Molecular Phylogenetics and Evolution
Vol. 19, No. 3, June, pp. 421–429, 2001
doi:10.1006/mpev.2001.0931, available online at http://www.idealibrary.com on IDEAL®

Phylogenetic Utility of the Glycerol-3-Phosphate Acyltransferase Gene: Evolution and Implications in *Paeonia* (Paeoniaceae)

David C. Tank[1] and Tao Sang[2]

Department of Botany and Plant Pathology, Michigan State University, East Lansing, Michigan 48824

Received June 27, 2000; revised December 19, 2000

The nuclear-encoded chloroplast-expressed glycerol-3-phosphate acyltransferase (GPAT) gene has been found to be single-copy in a number of angiosperm families. In this study we investigated the phylogenetic utility of the GPAT gene at the interspecific level using the genus *Paeonia* (Paeoniaceae) as an example. An approximately 2.3- to 2.6-kb fragment of the GPAT gene, containing a large intron of more than 2 kb, was amplified, cloned, and sequenced from 19 accessions representing 13 *Paeonia* species. The GPAT gene phylogeny inferred by parsimony analysis supported interspecific relationships that were previously unresolved, suggesting that large introns of low-copy nuclear genes are particularly informative in the resolution of close relationships at low taxonomic levels. Whereas the GPAT phylogeny is largely congruent with the previous phylogenetic hypothesis of *Paeonia,* it shows a significant discordance involving the paraphyly of section *Paeonia.* Given evidence of an ancient duplication and the subsequent silencing of one GPAT locus in *P. anomala,* this discordance is most likely the result of paralogy. Two distinct genomic clones containing partial GPAT genes were isolated from *P. anomala.* The GPAT sequence from one clone corresponded to the functional copy of the gene, and the second genomic clone was determined to contain a GPAT pseudogene. The insertion of a retrotransposon in an intron of this pseudogene may have been responsible for the silencing of this GPAT locus in *P. anomala.* This study suggests that, although it is unlikely that universal nuclear gene markers free from paralogy are usually available, low-copy nuclear genes can be very useful in plant phylogenetic reconstruction, especially at low taxonomic levels, as long as the evolutionary dynamics of the genes are carefully examined. © 2001 Academic Press

[1] Current address: Department of Botany, University of Washington, Box 355325, Seattle, WA 98195.
[2] To whom correspondence should be addressed. Fax: (517) 355-1926. E-mail: sang@msu.edu.

INTRODUCTION

Low-copy nuclear genes have the enormous potential to provide numerous independent estimates of the underlying organismal phylogeny and a wealth of phylogenetic information for the resolution of close relationships. Because sequences of both chloroplast and mitochondrial genomes evolve more slowly than those of the nuclear genome in plants, nuclear gene markers are especially needed for phylogeny reconstruction at low taxonomic levels (Wolfe *et al.,* 1987; Palmer, 1992). Recent studies have demonstrated that rapidly evolving introns of low-copy nuclear genes can provide sufficient phylogenetic information to resolve interspecific relationships previously unresolved, or poorly resolved, by chloroplast DNA or nuclear ribosomal DNA (e.g., Doyle *et al.,* 1996; Sang *et al.,* 1997b; Small *et al.,* 1998; Emshwiller and Doyle, 1999).

However, many theoretical and practical questions concerning the phylogenetic utility of low-copy nuclear genes remain open. Consequently, the broad application of low-copy nuclear gene markers in plant phylogenetic reconstruction has been severely impeded. The lack of universal gene markers and the corresponding universal PCR primers has also limited the utility of low-copy nuclear genes in plant phylogenetic studies (Strand *et al.,* 1997). Another significant limiting factor seems to be the extra lab work and analytical difficulties involved in the determination of orthology of low-copy nuclear genes among the studied taxa. The ideal solution to these problems would be to identify low-copy nuclear genes that are generally less dynamic in their history of duplication and deletion and thus less susceptible to the problem of paralogy. This reasoning has led us to search for genes that are reported to be single-copy among a diverse group of angiosperms and to evaluate the potential of developing universal nuclear markers for phylogenetic studies at low taxonomic levels in flowering plants.

Glycerol-3-phosphate acyltransferase (GPAT) is an essential enzyme utilized in the catalysis of the initial step of glycerolipid synthesis, specifically, the formation of lysophosphotidic acid from glycerol 3-phosphate

1055-7903/01 $35.00
Copyright © 2001 by Academic Press
All rights of reproduction in any form reserved.

and acylthioesters, in the cells of all higher organisms (Nishida *et al.,* 1993). GPAT exists in plants in at least three forms that differ in their subcellular location and substrate specificity. The nuclear-encoded and chloroplast-expressed GPAT gene was selected for this study because it has been previously characterized in plants belonging to five angiosperm families, including *Arabidopsis thaliana* of Brassicaceae (Nishida *et al.,* 1993), *Pisum sativum* of Fabaceae (Weber *et al.,* 1991), *Carthamus tinctorius* of Asteraceae (Bhella and MacKenzie, 1994), *Spinacia oleracea* of Chenopodiaceae (Wolter, unpublished), and *Cucurbita moschata* of Cucurbitaceae (Ishizaki *et al.,* 1988). Based on enzyme activity in *A. thaliana* (Kunst *et al.,* 1988) and Southern blot hybridizations in the remaining plants (Ishizaki *et al.,* 1988; Weber *et al.,* 1991; Bhella and MacKenzie, 1994), the GPAT gene has been reported as single-copy in these distantly related plant species. The presence of the gene at a single locus in the five eudicot families suggests that the GPAT gene may not be subject to dynamic cycles of duplication and deletion and could potentially become a useful phylogenetic marker for eudicot plants.

In this study, we chose the genus *Paeonia* to test the phylogenetic utility of the GPAT gene, because the phylogenetic relationships within the genus have been well established based on both morphological and molecular data. *Paeonia* is placed in the monotypic family Paeoniaceae. Although Paeoniaceae has been considered enigmatic with respect to its phylogenetic position within angiosperms and has often been referred to the Ranunculales, recent molecular phylogenetic analyses have resolved Paeoniaceae in a well-supported Saxifrigales clade within the core eudicots (Soltis and Soltis, 1997; Soltis *et al.,* 1997, 2000). The five GPAT sequences reported in GenBank represent a sampling from the core eudicots, including both rosid and asterid families and the order Caryophyllales (Bremer *et al.,* 1998). Therefore, the conserved gene regions shared by these sequences should be appropriate for the design of PCR primers to amplify the GPAT gene from *Paeonia.*

Paeonia consists of ~35 species of herbaceous and woody habit disjunctly distributed in the northern temperate region. *Paeonia* has been further divided into three sections: *Moutan, Onaepia,* and *Paeonia.* The largest section, *Paeonia,* contains ~27 herbaceous diploid and tetraploid species distributed in eastern and central Asia, the western Himalayas, and the European Mediterranean region. Based on leaf morphology, section *Paeonia* has been further partitioned into two subsections: *Paeonia* and *Foliolatae.* Section *Moutan* consists of five diploid woody species in two subsections, *Delavayanae* and *Vaginitae,* distributed in central and western China. The smallest section, *Onaepia,* consists of only two diploid herbaceous species endemic to Pacific North America (Stern, 1946; Pan, 1979; Tzanoudakis, 1983; Pei, 1993).

The phylogeny of *Paeonia* was previously inferred based on nucleotide sequences from multiple genic and intergenic regions, including two loci of the low-copy nuclear gene alcohol dehydrogenase (*Adh1* and *Adh2*), the cpDNA gene *mat*K, two intergenic cpDNA spacers (*trn*L–*trn*F and *psb*A–*trn*H), and the nrDNA ITS region (Sang *et al.,* 1995, 1997a,b; Sang and Zhang, 1999). The molecular phylogenies support the monophyly of each of the three taxonomic sections of *Paeonia* and each subsection of section *Moutan.* Despite the large amount of sequence data analyzed, relationships within subsection *Vaginitae* of section Moutan remain unresolved. It was previously hypothesized that this group of shrubby species, including *P. suffruticosa,* known as the king of flowers in China (Hong *et al.,* 1998a), may have experienced either rapid morphological divergence or slowed molecular evolution (Sang *et al.,* 1997a). In section *Paeonia,* which contains numerous hybrid species, relationships among diploid species of presumably nonhybrid origin have never been resolved with strong statistical support (Sang *et al.,* 1997b). Therefore, new gene markers with higher rates of sequence divergence are needed to further improve resolution of the *Paeonia* phylogeny.

MATERIALS AND METHODS

Sampling of *Paeonia* species for this study included 19 accessions of 13 species representing each of the three sections and four subsections. Seven species were sampled from section *Paeonia,* including, the 4 diploid members of subsection *Paeonia, P. anomala, P. lactiflora, P. tenuifolia,* and *P. veitchii* (2 accessions), and 3 species from subsection *Foliolatae,* the diploid *P. japonica,* 3 accessions of both diploid and tetraploid *P. obovata,* and the tetraploid *P. mairei.* We sampled all 5 species of section *Moutan, P. delavayi, P. lutea* (2 accessions), *P. rockii* (2 accessions), *P. suffruticosa* ssp. *spontanea,* and *P. szechuanica.* To represent section *Onaepia,* two populations of *P. californica* were sampled. For the majority of accessions, the same DNAs studied previously for ITS, *mat*K, and *Adh* sequences were used here, and the voucher information can be found in Sang *et al.* (1997a). Additional sampling for this study includes two new populations of *P. obovata, P. obovata2* and *P. obovata3,* with collection numbers Hong 85008-4 and WLS01, respectively. Voucher specimens for these new accessions are deposited in the Herbarium of the Institute of Botany, Beijing (PE). Total DNAs were isolated by the CTAB method (Doyle and Doyle, 1987) from fresh leaves collected in the field.

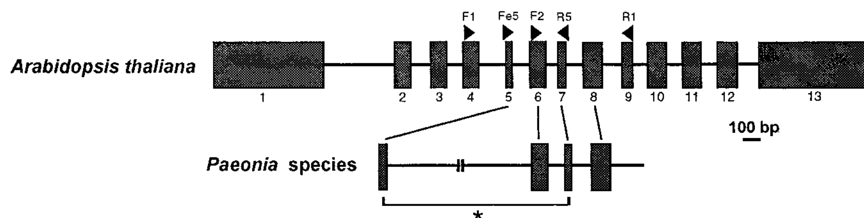

FIG. 1. Diagram of the full-length GPAT gene in *Arabidopsis thaliana* and a portion of the gene in peonies. Boxes, numbered from the 5′ to the 3′ end, represent exons, and lines between the exons represent introns. Lines connecting exons between *A. thaliana* and *Paeonia* species indicate homologous exons. Arrows above exons indicate the location and direction of PCR primers used in this study. The bracketed region, indicated with an asterisk, represents the region used for the phylogenetic analyses. The size of each region is measured by the scale bar except where indicated.

PCR and Sequencing

Two PCR primers, GAF1 (5′-TTTGGYCAAAATTATATTCGKCC) and GAR1 (5′-CCACCACTKGGTGCAATCCA; Fig. 1), were designed in the most conserved regions across GPAT DNA and mRNA sequences of the five eudicots. The intron and exon boundaries were determined based on the gene sequence of *A. thaliana* (Fig. 1), because this was the only available DNA sequence of the GPAT gene (the remaining four were mRNA sequences). These primers failed to amplify the gene from any of the peony species screened. Instead, a fragment of approximately 300 bp was amplified from *P. californica,* which upon sequencing was determined to be a retrogene containing only exon sequences. Based on the sequence of the retrogene fragment, three new PCR primers, GAF2 (5′-AGCAGACCCTGCTATCATTGC), GAFe5 (5′-CCCTGTTCTCTGGAATGGAAG), and GAR5 (5′-CATGCTGAATGGCTTGCAAAG) were designed (Fig. 1). A fragment of approximately 2.5 kb was amplified from all peony accessions with the primers GAFe5 and GAR5 and was used for the phylogenetic analyses.

For all PCR amplifications in this study, the GPAT gene was amplified through the following PCR cycles: (1) 70°C, 4 min; (2–4) 94°C, 1 min; 52–55°C, 30 s; 72°C, 2 min; (5–7) 94°C, 20 s; 52–55°C, 30 s; 72°C, 2 min; (8) repeat steps 5–7 29 times; (9) 72°C, 10 min. All resulting PCR products were cloned with a TOPO TA cloning kit (Invitrogen). For each species, 10 to 20 clones were screened by examination of restriction-site or sequence (from one primer) variation (Sang *et al.,* 1997b; Wang *et al.,* 2000). Restriction-site variation was determined with the restriction enzymes *Hae*III and/or *Hin*fI. Clones with a distinct restriction fragment profile, or those showing sequence variation, were fully sequenced and included in the phylogenetic analyses. Sequencing was done on an ABI 373 automated DNA sequencer with either the Dye Terminator Cycle Sequencing reaction kit (PE Applied Biosystems) or the DYEnamic ET Terminator Cycle Sequencing reaction kit (Amersham Pharmacia Biotech). Three peony-specific primers (GAFI6, 5′-GCTTTATGTTACGTGTACTC; GAR6, 5′-GTGGTATTTCATCATCACCG; and GAR7, 5′-CAGGAAAAAGAGTTGGAAGTG) designed in the large intron flanked by exons 5 and 6 (Fig. 1) were used as additional primers for sequencing. Sequences were edited with the program SeqEd and aligned manually. A few regions in the GPAT introns that could not be unambiguously aligned were excluded from the phylogenetic analyses.

Genomic Library Screening

A genomic library of *P. anomala* was constructed with the Lambda FIX II partial fill-in vector kit and Gigapack III package extract kit (Stratagene). The genomic library was screened with a ^{32}P-labeled probe following protocols of Sambrook *et al.* (1989). The probe was labeled by a random priming reaction using a PCR fragment amplified by the primers GAF2 and GAR1 from *P. veitchii.* After a total of ~500,000 bacteriophage plaques were screened, two positive clones (C2 and C3) were isolated and purified with a Qiagen Lambda DNA Mini Kit (Qiagen Inc.). Restriction-site analysis indicated that the two clones were identical. Screening of the genomic library with a probe amplified by the PCR primers GAF2 and GAR5 yielded a second distinct genomic clone containing the GPAT gene (C7).

The two distinct genomic clones were characterized by restriction mapping (Fig. 2) and subcloned with a Zero Background/Kan Cloning Kit (Invitrogen). Subcloned fragments were sequenced with the M13 forward and reverse primers located on the plasmid vector and with sequencing primers designed subsequently. BLAST searches were performed to locate the GPAT gene within the genomic clones and to determine sequence similarity to other sequences in GenBank.

Phylogenetic Analyses

Parsimony, as executed in PAUP* 4.0 (Swofford, 1998), was used to infer the gene phylogeny based on nucleotide substitutions in the aligned GPAT sequences. Parsimony analysis was performed by heuristic search with tree bisection-reconnection (TBR) branch swapping, the MULTREES option, ACCTRAN optimization, and 10,000 random-addition replicates. Bootstrap analysis was carried out with 10,000 repli-

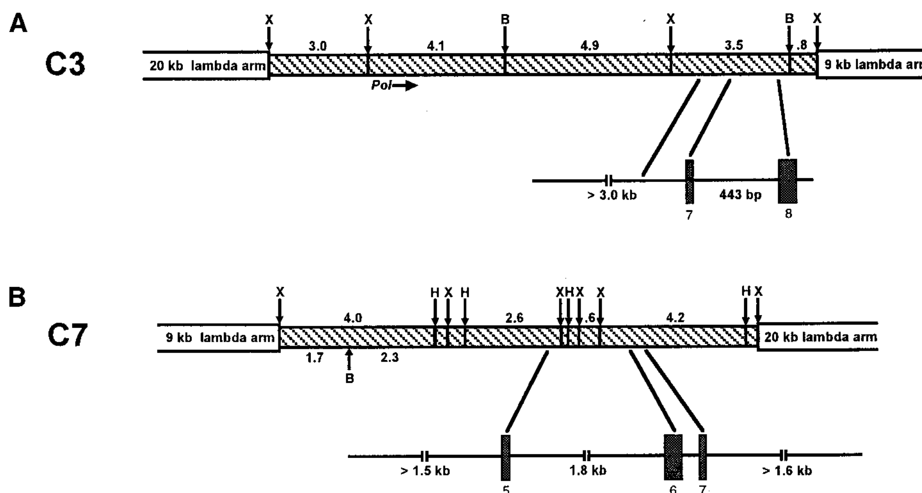

FIG. 2. Characterization of two genomic clones isolated from *Paeonia anomala* genomic library screening. Arrows indicate the position of restriction endonuclease cut sites (B, *Bam*HI; H, *Hin*dIII; X, *Xba*I) used for restriction mapping and subcloning. Sizes of the resulting fragments are given in kilobases. Underneath each genomic clone characterization is an enlargement of the portion of the GPAT gene identified in each with sizes of introns given in base pairs, numbers of exons indicated, and lines indicating the corresponding region of the genomic clone from which they were identified. (A) Genomic clone C3; *Pol* indicates the position and orientation of the *Pol* gene; (B) genomic clone C7.

cates of heuristic search with TBR branch swapping, ACCTRAN optimization, and simple taxon addition. Section *Moutan* was used as a functional outgroup based on previous phylogenetic hypotheses (Sang *et al.*, 1997b). Topological congruence to previous phylogenetic hypotheses was assessed with the Templeton test (Templeton, 1983), as implemented in PAUP* 4.0.

In addition to the parsimony analysis, a maximum-likelihood (ML) analysis was conducted on a reduced data set of 15 taxa. The data set was reduced to include one clone from each monophyletic species on the GPAT phylogeny (Fig. 3) from sections *Paeonia* and *Onaepia* and two randomly selected clones from both *P. mairei* and *P. obovata.* Section *Moutan* was reduced to one clone from each of the five species and was used as a functional outgroup to root the resulting ML tree. The program Modeltest 3.0 (Posada and Crandall, 1998) was used to determine the model of sequence evolution best fit to the GPAT data set by the hierarchical likelihood ratio test. Maximum-likelihood, as implemented in PAUP* 4.0 (Swofford, 1998), was used to infer the gene phylogeny based on the HKY (Hasegawa *et al.*, 1985) model with rate heterogeneity. Under this model, rates were assumed to follow a gamma distribution with the shape parameter estimated via maximum-likelihood.

RESULTS

Phylogenetic Reconstruction

After 10–20 GPAT clones from each accession of the *Paeonia* species were screened, one type of sequence was isolated from accessions *P. lactiflora, P. obovata1, P. obovata3, P. rockii2,* and *P. veitchii2.* Two to four types of GPAT sequences were isolated from all other accessions sampled for this study. In total, 41 distinct GPAT clones (GenBank Accession Nos. AY016247–AY016286, AF325203–AF325204) were included in the phylogenetic analysis. The resulting data set contained 2674 bp of alignable sequence, spanning three exons and two introns (Fig. 1), of which 141 bp were of exon regions and 2533 bp were of alignable intron regions. Of the alignable intron sequence, 2438 bp were from the large intron flanked by exons 5 and 6 (Fig. 1).

The aligned GPAT sequences were variable at 445 nucleotide sites, of which 429 were from intron regions and 16 were from exon regions. A total of 308 nucleotide sites were parsimony informative, of which 303 were from introns and 5 were from exons. The large intron alone provided 290 parsimony-informative characters. Parsimony analysis resulted in 45 most parsimonious trees (tree length = 530; consistency index = 0.88, 0.84 excluding uninformative characters; retention index = 0.96). One of the 45 most parsimonious trees was randomly selected to represent the GPAT gene phylogeny with nodes that collapse on the strict consensus indicated (Fig. 3).

The GPAT gene phylogeny (Fig. 3) is well resolved, and the interspecific relationships therein are strongly supported. Only one node is collapsed on the strict consensus as a result of topological discordance between the equally most parsimonious trees. Nearly every node on the gene tree has bootstrap support >50%, and most are supported by bootstrap values >90%.

Only one type of sequence was identified for each of

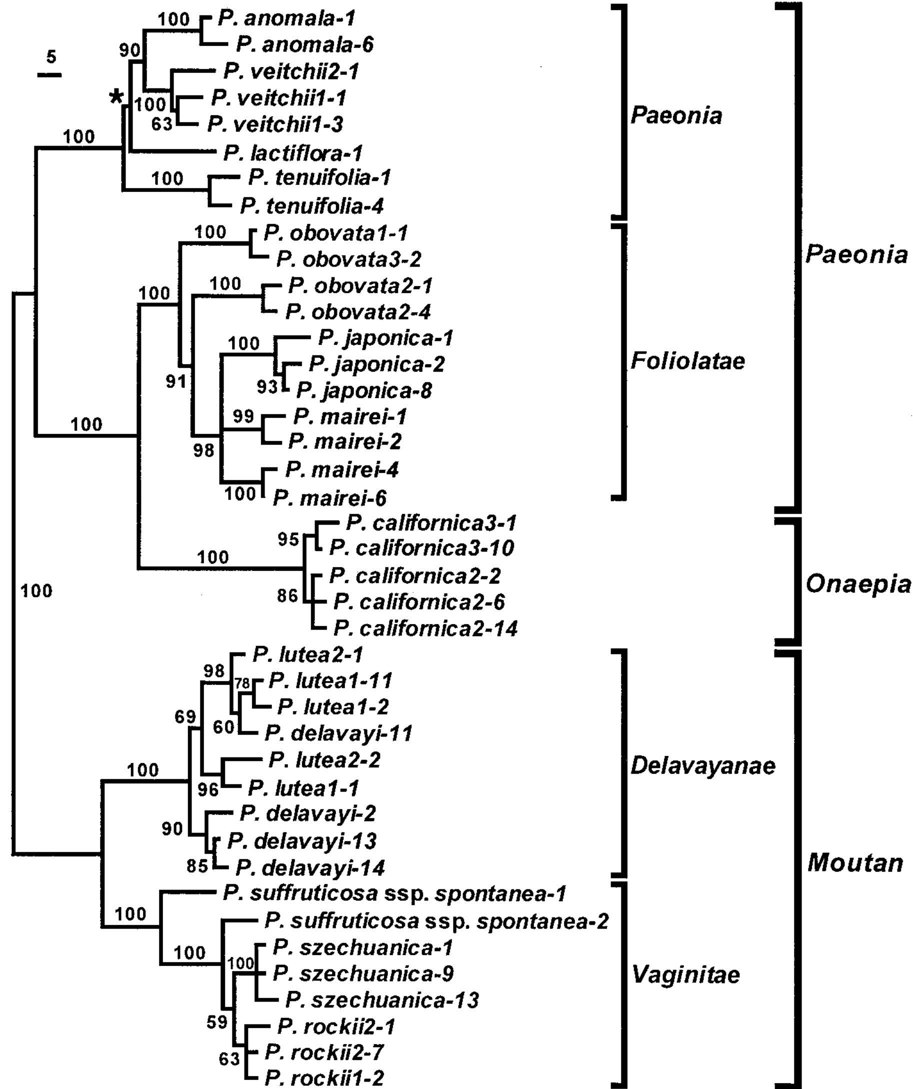

FIG. 3. Phylogeny of the GPAT gene of *Paeonia.* One randomly selected tree of 45 most parsimonious trees (tree length = 530, consistency index = 0.88, retention index = 0.96). Species represented by more than one population are indicated with accession numbers following the species name. Numbers following a hyphen indicate the clone numbers. Numbers associated with the branches are bootstrap percentages greater than 50%. *Branch collapses on the strict consensus. Branch lengths are proportional to the numbers of nucleotide substitutions and are measured by the scale bar.

the following accessions: *P. lactiflora, P. obovata1, P. obovata3, P. veitchii2,* and *P. rockii2* (Fig. 3). Two different sequences were cloned from each of the accessions *P. anomala, P. veitchii1, P. tenuifolia, P. obovata2, P. californica3, P. lutea2,* and *P. suffruticosa* ssp. *spontanea.* Three different GPAT clones were isolated from each of the accessions *P. japonica, P. californica2, P. lutea1, P. rockii2,* and *P. szechuanica.* Clones from these accessions except *P. lutea1* form a monophyletic group within each accession. Four distinct clones were identified from *P. mairei* and *P. delavayi* and did not form a monophyletic group within an accession.

The topology of the GPAT phylogeny is largely congruent with the previous phylogenetic hypotheses with two exceptions; the sister relationship of *P. anomala* and *P. veitchii* in subsection *Paeonia* of section *Paeonia* and the sister relationship of *P. californica* of section *Onaepia* and subsection *Foliolatae* of section *Paeonia* (Fig. 3). These associations were not congruent with the previous phylogenetic hypotheses which identified *P. anomala* as the sister species of *P. tenuifolia* and strongly support the monophyly of both section *Paeonia* and section *Onaepia* (Sang *et al.,* 1995, 1997a,b). The Templeton test was performed on the GPAT phylogeny separately for each of these two incongruencies while previous phylogenetic relationships were used as a topological constraint. These analyses indicate that

the relationships involving *P. anomala, P. veitchii,* and *P. tenuifolia* within subsection *Paeonia* are not significantly incongruent ($P = 0.51$). However, the sister relationship of subsection *Foliolatae* of section *Paeonia* and *P. californica,* representing section *Onaepia,* is significantly incongruent ($P < 0.0001$), with the previous phylogenetic hypotheses supporting the monophyly of each of the two sections.

Maximum-likelihood analyses of the reduced GPAT data set resulted in a tree (not shown) identical in overall topology to those obtained from the parsimony analysis. The model of sequence evolution best fit to the GPAT data set, as determined by the likelihood-ratio test, was the HKY model with rate heterogeneity (HKY + Γ). Parameters estimated via maximum-likelihood include base frequencies (A: 0.28, C: 0.18, G: 0.18, and T: 0.36), transition/transversion ratio (1.82), and the value of the gamma shape parameter (1.45).

Characterization of Genomic Clones

Genomic library screening isolated two distinct genomic clones containing GPAT genes in *P. anomala,* clones C3 and C7 (Fig. 2). Based on sequence comparisons to the two PCR clones from *P. anomala, P. anomala*-1 and *P. anomala*-6 (Fig. 3), sequence of the C7 clone was determined to be nearly identical to *P. anomala*-1 in both exon and intron regions. The two sequences differed by only one nucleotide of more than 1700 bp, and this difference is likely due to PCR error. The C3 genomic clone was found to contain a portion of the GPAT gene including exons 7 and 8 and their flanking intron regions (Figs. 1 and 2A). This clone was determined to be a pseudogene based on multiple insertions and deletions that resulted in stop codons within each of the two exons. Genomic library screening, therefore, identified a pseudogene and one functional locus of the GPAT gene in *P. anomala.*

Despite position conservation, the intron sizes of the GPAT gene differ drastically between *Paeonia* and *A. thaliana.* The intron between exons 5 and 6 was >2 kb long in all *Paeonia* species, whereas the corresponding intron in *A. thaliana* is only 103 bp long. Characterization of genomic clone C7 revealed that the intron flanked by exons 4 and 5 is also much larger than the corresponding intron of *A. thaliana* (Figs. 1 and 2B). Sequencing the portion of the genomic clone upstream from exon 5 failed to identify any exon sequence, suggesting that the intron may be larger than 7 kb. The presence of this large intron was likely responsible for preventing PCR amplification of the portion of the gene between the initial PCR primers GAF1 and GAR1 (Fig. 1). For the pseudogene identified in genomic clone C3, sequencing the region upstream of exon 7 failed to locate exon 6. Instead, BLAST searches identified sequence with a high identity to the *Pol* gene, a gene involved in replication and transposition of retrotransposable elements (Li, 1997).

DISCUSSION

The interspecific relationships of *Paeonia* were well resolved and strongly supported on the GPAT gene phylogeny (Fig. 3). Relationships within subsection *Vaginitae* of the shrubby section *Moutan* were previously never resolved by ITS, *mat*K, or *Adh* genes or by the combined analysis of these genes (Sang *et al.,* 1997b). The GPAT gene phylogeny suggests a sister group relationship between *P. rockii* and *P. szechuanica.* Relationships within subsection *Paeonia* of section *Paeonia* were either unresolved or poorly supported by the previous gene trees. The GPAT phylogeny strongly supports (90% of bootstrap replicates) the sister group relationship of *P. anomala* and *P. veitchii.*

The better resolution of the GPAT gene phylogeny apparently stems from a sufficient amount of phylogenetic information provided by the large, rapidly evolving intron flanked by exons 5 and 6 (Fig. 1). The average sequence divergences of the intron between species of subsection *Paeonia* and between species of section *Moutan,* estimated by the one-parameter model (Jukes and Cantor, 1969), were 0.0185 and 0.0210, respectively. These values are higher than corresponding sequence divergences of introns of the *Adh1* gene (0.0154 and 0.0141; Sang *et al.,* 1997b). The previous study has already demonstrated that introns of the *Adh* genes in *Paeonia* evolve more rapidly than the ITS region of nrDNA, a commonly used phylogenetic marker at lower taxonomic levels in plants. Thus, these results suggest that large introns of low-copy nuclear genes are particularly informative in the resolution of close relationships at low taxonomic levels.

Since the GPAT gene is present at a single locus in plants from the five eudicot families, it is of great interest to examine the assumption that the gene is also likely to be single-copy in *Paeonia* and thus to have a less dynamic history of duplication and deletion. The five accessions with only one type of sequence identified are likely to have only one copy of the gene. The seven accessions with two different sequences cloned may also have only one GPAT locus, with the different sequences likely representing allelic variation. For the five diploid accessions with three different GPAT sequences identified, the amount of sequence polymorphism apparently indicates at least one recent gene duplication, although the possibility of PCR errors cannot be ruled out.

Two species, *P. delavayi* and *P. mairei,* are likely to have more than one copy of the GPAT gene, because each of them has four different GPAT clones that are not monophyletic. For *P. delavayi,* three of the four clones are monophyletic and one clone, *P. delevayi*-11, is nested within sequences from the two *P. lutea* accessions (Fig. 3). *P. delavayi* and *P. lutea* are sister species with overlapping distributions in southwestern China.

In previous taxonomic treatments of *Paeonia,* these two species were sometimes recognized as two subspecies of *P. delavayi* (Hong *et al.,* 1998b). Paralogy, lineage sorting, or introgression could explain the nesting of a GPAT clone from *P. delavayi* within clones from *P. lutea* (Doyle, 1997; Maddison, 1997). However, the identification of the mechanism responsible for the observed pattern is impossible without additional evidence. *P. mairei* is a tetraploid species and has even been hypothesized to be of allotetraploid origin (Sang *et al.,* 1997a). Thus, the two groups of GPAT clones of *P. mairei* may have resulted from polyploidization rather than gene duplication.

The fact that the majority of the diploid accessions have only one or two different GPAT sequences, and most clones from an accession are monophyletic, suggests that the GPAT gene is probably single-copy in the majority of peony species. Furthermore, the majority of clades in the GPAT phylogeny are congruent with the corresponding sections and subsections of the genus, implying that paralogy is not a prevalent problem in the GPAT phylogeny. However, a significant topological incongruence has been identified between the GPAT phylogeny and the previous phylogenetic hypotheses. This involves the sister relationship of section *Onaepia* and subsection *Foliolatae* of section *Paeonia* on the GPAT tree (Fig. 3). The monophyly of both section *Paeonia* and section *Onaepia* was strongly supported by nrDNA ITS, chloroplast DNA, and *Adh* gene phylogenies (Sang *et al.,* 1997a,b). The question becomes whether the incongruence of this intersectional relationship is the result of paralogy, lineage sorting, or hybridization (Doyle, 1997; Maddison, 1997; Wendel and Doyle, 1998).

Hybridization is almost certainly not the cause of the observed incongruence. Hybrid speciation has been documented within section *Paeonia* based on morphological, cytogenetic, and molecular data. No evidence has ever indicated the possibility of gene flow between section *Paeonia* and section *Onaepia.* The likelihood of intersectional hybridization is further reduced by the geographic isolation between the two sections. The disjunct distributions of the two sections in Pacific North America and Eurasia may have been present as early as the Miocene when the two subsections of section *Paeonia* had not yet diverged from one another (Sang *et al.,* 1997a).

It can be quite difficult, if not impossible, to distinguish paralogy and lineage sorting as the cause of incongruence between gene trees at the infrageneric level. Lineage sorting is the failure of alleles to coalesce at the time of speciation, but rather, for coalescence to occur deeper than the speciation event. Given that an ancestral allelic polymorphism is more likely to be lost the longer that two species have been divergent, lineage sorting tends to cause more problems at lower taxonomic levels. Paralogy, resulting from the random sampling or retention of duplicate loci, may pose problems at all taxonomic levels. Teasing apart these two processes at the interspecific level largely depends on an understanding of the molecular background and evolutionary dynamics of the nuclear genes in question.

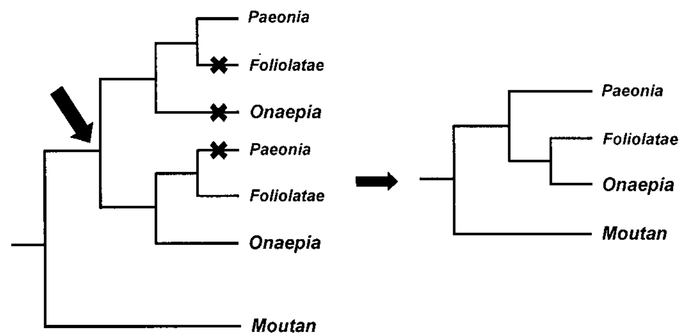

FIG. 4. Trees depicting the paralogous relationship of the GPAT gene between section *Paeonia* (subsections *Paeonia* and *Foliolatae*) and section *Onaepia.* The large arrow indicates the gene duplication event, Xs represent independent deletion events, and the small arrow indicates the resulting GPAT gene tree.

In comparison with paralogy, the hypothesis of lineage sorting is equally likely, if not favored, in the earlier discussion of the paraphyly of *P. delavayi* sequences. However, to explain the paraphyly of section *Paeonia,* paralogy becomes a more appealing alternative. This is because the lineage sorting hypothesis requires that the polymorphic alleles arose in the common ancestor of section *Paeonia* and section *Onaepia* and were maintained throughout the diversification of both subsections *Paeonia* and *Foliolatae* of section *Paeonia.* The probability of the maintenance of ancestral allelic polymorphisms decreases as the branch length that they travel through increases with time (Maddison, 1997).

The paralogy hypothesis assumes that the GPAT gene was duplicated before the diversification of sections *Paeonia* and *Onaepia,* and the duplicated genes then underwent three independent deletions (Fig. 4). To support this hypothesis, it is necessary to have evidence for an ancient duplication of the GPAT gene. Indeed, genomic clone C3 (Fig. 2A) was determined to contain a GPAT pseudogene. Parsimony analysis based on sequence alignment of the normal GPAT genes with the recovered pseudogene exons and portions of the flanking introns that could be aligned unambiguously illustrates the degree of sequence divergence of the GPAT pseudogene (Fig. 5). Of a total of ~250 bp of alignable sequence, there were over 81 substitutions along the branch leading to the GPAT pseudogene. This indicates that the GPAT gene in *P. anomala* has undergone an ancient duplication, followed by the subsequent silencing of one of the duplicate loci.

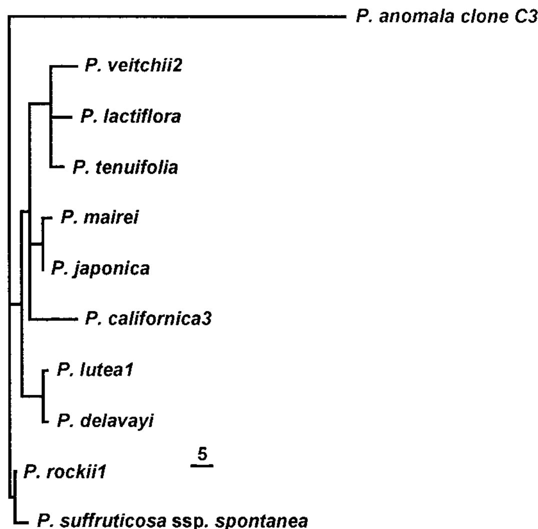

FIG. 5. GPAT gene phylogeny illustrating the long branch leading to the pseudogene from genomic clone C3. This tree is the strict consensus of four most parsimonious trees. Branch lengths are proportional to the numbers of nucleotide substitutions and are measured by the scale bar.

As indicated by high identity to sequences available in GenBank, the *Pol* gene, a gene found in retrotransposable elements (Li, 1997), was located upstream of exon 7 of the GPAT pseudogene (Fig. 2A). Furthermore, sequencing upstream of exon 7 of the pseudogene failed to identify the upstream exons of the GPAT gene. This suggests that the insertion of a retrotransposon-like element, interrupting the GPAT gene, was the mechanism responsible for the silencing of the GPAT locus. Thus, the evidence for the ancient gene duplication and silencing in peonies supports the paralogy hypothesis to explain the paraphyly of section *Paeonia* on the GPAT phylogeny. These results further suggest that the presence of only one functional copy of a gene in the genome does not necessarily rule out the possibility of historical duplication and deletion of the gene. Consequently, it is not impossible to yield paralogous relationships by use of single-copy nuclear genes for phylogeny reconstruction.

Taken together, the GPAT gene provided better resolution and stronger support of close interspecific relationships within subsection *Paeonia* and subsection *Vaginitae* of the genus *Paeonia.* This demonstrates the potential utility of low-copy nuclear genes, especially their large introns, in resolving close relationships at low taxonomic levels in plants. However, to correctly infer organismal phylogenies from low-copy nuclear gene phylogenies, it is important that the mechanisms underlying low-copy nuclear gene evolution are explored. Although the GPAT gene is present in a single copy in several eudicot families and is likely to be single-copy in the majority of diploid *Paeonia* species, screening of the genomic library of *P. anomala* has isolated a pseudogene resulting from an ancient gene duplication. The history of duplication and deletion of low-copy nuclear genes, thus, may not be accurately predicted by the number of functional loci currently identified in the genome. Depending on where and when gene duplication and/or deletion has occurred, paralogy will have differential impacts on the resulting gene phylogeny at various taxonomic levels of the study group. Therefore, it is unlikely that "universal" low-copy nuclear gene markers that are free from paralogy can be developed from single-copy genes for plant phylogenetic studies. While genes characterized as having a small number of copies may be preferred, caution must be exercised to evaluate their phylogenetic utility on a group to group basis and within each group studied.

ACKNOWLEDGMENTS

We thank Fang Wang for technical assistance and two anonymous reviewers for helpful comments on the manuscript. This research was supported by NSF DEB-9726838 to T.S.

REFERENCES

Bhella, R. S., and Mackenzie, S. L. (1994). Nucleotide sequence of a cDNA from *Carthamus tinctorius* encoding a glycerol-3-phosphate acyltransferase. *Plant Physiol.* **106:** 1713–1714.

Bremer, K., Chase, M. W., Stevens, P. F., Anderberg, A. A., Backlund, A., Bremer, B., Briggs, B. G., Endress, P. K., Fay, M. F., Goldblatt, P., Gustafsson, M. H. G., Hoot, S. B., Judd, W. S., Kallersjo, M., Kellogg, E. A., Kron, K. A., Les, D. H., Morton, C. M., Nickrent, D. L., Olmstead, R. G., Price, R. A., Quinn, C. J., Rodman, J. E., Rudall, P. J., Savolainen, V., Soltis, D. E., Soltis, P. S., Sytsma, K. J., and Thulin, M. (1998). An ordinal classification for the families of flowering plants. *Ann. Bot. Gard.* **85:** 531–553.

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19:** 11–15.

Doyle, J. J. (1997). Trees within trees: Genes and species, molecules and morphology. *Syst. Biol.* **46:** 537–553.

Doyle, J. J., Kanazin, V., and Shoemaker, R. C. (1996). Phylogenetic utility of histone H3 intron sequences in the perennial relatives of soybean (*Glycine:* Leguminosae). *Mol. Phylogenet. Evol.* **6:** 438–447.

Emshwiller, E., and Doyle, J. J. (1999). Chloroplast-expressed glutamine synthetase (ncpGS): Potential utility for phylogenetic studies with an example from *Oxalis* (Oxalidaceae). *Mol. Phylogenet. Evol.* **12:** 310–319.

Hasegawa, M., Kishino, H., and Yano, T. A. (1985). Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22:** 160–174.

Hong, D.-Y., Pan, K.-Y., and Yu, H. (1998a). Yingpingmudan, the wild relative of the king of flowers, *Paeonia suffruticosa* Andrews. *Acta Phytotax. Sin.* **36:** 515–520.

Hong, D.-Y., Pan, K.-Y., and Yu, H. (1998b). Taxonomy of the *Paeonia delavayi* complex (Paeoniaceae). *Ann. Missouri Bot. Gard.* **85:** 554–564.

Ishizaki, O., Nishida, I., Agata, K., Eguchi, G., and Murata, N. (1988). Cloning and nucleotide sequence of cDNA for the plastid glycerol-3-phosphate acyltransferase from squash. *FEBS Lett.* **238:** 424–430.

Jukes, T. H., and Cantor, C. R. (1969). Evolution of protein molecules. *In* "Mammalian Protein Metabolism" (H. N. Munro, Ed.), pp. 21–132, Academic Press, New York.

Kunst, L., Browse, J., and Somerville, C. (1988). Altered regulation of lipid biosynthesis in a mutant of *Arabidopsis* deficient in chloroplast glycerol-3-phosphate acyltransferase activity. *Proc. Natl. Acad. Sci. USA* **85:** 4143–4147.

Li, W.-H. (1997). "Molecular Evolution," Sinauer, Sunderland, MA.

Maddison, W. P. (1997). Gene trees in species trees. *Syst. Biol.* **46:** 523–536.

Nishida, I., Tasaka, Y., Shiraisi, H., and Murata, N. (1993). The gene and the RNA for the precursor to the plastid-located glycerol-3-phosphate acyltransferase of *Arabidopsis thaliana. Plant Mol. Biol.* **21:** 267–277.

Palmer, J. D. (1992). Mitochondrial DNA in plant systematics: Applications and limitations. *In* "Molecular Systematics of Plants" (P. S. Soltis, D. E. Soltis, and J. J. Doyle, Eds.), pp. 36–49. Chapman & Hall, New York.

Pan, K.-Y. (1979). *Paeonia. In* "Flora Reipublicae Siniccae," Vol. 27, pp. 37–59. Science Press, Beijing.

Pei, Y.-L. (1993). "Studies on the *Paeonia suffruticosa* Andr. Complex," Ph.D. dissertation, Institute of Botany, Chinese Academy of Sciences, Beijing.

Posada, D., and Crandall, K. A. (1998). MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sang, T., Crawford, D. J., and Stuessy, T. F. (1995). Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: Implications for biogeography and concerted evolution. *Proc. Natl. Acad. Sci. USA* **92:** 6813–6817.

Sang, T., Crawford, D. J., and Stuessy, T. F. (1997a). Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *Am. J. Bot.* **84:** 1120–1136.

Sang, T., Donoghue, M. J., and Zhang, D. (1997b). Evolution of alcohol dehydrogenase genes in peonies (*Paeonia*): Phylogenetic relationships of putative nonhybrid species. *Mol. Biol. Evol.* **14:** 994–1007.

Sang, T., and Zhang, D. (1999). Reconstructing hybrid speciation using sequences of low copy nuclear genes: Hybrid origins of five *Paeonia* species based on *Adh* gene phylogenies. *Syst. Bot.* **24:** 148–163.

Small, R. L., Ryburn, J. A., Cronn, R. C., Seelanan, T., and Wendel, J. F. (1998). The tortoise and the hare: Choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *Am. J. Bot.* **85:** 1301–1315.

Soltis, D. E., and Soltis, P. S. (1997). Phylogenetic relationships in Saxifragaceae sensu lato: A comparison of topologies based on 18S rDNA and *rbc*L sequences. *Am. J. Bot.* **84:** 504–522.

Soltis, D. E., Soltis, P. S., Nickrent, D. L., Johnson, L. A., Hahn, W. J., Hoot, S. B., Sweere, J. A., Kuzoff, R. K., Kron, K. A., Chase, M. W., Swensen, S. M., Zimmer, E. A., Chaw, S. M., Gillespie, L. J., Kress, W. J., and Sytsma, K. J. (1997). Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Ann. Bot. Gard.* **84:** 1–49.

Soltis, D. E., Soltis, P. S., Chase, M. W., Mort, M. E., Albach, D. C., Zanis, M., Savolainen, V., Hahn, W. H., Hoot, S. B., Fay, M. F., Axtell, M., Swensen, S. M., Prince, L. M., Kress, W. J., Nixon, K. C., and Farris, J. S. (2000). Angiosperm phylogeny inferred from 18S rDNA, *rbc*L, and *atp*B sequences. *Bot. J. Linn. Soc.* **133:** 381–461.

Stern, F. C. (1946). "A Study of the Genus *Paeonia,*" Royal Horticult. Soc., London.

Strand, A. E., Leebens-Mack, J., and Milligan, B. G. (1997). Nuclear DNA-based markers for plant evolutionary biology. *Mol. Ecol.* **6:** 113–118.

Swofford, D. L. (1998). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other methods). Version 4. Sinauer, Sunderland, MA.

Templeton, A. R. (1983). Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* **37:** 221–244.

Tzanoudakis, D. (1983). Karyotypes of four wild *Paeonia* species from Greece. *Nordic J. Bot.* **3:** 307–318.

Wang, X.-Q., Tank, D. C., and Sang, T. (2000). Phylogeny and divergence times in Pinaceae: Evidence from three genomes. *Mol. Biol. Evol.* **17:** 773–781.

Weber, S., Wolter, F. P., Buck, F., Frentzen, M., and Heinz, E. (1991). Purification and cDNA sequencing of an oleate-selective acyl-ACO: sn-glycerol-3-phosphate acyltransferase from pea chloroplasts. *Plant Mol. Biol.* **17:** 1067–1076.

Wendel, J. F., and Doyle, J. J. (1998). Phylogenetic incongruence: Window into genomes history and molecular evolution. *In* "Molecular Systematics of Plants, II: DNA Sequencing" (D. E. Soltis, P. S. Soltis, and J. J. Doyle, Eds.), pp. 265–296. Kluwer Academic, Boston.

Wolfe, K. H., Li, W.-H., and Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* **84:** 9054–9058.