, and (3) codon position. The lengths of particular alternative topologies were investigated either by manual branch-swapping or by using the PAUP 3.0 "constraints" option, which finds the most parsimonious tree topologies under the constraint of keeping selected taxa monophyletic. The PAUP 3.0 implementation of Lake's (1987) invariants method ("evolutionary parsimony") was utilized.

Distance Methods

Pairwise nucleotide substitutions were calculated by using the method of Li et al. (1985) to estimate both nonsynonymous or synonymous substitution values and substitutions at each codon position. Trees were constructed from distance matrices by using either the neighbor-joining method (NJ; Saitou and Nei 1987) or the unweighted pair-group method (UPGMA). Both methods were implemented by using the RESTSITE computer package (Miller 1989), which contains NJ and UPGMA programs written by O. Gotoh.

Results

Cladistic Analyses

All Characters

Initial analysis of the entire data set by heuristic options produced a single shortest tree of 2,190 steps, with a consistency index (CI; Kluge and Farris 1969) of 0.49 when uninformative (autapomorphic) characters were excluded. In this tree (fig. 2), with the rat sequence as an out-group, all chloroplast GS sequences formed a clade separate from all cytosolic sequences; in both clades, sequences from legume genera formed subclades. Within the legume cytosolic clade, multiple sequences from individual genera did not form a clade; nor did genes whose sole predominant expression site is the nodule. The pea and lupine nodule genes were part of a clade that included the constitutively expressed bean-cyt1, while the pea nodule sequence had its closest cladistic relationship with the bean-cyt2 sequence, which is not expressed at all in the nodule (Bennett et al. 1989).

Strict-consensus analysis of suboptimal trees suggested several of the same areas of strong and weaker support as did the bootstrap replicate analysis shown in figure 2. At +2 steps (three trees), the cytosolic clade collapsed to a trichotomy in which

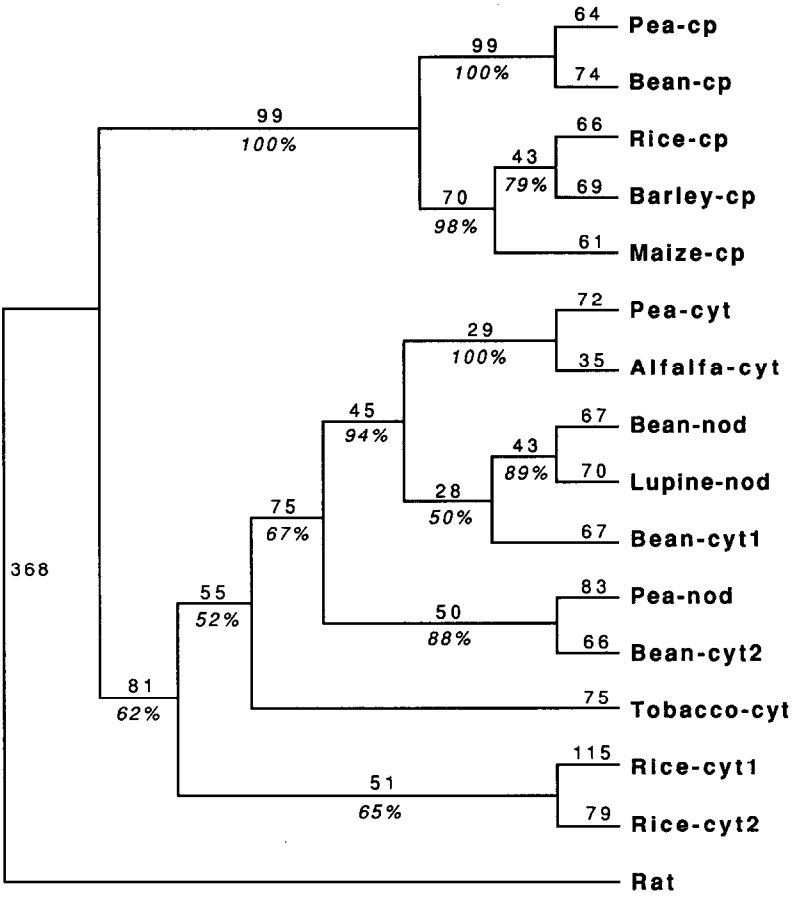

FIG. 2.—Single most parsimonious tree (2,190 steps, CI 0.49) for all taxa, when all codon positions and substitution types are used. Gene abbreviations are as in fig. 1. Numbers of steps are given above the branches, and bootstrap replicate percentages are given below the branches.

each of the three plant families formed a clade, and the major legume cytosolic subclade collapsed to a trichotomy consisting of bean-cyt1, (pea-cyt, alfalfa-cyt), and (bean-nod, lupine-nod). At +4 steps (five trees), the grass chloroplast sequences formed a trichotomy, while at +6 steps (12 trees), the (bean-nod, lupine-nod) clade was lost. At +8 steps (21 trees) several of the cytosolic clades formed a polytomy together with the chloroplast clade, but the two major legume cytosolic groups were supported as separate clades, maintaining the split among nodule sequences. The shortest tree in which the nodule sequences of pea, lupine, and common bean formed a unique clade was 2,231 steps long, 41 steps longer than the shortest tree. Inclusion of the partial alfalfa nodule GS cDNA sequence (Dunn et al. 1988) produced a single most parsimonious tree with the same general topology as in figure 2, but with alfalfa-nod as sister to the pea-nod sequence.

Analysis of only the chloroplast sequences with pea-cp as out-group produced a single shortest tree of 537 steps (CI 0.83). This tree had the same topology as that obtained with the full data set (fig. 2). The (barley-cp, rice-cp) clade appeared in 90 of 100 bootstrap replicates, with a (maize-cp, barley-cp) relationship being supported by the other 10 replicates. A tree with this alternative topology was nine steps longer

than the preferred tree (546 steps), while the third grass permutation required 552 steps. Similar analyses using only cytosolic sequences were not considered significant, because of synonymous-site saturation in many pairwise comparisons (see below).

Character Subsets

Removal of third-codon positions produced three shortest trees at 704 steps (CI 0.55), which differed only in the arrangement of legume cytosolic sequences and retained the major features of the figure 2 tree. Monophyletic groups of chloroplast and cytosolic GS sequences occurred in 100% of bootstrap replicates. A legume cytosolic subclade appeared in the bootstrap majority-rule tree but was weakly supported (34%); a (pea-nod, bean-cyt2) clade appeared in 97% of replicates. Removal of third-codon positions resulted in a change in the topology within the grass cytosolic clade, favoring a (barley-cp, maize-cp) grouping; however, this topology had only weak bootstrap support (57% of replicates).

Analysis of third-codon positions alone produced a single tree with no increase in homoplasy, relative to the analysis using all positions (1,432 steps; CI 0.49), but with a strikingly different topology having the chloroplast clade nested within the plant cytosolic sequences, with the (pea-nod, bean-cyt2) clade as its sister group. Analyses using only third-codon positions were also conducted for the chloroplast sequences alone and for the legume cytosolic clade excluding pea-nod and bean-cyt2, since pairwise comparisons within these groups indicated that third-codon positions were not saturated. In both cases the topologies shown in figure 2 were preferred. The (rice-cp, barley-cp) clade occurred in 72% of chloroplast bootstrap replicates. Some 97% of the cytosolic replicates included a (pea-cyt, alfalfa-cyt) clade, while (bean-nod, lupine-nod) occurred in 84%.

When only transversional changes were considered for the full data set (including all taxa and codon positions), three trees at 1,058 steps (CI 0.38) were identified which differed from one another only in the placement of tobacco-cyt relative to the two rice cytosolic sequences. The strict-consensus tree was in most respects identical to that obtained with the full data set including transitions (fig. 2), except for the unresolved positions of the tobacco and rice cytosolic sequences. Separate analyses in which chloroplast or cytosolic sequences were considered separately gave similar results; various out-groups were utilized for subsets of cytosolic sequences, with little effect on topologies. Results supported by the appearance of clades in ≥90% of bootstrap replicates of these data subsets were (1) grass vs. legume chloroplast sequences, (2) (barley-cp, rice-cp), (3) (pea-nod, bean-cyt2) vs. remaining legume cytosolic sequences, and (4) (pea-cyt, alfalfa-cyt).

Evolutionary Parsimony

The PAUP implementation of Lake's (1987) invariants method was used to study the three grass chloroplast sequences, by using either legume chloroplast sequences as the fourth taxon and analyzing either (1) all codon positions, (2) first- and second-codon positions only, or (3) third-codon positions only. In five of the six tests, the (maize-cp, barley-cp) topology was preferred by this method, but in no case did nonnegative values for preferred trees have significant binomial test probabilities. When pea-cp was used with the three grass sequences and when all codon positions were considered, the (rice-cp, barley-cp) topology was favored, but the binomial test gave $P = 0.82$.

Distance Analyses

Pairwise comparisons showed that in no case were nonsynonymous substitutions saturated (table 1). K_A values varied over a 10-fold range, from <0.04 among grass chloroplast sequences to >0.40 between rat and most chloroplast genes. Both NJ and UPGMA analyses of nonsynonymous divergences supported a dichotomy between chloroplast and cytosolic plant sequences (e.g., see fig. 3). Both NJ and UPGMA produced a grass chloroplast group and joined rice-cp and barley-cp within it. Relationships among cytosolic sequences were poorly resolved, with polytomies or short branch lengths relative to standard errors; topologies were also sensitive to inclusion or exclusion of either the rat sequence or chloroplast sequences. In no analysis were nodule sequences grouped together; pea-nod and bean-cyt2 were always paired.

Silent substitutions were found to be saturated in many pairwise comparisons (table 1), and therefore the entire data set was not used to construct trees. UPGMA and NJ trees constructed from K_S values for the chloroplast sequences agreed both with each other and with the cladistic analyses in separating grass sequences from those of legumes and in supporting a (rice-cp, barley-cp) grouping. NJ topologies of the legume sequences varied as more diverged taxa (rice-cyt2, tobacco-cyt) were progressively deleted; all, however, strongly supported a (pea-nod, bean-cyt2) pairing. UPGMA analyses all retained a very similar topology, with a major legume cluster separated from a (pea-nod, bean-cyt2) pair.

Discussion

Construction of Gene Trees: Comparisons of Methods and Subsets of Data

Results from parsimony, UPGMA, and NJ were generally congruent, except in areas where all methods showed only weak support for particular topologies, and

Table 1
Divergences among GS Genes

	GENE															
GENE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 ..		48	229	264	220	>400	>400	>400	>400	>400	>400	>400	>400	>400	>400	>400
2 ..	7		204	210	236	>400	>400	>400	>400	>400	>400	>400	>400	>400	>400	>400
3 ..	7	8		63	86	231	220	254	253	244	>400	287	>400	>400	>400	>400
4 ..	9	10	4		72	298	258	398	>400	>400	>400	>400	>400	191	>400	>400
5 ..	8	9	4	6		245	193	>400	247	>400	261	313	400	>400	>400	>400
6 ..	14	15	14	16	14		38	82	75	73	141	128	154	>400	232	>400
7 ..	15	16	16	17	15	5		62	62	67	98	92	131	>400	196	>400
8 ..	16	17	17	18	16	8	7		59	76	124	137	158	>400	242	>400
9 ..	17	18	17	18	15	8	6	7		69	121	130	136	>400	227	>400
10 ..	16	16	17	17	15	6	4	6	7		123	141	123	194	158	>400
11 ..	14	16	15	16	15	7	6	8	8	7		74	142	>400	226	>400
12 ..	14	16	16	16	15	8	7	8	8	8	4		154	>400	>400	>400
13 ..	16	17	16	18	16	8	5	8	7	8	7	7		>400	136	365
14 ..	15	17	16	18	15	10	9	10	10	10	9	10	9		81	>400
15 ..	17	18	17	17	16	8	8	7	8	8	8	8	7	10		252
16 ..	44	46	40	42	40	38	40	41	39	40	40	40	41	40	42	

NOTE.—Each gene is denoted by a number, as follows: 1, pea-cp; 2, bean-cp; 3, rice-cp; 4, barley-cp; 5, maize-cp; 6, pea-cyt; 7, alfalfa-cyt; 8, bean-nod; 9, lupine-nod; 10, bean-cyt1; 11, bean-cyt2; 12, pea-nod; 13, tobacco-cyt; 14, rice-cyt1; 15, rice-cyt2; and 16, rat. Divergences were calculated by using formulas of Li et al. (1985). Replacement-sites ($K_A \times 100$) are given below the null diagonal; silent-site divergences ($K_S \times 100$) are above the diagonal.

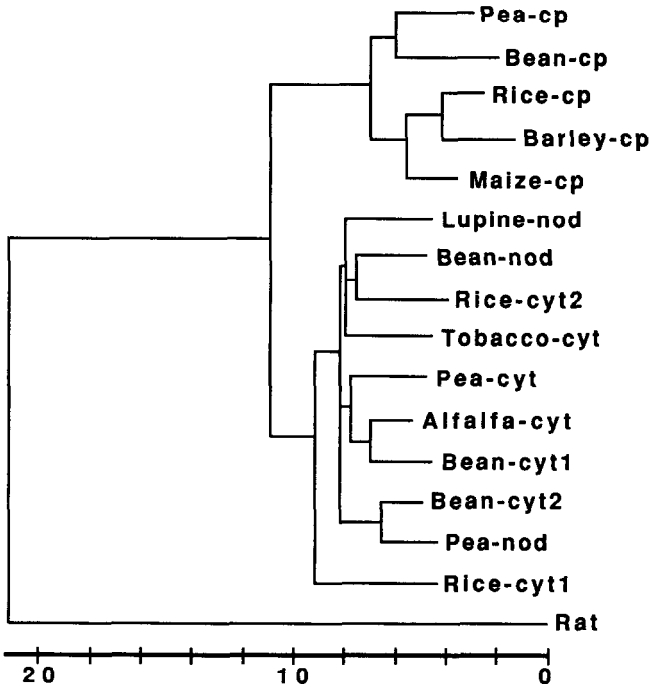

FIG. 3.—NJ dendrogram of replacement divergences (K_A; table 1)

suggest the following conclusions: (1) chloroplast and cytosolic GS sequences are paralogous; (2) among chloroplast sequences, grasses and legumes form separate GS lineages, with the genes of barley and rice as closest relatives; and (3) legume cytosolic GS sequences are supported as a monophyletic group, albeit weakly, with a major dichotomy between (pea-nod, bean-cyt2) and the remaining complete sequences (the partial alfalfa-nod sequence is most closely related to pea-nod). Most analyses supported the recognition of (bean-nod, lupine-nod) and (pea-cyt, alfalfa-cyt) as pairs of orthologues.

Because of problems with saturation of transitional substitutions and at synonymous sites over long evolutionary periods, all positions in a coding sequence are not expected to be equally useful as phylogenetic characters. Characters derived from saturated positions may be expected to add homoplasy and perhaps to obscure true relationships. Removing third-codon positions from the GS analysis did result in some decrease in homoplasy (table 2); however, the tree favored when *only* third-codon positions were used was not more homoplasious than the topologically quite different tree favored by the full data set. Similar results were obtained when chloroplast or cytosolic sequences were considered separately (table 2). Thus, although saturated third-codon positions do not appear to provide reliable characters for GS sequences over long evolutionary periods, this is not reflected in an increase of homoplasy. For these characters the amount of homoplasy alone does not appear to be an adequate measure of phylogenetic utility; this is a specific instance of what may be a general phenomenon (Archie 1989; Sanderson and Donoghue 1989).

Transversions, as rarer substitution events, are expected to be more reliable over long evolutionary periods (Lake 1987). However, homoplasy *increases* were observed in GS comparisons for which transitions were excluded (table 2). The shortest trees

Table 2
CI Values of Most Parsimonious Trees of Subsets of GS Data

	All Taxa	Chloroplast Taxa Only	Cytosolic Taxa Only[a]
All characters	0.49	0.83	0.58
No third-codon positions	0.56	0.78	0.56
Only third-codon positions	0.49	0.85	0.60
Transversions only	0.38	0.80	0.45

NOTE.—CI is calculated by dividing the number of steps by the total number of character states in the data set; in a data set with no homoplasy, each nucleotide substitution would occur only once on the tree and CI would be 1.0.

[a] All plant cytosolic sequences, with rat as out-group.

obtained in the various analyses using only transversions did not produce any unexpected results; however, as considerable resolution was lost in these analyses, it is difficult to see any advantage in excluding transitions, particularly as this class of mutations appears neither especially homoplasious nor phylogenetically misleading.

Predicting Orthology from GS Tissue Specificity

The orthologous genes encoding chloroplast GS isozymes are much more similar to cytosolic genes from both plants and animals than they are to GS genes of prokaryotes, and thus they presumably originated by gene duplication in the plant nucleus rather than by transfer from the chloroplast (Tingey et al. 1988). The chloroplast-specific expression (e.g., see Edwards et al. 1990) of this class of GS genes seems an accurate predictor of orthology, as are structural features of their mRNAs, such as the presence of sequences encoding transit peptides.

While higher-plant genomes appear to contain only a single gene for the chloroplast GS isozyme, multiple genes encode the various cytosolic enzymes (Lightfoot et al. 1988; Snustad et al. 1988; Sakamoto et al. 1989; Edwards et al. 1990; Freeman et al. 1990). The occurrence of multiple genes for cytosolic GS in both rice and various legumes suggests that duplication of this class of gene predated the divergence of monocots and dicots. However, there is little evidence for orthologous relationships across plant familial boundaries, suggesting either that there have been several waves of duplication or that concerted evolution is operating to obscure orthologous relationships. At present it is not possible to distinguish between these two possibilities.

Groups of orthologous legume cytosolic genes are not predicted by their patterns of expression. Genes whose predominant expression is in the leguminous nodule belong to two strongly supported clades, each of which includes genes not expressed in nodules. The similar regulation (Bennett et al. 1989; Edwards et al. 1990) of these paralogous genes, coupled with divergent regulation of orthologues, could be explained by recombination between upstream regulatory sequences and coding regions.

Estimation of Species Trees from GS Gene Trees

Considered separately, topologies within chloroplast or cytosolic GS gene groups are broadly congruent with currently accepted views of flowering-plant phylogeny (e.g., see Cronquist 1988), as depicted in figure 1. The consistent occurrence of distinct legume and grass groups within the chloroplast sequences is expected, given the clear monophyly of these two families and their known phylogenetic affinities. Similarly,

recognition of clades within the cytosolic gene subfamily consisting of (1) sequences from all legume genera and (2) dicot genes (legumes + tobacco) is not surprising taxonomically. Nevertheless, these observations, as obvious as they may appear, are worth mentioning, if only because even such rudimentary consistency is not always found when plant nuclear sequences are studied. In a recent study of the nuclear multigene family encoding the small subunit of ribulose 1,5-bisphosphate carboxylase-oxygenase (*rbc*S), Meagher et al. (1989) presented distance trees with several instances of unexpected groupings, including a large separation between sequences from the legumes pea and soybean. Similar results have been obtained in a cladistic reanalysis of those data (J. J. Doyle, unpublished data). Plant protein sequence studies have long been notorious for providing unexpected phylogenetic results and have been the focus of recent analyses that suggest that the phylogenetic information content of some of the data sets is quite low (Bremer 1988; Archie 1989). The GS data appear to be both more robust and more consistent with phylogenetic expectations.

Sequences encoding chloroplast GS isozymes are available from three grasses, each representing a different subfamily whose phylogenetic relationships have been debated (Watson et al. 1985; Kellogg and Campbell 1987). Both cladistic and distance analyses of the GS chloroplast genes suggest that barley and rice shared a more recent common ancestor than either did with maize, but, as in other molecular studies of grasses (Hamby and Zimmer 1988; Wolfe et al. 1989; Doebley et al. 1990), taxonomic conclusions are not strongly supported.

The clearest group of cytosolic orthologues comprises bean-cyt2, pea-nod, and the partial alfalfa-nod gene, which groups with the pea-nod sequence. This topology is in agreement with (1) current taxonomic concepts (fig. 1; Polhill 1981), (2) with recent evidence from chloroplast DNA rearrangements (Lavin et al. 1990), and (3) with the observation that the pea and alfalfa nonnodular GS sequences are also closely related (fig. 2). Reliance on tissue-specific expression patterns as predictors of orthology can produce both incorrect phylogenetic topologies and incorrect branch lengths, as has been noted elsewhere for avian lysozymes (Arnheim et al. 1973; Hindenberg et al. 1974; Wilson et al. 1977). The assumption that all nodule GS sequences are orthologous would lead to the conclusion that lupine and common bean are more closely related than either is to pea, when the current view of legume phylogeny suggests that the pea and common-bean lineages share a more recent common ancestor (fig. 1; Polhill 1981).

A second nodule multigene family, leghemoglobin, is regulated differently in soybean (a close relative of common bean) and alfalfa (Barker et al. 1988). Such differences in nodule gene expression parallel the numerous physiological and morphological differences associated with nodulation in these lineages (Sprent 1981). Thus, although the shift in GS expression is a problem for predicting orthology, it may represent a phylogenetic character for a major legume lineage, one of potentially many synapomorphies associated with symbiotic nitrogen fixation.

Acknowledgments

I wish to thank Michael Sanderson for many fruitful discussions and for his assistance in running PAUP. I am grateful to Walter Fitch, Jerry Slightom, and an anonymous reviewer for many helpful suggestions on the manuscript. This research was supported by NSF grant BSR-8805630.

LITERATURE CITED

ARCHIE, J. W. 1989. Phylogenies of plant families: a demonstration of phylogenetic randomness in DNA sequence data derived from proteins. Evolution **43**:1796–1800.

ARNHEIM, N., A. HINDENBERG, G. S. BEGG, and F. J. MORGAN. 1973. Multiple genes for lysozyme in birds: studies on black swan egg white lysozymes. J. Biol. Chem. **248**:8036–8042.

BARKER, D. G., P. GALLUSCI, V. LULLIEN, H. KHAN, M. GHERARDI, and T. HUGUET. 1988. Identification of two groups of leghemoglobin genes in alfalfa (*Medicago sativa*) and a study of their expression during root nodule development. Plant Mol. Biol. **11**:761–772.

BENNETT, M. J., D. A. LIGHTFOOT, and J. V. CULLINORE. 1989. cDNA sequence and differential expression of the gene encoding the glutamine synthetase γ polypeptide of *Phaseolus vulgaris*. Plant Mol. Biol. **12**:553–565.

BREMER, K. D. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution **42**:795–803.

CRONQUIST, A. 1988. An integrated system of classification of flowering plants. Columbia University Press, New York.

DEMMIN, D. S., E. J. STOCKINGER, Y. C. CHANG, and L. L. WALLING. 1989. Phylogenetic relationships between the chlorophyll a/b binding protein (CAB) multigene family: an intra- and interspecies study. J. Mol. Evol. **29**:266–279.

DOEBLEY, J., M. DURBIN, E. M. GOLENBERG, M. T. CLEGG, and D. P. MA. 1990. Evolutionary analysis of the large subunit of carboxylase (*rbc*L) nucleotide sequence among the grasses (Gramineae). Evolution **44**:1097–1108.

DUNN, K., R. DICKSTEIN, R. FEINBAUM, B. K. BURNETT, T. K. PETERMAN, G. THOIDIS, H. M. GOODMAN, and F. M. AUSUBEL. 1988. Developmental regulation of nodule-specific genes in alfalfa root nodules. Mol. Plant-Microbe Interact. **2**:66–74.

EDWARDS, J. W., E. L. WALKER, and G. M. CORUZZI. 1990. Cell-specific expression in transgenic plants reveals nonoverlapping roles for chloroplast and cytosolic glutamine synthetase. Proc. Natl. Acad. Sci. USA **87**:3459–3463.

FARRIS, J. S. 1988. Hennig86 user's manual. State University of New York, Stony Brook.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

FITCH, W. M. 1970. Distinguishing homologous from analogous proteins. Syst. Zool. **19**:99–113.

FREEMAN, J., A. J. MARQUEZ, R. M. WALLSGROVE, R. SAARELAINEN, and B. G. FORDE. 1990. Molecular analysis of barley mutants deficient in chloroplast glutamine synthetase. Plant Mol. Biol. **14**:297–311.

GEBHARDT, C., J. E. OLIVER, B. G. FORDE, R. SAARELAINEN, and B. J. MIFLIN. 1986. Primary structure and differential expression of glutamine synthetase genes in nodules, roots and leaves of *Phaseolus vulgaris*. EMBO J. **5**:1429–1435.

GIBBS, P. E. M., K. B. STRONGIN, and A. MCPHERSON. 1989. Evolution of legume seed storage proteins—a domain common to legumins and vicilins is duplicated in vicilins. Mol. Biol. Evol. **6**:614–623.

GRANT, M. R., A. CARNE, D. F. HILL, and K. J. F. FARNDEN. 1989. The isolation and characterization of a cDNA clone encoding *Lupinus angustifolius* root nodule glutamine synthetase. Plant Mol. Biol. **13**:481–490.

HAMBY, R. K., and E. A. ZIMMER. 1988. Ribosomal RNA sequences for inferring phylogeny within the grass family. Plant Syst. Evol. **160**:29–37.

HINDENBERG, A., J. SPITZNAGEL, and N. ARNHEIM. 1974. Isozymes of lysozyme in leukocytes and egg whites: evidence for the species-specific control of egg-white lysozyme synthesis. Proc. Natl. Acad. Sci. USA **71**:1653–1657.

KELLOGG, E. A., and C. S. CAMPBELL. 1987. Phylogenetic analysis of the Gramineae. Pp. 310–

322 *in* T. R. SODERSTROM, K. W. HILU, C. S. CAMPBELL, and M. E. BARKWORTH, eds. Grass systematics and evolution. Smithsonian Institution Press, Washington, D.C.

KLUGE, A., and J. S. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. Syst. Zool. **18**:1–32.

LAKE, J. A. 1987. A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. Mol. Biol. Evol. **4**:167–191.

LAVIN, M., J. J. DOYLE, and J. D. PALMER. 1990. Evolutionary significance of the loss of the chloroplast DNA inverted repeat in the Leguminosae subfamily Papilionoideae. Evolution **44**:390–402.

LI, W.-H., C.-I. WU, and C.-C. LUO. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**:150–174.

LIGHTFOOT, D. A., N. K. GREEN, and J. V. CULLINORE. 1988. The chloroplast-located glutamine synthetase of *Phaseolus vulgaris* L.: nucleotide sequence, expression in different organs and uptake into isolated chloroplasts. Plant Mol. Biol. **11**:191–202.

MEAGHER, R. B., S. BERRY-LOWE, and K. RICE. 1989. Molecular evolution of the small subunit of ribulose bisphosphate carboxylase: nucleotide substitution and gene conversion. Genetics **123**:845–863.

MILLER, J. C. 1989. RESTSITE: a program for analyzing restriction site or fragment data and calculating the genetic distances between OTUs, for use in molecular systematic studies. University of Texas Health Center at Houston, Houston.

POLHILL, R. M. 1981. Papilionoideae. Pp. 191–208 *in* R. M. POLHILL and P. H. RAVEN, eds. Advances in legume systematics, part 1. Royal Botanic Gardens, Kew, England.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SAKAMOTO, A., M. OGAWA, T. MASUMURA, D. SHIBATA, G. TAKEBA, K. TANAKA, and S. FUJII. 1989. Three cDNA sequences coding for glutamine synthetase polypeptides in *Oryza sativa* L. Plant Mol. Biol. **13**:611–614.

SANDERSON, M. J., and M. J. DONOGHUE. 1989. Patterns of variation in levels of homoplasy. Evolution **43**:1781–1795.

SNUSTAD, D. P., J. P. HUNSPERGER, B. M. CHERESKIN, and J. MESSING. 1988. Maize glutamine synthetase cDNAs: isolation by direct genetic selection in *Escherichia coli*. Genetics **120**: 1111–1124.

SPRENT, J. I. 1981. Functional evolution of some Papilionoid root nodules. Pp. 671–676 *in* R. M. POLHILL and P. H. RAVEN, eds. Advances in legume systematics, part 2. Royal Botanic Gardens, Kew, England.

SWOFFORD, D. 1989. PAUP version 3.0. Illinois Natural History Museum, Champaign.

TAGLE, D. A., B. F. KOOP, M. GOODMAN, J. L. SLIGHTOM, D. L. HESS, and R. T. JONES. 1988. Embryonic and globin genes of a prosimian primate (*Galago crassicaudatus*): nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. J. Mol. Biol. **203**:439–455.

TINGEY, S. V., and G. M. CORUZZI. 1987. Glutamine synthetase of *Nicotiana plumbaginifolia*. Plant Physiol. **84**:366–373.

TINGEY, S. V., F.-Y. TSAI, J. W. EDWARDS, E. L. WALKER, and G. M. CORUZZI. 1988. Chloroplast and cytosolic glutamine synthetase are encoded by homologous nuclear genes which are differentially expressed *in vivo*. J. Biol. Chem. **263**:9651–9657.

TINGEY, S. V., E. L. WALKER, and G. M. CORUZZI. 1987. Glutamine synthetase genes of pea encode distinct polypeptides which are differentially expressed in leaves, roots and nodules. EMBO J. **6**:1–9.

TISCHER, E., S. DASSARMA, and H. M. GOODMAN. 1986. Nucleotide sequence of an alfalfa glutamine synthetase gene. Mol. Gen. Genet. **203**:221–229.

VAN DE ZANDE, L., W. T. LABRUYERE, M. M. SMALING, A. F. M. MOORMAN, R. H. WILSON,

R. CHARLES, and W. H. LAMERS. 1988. Nucleotide sequence of rat glutamine synthetase mRNA. Nucleic Acids Res. **16**:7726.

WATSON, L., H. T. CLIFFORD, and M. J. DALLWITZ. 1985. The classification of Poaceae: subfamilies and supertribes. Aust. J. Botany **33**:433–484.

WILSON, A. C., S. S. CARLSON, and T. J. WHITE. 1977. Biochemical evolution. Annu. Rev. Biochem. **46**:573–639.

WOLFE, K. H., M. GOUY, Y.-M. YANG, P. M. SHARP, and W.-H. LI. 1989. Dates of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci USA **86**:6201–6205.

WALTER M. FITCH, reviewing editor

Received July 5, 1990; revision received November 19, 1990

Accepted December 4, 1990