Molecular Phylogenetics and Evolution 48 (2008) 1013–1026

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenetic utility of 141 low-copy nuclear regions in taxa at different taxonomic levels in two distantly related families of rosids

P. Roxanne Steele [a,b,*], Mary Guisinger-Bellian [a,b], C. Randal Linder [a], Robert K. Jansen [a,b]

[a] *Section of Integrative Biology, The University of Texas at Austin, 1 University Station C0930, Austin, TX 78712, USA*
[b] *Institute of Cellular and Molecular Biology, The University of Texas at Austin, 1 University Station C0930, Austin, TX 78712, USA*

ARTICLE INFO

Article history:
Received 19 December 2007
Revised 2 May 2008
Accepted 5 May 2008
Available online 21 May 2008

Keywords:
Low-copy nuclear markers
Rosids
Phylogenetic utility
MoBIoS
Psiguria
Geraniaceae
Cucurbitaceae
Molecular systematics
Angiosperms

ABSTRACT

Angiosperm systematics has progressed to the point where it is now expected that multiple, independent markers be used in phylogenetic studies. Universal primers for amplifying informative regions of the chloroplast genome are readily available, but in the faster-evolving nuclear genome it is challenging to discover priming sites that are conserved across distantly related taxa. With goals including the identification of informative markers in rosids, and perhaps other angiosperms, we screened 141 nuclear primer combinations for phylogenetic utility in two distinct groups of rosids at different taxonomic levels—*Psiguria* (Cucurbitaceae) and Geraniaceae. We discovered three phylogenetically informative regions in *Psiguria* and two in Geraniaceae, but none that were useful in both groups. Extending beyond rosids, we combined our findings with those of another recent effort testing these primer pairs in Asteraceae, Brassicaceae, and Orchidaceae. From this comparison, we identified 32 primer combinations that amplified regions in representative species of at least two of the five distantly related angiosperm families, giving some prior indication about phylogenetic usefulness of these markers in other flowering plants. This reduced set of primer pairs for amplifying low-copy nuclear markers along with a recommended experimental strategy provide a framework for identifying phylogenetically informative regions in angiosperms.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Sequences of rapidly evolving molecular loci from multiple, independent sources are required for inferring the relationships between species at low taxonomic levels. To address questions about evolutionary and biogeographic history, speciation, polyploidy, and hybridization, systematists and evolutionary biologists require well-resolved estimates of the phylogenetic histories of organisms. In taxa that have undergone recent speciation events or that evolve slowly, there is often little molecular variation detected between species. Because of this lack of variation, systematists often find it difficult to identify phylogenetically informative molecular markers at low taxonomic levels. The majority of molecular data used in plant systematics has come from two sources: chloroplast DNA (cpDNA) and nuclear ribosomal DNA (rDNA) (Small et al., 2004).

The greatest advantages of using the chloroplast genome are its highly conserved gene content and gene order throughout most flowering plants, allowing the development of numerous universal primers for amplifying noncoding regions of the chloroplast genome (Taberlet et al., 1991; Demesure et al., 1995; Shaw et al., 2005, 2007). However, disadvantages include biparental inheritance in some groups (Corriveau and Coleman, 1988) and the inability to detect hybridization and polyploidy in those taxa with uniparental inheritance (Small et al., 2004). Furthermore, because the genes are linked, the chloroplast genome provides only a single marker and because of its relatively slow rate of evolution, cpDNA often lacks sufficient variation to estimate species-level relationships.

Corroboration of phylogenetic hypotheses by independent data sets increases confidence in a resulting tree. Phylogenetic incongruence between data sets can provide insight into evolutionary phenomena relatively common in plants such as hybridization, introgression, or lineage sorting (Wendel and Doyle, 1998; Small et al., 2004; Mort et al., 2007). To provide independent support for phylogenetic trees inferred from cpDNA sequences, molecular systematists began utilizing nuclear rDNA. In many respects rDNA repeats have been very useful phylogenetic regions; however, they also have some problems.

Nuclear rDNA has largely been assumed to undergo concerted evolution (Baldwin et al., 1995). Not only do the abundant copies of rDNA facilitate amplification, but the nuclear genome typically

* Corresponding author. Address: Plant Biology Graduate Program, The University of Texas at Austin, 1 University Station, A6720, Austin, TX 78712, USA. Fax: +1 512 232 3402.

E-mail addresses: roxisteele@mail.utexas.edu (P. Roxanne Steele), mary.guisinger@mail.utexas.edu (M. Guisinger-Bellian), rlinder@mail.utexas.edu (C. Randal Linder), jansen@mail.utexas.edu (R.K. Jansen).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.05.017

Table 1
Taxa used in the study

Genus	Specific epithet	Authority	Voucher[a]	Abbreviation	Where utilized
Psiguria	*bignoniacea*	(Poepp. and Endl.) Wunderlin	Steele 1007	Pbig	*Psiguria*—ingroup
Psiguria	*pedata*	(L.) R.A. Howard	Steele 1035	Pped	*Psiguria*—ingroup
Psiguria	*racemosa*	C. Jeffrey	Steele 1018	Prace	*Psiguria*—ingroup
Psiguria	sp.	unpublished	Steele 1004	Ptab	*Psiguria*—ingroup
Psiguria	*umbrosa*	(Kunth) C. Jeffrey	Steele 1017	Pumb	*Psiguria*—ingroup
Psiguria	*warscewiczii*	(Hook. f.) Wunderlin	Steele 1006	Pwars	*Psiguria*—ingroup
Gurania	*costaricensis*	Cogn.	Steele 1009	Gcost	*Psiguria*—ingroup/outgroup
Ibervillea	*lindheimeri*	Greene	Steele 1019	Ilin	*Psiguria*—outgroup
Geranium	*macrorrhizum*	L.	Guisinger 1002	Gmac	Geraniaceae—ingroup
Pelargonium	*cotyledonis*	(L.) L'Hér.	Guisinger 1014	Pcoty	Geraniaceae—outgroup
Monsonia	*speciosa*	L.	Guisinger 1005	Mspec	Geraniaceae—ingroup
Erodium	*chrysanthum*	L'Hér. ex DC.	Guisinger 1004	Echry	Geraniaceae—ingroup
Sarcocaulon	*crassicaule*	Rehm	Guisinger 1006	Scras	Geraniaceae—ingroup

[a] Vouchers for all specimens were deposited at TEX-LL.

evolves faster than cpDNA, providing greater variation at lower taxonomic levels (Small et al., 2004). Moreover, coding sequences are often conserved such that universal primers amplify the internal transcribed spacers (ITS) in most angiosperms. These features have long supported the claim that ITS can contribute positively to phylogenetic reconstructions (Baldwin et al., 1995; Álvarez and Wendel, 2003). However, Álvarez and Wendel (2003) pointed out that several of these advantages may be counterbalanced by molecular evolutionary phenomena that can confound phylogenetic analyses. Although rDNA sequences continue to provide insights into phylogenetic history, if issues raised by Álvarez and Wendel (2003) are experimentally addressed, most systematists agree that alternatives to the ITS region are needed (Sang, 2002; Álvarez and Wendel, 2003; Small et al., 2004).

Systematists have long realized that there is a fundamental distinction between gene trees and species trees, and that a tree based on only one gene may be fundamentally incongruent with the species phylogeny, due to introgression, lineage sorting, or mistaken orthology (Doyle, 1992). The potential for low-copy nuclear (LCN or lcnDNA) markers to improve the robustness of phylogenetic reconstructions at all taxonomic levels is recognized (Small et al., 2004; Mort and Crawford, 2004).

The nuclear genome of angiosperms contains a large number of potential genes for phylogenetic analysis—predicted to be from 26,500 nuclear genes in *Arabidopsis thaliana* (L.) Heynh. to 41,000 in *Oryza sativa* L. (Sterck et al., 2007). As with rDNA, evolutionary phenomena present challenges to using LCN sequences. For example, concerted evolution may complicate analyses for similar reasons as described for rDNA leading to three possible scenarios: (1) if concerted evolution among members of a gene family is absent, then orthologous and paralogous copies would be indistinguishable despite even a complete sampling of genes from all species; (2) if concerted evolution results in complete homogenization of members of a gene family, then sampling of any gene of a gene family will result in its correct phylogenetic placement; or (3) if concerted evolution occurs but is incomplete, then sampled genes may represent a mixture of orthologous and non-homogenized, paralogous sequences or even different alleles, and accurate reconstruction of organismal phylogenies would be problematic (Sanderson and Doyle, 1992). However, in many cases, sequencing an adequate number of clones can provide the opportunity to determine which of these scenarios is occurring. Furthermore, the higher level of sequence variation characteristic of LCN genes often compensates for the added cost and effort (briefly described by Small et al., 2004). Advantages of LCN sequences include not only a higher rate of evolution than for organellar sequences, but also the potential to accumulate data sets from multiple, unlinked loci. The application of multiple LCN markers may present the only viable approach for teasing apart temporally compressed divergence events (Small et al., 2004).

The primary difficulty with using LCN sequences for low-level phylogenetic studies is the identification of easily amplifiable, relatively rapidly evolving, unambiguously alignable DNA regions that can provide sufficient variation (Baldwin et al., 1995). Examples of plant lineages that have been broadly sampled for nuclear genes are relatively few, and in those that do exist, a dissimilarity in the phylogenetic utility between the nuclear genes is evident, highlighting the need for preliminary studies to determine the most appropriate locus (or loci) (Sang, 2002; Small et al., 2004; Hughes et al., 2006).

Over the last 10 years, systematists have attempted to include LCN regions in phylogenetic studies, primarily by focusing on well-characterized genes (see Mort and Crawford, 2004; Whittall et al., 2006). While it is not a prerequisite for a phylogenetic marker to have any known molecular function, functional genes whose exon–intron boundaries have been identified have certain advantages. Exons are typically conserved enough to provide suitable primer binding sites, while introns are desired for higher variability (Schlüter et al., 2005). Several of these regions proved useful at higher taxonomic levels (for examples see: Galloway et al., 1998;

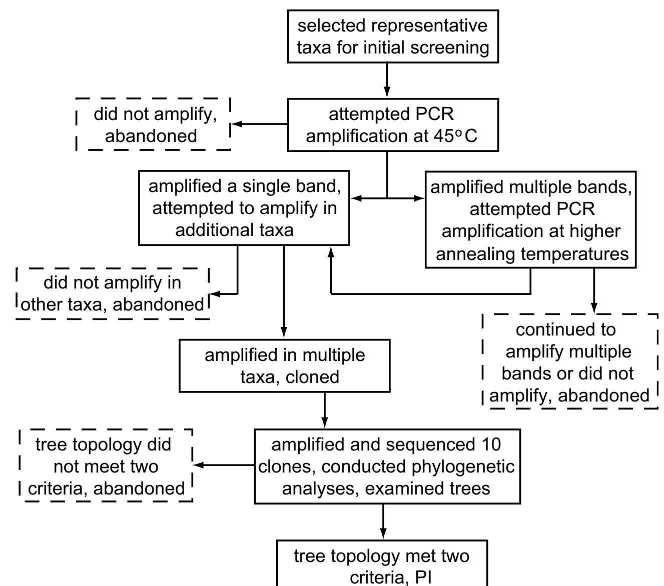

Fig. 1. Methodology used in screening 141 low-copy nuclear primer combinations in taxa at different taxonomic levels in two distantly related families of rosids. PI, potentially, phylogenetically informative.

Table 2
Eighteen primer combinations that amplified regions in both initial test species, and the number of bands visualized

Primer Pair[a]	*Psiguria bignoniacea*[b]	*Geranium macrorrhizum*[b]
5	Double	Double
13	Multiple	Multiple
15	Single	Single
31	Multiple	Multiple
46	Multiple	Multiple
51	Single	Single
56	Single	Single
61	Single	Single
64	Double	Single
69	Double	Double
73	Double	Single
85	Single	Single
86	Single	Single
90	Multiple	Multiple
125	Single	Single
126	Double	Single
129	Single	Single
133	Multiple	Multiple

[a] Primers are described in Padolina (2006); sequences online at www.bio.utexas.edu/faculty/linderr/website.
[b] "Multiple" means greater than 2.

Wang et al., 2000; Mason-Gamer et al., 1998), while few turned out to be phylogenetically informative at the species level (but see: Sang et al., 1997; Small et al., 1998; Small and Wendel, 2000; Bailey and Doyle, 1999). Further, these LCN regions do not always amplify across diverse plant groups, and if they do, they do not always provide resolution at the desired taxonomic level. Such genes often have multiple copies, which can confound phylogenetic analyses; therefore, ideally one would use single-copy genes.

The limited success in using genes characterized by molecular biologists has prompted other strategies for identifying phylogenetically useful nuclear regions. Schlüter et al. (2005) described four alternative approaches to identifying LCN markers when little or no sequence information is available: (1) design of new primers from information in sequence databases; (2) isolation of homologous DNA using a gene probe from another organism; (3) characterization of sequence markers from DNA fingerprints; and (4) obtaining novel sequences via cDNA cloning. Unfortunately, many practical applications of these four strategies are not yet available for comparison. One exception includes the search through and comparison between the genomic libraries of those organisms well represented in public nucleotide databases in order to identify informative regions (Álvarez et al., 2008). However, whole genome sequences for most groups of flowering plants are still nonexistent. Another exception might be production of a project-specific cDNA library. For example, Whittall et al. (2006) used cDNAs to design primers that could amplify the 3′UTR region of various genes for investigating adaptive radiation in *Aquilegia*.

The problem with identifying "universal" primers to amplify LCN markers across flowering plants results from variation in rates of sequence evolution and gene histories that exist between groups of angiosperms (Small et al., 2004). Indeed many systematists

Table 3
Twelve regions cloned and sequenced in one or both angiosperm groups, and the results of phylogenetic analyses including additional taxa

Primer pair[f]	*Psiguria*—11 LCN regions		Geraniaceae—6 LCN regions	
	Results of phylogenetic analyses[a]	Taxa included in cloning[g]	Results of phylogenetic analyses[a]	Taxa included in cloning[g]
6	MC, NR (42); Supplementary Fig. 4	Gcost, Pped, Prace, Ptab, Pumb (EU648194–EU648235)	–	–
15	MC (35); Supplementary Fig. 5	Gcost, Pbig, Prace, Pumb, Pwars (EU293009–EU293043)	PI (35); Fig. 4	Echry, Gmac, Mspec, Pcoty, Scras (EU648724–EU648758)
51	PI cpy1(42), MC cpy2(40); Figs. 5 and 6, & Supplementary Fig. 6	Gcost, Ilin, Pbig, Pped, Prace, Pumb (EU648514–EU648595)	MC (43); Supplementary Fig. 13	Echry, Gmac, Mspec, Pcoty, Scras (EU648681–EU648723)
57	MC cpy1 (40), cpy2 no phyl[c] (13); Supplementary Figs. 7 and 8	Gcost, Pbig, Pped, Prace, Ptab, Pumb (EU648808–EU648860)	–	–
61	PI cpy1 (48), cpy2 no Ilin (17)[c]; Supplementary Figs. 2 and 3	Gcost, Ilin, Pbig, Pped, Prace, Pumb (EU648449–EU648513)	–	–
64	–	–	MC (40); Supplementary Fig. 14	Echry, Gmac, Scras[d] (EU648596–EU648635)
64s[b]	MC (55); Supplementary Fig. 9	Gcost, Ilin, Pbig, Pped, Prace, Pumb, Pwars (EU648394–EU648448)	–	–
85s[b]	MC (48); Supplementary Fig. 10	Gcost, Ilin, Pbig, Pped, Prace, Pumb, Pwars (EU648146–EU648193)	–	–
86	NR (55); Fig. 3	Gcost, Pbig, Prace, Pumb, Pwars (EU648339–EU648393)	MC (49); Supplementary Fig. 15	Echry, Gmac, Mspec, Pcoty, Scras (EU648861–EU648909)
91	–	–	PI (45); Supplementary Fig. 1	Echry, Gmac, Mspec, Pcoty, Scras (EU648636–EU648680)
96	PI (46); Fig. 7	Gcost, Ilin, Pbig, Pped, Prace (EU648293–EU648338)	–	–
125	MC, NR (49); Supplementary Fig. 11	Gcost, Pbig, Prace, Pumb, Pwars (EU648759–EU648807)	MC (49); Fig. 2	Echry, Gmac, Mspec, Pcoty, Scras (EU648910–EU648958)
129	MC (57); Supplementary Fig. 12	Gcost, Ilin, Pbig, Pped[e], Prace, Pumb (EU648236–EU648292)	–	–

The total number of clones amplified and sequenced for each region is shown in parentheses.
[a] MC, multiple, indistinguishable copies of the amplified region; NR, little or no resolution between taxa; PI(shaded), potentially, phylogenetically informative; "–" indicates region was not pursued with cloning, and "cpy" distinguishes one of the orthologous copies.
[b] An "s" was added to those names for which primers specific to *Psiguria* were designed (64s-forward—5′-GAA TAG CCA AGG ATA CGC-3′, 64s-reverse—5′-CGA CTT CCT CAA GCA AGC-3′, 85s-forward—5′-TTT GCT GGG GAT GAT GCT CC-3′, 85s-reverse—5′-AAT CCA GAC ACT GTA TTT CC-3′).
[c] Where less than 20 clones of a copy were recovered, a phylogenetic analysis was not conducted ("no phyl").
[d] Taxa "Mspec and Pcoty" did not amplify.
[e] A different collection of *P. pedata* was used (Steele 1036).
[f] Primers are described in Padolina (2006); sequences online at www.bio.utexas.edu/faculty/linderr/website.
[g] Accession numbers for sequences uploaded to GenBank are shown in parentheses.

question whether universal primers even exist. Sang (2002) predicted that it would be unlikely that there would be universal primers for the majority of LCN genes used in plant phylogenetic studies because they would have too many degenerative sites, reducing amplification specificity and efficiency.

Padolina (2006) took a computational approach to designing universal LCN angiosperm primers by utilizing the database engine, MoBIoS, designed specifically for biological inquiries of DNA and protein sequences (Xu et al., 2004). She queried this database by comparing the nuclear genomes of the monocot *O. sativa* and the eudicot *A. thaliana* and searching for primer combinations that occurred only once in each of the two genomes. After imposing additional criteria as described in Padolina (2006), this search resulted in 141 primer combinations, which she screened in *A. thaliana*, the monocot *Phalaenopsis* Blume (Orchidaceae), and in the asterid/eudicot *Helianthus* L. (Asteraceae).

In the current study we screened these 141 primer combinations at different taxonomic levels in two distinct groups of rosids—*Psiguria* Neck ex. Arn. (Cucurbitaceae) and Geraniaceae. Our goals included: (1) identifying potential LCN markers for examining phylogenetic relationships in rosids, one of the largest clades of angiosperms; (2) providing new phylogenetic markers for expanded investigations of the two lineages; (3) investigating the hypothesis that universal primers for amplifying phylogenetically informative regions across angiosperms may not exist; and (4) developing an experimental strategy that other groups can utilize to identify phylogenetically informative regions.

2. Materials and methods

2.1. Study organisms

Rosids make up one of several large clades of angiosperms within the core eudicots, including approximately 140 families, perhaps one-third of all angiosperm species, and roughly 39% of eudicots

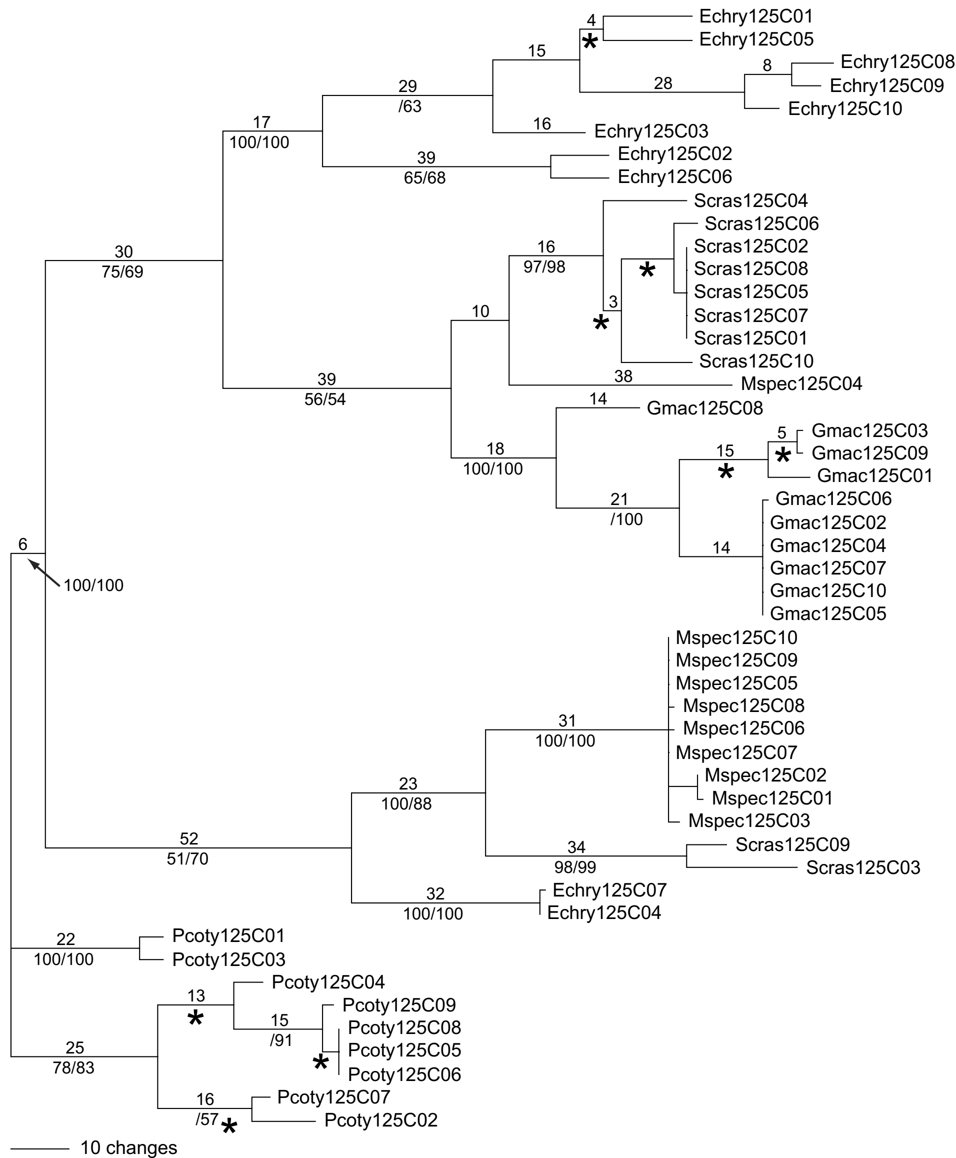

Fig. 2. Example of phylogenetic tree with multiple, indistinguishable copies (MC) of the region amplified with primer pair 125 in Geraniaceae. One of 2,772,110 MP trees (length = 720; CI = 0.53; RI = 0.84) of 49 clones of the LCN gene, *ADP/ATP translator*, for five species of Geraniaceae. ML tree topology ($-\ln L = 4189.00$) was identical to this MP tree except for minor variations within the large Echry clade that do not alter the overall tree topology. Nodes marked with an "*" collapse in strict consensus trees. Numbers above the lines indicate the number of changes, and numbers below the lines indicate ML/MP bootstrap values >50%. "C" followed by a numerical value in the taxon names indicates the clone copy number.

(Soltis et al., 2005). The two groups used in this study are currently placed in distantly related positions in the rosid clade—*Psiguria* (Cucurbitaceae) in Cucurbitales within the nitrogen-fixing clade of eurosids I, and Geraniaceae in Geraniales sister to eurosids II (Soltis et al., 1995, 2000, 2003; Savolainen et al., 2000a,b; Jansen et al., 2007).

All taxa used in the study are shown in Table 1 along with voucher numbers and abbreviations. These include one taxon from each of the five genera of Geraniaceae (Parkinson et al., 2005), and several *Psiguria* spanning the morphological and geographical range of the genus. All listed species of Geraniaceae were included in cloning the amplicons from each successful primer pair, with one exception noted below. Not all listed species of *Psiguria* were cloned for every primer pair investigated. In the interest of time, those that successfully amplified and cloned quickly and with the least difficulty were included (at least three in each ingroup). *Gurania costaricensis* is a member of the genus that is putatively sister to *Psiguria*, and *Ibervillea lindheimeri* represents a distantly related Cucurbitaceae that was readily available for collection.

2.2. Screening primer combinations

One hundred and forty-one primer combinations (described in Padolina, 2006; sequences online at www.bio.utexas.edu/faculty/linderr/website) were investigated following the strategy outlined in Fig. 1. Representative taxa were selected, initial screening with polymerase chain reaction (PCR) was conducted at 45 °C, successful markers were cloned and sequenced, sequences of multiple clones were analyzed phylogenetically, and potentially useful regions were identified based on resulting tree topologies.

Initially, one taxon was investigated in each study group—*Geranium macrorrhizum* (Geraniaceae) and *Psiguria bignoniacea* (Cucurbitaceae). Leaf material was field-collected and dried over silica, or it was harvested from cultivated plants grown in the greenhouse at The University of Texas at Austin. Total DNA was extracted using either the CTAB protocol of Doyle and Doyle (1987) or the DNeasy Plant DNA Extraction Kit (Qiagen Inc., GmbH, Germany). CTAB products were purified by ultracentrifugation in cesium chloride and ethidium bromide gradients (Sambrook et al., 1989).

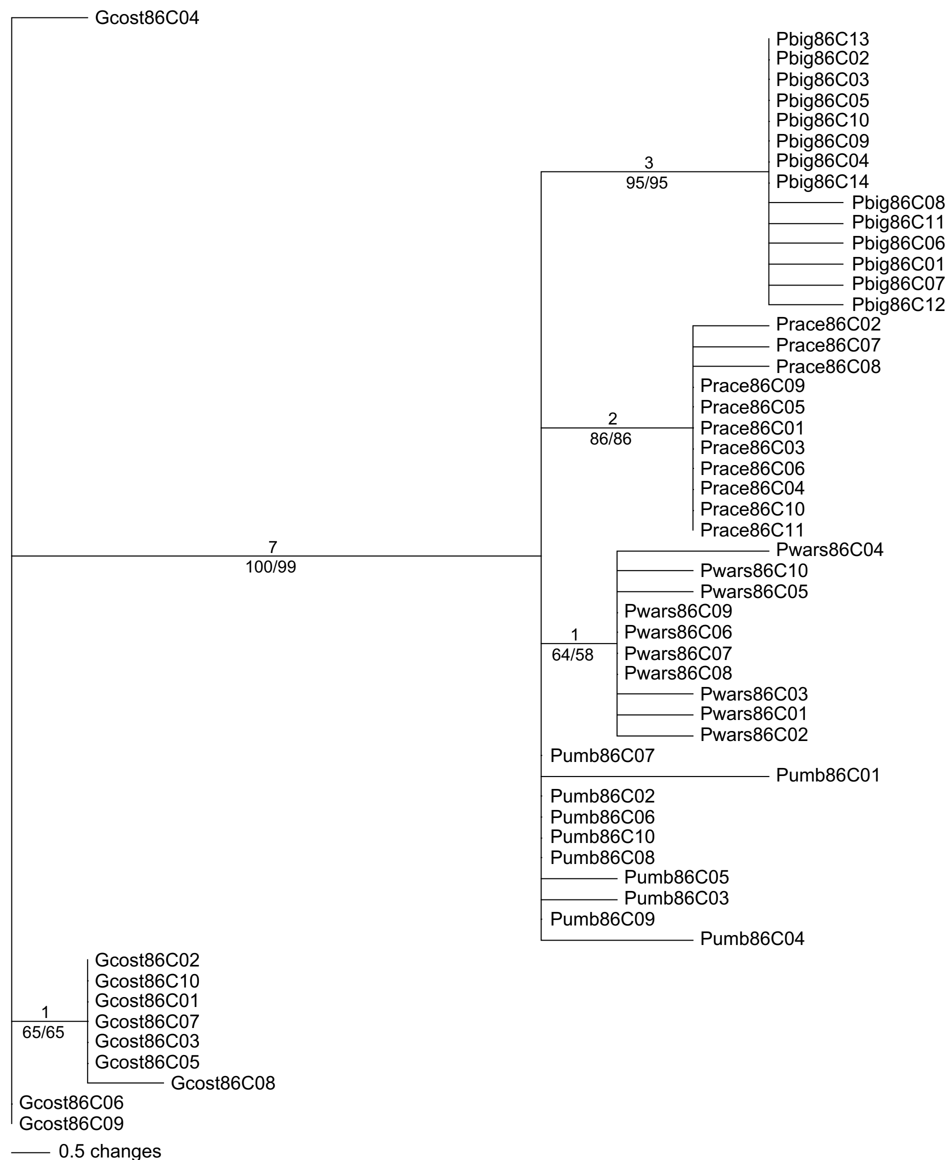

Fig. 3. Example of phylogenetic tree with little or no resolution between species (NR) in the region amplified with primer pair 86 in *Psiguria*. One of two MP trees (length = 39; CI = 0.94; RI = 0.99) of 55 clones of the LCN gene, *methionine synthase*, for four species of *Psiguria* and the outgroup Gcost. ML tree topology (−ln *L* = 993.30) was identical to this MP tree. Numbers above the lines indicate the number of changes, and numbers below the lines indicate ML/MP bootstrap values >50%. "C" followed by a numerical value in the taxon names indicates the clone copy number.

Amplification was performed using PCR in 25 μL volumes containing 15.2 μL of ddH_2O, 6 μL of FailSafe buffer—PreMix E for *Psiguria* and PreMix D for *Geranium* L. (EPICENTRE® Biotechnologies, Madison, WI), 0.4 μL of a 10-μM solution of each forward and reverse primer, 1.0 μL of *Taq* polymerase (produced in the laboratory of R.K.J. following the protocol of Pluthero (1993) and diluted to 1 U), and 2 μL of unquantified DNA template. Initial reaction conditions were as follows: one round of amplification consisting of denaturation at 94 °C for 1 min 30 s, annealing at 45 °C for 30 s, and extension at 72 °C for 1 min 30 s, followed by 29 cycles of 94 °C for 30 s, 45 °C for 30 s, and 72 °C for 1 min 30 s, with a final extension step of 72 °C for 7 min. Amplifications were visualized on 1% agarose gels with ethidium bromide and a size standard to estimate fragment sizes and quantities. If one or both test species were not observed to amplify in the first attempt, that primer combination was abandoned in this study, but it may have been pursued as part of the systematics studies of Geraniaceae or *Psiguria* discussed below. If two or more bands were observed, PCR amplification of the region was attempted at higher annealing temperatures (progressively ranging from 45 to 55 °C) until a single fragment was observed or the amplification failed.

For the regions that amplified as described above, taxon sampling was expanded, and amplicons were cloned. Cloning was conducted using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA). We attempted to amplify at least 10 colonies from each individual using PCR in 25 μL volumes containing 17.5 μL of ddH_2O, 2.5 μL of 10× buffer, 1.5 μL of 25 μM $MgCl_2$, 2 μL of 0.25 μM dNTPs (or 17.0 μL of ddH_2O and 6.5 μL of FailSafe buffer PreMix E) plus 0.25 μL of a 20-μM solution of each pUC-18 plasmid primer (forward 5′-CCA GTC ACG ACG TTG TAA-3′ and reverse 5′-CGG ATA ACA ATT TCA CAC-3′), 1.0 μL of *Taq* polymerase, and one colony. Reaction conditions were as follows: one hot-start cycle at 95 °C for 3 min 30 s (after which the *Taq* polymerase was added to each reaction), followed by 35 cycles consisting of denaturation at 95 °C for 45 s, annealing at 56 °C for 45 s, and extension at 72 °C for 50 s, with a final extension step of 72 °C for 3 min. PCR amplicons were cleaned using Exo-SAP by adding a 3-μL solution of 2.25 μL of ddH_2O, 0.25 μL of Exonuclease I (New England Biolabs, Inc., Ipswich, MA), and 0.50 μL of Shrimp Alkaline Phosphotase (Promega Corp., Madison, WI) to each product, and processing on a thermocycler at 37 °C for 30 min, followed by 80 °C for 15 min. Sequencing was conducted at the ICMB Core Facilities at The University of Texas at Austin, using ABI Big Dye chemistry.

2.3. Phylogenetic analyses

Forward and reverse sequences of clones were assembled and edited with Sequencher 4.5 (Gene Codes Corp., 2000), then multiple clones of the same primer pair amplicon were aligned with ClustalX (Thompson et al., 1997) followed by manual adjustments. Maximum Parsimony (MP) analyses were performed on each data set with PAUP* 4.0b10 (Swofford, 2002). For each data set, heuristic searches were conducted using 100 random addition replicates with tree-bisection–reconnection (TBR) branch swapping, characters equally weighted, gaps treated as missing, and the MulTrees option. Support for internal nodes was assessed using bootstrap analysis (Felsenstein, 1985) of 100 replicates with one random addition per replicate. For five data sets (Geraniaceae—primer pairs 125, 51, and 86, and *Psiguria*—primer pairs 51 and 129), a time limit of 10 min per replicate was enforced. Maximum Likelihood (ML) analyses were performed using Garli 0.951 with the default model ($GTR + \Gamma + I$) of evolution (Zwickl, 2006). The ML analyses utilized the automated stopping criterion, terminating a search when the $\ln L$ score remained constant for 20,000 consecutive generations. Likelihood scores were calculated using PAUP*, which better optimizes branch lengths (Zwickl, 2006). ML bootstrap analyses were performed in Garli with 100 replicates using an automated stopping criterion set at 10,000 generations.

In analyses of Geraniaceae, *Pelargonium cotyledonis* was used to root trees based on Parkinson et al. (2005). In analyses for *Psiguria*, *Gurania* Cogn. was initially included as the outgroup (Jeffrey, 2005), and for the most promising primer combinations, a more distant outgroup, *Ibervillea lindheimeri*, was used.

For each LCN marker, the topologies of the resulting trees were examined for evidence of phylogenetic utility based on two criteria. First, a single copy or multiple, orthologous copies were observed. If the clones of each taxon formed a monophyletic group without any clones outside of that group, the region was considered to be a single copy even if there was an inparalog in a taxon such as described by Koonin (2005). If clones were not all in monophyletic groups by taxon, but a monophyletic clade of the tree included multiple clones of all the species included in the study group, and all of those clones were in monophyletic groups by taxon, then it was concluded that this was an orthologous copy that could be distinguished from other paralogous copies. Second, there had to be sufficient variation based on moderate to strong (>80%) bootstrap support for individual clades.

3. Results

Amplification results at the 45 °C annealing temperature for all 141 primer combinations in the two initial taxa appear in Appendix A. Ninety-five primer pairs (67.4%) did not amplify regions in either *Geranium* or *Psiguria*, 28 pairs (19.8%) amplified in only one of the two taxa, and 18 pairs (12.8%) amplified in both species. For those 18 regions, the number of observed bands is listed in

Table 4
Tree characteristics for five potentially, phylogenetically informative markers—two in Geraniaceae, three in *Psiguria*

Primer pair[a]	BLAST match	Study group	Number of clones	Aligned characters (exon/intron)[d]	Parsimony-informative characters[e]	No. of MP trees	Tree length	CI[f]	RI[g]	ML score
15	*hsp90*	Geraniaceae	35	1093 (1093/0)	289 (26.4%), 289 (100%), 0 (0%)	7	411	0.73	0.94	−4224.10
51 copy1[b]	*atpB*	*Psiguria*	42	622 (300/322)	30 (4.8%), 11 (3.7%), 19 (5.9%)	3	75	0.91	0.99	−1355.60
61 copy1[b]	actin	*Psiguria*	48	1132 (1000/132)	61 (5.4%), 52 (5.2%), 9 (6.8%)	16	108	0.86	0.97	−2339.91
91	*hsp70*	Geraniaceae	45	1173 (1173/0)	424 (36.1%), 424 (100%), 0 (0%)	8	1009	0.57	0.91	−7239.94
96	*s/t phos*[c]	*Psiguria*	46	1384 (295/1089)	104 (7.5%), 12 (0.3%), 92 (8.4%)	2	168	0.95	0.99	−3028.24

Results are from maximum parsimony (MP) analyses using PAUP* and maximum likelihood (ML) analyses conducted with Garli. ML scores calculated in PAUP*.
[a] Primers are described in Padolina (2006); sequences online at www.bio.utexas.edu/faculty/linderr/website.
[b] Data for these regions given only for the orthologous copy detected in *Psiguria* (shown as copy 1).
[c] *s/t phos*, serine/threonine phosphoesterase.
[d] Total aligned characters and aligned characters that matched exon/intron in BLAST search.
[e] Total parsimony-informative characters (and %), number of parsimony-informative characters in exon (and %), number of parsimony-informative characters in intron (and %).
[f] CI, consistency index (excluding autapomorphies).
[g] RI, retention index.

Table 2. With six primer pairs (5, 13, 46, 69, 73, and 90), the initial amplifications resulted in multiple bands for one or both species, and at higher annealing temperatures multiple bands persisted or amplification was unsuccessful. No further adjustments were made to the PCR routines, and these primer combinations were abandoned. Of the remaining 12 primer pairs, six regions (31, 61, 85, 126, 129, and 133) were abandoned for one or more of the following reasons: (1) did not successfully amplify at higher annealing temperatures in *Geranium*; (2) cloning showed multiple size copies; (3) did not amplify in the other taxa within Geraniaceae; or (4) the region was not informative for *Psiguria*. Because they looked promising only for *Psiguria*, some of these six were pursued in *Psiguria* and will be discussed below. In both initial species, primer pair 56 amplified *nad5*—a gene encoded by the mitochondrial genome. Sequences of PCR products were very clean, and were compared between *Psiguria bignoniacea*, *Psiguria umbrosa,* and *Gurania costaricensis.* For *nad5*, there was no variation between these taxa, and there was little variation between *P. bignoniacea* and *Geranium macrorrhizum*, so this marker was also abandoned.

Results of cloning, sequencing, and assessment of phylogenetic utility for the remaining five regions (15, 51, 64, 86, and 125) in both rosid groups are shown in Table 3. Additionally, results are given for six regions pursued individually in *Psiguria* (6, 57, 61, 85, 96, and 129) and one region in Geraniaceae (91). For primer pairs 64 and 85, amplification in *Psiguria* was not successful in all taxa, so specific primers (names shown with an "s" in Table 3) were designed to amplify problematic taxa. Three different results of cloning were observed in phylogenetic analyses: (1) multiple, indistinguishable copies of the amplified region (i.e. those for which orthology/paralogy relationships could not be assessed) (MC); (2) little or no resolution between taxa (NR); or (3) the region was potentially, phylogenetically informative (PI). Examples of these include the MC result for primer pair 125 in Geraniaceae (Fig. 2). Two copies were detected in *Monsonia speciosa, Erodium chrysanthum*, and *Sarcocaulon crassicaule*, and these copies were not congruent with previous phylogenetic estimates of Geraniaceae (Parkinson et al., 2005). An analysis that resulted in an NR result is shown in Fig. 3 with primer pair 86 in *Psiguria*. Of the 509

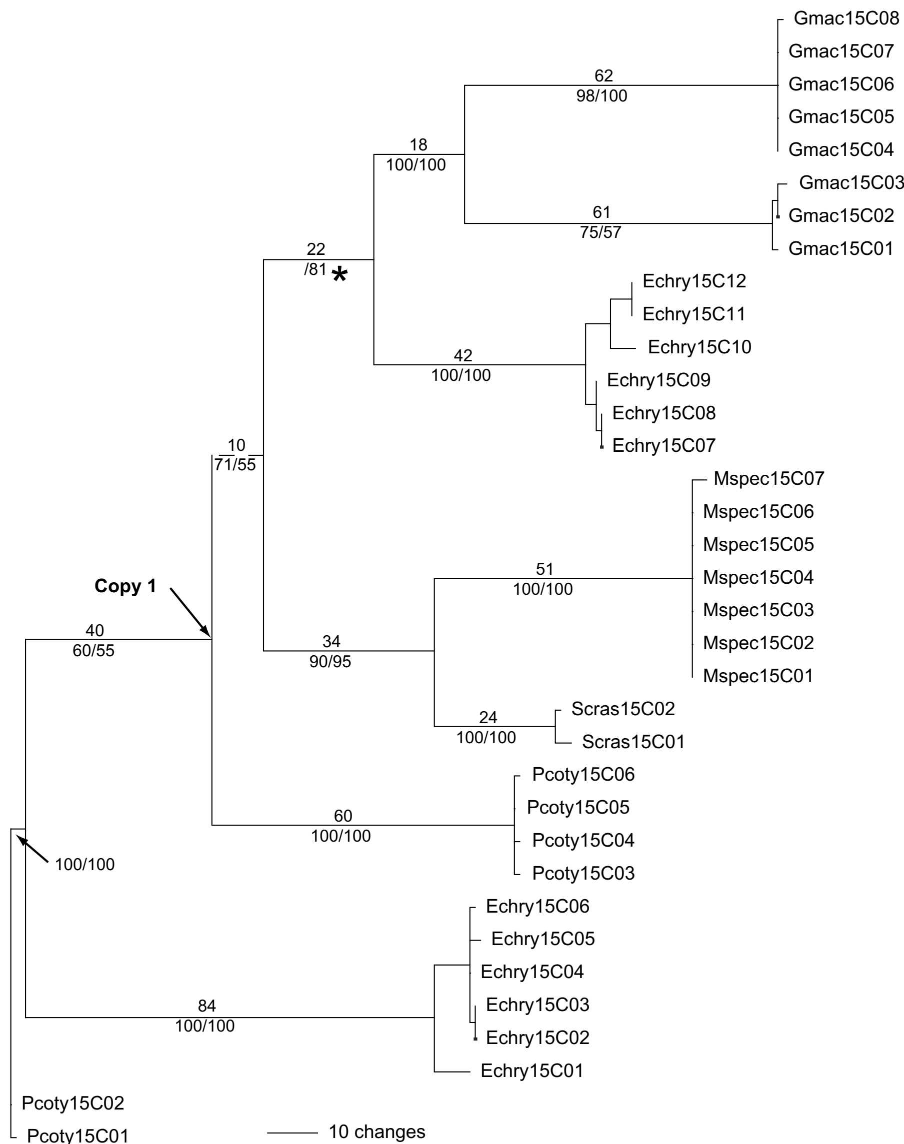

Fig. 4. One of seven MP trees (length = 411, CI = 0.73, and RI = 0.94) of 35 clones of primer pair 15 (*hsp90*) for five species of Geraniaceae. ML tree ($-\ln L = 3160.31$) was congruent with this MP tree. Nodes marked with an "*" collapse in strict consensus trees. Numbers above the lines indicate the number of changes, and numbers below the lines indicate ML/MP bootstrap values >50%. "C" followed by a numerical value in the taxon names indicates the clone copy number.

aligned characters, only 15 were parsimony-informative, and only 1- to 3-bp differences were detected between taxa and between clones of the same taxon.

No markers were found to be phylogenetically informative in both Geraniaceae and *Psiguria*, but two regions were PI in Geraniaceae (15 and 91) and three in *Psiguria* (51, 61, and 96) (shaded in Table 3). Tree statistics for both MP and ML analyses are shown for these regions in Table 4. In Geraniaceae, primer pair 15 amplified heat shock protein 90 (*hsp90*), and primer pair 91 amplified heat shock protein 70 (*hsp70*). Both regions completely matched exons in BLAST searches. Multiple copies of both *hsp90* and *hsp70* were detected in some taxa, but in each case, one copy was distinguishable from the others, and all clones from the same species for this orthologous copy (indicated as copy 1) formed monophyletic groups (see Fig. 4 for *hsp90*—primer pair 15). The relationships for copy 1 clones are congruent with previous phylogenetic trees with *Pelargonium* sister to the remaining four genera, *Erodium* sister to *Geranium*, and *Monsonia* sister to *Sarcocaulon* (Parkinson et al., 2005). This tree also revealed a duplication of copy 1 in *Geranium* as evidenced by the presence of an inparalog (two divergent clades of clones). The phylogenetic analysis of the *hsp70* (primer pair 91) data set revealed a slightly more complicated situation. There were duplicate copies present in several taxa, yet one distinguishable, orthologous copy grouped together. Again, the relationships of the genera are congruent with those from previous studies (Supplementary Fig. 1) (Parkinson et al., 2005).

In *Psiguria*, primer pair 51 amplified *atp synthase subunit β (atpB)* consisting of approximately half exon and half intron. An orthologous copy with a 29% divergence from all other copies was found. Phylogenetic analyses were conducted for all sequences together (Fig. 5) and for the orthologous copy (copy 1) alone (Fig. 6). As seen in both trees, it is clear that "copy 1" is quite different from the other copies; the clones for each taxon group together, *Psiguria* is monophyletic, *Gurania* is sister, and branches have mod-

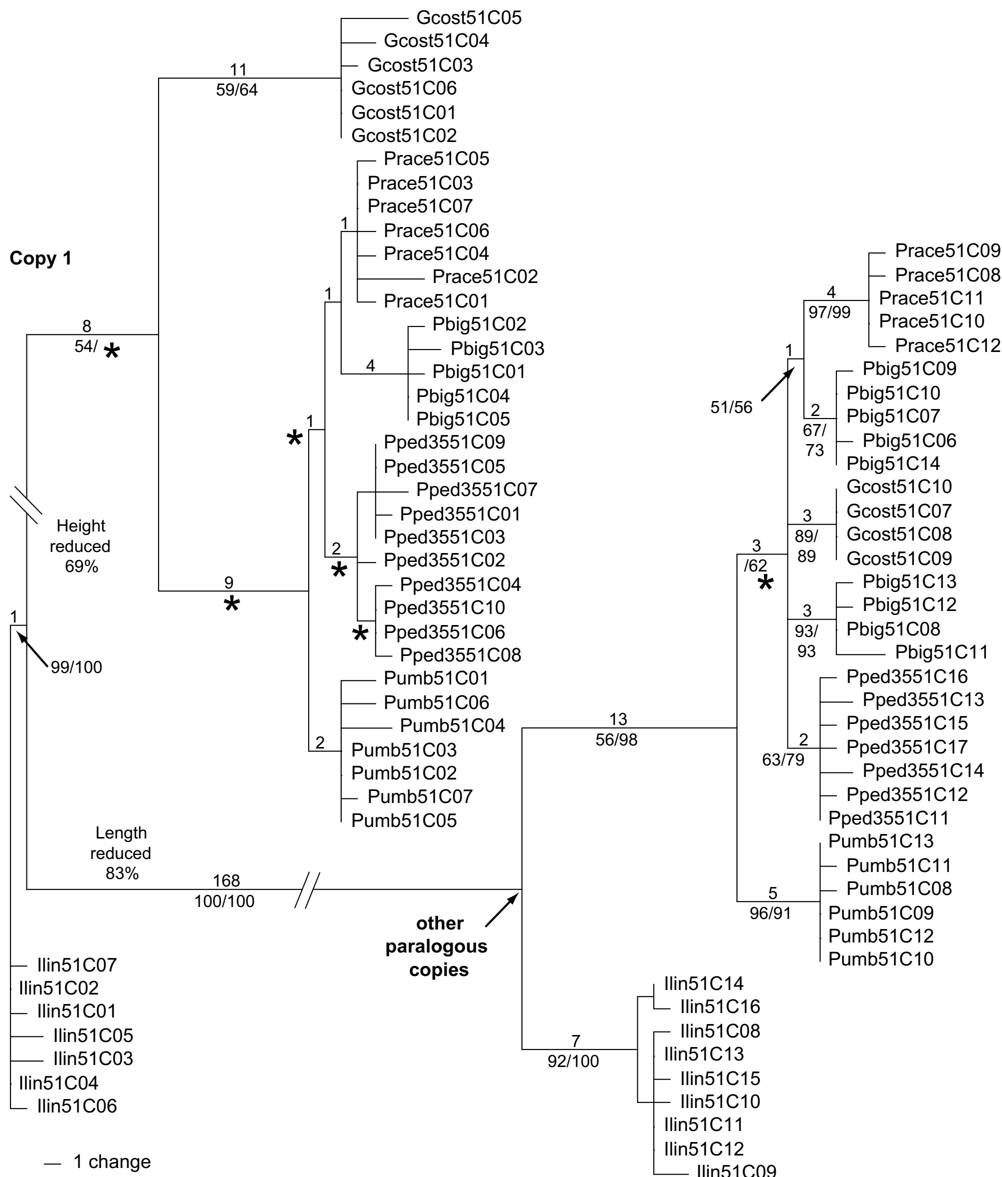

Fig. 5. One of 268 MP trees (length = 336; CI = 0.78; RI = 0.99) with all 82 clone sequences of primer pair 51 (*atpB*) in 6 species of *Psiguria* and the outgroup Ilin, showing one orthologous copy (copy 1) that is 29% divergent from all other copies. ML tree (−ln*L* = 2376.16) was congruent with this MP tree except that the Pumb clade in copy 1 was sister to the Gcost clade. Nodes marked with an "*" collapse in strict consensus trees. Numbers above the lines indicate the number of changes, and numbers below the lines indicate ML/MP bootstrap values >50%. "C" followed by a numerical value in the taxon names indicates the clone copy number.

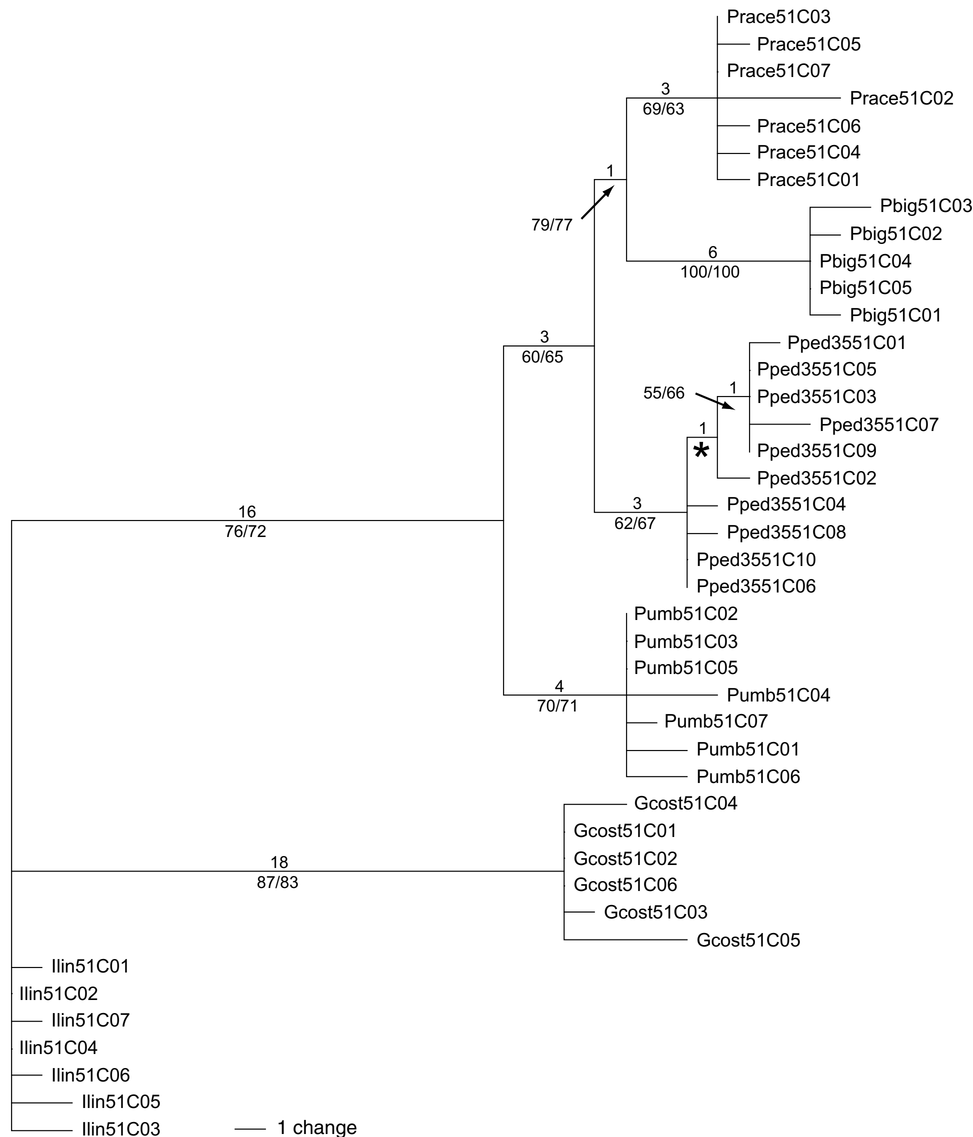

Fig. 6. One of three MP trees (length = 75; CI = 0.91; RI = 0.99) for the 40 orthologous, copy 1 clones of the LCN gene, *atpB* (primer pair 51) for four species of *Psiguria*, Gcost, and the outgroup Ilin. ML tree ($-\ln L = 1355.58$) was congruent with this MP tree. Nodes marked with an "*" collapse in strict consensus trees. Numbers above the lines indicate the number of changes, and numbers below the lines indicate ML/MP bootstrap values >50%. "C" followed by a numerical value in the taxon names indicates the clone copy number.

erate to strong bootstrap support. The paralogous group of clones reveals two similar copies of *Psiguria bignoniacea,* and, since this topology did not meet the criteria above, these sequences were excluded from the phylogenetic analyses. In the analysis that excludes the paralogs, (Fig. 6), bootstrap support values are higher, and only one branch collapses in the strict consensus tree.

Also in *Psiguria*, primer pair 61 amplified actin (88% exon) and primer pair 96 amplified a small portion of serine/threonine phosphoesterase (*s/t phos*) with a high percentage of intron (79%). Phylogenetic analyses of the actin (61) data set revealed two copies in *Psiguria* with 20% divergence. When all of the sequences were analyzed together (Supplementary Fig. 2), in both copies the *Gurania* clade was sister to a monophyletic clade of *Psiguria*, but this relationship collapsed in the strict consensus. The same result was obtained when copy 1 was analyzed alone (Supplementary Fig. 3). Cloning and sequencing of *s/t phos* (primer pair 96) detected only one copy in *Psiguria.* The clones formed monophyletic groups by taxon, *Psiguria* formed a monophyletic group, and the bootstrap analyses resulted in relatively strong support for individual clades (Fig. 7). The phylogenetic analysis revealed inparalogs in both *G. costaricensis* and *P. racemosa* as evidenced by the presence of two divergent clades of clones in each taxon.

4. Discussion

The need for well-resolved evolutionary histories of organisms at low taxonomic levels and multiple, independent markers for inferring hybrid speciation events fuels the effort to identify LCN markers useful for estimating phylogenetic relationships. A multitude of problems can hinder successful phylogenetic interpretation when using only chloroplast and nuclear ribosomal sequences, such as uniparental inheritance, lack of variation, unreliability of concerted evolution, secondary structure, the presence of major and minor rDNA arrays, and pseudogenes of various ages (Corriveau and Coleman, 1988; Sang, 2002; Álvarez and Wendel, 2003; Small et al., 2004). Furthermore, since gene trees may not reflect species trees, basing final estimates of phylogenetic relationships

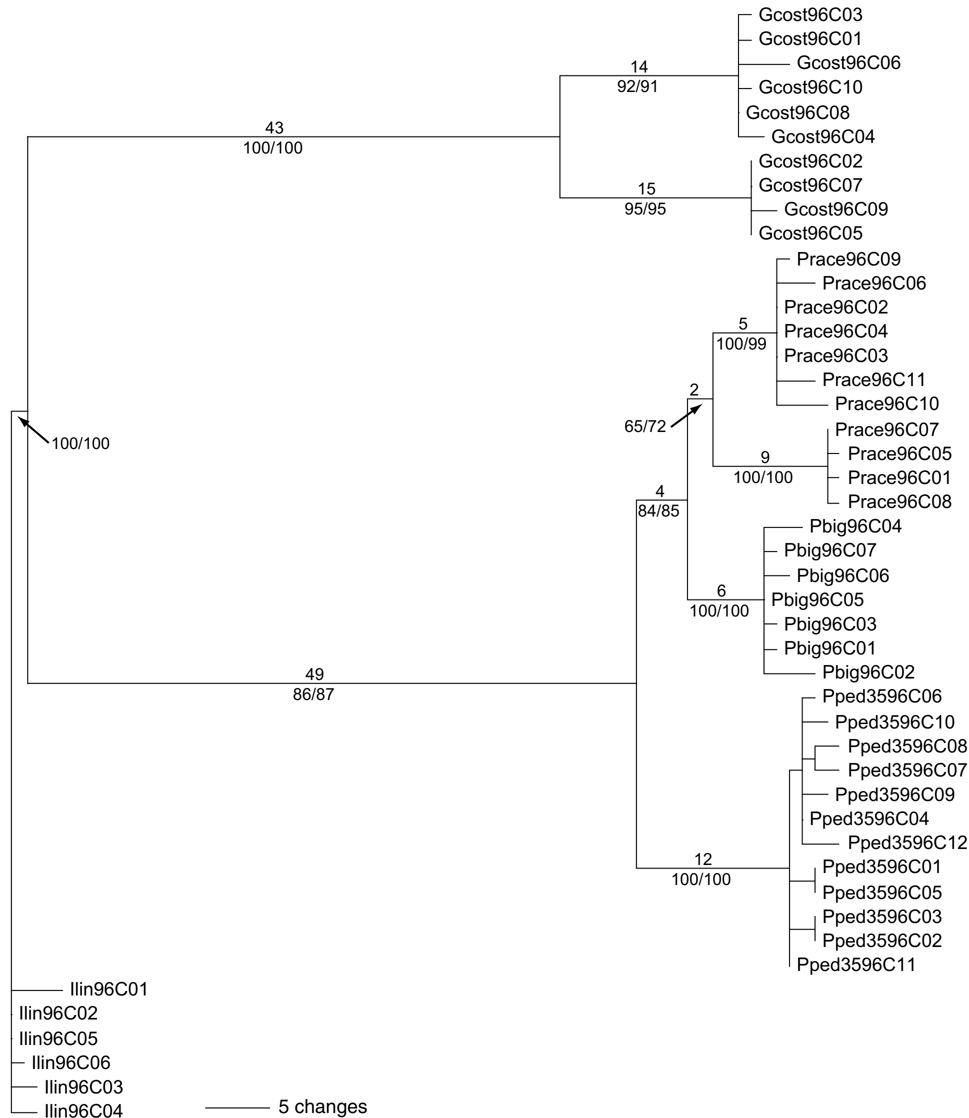

Fig. 7. One of two MP trees (length = 168; CI = 0.95; RI = 0.99) for 46 clones of the LCN gene, *serine/threonine phosphoesterase* (primer pair 96) for three species of *Psiguria*, Gcost, and the outgroup Ilin. ML tree ($-\ln L$ = 3028.24) was congruent with this MP tree. Numbers above the lines indicate the number of changes, and numbers below the lines indicate ML/MP bootstrap values >50%. "C" followed by a numerical value in the taxon names indicates the clone copy number.

on multiple, independent markers can strengthen results (Doyle, 1992). However, the effort to identify such markers has been impeded by difficulties in identifying nuclear primer combinations that amplify orthologous regions in multiple groups (i.e. "universal" primers). And, even if primers do amplify, lack of variation and problems confirming orthology can further hinder efforts (Baldwin et al., 1995; Small et al., 2004; Schlüter et al., 2005).

Several strategies (with various levels of application) for identifying phylogenetically informative regions of the nuclear genome have been reported. Although Strand et al. (1997) published a set of nuclear DNA-based markers over 10 years ago, they were only identified by their ability to amplify and not for their phylogenetic utility. The use of random genomic regions identified by RAPD or AFLP primers has been promoted by some researchers but is problematic for several reasons including the lack of knowledge about inheritance and the possibility of amplifying regions of junk repeat DNA, with consequent paralogy problems (Hughes et al., 2006). For a few angiosperm groups (including Brassicaceae, Asteraceae, and Fabaceae in eudicots, and several monocots—*Oryza* (rice), *Allium* L. (onion), and *Musa* L. (banana)), the whole genomes of representative taxa have been sequenced. For taxa closely related to these, expressed sequenced tag (EST) collections have been quite useful for intrafamilial phylogenetic studies (for recent examples see Chapman et al., 2007; Álvarez et al., 2008; Bacon et al., 2008), but there is no indication that they would be useful in distantly related groups. Finally, the use of LCN genes identified by molecular biologists has strengthened some phylogenetic reconstructions, but the results are highly variable (Sang, 2002; Mort and Crawford, 2004; Small et al., 2004; Whittall et al., 2006). Few plant lineages have been thoroughly sampled for phylogenetic utility of these nuclear genes; therefore, there is little evidence that these regions will be useful in other groups.

The primary purpose of this study was to test the phylogenetic utility of 141 primer pairs identified by Xu et al. (2004) and Padolina (2006), in two taxonomically disparate groups of rosids. Because these primer pairs were designed from conserved regions in distantly related angiosperms (a monocot and a eudicot), the hope was that they might be universally amplifiable as well as phylogenetically informative. Our results show that "universal" primers may not exist across angio-

Table 5
Thirty-two primer combinations that amplified in representative species of at least two of five distantly related angiosperm families in the combined data sets of this screening study and that conducted by Padolina (2006)

Primer pair	BLAST match[a]	*Psiguria* (Cucurbitaceae)	*Geranium* (Geraniaceae)	*Helianthus* (Asteraceae)	*Phalaenopsis* (Orchidaceae)	*Arabidopsis* (Brassicaceae)
1	No match		X	X	X	
5	Ubiquitin, polyubiquitin	X	X			
6	Actin	X			**X** (shaded)	
13	No match	X	X		X	X
15	Heat-shock protein 90	X	**X** (shaded)		X	X
30	No match	X		X	X	X
31	No match	X	X			
46	Ubiquitin, polyubiquitin precursor	X	X		X	X
50	Heat-shock protein	X		X	X	X
51	atpB	**X** (shaded)	X	X	**X** (shaded)	X
56	*nad*5[b]	X	X	X	X	X
57	Heat-shock protein	X		X	**X** (shaded)	X
58	β-Tubulin		X		X	X
61	Actin	**X** (shaded)	X	X	X	X
64	Transitional endoplasmic reticulum ATPase	X	X		X	X
65	Cellulose synthase		X			X
68	ATPase B subunit		X		X	X
69	Plasma membrane H^+-ATPase	X	X		X	X
73	Ribosomal protein L11 (RL5)	X	X		X	X
81	Elongation factor 1-α		X			X
82	No match		X			X
85	Actin	X	X	**X** (shaded)		X
86	Methionine synthase	X	X	X	**X** (shaded)	X
87	Cellulose synthase				X	X
90	No match	X	X		X	
91	Heat-shock protein		**X** (shaded)		X	X
94	No match				X	X
96	s/t phos	**X** (shaded)				X
125	ADP/ATP translator	X	X			
126	Ribosomal protein S4 type I (rps4)	X	X			
129	No match	X	X			
133	Ammonium/proton antiporter	X	X			

Successful amplifications are indicated with an "X". Shaded boxes indicate regions that were determined to be potentially, phylogenetically informative.
[a] If the amplified region was sequenced, BLAST searches were conducted with the entire sequence, but if not, the region was identified as described in Padolina (2006).
[b] "*nad*5" is a mitochondrial-encoded gene.

sperms. However, our results do provide a reduced set of primer combinations ideal for preliminary investigations into identifying phylogenetically informative nuclear markers for studies of other rosids and, perhaps, more distantly related angiosperms. Furthermore, we provide a strategy (Fig. 1) for exploring independent LCN markers in phylogenetic investigations across angiosperms.

In contrast to using genes identified by molecular biologists, our test results, when combined with those of Padolina (2006) for *Helianthus* and *Phalaenopsis*, provide some indication of phylogenetic usefulness across a taxonomically broad range of flowering plants. Thirty-two primer pairs successfully amplified regions in representative species in at least two of five distantly related angiosperm families (Table 5). In Brassicaceae, *A. thaliana* was used as a positive control; it was not investigated for phylogenetic utility. In the other four groups, some regions were found to be PI (shaded). Primer pair 51 (*atpB*) amplified in all five groups and is PI for two distantly related angiosperms at the species level—a monocot *Phalaenopsis* (Orchidaceae) and a eudicot *Psiguria* (Cucurbitaceae). Furthermore, 28 of these primer pairs amplified regions in at least two of the rosids, 11 of which (primer pairs 13, 15, 46, 51, 56, 61, 64, 69, 73, 85, and 86) amplified in all three (*Psiguria, Geranium*, and *Arabidopsis*).

Systematists interested in resolving angiosperm relationships at low taxonomic levels may begin searching for useful LCN markers with this reduced set of primer pairs following the strategy in Fig. 1. For groups where finding markers proves more difficult, there are several opportunities to modify the method and adjust reaction conditions to more exhaustively investigate a region. In contrast to the steps taken in our study, if success is not achieved the first time through the process, adjustments can be made rather than abandoning failures at each step (for suggestions see Don et al., 1991; Kellogg et al., 1994; Siebert et al., 1995; Padegimas and Reichert, 1998; Ochman et al., 1988; Triglia et al., 1988; Small et al., 2004). For example, for those primer combinations that consistently amplified strong, distinct, double bands, the region could be cloned and sequenced for only those fragments of the same length. Additionally, the early decision to identify a region as potentially, phylogenetically informative versus abandoning it may be subjective. The choice to pursue cloning and expand taxon sampling for a particular marker will depend on time, money, and previous success with other regions. As an example, the tree generated for primer pair 61 (actin), copy1 in *Psiguria* has only moderate resolution between some species (Supplementary Fig. 3), but the addition of both ingroup and outgroup taxa may help resolve relationships within the group. Exhaustive investigations of all PI regions will reveal the ultimate value of each marker.

Furthermore, if the reduced set of 32 primer combinations does not result in useful nuclear markers for a particular study group, the full set of 141 primer pairs may be explored. In any case, identifying phylogenetically informative LCN markers remains a time-consuming endeavor, but this study provides a convenient place to begin a search.

Acknowledgments

We thank Beryl Simpson, Debra Hansen, Tim Chumley, and two anonymous reviewers for valuable comments on the manuscript,

and Cai Zhengqiu for assisting with the preparation of sequences for submission to GenBank. This work was supported by NSF Grant DEB-0717372 and the Blake Centennial Professorship in Systematic Botany to R.K.J. An assistantship supporting P.R.S. was provided from the NSF GK-12 program through the Environmental Science Institute (DGE-0638740).

Appendix A

Initial results of screening 141 primer combinations in two taxa at different taxonomic levels in two distantly related families of rosids

Primer Pair[a]	Results of initial PCR attempt at 45 °C annealing temperature		Results of further amplification/ cloning/analyses[b]
	Psiguria bignoniacea	*Geranium macrorrhizum*	
1	No amp.	Single	Abandoned
2	No amp.	No amp.	Abandoned
3	No amp.	No amp.	Abandoned
4	Double	No amp.	Multiple bands persisted in *Psiguria*—abandoned
5	Double	Double	Double bands persisted in both species—abandoned
6	Single	No amp.	No resolution between species and multiple, indistinguishable copies for *Psiguria*—abandoned
7	No amp.	Double	Abandoned
8	Double	No amp.	Multiple bands persisted in *Psiguria*—abandoned
9	No amp.	No amp.	Abandoned
10	No amp.	Single	Abandoned
11	No amp.	No amp.	Abandoned
12	Triple	No amp.	Multiple bands persisted in *Psiguria*, and it did not amplify at higher annealing temperatures—abandoned
13	Multiple	Multiple	Multiple bands persisted in both species—abandoned
14	No amp.	No amp.	Abandoned
15	Single	Single	Phylogenetically informative for Geraniaceae but multiple, indistinguishable copies for *Psiguria*
16	No amp.	No amp.	Abandoned
17	Double	No amp.	*Psiguria* did not amplify at higher annealing temperatures—abandoned
18	No amp.	No amp.	Abandoned
19	No amp.	No amp.	Abandoned
20	No amp.	No amp.	Abandoned
21	Multiple	No amp.	*Psiguria* did not amplify at higher annealing temperatures—abandoned
22	No amp.	Single	Abandoned
23	No amp.	No amp.	Abandoned
24	No amp.	No amp.	Abandoned
25	No amp.	No amp.	Abandoned
26	No amp.	No amp.	Abandoned
27	No amp.	No amp.	Abandoned
28	No amp.	No amp.	Abandoned
29	Multiple	No amp.	*Psiguria* did not amplify at higher annealing temperatures—abandoned
30	Single	No amp.	Clones showed various size bands in *Psiguria*—abandoned
31	Multiple	Multiple	Clones showed various size bands in *Psiguria*, and *Geranium* did not amplify at higher annealing temperatures—abandoned
32	No amp.	No amp.	Abandoned
33	No amp.	No amp.	Abandoned
34	Multiple	No amp.	*Psiguria* did not amplify at higher annealing temperatures—abandoned
35	No amp.	No amp.	Abandoned
36	No amp.	No amp.	Abandoned
37	No amp.	No amp.	Abandoned
38	No amp.	No amp.	Abandoned
39	No amp.	No amp.	Abandoned
40	No amp.	No amp.	Abandoned
41	No amp.	No amp.	Abandoned
42	No amp.	No amp.	Abandoned
43	No amp.	No amp.	Abandoned
44	No amp.	Single	Abandoned
45	No amp.	No amp.	Abandoned
46	Multiple	Multiple	multiple bands persisted in both species—abandoned
47	No amp.	No amp.	Abandoned
48	No amp.	No amp.	Abandoned
49	No amp.	No amp.	Abandoned
50	Single	No amp.	Clones showed various size bands in *Psiguria*—abandoned
51	Single	Single	Phylogenetically informative for *Psiguria* but multiple, indistinguishable copies for Geraniaceae
53	No amp.	No amp.	Abandoned
54	No amp.	No amp.	Abandoned
55	No amp.	No amp.	Abandoned
56	Single	Single	No resolution between *Psiguria bignoniacea* and *Geranium macrorrhizum*—abandoned
57	Single	No amp.	No resolution between species and multiple, indistinguishable copies for *Psiguria* - abandoned
58	No amp.	Single	Abandoned
59	No amp.	No amp.	Abandoned
60	No amp.	No amp.	Abandoned
61	Single	Single	Phylogenetically informative for *Psiguria*, but clones showed various size bands in Geraniaceae
62	No amp.	No amp.	Abandoned
63	No amp.	No amp.	Abandoned
64	Double	Single	Multiple, indistinguishable copies for *Psiguria* (primers designed for group), multiple, indistinguishable copies for Geraniaceae—abandoned
65	No amp.	Double	*Geranium* did not amplify at higher annealing temperatures—abandoned
66	No amp.	No amp.	Abandoned
67	No amp.	No amp.	Abandoned
68	No amp.	Single	Abandoned
69	Double	Double	Double bands persisted in both species—abandoned
70	No amp.	No amp.	Abandoned
71	No amp.	No amp.	Abandoned
72	No amp.	No amp.	Abandoned
73	Double	Single	*Psiguria* did not amplify at higher annealing temperatures—abandoned
74	No amp.	No amp.	Abandoned
75	No amp.	No amp.	Abandoned
76	No amp.	No amp.	Abandoned
77	No amp.	No amp.	Abandoned
78	No amp.	No amp.	Abandoned
79	No amp.	No amp.	Abandoned
80	No amp.	No amp.	Abandoned

Appendix A (*continued*)

Primer Pair[a]	Results of initial PCR attempt at 45 °C annealing temperature		Results of further amplification/ cloning/analyses[b]
	Psiguria bignoniacea	*Geranium macrorrhizum*	
81	No amp.	Single	Abandoned
82	No amp.	Single	Abandoned
83	No amp.	No amp.	Abandoned
84	No amp.	No amp.	Abandoned
85	Single	Single	Did not amplify in other genera of Geraniaceae and multiple, indistinguishable copies for *Psiguria* (primers designed for group)
86	Single	Single	Multiple, indistinguishable copies in Geraniaceae and no resolution between species of *Psiguria*
87	No amp.	No amp.	Abandoned
88	No amp.	No amp.	Abandoned
89	No amp.	No amp.	Abandoned
90	Multiple	Multiple	Neither amplified at higher annealing temperatures—abandoned
91	No amp.	Single	Phylogenetically informative for Geraniaceae with multiple, orthologous copies
92	No amp.	No amp.	Abandoned
93	No amp.	No amp.	Abandoned
94	No amp.	No amp.	Abandoned
95	No amp.	No amp.	Abandoned
96	Single	No amp.	Phylogenetically informative for *Psiguria*
97	No amp.	No amp.	Abandoned
98	No amp.	No amp.	Abandoned
99	No amp.	No amp.	Abandoned
100	No amp.	No amp.	Abandoned
101	No amp.	No amp.	Abandoned
102	No amp.	No amp.	Abandoned
103	No amp.	No amp.	Abandoned
104	No amp.	No amp.	Abandoned
105	No amp.	No amp.	Abandoned
106	Single	No amp.	Very short—abandoned
107	No amp.	No amp.	Abandoned
108	No amp.	No amp.	Abandoned
109	No amp.	No amp.	Abandoned
110	No amp.	No amp.	Abandoned
111	No amp.	No amp.	Abandoned
112	No amp.	No amp.	Abandoned
113	Single	No amp.	Very short—abandoned
114	No amp.	No amp.	Abandoned
115	No amp.	No amp.	Abandoned
116	Single	No amp.	Very short—abandoned
117	No amp.	No amp.	Abandoned
118	No amp.	No amp.	Abandoned
119	No amp.	No amp.	Abandoned
120	No amp.	No amp.	Abandoned
121	No amp.	No amp.	Abandoned
122	No amp.	No amp.	Abandoned
123	No amp.	No amp.	Abandoned
124	No amp.	No amp.	Abandoned
125	Single	Single	Multiple, indistinguishable copies in Geraniaceae and no resolution between species of *Psiguria*
126	Double	Single	Clones showed various size bands in *Psiguria,* did not amplify in other genera of Geraniaceae—abandoned
128	No amp.	No amp.	Abandoned
129	Single	Single	Multiple, indistinguishable copies in *Psiguria,* did not amplify in other genera of Geraniaceae—abandoned
130	No amp.	No amp.	Abandoned
131	No amp.	No amp.	Abandoned
132	No amp.	No amp.	Abandoned
133	Multiple	Multiple	Clones showed various size bands in *Psiguria*, *Geranium* did not amplify at higher annealing temperatures—abandoned
134	No amp.	No amp.	Abandoned
135	Single	No amp.	Clones showed various size bands in *Psiguria*—abandoned
136	No amp.	No amp.	Abandoned
137	No amp.	No amp.	Abandoned
138	No amp.	No amp.	Abandoned
139	No amp.	No amp.	Abandoned
140	No amp.	No amp.	Abandoned
141	No amp.	No amp.	Abandoned
142	Double	No amp.	Very short—abandoned
144	No amp.	No amp.	Abandoned

A total of 848 sequences was uploaded to GenBank.

[a] Primers are described in Padolina (2006); sequences online at www.bio.utexas.edu/faculty/linderr/website.

[b] "abandoned" indicates that no more attempts were made to amplify the region for the rosid comparison, but it may have been pursued in the individual systematics studies of Geraniaceae or *Psiguria* reported in the text. If so, results of further testing are given.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.05.017.

References

Álvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29, 417–434.

Álvarez, I., Costa, A., and Feliner, G.N. 2008. Selecting single-copy nuclear genes for plant phylogenetics: a preliminary analysis for the Senecioneae (Asteraceae). J. Mol. Evol. doi:10.1007/s00239-008-9083-7.

Bacon, C.D., Feltus, F.A., Paterson, A.H., Bailey, C.D., 2008. Novel nuclear intron-spanning primers for Arecaceae evolutionary biology. Mol. Ecol. Res. 8, 211–214.

Bailey, C.D., Doyle, J.J., 1999. Potential phylogenetic utility of the low-copy nuclear gene *pistillata* in dicotyledonous plants: comparison to rDNA ITS and *trnL* intron in *Sphaerocardamum* and other Brassicaceae. Mol. Phylogenet. Evol. 13, 20–30.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., Donoghue, M.J., 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Missouri Bot. Gard. 82, 247–277.

Chapman, M.A., Chang, J-C., Weisman, D., Kesseli, R.V., Burke, J.M., 2007. Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). Theor. Appl. Genet. 115, 747–755.

Corriveau, J.L., Coleman, A.W., 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. Am. J. Bot. 75, 1443–1458.

Demesure, B., Sodzi, N., Petit, R.J., 1995. A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. Mol. Ecol. 4, 129–131.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., Mattick, J.S., 1991. "Touchdown" PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. 27, 3219–3228.

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

Doyle, J.J., 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. Syst. Bot. 17, 144–163.

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

Galloway, G.L., Malmberg, R.L., Price, R.A., 1998. Phylogenetic utility of the nuclear gene arginine decarboxylase: an example from Brassicaceae. Mol. Biol. Evol. 15, 1312–1320.

Hughes, C.E., Eastwood, R.J., Bailey, C.D., 2006. From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. Philos. Trans. R. Soc. B 361, 211–225.

Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., dePamphilis, C.W., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., Lee, S.-B., Peery, R., McNeal, J., Kuehl, J.V., Boore, J.L., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl. Acad. Sci. USA 104, 19369–19374.

Jeffrey, C., 2005. A new system of Cucurbitaceae. Bot. Zhurn. 90, 332–335.

Kellogg, D.E., Rybalkin, I., Chen, S., Mukhamedova, N., Vlasik, T., Siebert, P.D., Chenchik, A., 1994. *Taq* Start antibody: "hot start" PCR facilitated by a neutralizing monoclonal antibody directed against *Taq* DNA polymerase. BioTechniques 16, 1134–1137.

Koonin, E.V., 2005. Orthologs, paralogs, and evolutionary genomics. Annu. Rev. Genet. 39, 309–338.

Mason-Gamer, R.J., Well, C.F., Kellogg, E.A., 1998. Granule-bound starch synthase: structure, function, and phylogenetic utility. Mol. Biol. Evol. 15, 1658–1673.

Mort, M.E., Crawford, D.J., 2004. The continuing search: low-copy nuclear sequences for lower-level plant molecular phylogenetic studies. Taxon 53, 257–261.

Mort, M.E., Archibald, J.K., Randle, C.P., Levsen, N.D., O'Leary, T.R., Topalov, K., Wiegand, C.M., Crawford, D.J., 2007. Inferring phylogeny at low taxonomic levels: utility of rapidly evolving cpDNA and nuclear ITS loci. Am. J. Bot. 94, 173–183.

Ochman, H., Gerber, A.S., Hartl, D.L., 1988. Genetic applications of an inverse polymerase chain reaction. Genetics 120, 621–623.

Padegimas, L.S., Reichert, N.A., 1998. Adapter ligation-based polymerase chain reaction-mediated walking. Anal. Biochem. 260, 149–153.

Padolina, J.M. 2006. Phylogenetic Reconstruction of *Phalaenopsis* Using Nuclear and Chloroplast DNA Sequence Data and Using *Phalaenopsis* as a Natural System for Assessing Methods to Reconstruct Hybrid Evolution in Phylogenetic Analyses. Ph.D. Dissertation, The University of Texas at Austin.

Parkinson, C.L., Mower, J.P., Qiu, Y-L., Shirk, A.J., Song, K., Young, N.D., dePamphilis, C.W., Palmer, J.D., 2005. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol. Biol. 5, 73.

Pluthero, F.G., 1993. Rapid purification of high-activity *Taq* DNA polymerase. Nucleic Acids Res. 21, 4850–4851.

Sambrook, J.E., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, a Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.

Sanderson, M.T., Doyle, J.J., 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy and confidence. Syst. Biol. 41, 4–17.

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

Sang, T., 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit. Rev. Biochem. Mol. Biol. 37, 121–147.

Savolainen, V., Chase, M., Morton, C.M., Soltis, D.E., Bayer, C., Fay, M.F., de Bruijn, A., Sullivan, S., Qiu, Y-L., 2000a. Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences. Syst. Biol. 49, 306–362.

Savolainen, V., Fay, M.F., Albach, D.C., Backlund, A., van der Bank, M., Cameron, K.M., Johnson, S.A., Lledo, M.D., Pintaud, J-C., Powell, M., Sheahan, M.C., Soltis, D.E., Soltis, P.S., Weston, P., Whitten, W.M., Wurdack, J., Chase, M.W., 2000b. Phylogeny of the eudicots: a nearly complete familial analysis based on *rbcL* gene sequences. Kew Bull. 55, 257–309.

Schlüter, P.M., Stuessy, T.F., Paulus, H.F., 2005. Making the first step: practical considerations for the isolation of low-copy nuclear sequence markers. Taxon 54, 766–770.

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E., Small, R.L., 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. 92, 142–166.

Shaw, J., Lickey, E.B., Schilling, E.E., Small, R.L., 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare. III. Am. J. Bot. 94, 275–288.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A., Lukyanov, S.A., 1995. An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res. 23, 1087–1088.

Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., 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.

Small, R.L., Wendel, J.F., 2000. Copy number lability and evolutionary dynamics of the *Adh* gene family in diploid and tetraploid cotton (*Gossypium*). Genetics 155, 1913–1926.

Small, R.L., Cronn, R.C., Wendel, J.F., 2004. Use of nuclear genes for phylogeny reconstruction in plants. Austral. Syst. Bot. 17, 145–170.

Soltis, D.E., Soltis, P.S., Morgan, D.R., Swensen, S.M., Mullin, B.C., Dowd, J.M., Martin, P.G., 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc. Natl. Acad. Sci. USA 92, 2647–2651.

Soltis, D.E., Soltis, P.S., Chase, M.W., Mort, M., Albach, D., Zanis, M., Savolainen, V., Hahn, W., Hoot, S., Fay, M., Axtell, M., Swensen, S., Nixon, K., Farris, J., 2000. Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcL* and *atpB* sequences. Bot. J. Linn. Soc. 133, 381–461.

Soltis, D.E., Senters, A.E., Zanis, M., Kim, S., Thompson, J.D., Soltis, P.S., Ronse De Craene, L.P., Endress, P.K., Farris, J.S., 2003. Gunnerales are sister to other core eudicots: implications for the evolution of pentamery. Am. J. Bot. 90, 461–470.

Soltis, D.E., Soltis, P.S., Endress, P.K., Chase, M.W., 2005. Phylogeny and Evolution of Angiosperms. Sinauer Associates, Inc., Sunderland, Massachusetts, USA. p.23.

Sterck, L., Rombauts, S., Vandepoele, K., Rouzé, P., Van de Peer, Y., 2007. How many genes are there in plants (...and why are they there)? Curr. Opin. Plant Biol. 10, 199–203.

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

Swofford, D.L., 2002. PAUP*:Phylogenetic Analysis Using Parsimony (* and Other Methods), version 4.0b10. Sinauer Associates, Sunderland, MA, USA.

Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of 3 noncoding regions of chloroplast DNA. Plant Mol. Biol. 17, 1105–1109.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Triglia, T., Peterson, M.G.P., Kemp, D.J., 1988. A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. Nucleic Acids Res. 16, 8186.

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

Wendel, J.F., Doyle, J.J., 1998. Phylogenetic incongruence: window into genomes history and molecular evolution. In: Soltis, D., Soltis, P., Doyle, J. (Eds.), Molecular Systematics of Plants. II. DNA Sequencing. Kluwer Academy Publications, Boston, MA, USA.

Whittall, J.B., Medina-Marino, A., Zimmer, E.A., Hodges, S.A., 2006. Generating single-copy nuclear gene data for a recent adaptive radiation. Mol. Phylogenet. Evol. 39, 124–134.

Xu, W., Briggs, W.J., Padolina, J., Timme, R.E., Liu, W., Linder, C.R., Miranker, D.P., 2004. Using MoBIoS' scalable genome join to find conserved primer pair candidates between two genomes. Bioinformatics 20 (Suppl. 1), i355–i362.

Zwickl, D.J. 2006. Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Data Sets Under The Maximum Likelihood Criterion. Ph.D. Dissertation, The University of Texas at Austin. Available at: <www.bio.utexas.edu/faculty/antisense/garli/Garli.html>.