American Journal of Botany 92(1): 142–166. 2005.

THE TORTOISE AND THE HARE II: RELATIVE UTILITY OF 21 NONCODING CHLOROPLAST DNA SEQUENCES FOR PHYLOGENETIC ANALYSIS[1]

JOEY SHAW,[2] EDGAR B. LICKEY, JOHN T. BECK, SUSAN B. FARMER, WUSHENG LIU, JERMEY MILLER, KUNSIRI C. SIRIPUN, CHARLES T. WINDER, EDWARD E. SCHILLING, AND RANDALL L. SMALL

Department of Botany, 437 Hesler Biology, University of Tennessee, Knoxville, Tennessee 37996 USA

Chloroplast DNA sequences are a primary source of data for plant molecular systematic studies. A few key papers have provided the molecular systematics community with universal primer pairs for noncoding regions that have dominated the field, namely *trnL-trnF* and *trnK/matK*. These two regions have provided adequate information to resolve species relationships in some taxa, but often provide little resolution at low taxonomic levels. To obtain better phylogenetic resolution, sequence data from these regions are often coupled with other sequence data. Choosing an appropriate cpDNA region for phylogenetic investigation is difficult because of the scarcity of information about the tempo of evolutionary rates among different noncoding cpDNA regions. The focus of this investigation was to determine whether there is any predictable rate heterogeneity among 21 noncoding cpDNA regions identified as phylogenetically useful at low levels. To test for rate heterogeneity among the different cpDNA regions, we used three species from each of 10 groups representing eight major phylogenetic lineages of phanerogams. The results of this study clearly show that a survey using as few as three representative taxa can be predictive of the amount of phylogenetic information offered by a cpDNA region and that rate heterogeneity exists among noncoding cpDNA regions.

Key words: angiosperms; cpDNA; intergenic spacers; introns; molecular systematics; noncoding chloroplast DNA; phylogeny; seed plants.

Chloroplast DNA sequences are the primary source of data for inferring plant phylogenies, rivaled only perhaps by nuclear ribosomal ITS sequences in recent years (Baldwin, 1992; Baldwin et al., 1995 Álvarez and Wendel, 2003). Early in the plant molecular systematics era chloroplast DNA (cpDNA) was surveyed through restriction site polymorphism studies (see Olmstead and Palmer [1994] for a review of cpDNA studies through the early 1990s). As DNA sequencing technology became available, comparative studies of cpDNA gene sequences began to accumulate sparked by the observations of Ritland and Clegg (1987) and Zurawski and Clegg (1987). A landmark publication, the angiosperm *rbcL* study of Chase et al. (1993), set the stage for the increased use of cpDNA sequences for phylogenetic studies. Most early publications employed sequences of *rbcL* and were focused on suprageneric taxonomic questions (e.g., Chase et al., 1993). Subsequent workers began to explore additional gene sequences such as *ndhF* (Olmstead and Sweere, 1994; Olmstead and Reeves, 1995; Clark et al., 1995; Kim and Jansen, 1995), *atpB* (Hoot et al., 1995; Jensen et al., 1995; Wolf, 1997), and *matK* (Johnson and Soltis, 1994; Steele and Vilgalys, 1994). Simultaneously, noncoding regions of the chloroplast were being explored for lower level taxonomic studies under the assumption that noncoding regions should be under less functional constraint than coding regions and should provide greater levels of variation for phylogenetic analyses (Gielly and Taberlet, 1994). Among the first regions to be exploited were the *trnT-trnL-trnL-trnF* region (Taberlet et al., 1991), the *atpB-rbcL* intergenic spacer (Golenberg et al., 1993; Ehrendorfer et al., 1994; Hodges and Arnold, 1994; Manen et al., 1994), and the noncoding intron portions of the *trnK/matK* region (Johnson and Soltis, 1994; Steele and Vilgalys, 1994). Following these pioneering studies, the use of noncoding cpDNA regions has continually increased and is now routinely employed for studies of phylogeny at intergeneric and interspecific levels. Even though many noncoding regions have been explored by different workers (e.g., Taberlet et al., 1991; Johnson and Soltis, 1994; Demesure et al., 1995; Dumolin-Lapegue et al., 1997; Sang et al., 1997; Small et al., 1998) many investigators continue to use a limited number of regions. A survey of papers published from 1995 through 2002 in *American Journal of Botany*, *Systematic Botany*, *Molecular Phylogenetics and Evolution*, and *Plant Systematics and Evolution* illustrates that the number of investigations employing noncoding cpDNA is rapidly increasing (Fig. 1). However, of 445 studies, 342 (77%) used some portion of either *trnK-matK-trnK*, the *trnL* intron, and/or the *trnL-trnF* spacer. Two other relatively popular regions are the *rpS16* and *rpL16* introns. Studies that employed *rpS16*, *rpL16*, *trnK-matK-trnK*, or *trnL-trnL-trnF* (either alone or in combination with other regions) account for approximately 84% of all noncoding cpDNA-based phylogenetic investigations since 1995 and approximately 83% of the studies in 2002. This illustrates that, although the number of phylogenetic investigations using noncoding cpDNA is increasing every year, so too is the continued reliance on a few regions. Figure 1 also shows the slow increase in the use of other noncoding cpDNA regions, such as the *trnH-psbA* and *trnS-trnG* intergenic spacers, which have nearly always been added to supplement data collected from *trnL-trnL-trnF* or *trnK-matK-trnK*. It is important to note this apparent reliance on a

[1] Manuscript received 13 January 2004; revision accepted 2 September 2004.

The authors acknowledge funding for this project from the National Science Foundation (DEB-0108231 to RLS), the Hesler Endowment Fund from the University of Tennessee Herbarium, and the Department of Botany, University of Tennessee. The authors thank Karen Hughes for her initial encouragement and Stephen Downie and an anonymous reviewer for their helpful comments.

[2] Author for correspondence: e-mail: joeyshaw@aol.com.

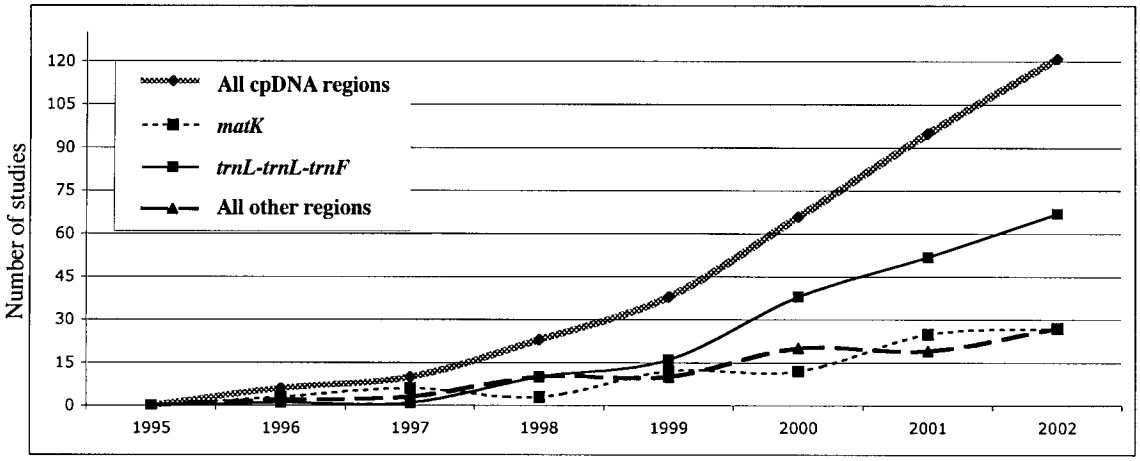

Fig. 1. Survey of 445 phylogenetic studies using cpDNA *matK* and noncoding regions published in *American Journal of Botany*, *Systematic Botany*, *Molecular Phylogenetics and Evolution*, and *Plant Systematics and Evolution* from 1995 through 2002. Lines and symbols represent the number of studies appearing each year (not cumulative). Solid gray line with diamonds = total number of papers published for that year, solid black line with square symbols = studies using some part or all of *trnL-trnL-trnF*, dashed thin line with square symbols = those using some part or all of *trnK-matK-trnK*, and dashed thick line with triangles = studies using regions other than *trnL-trnL-trnF* and *trnK-matK-trnK*.

few regions is in spite of the fact that in comparative studies, the phylogenetic utility of *trnL-trnL-trnF* and *trnK*/*matK* is often limited with respect to other regions (Sang et al., 1997; Small et al., 1998; see below for others).

As the majority of current phylogenetic investigations are focused at shallower phylogenetic levels, regions like the *trnL* intron, the *trnL-trnF* spacer, and the *trnK* intron/*matK* gene have provided satisfactory information in some groups (Bellstedt et al., 2001; Ge et al., 2002), but often yield poor resolution in other groups, at least in some clades (Bell and Patterson, 2000; Cuénoud et al., 2000; Hardig et al., 2000; Goldblatt et al., 2002; Klak et al., 2003; Muellner et al., 2003; Samuel et al., 2003). To obtain additional data and provide better phylogenetic resolution, sequences from these popular regions are often coupled with other sequence data, cpDNA or otherwise (Sang et al., 1997; Wang et al., 1999; Hardig et al., 2000; Kusumi et al., 2000; Azuma et al., 2001; Bortiri et al., 2001; Soltis et al., 2001; Bayer et al., 2002; Cronn et al., 2002; Hartmann et al., 2002; Mast and Givnish, 2002; Nyffeler, 2002; Schönenberger and Conti, 2003; Yamane et al., 2003), because additional data are often required to generate a phylogenetic hypothesis with acceptable resolution.

It has been clearly shown that the phylogenetic utility of different noncoding cpDNA regions within a given taxonomic group can vary tremendously (Sang et al., 1997; Small et al., 1998; Xu et al., 2000; Hartmann et al., 2002; Mast and Givnish, 2002; Cronn et al., 2002; Hamilton et al., 2003; Perret et al., 2003; Sakai et al., 2003), but choosing an appropriate cpDNA region for phylogenetic investigation is often difficult because of the paucity of information about the relative tempo of evolution among different noncoding cpDNA regions. Gielly and Taberlet (1994, p. 774) wrote: ''it is not easy, for many reasons, to establish a rule for the choice of a particular region of the chloroplast genome for resolving phylogenies.'' While many authors have compared relative rates of evolution among a few noncoding regions (Sang et al., 1997; Small et al., 1998; Wang et al., 1999; Kusumi et al., 2000; Xu et al., 2000; Soltis et al., 2001; Cronn et al., 2002; Mast and Givnish, 2002; Hamilton et al., 2003; Perret et al., 2003; Sakai et al., 2003; Yamane et al., 2003), these studies are all of a relatively narrow phylogenetic context and there is no consensus as to variability in evolutionary rates among noncoding cpDNA regions across a broad phylogenetic range. To our knowledge, the only work that has attempted to compare levels of variation among several different noncoding cpDNA regions across a wide range of lineages is Aoki et al. (2003). However, their results are equivocal because of insufficient data. Therefore, for most investigators, choosing the appropriate region for phylogenetic investigation at a particular taxonomic level is often guesswork.

We present a comparison of 21 noncoding cpDNA regions sampled across all of the major lineages of phanerogams sensu APG II (2003) (Fig. 2). Sequence divergence and, more importantly, the amount of information offered to phylogenetic investigations by the various noncoding cpDNA regions is compared across lineages to assess the phylogenetic utility of each. In this investigation, we determine whether there is any predictable rate heterogeneity among different noncoding chloroplast regions that have been employed in the field of molecular systematics. We will also provide a discussion of the often used noncoding cpDNA regions and present a general protocol for selecting potential noncoding cpDNA regions useful to systematic investigations.

MATERIALS AND METHODS

Taxonomic sampling—Species and lineages sampled in this study are listed in Table 1. Sampling focused on representing all major phanerogam lineages sensu APG II (2003) (Fig. 2, Table 1) in addition to representing different habits and life strategies (e.g., woody perennials, herbaceous perennials, and herbaceous annuals). Three fairly closely related species were chosen to represent each of 10 lineages. Earlier workers have shown that analysis of very closely related species, or even accessions of the same species, is likely to yield little or no information (e.g., Aoki et al., 2003) which would limit a comparison of different noncoding cpDNA regions. We therefore chose three species within each lineage that we knew from other studies, or our own unpublished data, were from separate but closely related clades. For each lineage two species were chosen to represent ingroup taxa of different clades, while the third was chosen as a closely related outgroup taxon (O.G.). Voucher information and GenBank accession numbers are listed in Table 1.

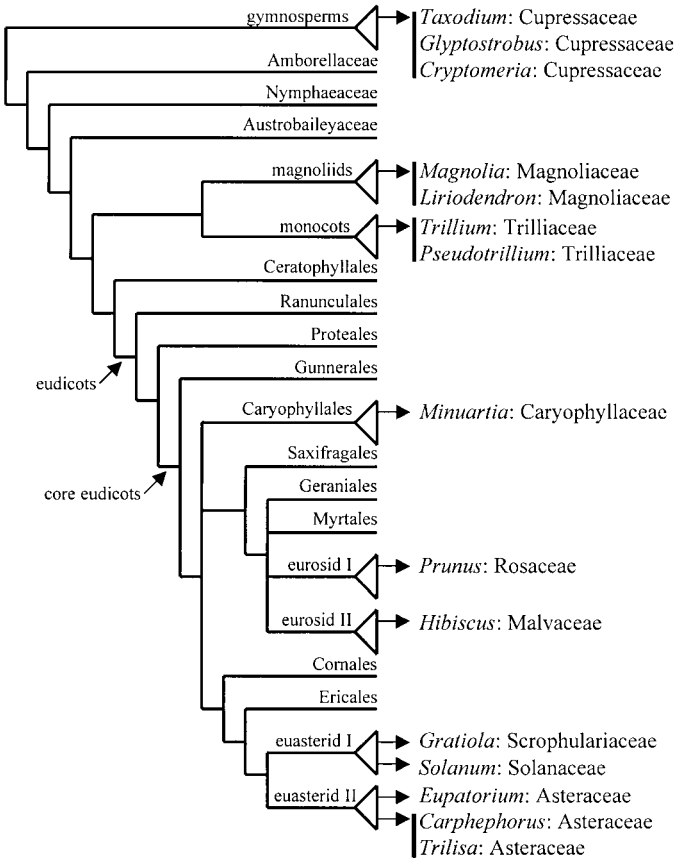

Fig. 2. Simplified phylogenetic representation, modified from APG II (2003), of the 10 lineages used in the survey of 21 noncoding cpDNA regions.

cpDNA sampling—After extensive literature review, 21 noncoding cpDNA regions were identified that had been previously used in interspecific or intraspecific phylogenetic investigations (Fig. 3). Some additional regions were added because they flank these previously utilized regions. All regions surveyed in this study occur in the Large Single Copy (LSC) region of the chloroplast genome. Listed as they occur on the Wakasugi et al. (1998) *Nicotiana* cpDNA map starting at the junction of Inverted Repeat A, they include: *trnH-psbA*; *psbA-3′trnK*; *3′trnK-matK*; *matK-5′trnK*; *rpS*16 intron; *trnS-trnG*; *trnG* intron; *rpoB-trnC*; *trnC-ycf6*; *ycf6-psbM*; *psbM-trnD*; *trnD-trnT*; *trnS-trnfM*; *trnS-rpS4*; *rpS4-trnT*; *trnT-trnL*; *trnL* intron; *trnL-trnF*; *5′rpS12-rpL20*; *psbB-psbH*; and *rpL*16 intron. Based on the Wakasugi et al. (1998) *Nicotiana* chloroplast map, these 21 regions comprise 14321 bp (35%) of the 40732 bp of the noncoding LSC.

Because the main focus of this investigation was to highlight cpDNA regions that may be the most beneficial to low-level systematic studies, coding cpDNA regions were excluded because they tend to provide fewer variable characters than their noncoding counterparts. Although the *rbcL* gene has arguably been the largest contributor to our phylogenetic understanding of chloroplast-containing life forms and has even been suggested to be nearly as informative as some noncoding regions of the cpDNA molecule (Chase et al., 2000), it was not included here. This is because *rbcL* is ''sometimes too conserved to clarify relationships between closely related genera'' (Gielly and Taberlet, 1994, p. 769) and other studies have shown it to provide fewer variable characters than several different noncoding regions (e.g., Renner, 1999; Richardson et al., 2000; Asmussen and Chase, 2001; Stefanovic et al., 2002; Salazar et al., 2003).

The *atpB-rbcL* spacer, perhaps one of the first intergenic spacers to be widely used, was excluded from our analysis because it is apparently of little infrageneric phylogenetic utility. It has consistently provided fewer variable characters compared to the entire *trnK* intron (Azuma et al., 2001), *trnH-psbA* (Azuma et al., 2001; Schönenberger and Conti, 2003; Hamilton et al., 2003), *5′rpS12-rpL20* (Hamilton et al., 2003), *rpL16* (Renner, 1999; Schönenberger and Conti, 2003), *rpS16* (Schönenberger and Conti, 2003), or *trnL-trnL-trnF* (Mayer et al., 2003).

Another well-characterized region found in the literature but excluded from this study is the *rpoC1* intron. The *rpoC1* intron was excluded here because it was shown to be less informative in cotton (*Gossypium*) than *atpB-rbcL*, *trnL-trnF*, *ndhA*, and *rpL16* (Small et al., 1998) and it yielded fewer characters than *rpL16*, *rpS16*, and *matK* in a study of the Apiaceae subfamily Apioideae (Downie et al., 2001). Although this region appears to show appropriate levels of variation for studies above the family level, it was noted as being ''largely inappropriate to infer phylogeny among closely related taxa'' (Downie et al., 1996, p. 14).

For the sake of clarity, we wish to point out that it is important to use specific terminology to describe a region of interest. For example, authors have used ''*trnL-trnF*'' to mean either the *trnL* intron plus *trnL-trnF* spacer or just the *trnL-trnF* spacer. To be precise we will use, for example, ''*trnL-trnF*'' to indicate the intergenic spacer alone, but ''*trnL-trnL-trnF*'' to indicate the intron plus the intergenic spacer. In addition, because there are multiple tRNA genes in the chloroplast genome that encode tRNAs for the same amino acid, it is desirable to denote the specific tRNA gene by the addition of the anti-codon as a superscript. For example, one of the regions we found to be highly variable is the $trnS^{GCU}$-$trnG^{UUC}$ intergenic spacer, which is different than the $trnS^{UGA}$-$trnG^{GCC}$ intergenic spacer that lies within the $trnS^{UGA}$-$trnfM^{CAU}$ region (Fig 3).

Molecular techniques—Because the genes surrounding noncoding regions are highly conserved across seed plants (and especially within angiosperms), many polymerase chain reaction (PCR) primers for amplification and sequencing could be used across the diverse taxonomic groups of this study. Nearly all of the primer regions used here were published in other studies. However, alignment of GenBank sequences from a wide array of phanerogam lineages was used to determine the universality of the previously published primers, modify problematic primers, and aid in the construction of new primers. In some cases, we designed new primers for regions not previously surveyed, or to help sequence through difficult regions (e.g., polynucleotide runs). Unless otherwise noted, all of the primers listed below and in Fig. 3 were successfully used for both amplification and sequencing reactions in all taxonomic groups.

DNA was extracted from leaf tissue using either the DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA) or the CTAB method (Doyle and Doyle, 1987). PCR was performed using either Eppendorf or MJ Research thermal cyclers in 20–50 μL volumes with the following reaction components: 1 μL template DNA (~10–100 ng), 1X buffer (PanVera/TaKaRa, Madison, Wisconsin, USA or Promega, Madison, Wisconsin, USA), 200 μmol/L each dNTP, 3.0 mmol/L $MgCl_2$, 0.1 μmol/L each primer, and 1.25 units *Taq* (PanVera/TaKaRa or Promega). Some reactions included bovine serum albumin with a final concentration of 0.2 μg/μL to improve amplification of difficult templates. In a few cases, 10 μmol/L tetramethyl ammonium chloride (TMACl) was included in the PCR solution because it is reported to reduce problems associated with long polynucleotide runs (Oxelman et al., 1997). However, we did not perform a comparative study to determine whether or not its presence actually improved our sequences. PCR amplification protocols and reaction conditions were continuously optimized throughout this investigation for all regions across all lineages. Material and methodological information and primer sequences specific to each of the different noncoding cpDNA regions are described below. All primer sequences are written in standard 5′ to 3′ orientation and their relative positions and orientations are illustrated in Fig. 3. A key to the shorthand for the following PCR parameters is as follows: initial denaturing step (temperature, time); number of repetitions of the amplification cycle [#× (denaturing temperature, time; primer annealing temperature, time; chain extension temperature, time)]; final extension step (temperature, time). All reactions ended with a final 4°C hold step.

PCR products were purified prior to sequencing with either the QIAquick PCR Purification Kit (Qiagen, Valencia, California, USA) or ExoSAP-IT (USB, Cleveland, Ohio, USA). All DNA sequencing was performed with the

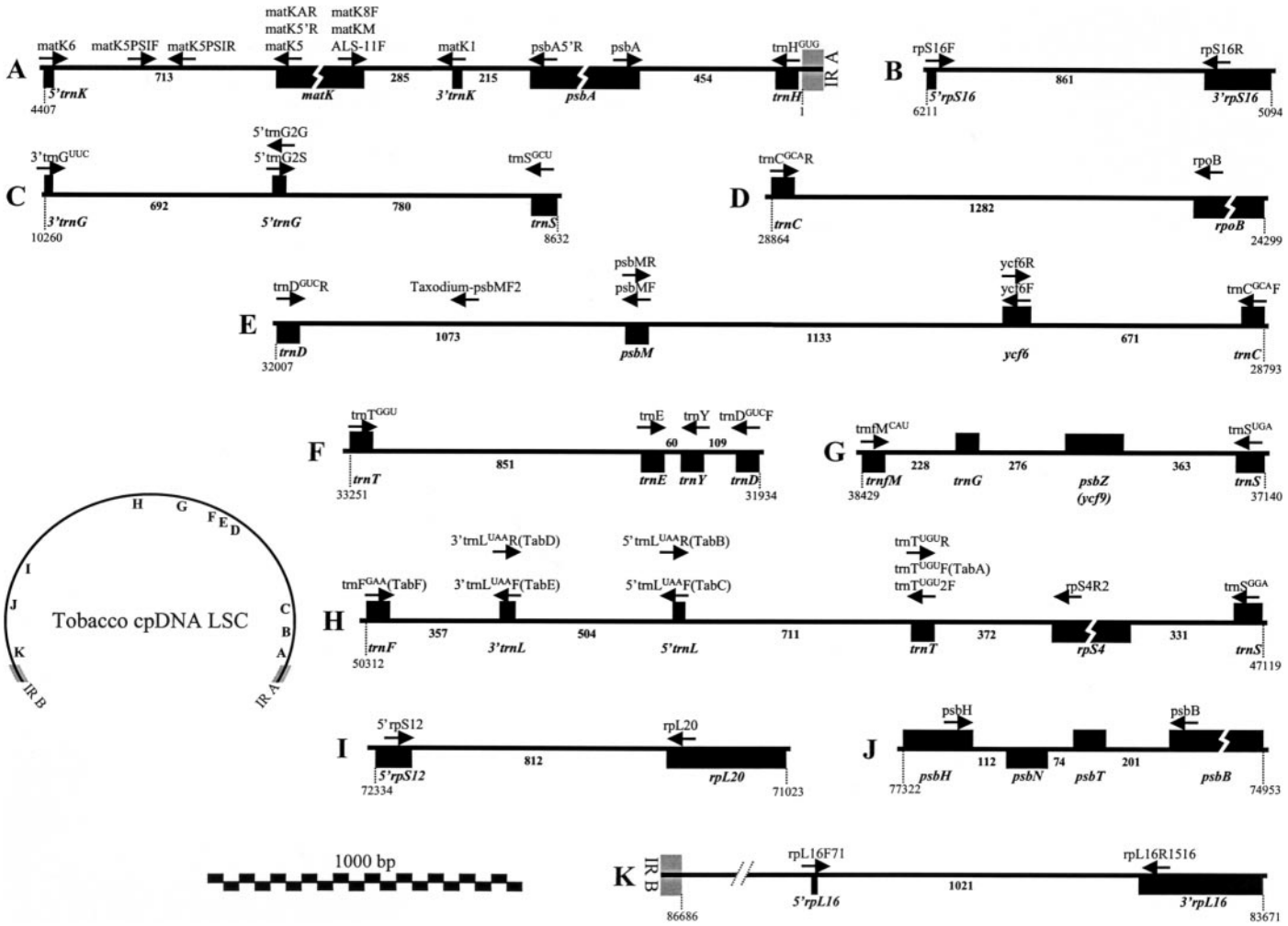

Fig. 3. Scaled map of the 21 noncoding cpDNA regions surveyed in this investigation (based on the *Nicotiana* chloroplast genome [Wakasugi et al., 1998]. The orientation and relative positions of the genes are identified (A–K) along the Large Single Copy (LSC) portion with specific positions denoted by offset numbers at the beginning and end of each region. Gene names are italicized below and amplification and sequencing primer names are in roman typeface above with directional arrows. Lengths of noncoding regions are centered below each intergenic spacer and intron.

ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit, v. 2.0 or 3.1 (Perkin-Elmer/Applied Biosystems, Foster City, California, USA), using the thermal cycle parameters 80°C, 5 min; 30× (96°C, 10 s; 50°C, 5 s; 60°C, 4 min). The products were electrophoresed and detected on an ABI Prism 3100 automated sequencer (University of Tennessee Molecular Biology Resource Facility). All sequences have been deposited in GenBank, and accession numbers are provided in Table 1.

trnH^{GUG}*-psbA*—The PCR parameters for this region were 80°C, 5 min; 35× (94°C, 30 s; 50–56°C, 30 s; 72°C, 1 min); 72°C, 10 min with primers **trnH**^{GUG} (CGC GCA TGG TGG ATT CAC AAT CC) (Tate and Simpson, 2003) and **psbA** (GTT ATG CAT GAA CGT AAT GCT C) (Sang et al., 1997). This region amplified and sequenced easily for all lineages. Because the average length of this region is relatively short (~500 bp), only the *trnH* primer was used in sequencing in most cases.

psbA-3′trnK^{UUU}*-[matK]-5′trnK*^{UUU}—These regions were the most problematic of any in this investigation. A variety of previously published and newly designed primers were required to amplify and sequence these regions, and very few completely universal primers were identified. We included only the noncoding portions of this region: *psbA*-3′*trnK* spacer, 3′*trnK-matK* intron, and *matK*-5′*trnK* intron. The *matK* gene was excluded primarily because it is a coding region, but also because of the inefficiency in designing the many primers that would be necessary to obtain this region for all lineages. In many cases, after amplifying the entire *trnK-matK-trnK* fragment, we were unable to sequence the PCR product with either the amplification or internal primers. However, if the region was PCR amplified in smaller sections using internal primers we were able to sequence these amplicons using the same primers that had previously failed. This phenomenon was observed independently in the laboratories of both E. E. Schilling and R. L. Small, as well as by J. Panero (University of Texas, personal communication) and R. Rapp (Iowa State University, personal communication) who suggested that dimethylsulfoxide might help during sequencing. Different primer combinations were often required for different taxa. The gymnosperm lineage is not represented in this data set because gymnosperm-specific primers were not obtained (Kusumi et al., 2000). The primers used in this study include: **psbA5′R** (AAC CAT CCA ATG TAA AGA CGG TTT), **ALS-11F** (ATC TTT CGC ATT ATT ATA G) (M. Nepokroeff, University of South Dakota, personal communication), **matKAR** (CTG TTG ATA CAT TCG A) (Kazempour Osaloo et al., 1999), **matKM** (TCG ACT TTC TGG GCT ATC) (Tate and Simpson, 2003), **matK1** (AAC TAG TCG GAT GGA GTA G) (Johnson and Soltis, 1994), **matK5** (TGT CAT AAC CTG CAT TTT CC) (Panero and Crozier, 2003), **matK5′R** (GCA TAA ATA TAY TCC YGA AAR ATA AGT GG), **matK6** (TGG GTT GCT AAC TCA ATG G) (Johnson and Soltis, 1994), **matK8F** (TCG ACT TTC TTG TGC TAG AAC TTT) (Steele and Vilgalys, 1994),

TABLE 1. Sources of cpDNA sequences employed in this study. GenBank accession numbers for each region are shown. O.G. = outgroup taxon.

Species	Source and voucher	*trnH-psbA*	*psbA-3′trnK*	*3′trnK-matK*	*matK-5′trnK*	*rpS16*	*trnS-trnG*	*trnG*	*rpoB-trnC*	*trnC-ycf6*	*ycf6-psbM*
Gymnosperm (Cupressaceae)											
Taxodium distichum var. *imbricarium* (Nutt.) Croom	S. Bacchus TA-CF-TSRE-S2; USA, FL; TENN	AY727188	AY727078	—	—	AY727428	AY727521	AY727521	—	AY727131	AY727131
Glyptostrobus pensilis (Staunton) K. Koch	USDA National Arboretum 70.0169; Hong Kong, China	AY727190	AY727076	—	—	AY727429	AY727520	AY727520	—	AY727130	AY727130
O.G. = *Cryptomeria japonica* (L.f.) Don.	E.B. Lickey T253; cultivated; TENN	AY727189	AY727077	—	—	—	AY727519	AY727519	—	AY727132	AY727132
Magnoliid (Magnoliaceae)											
Magnolia acuminata L.	J.T. Beck 6000; USA, TN; TENN	AY727183	AY727069	AY727069	AY727333	AY727447	AY727516	AY727516	AY727422	AY727149	AY727291
Magnolia tripetala L.	J.T. Beck 6001; USA, TN; TENN	AY727184	AY727068	AY727068	AY727338	AY727436	AY727517	AY727517	AY727423	AY727148	AY727290
O.G. = *Liriodendron tulipifera* L.	J.T. Beck 6002; USA, TN; TENN	AY727182	AY727070	AY727070	AY727340	AY727440	AY727518	AY727518	AY727424	AY727147	AY727289
Monocot (Trilliaceae)											
Trillium ovatum Pursh	S. Farmer s.n.; USA, OR; TENN	AY727187	AY727059	AY727059	AY727336	AY727437	—	AY727225	AY727421	AY727137	AY727288
Trillium texanum Buckl.	S. Farmer and Singhurst s.n.; USA, TX; TENN	AY727186	AY727060	AY727060	AY727335	AY727448	—	AY727224	AY727420	AY727136	AY727286
O.G. = *Pseudotrillium rivale* (S. Wats.) S.B. Farmer	Graham s.n.: cult. from USA, OR; TENN	AY727185	AY727061	AY727061	AY727334	AY727430	—	AY727223	AY727419	AY727138	AY727287
Caryophyllid (Caryophyllaceae)											
Minuartia cumberlandensis (B.E. Wofford & Kral) McNeill	C.T. Winder s.n.; USA, TN; TENN	AY727177	AY727065	AY727065	AY727332	AY727452	AY727514	AY727514	AY727418	AY727127	AY727127
Minuartia glabra (Michx.) Mattf.	C.T. Winder s.n.; USA, TN; TENN	AY727176	AY727066	AY727066	AY727337	AY727451	AY727515	AY727515	AY727417	AY727129	AY727129
O.G. = *Minuartia uniflora* (Walt.) Mattf.	C.T. Winder s.n.; USA, GA; TENN	AY727178	AY727067	AY727067	AY727339	AY727453	AY727513	AY727513	AY727416	AY727128	AY727128
Eurosid I (Rosaceae)											
Prunus hortulata Bailey	J. Shaw JSh821-017; USA, TN; TENN	AY500600	AY727082	AY727050	AY727329	AY500686	AY500705	AY500705	AY727413	AY727141	AY727294
Prunus nigra Ait.	J. Shaw JSh979-125; USA, VT; TENN	AY500605	AY727083	AY727051	AY727330	AY500691	AY500710	AY500710	AY727414	AY727142	AY727293
O.G. = *Prunus virginiana* L.	J. Shaw JSh871-040; USA, NH; TENN	AY500634	AY727084	AY727052	AY727331	AY727450	AY500739	AY500739	AY727415	AY727143	AY727292
Eurosid II (Malvaceae)											
Hibiscus cannabinus L.	R.L. Small s.n.; USA, FL (cultivar); TENN	AY727166	AY727062	AY727062	AY727327	AY727444	AY727504	AY727504	AY727410	AY727134	AY727283
Hibiscus mechowii Garcke	R.L. Small s.n.; Zambia; TENN	AY727167	AY727064	AY727064	AY727326	AY727435	AY727506	AY727506	AY727412	AY727133	AY727284
O.G. = *Hibiscus macrophyllus* Roxb.	L. Craven 10202; Indonesia: CANB	AY727165	AY727063	AY727063	AY727328	AY727442	AY727505	AY727505	AY727411	AY727135	AY727285
Euasterid I (Scrophulariaceae, Solanaceae)											
Gratiola brevifolia Raf.	D. Estes 02513; USA, TN; EKU	AY727170	AY727081	AY727054	AY727346	AY727441	AY727510	AY727510	AY727426	AY727145	AY727297
Gratiola virginiana L.	D. Estes 04608; USA, TN; TENN	AY727169	AY727080	AY727053	AY727345	AY727445	AY727512	AY727512	AY727427	AY727144	AY727298

TABLE 1. Extended.

Species	*psbM-trnD*	*trnD-trnT*	*trnS-trnfM*	*trnS-rpS4*	*rpS4-trnT*	*trnT-trnL*	*trnL*	*trnL-trnF*	*5′rpS12-rpL20*	*psbB-psbH*	*rpL16*
Gymnosperm (Cupressaceae)											
Taxodium distichum var. *imbricarium* (Nutt.) Croom	AY727095	—	AY727490	AY727243	AY727151	—	AY727215	AY727215	AY727312	AY727380	AY727401
Glyptostrobus pensilis (staunton) K. Koch	AY727094	—	AY727492	AY727242	AY727150	—	AY727217	AY727217	AY727323	AY727378	AY727402
O.G. = *Cryptomeria japonica* (L.f.) Don.	AY727096	—	AY727491	AY727244	AY727152	—	AY727216	AY727216	AY727325	AY727379	AY727403
Magnoliid (Magnoliaceae)											
Magnolia acuminata L.	AY727107	AY727455	AY727484	AY727249	AY727154	AY727277	AY727197	AY727230	AY727314	AY727371	AY727381
Magnolia tripetala L.	AY727110	AY727454	AY727486	AY727248	AY727155	AY727271	AY727195	AY727231	AY727308	AY727372	AY727382
O.G. = *Liriodendron tulipifera* L.	AY727106	AY727456	AY727485	AY727250	AY727153	AY727272	AY727196	AY727229	AY727306	AY727370	AY727383
Monocot (Trilliaceae)											
Trillium ovatum Pursh	AY727118	AY727466	AY727479	AY727253	AY727160	AY727275	AY727192	AY727233	—	AY727361	AY727405
Trillium texanum Buckl.	AY727120	AY727467	AY727480	AY727252	AY727161	AY727273	AY727191	AY727234	—	AY727362	AY727406
O.G. = *Pseudotrillium rivale* (S. Wats.) S.B. Farmer	AY727119	AY727468	AY727478	AY727251	AY727159	AY727274	AY727193	AY727232	—	AY727363	AY727404
Caryophyllid (Caryophyllaceae)											
Minuartia cumberlandensis (B.E. Wofford & Kral) McNeill	AY727091	AY727460	AY727487	AY727266	AY727097	AY727097	AY727204	AY727240	AY727321	AY727368	AY727399
Minuartia glabra (Michx.) Mattf.	AY727093	AY727462	AY727489	AY727267	AY727101	AY727101	AY727206	AY727239	AY727301	AY727369	AY727398
O.G. = *Minuartia uniflora* (Walt.) Mattf.	AY727092	AY727461	AY727488	AY727265	AY727100	AY727100	AY727205	AY727241	AY727313	AY727367	AY727400
Eurosid I (Rosaceae)											
Prunus hortulata Bailey	AY727109	AY727464	AY727481	AY727269	AY727103	AY727103	AY500748	AY500767	AY727317	AY727359	AY500643
Prunus nigra Ait.	AY727111	AY727463	AY727482	AY727270	AY727105	AY727105	AY500753	AY500772	AY727300	AY727360	AY500648
O.G. = *Prunus virginiana* L.	AY727108	AY727465	AY727483	AY727268	AY727102	AY727102	AY727194	AY727235	AY727303	AY727358	AY500677
Eurosid II (Malvaceae)											
Hibiscus cannabinus L.	AY727114	AY727470	AY727475	AY727245	AY727162	AY727278	AY727198	AY727227	AY727310	AY727355	AY727397
Hibiscus mechowii Garcke	AY727113	AY727471	AY727477	AY727246	AY727163	AY727276	AY727200	AY727228	AY727316	AY727357	AY727395
O.G. = *Hibiscus macrophyllus* Roxb.	AY727112	AY727469	AY727476	AY727247	AY727164	AY727279	AY727199	AY727226	AY727319	AY727356	AY727396
Euasterid I (Scrophulariaceae, Solanaceae)											
Gratiola brevifolia Raf.	AY727116	AY727474	AY727498	AY727255	AY727156	AY727280	AY727201	AY727237	AY727305	AY727366	AY727384
Gratiolia virginiana L.	AY727117	AY727473	AY727496	AY727254	AY727157	AY727281	AY727202	AY727236	AY727302	AY727364	AY727385

TABLE 1. Continued.

Species	Source and voucher	*trnH-psbA*	*psbA-3′trnK*	*3′trnK-matK*	*matK-5′trnK*	*rpS16*	*trnS-trnG*	*trnG*	*rpoB-trnC*	*trnC-ycf6*	*ycf6-psbM*
O.G. = *Gratiola neglecta* Torr.	D. Estes 04609; USA, TN; TENN	AY727168	AY727079	AY727055	AY727344	AY727433	AY727511	AY727511	AY727425	AY727146	AY727299
Solanum americanum Mill.	E.E. Schilling S-543; USA, FL; TENN	AY727179	AY727058	AY727058	AY727341	AY727438	AY727508	AY727508	AY727408	AY727122	AY727122
Solanum ptychanthum Dunal	E.E. Schilling S-522; USA, TN; TENN	AY727181	AY727056	AY727056	AY727342	AY727431	AY727507	AY727507	AY727407	AY727123	AY727123
O.G. = *Solanum physalifolium* Rusby	E.E. Schilling S-548; USA, TN; TENN	AY727180	AY727057	AY727057	AY727343	AY727449	AY727509	AY727509	AY727409	AY727121	AY727121
Euasterid II (Asteraceae)											
Eupatorium rotundifolium L.	C. Fleming TN002; USA, TN; TENN	AY727173	AY727073	AY727073	AY727349	AY727446	—	AY727222	—	AY727126	AY727126
Eupatorium hyssopifolium L.	K.C. Siripun 02-Eup-157; USA, NC; TENN	AY727172	AY727071	AY727071	AY727351	AY727434	—	AY727219	—	AY727125	AY727125
O.G. = *Eupatorium capillifolium* (Lamarck) Small	K.C. Siripun 02-Eup-155; USA, NC; TENN	AY727175	AY727072	AY727072	AY727350	AY727432	—	AY727218	—	AY727124	AY727124
Carphephorus corymbosus (Nutt.) Torr. & A. Gray	E.E. Schilling 2036; USA, GA; TENN	AY727174	AY727074	AY727074	AY727347	AY727443	—	AY727220	—	AY727140	AY727295
Trilisa paniculata (Willd.) Cass.	J.B. Nelson 21688; USA, SC; USCH	AY727171	AY727075	AY727075	AY727348	AY727439	—	AY727221	—	AY727139	AY727296
O.G. = *Eupatorium capillifolium* (Lamarck) Small	K.C. Siripun 02-Eup-155: USA, NC; TENN	AY727175	AY727072	AY727072	AY727350	AY727432	—	AY727218	—	AY727124	AY727124

TABLE 1. Continued Extended.

Species	*psbM-trnD*	*trnD-trnT*	*trnS-trnfM*	*trnS-rpS4*	*rpS4-trnT*	*trnT-trnL*	*trnL*	*trnL-trnF*	*5′rpS12-rpL20*	*psbB-psbH*	*rpL16*
O.G. = *Gratiola neglecta* Torr.	AY727115	AY727472	AY727497	AY727256	AY727158	AY727282	AY727203	AY727238	AY727320	AY727365	AY727386
Solanum americanum Mill.	AY727086	AY727459	AY727494	AY727262	AY727097	AY727097	AY727208	AY727208	AY727322	AY727352	AY727387
Solanum ptychanthum Dunal	AY727087	AY727457	AY727495	AY727263	AY727099	AY727099	AY727209	AY727209	AY727309	AY727353	AY727389
O.G. = *Solanum physalifolium* Rusby	AY727085	AY727458	AY727493	AY727264	AY727104	AY727104	AY727207	AY727207	AY727324	AY727354	AY727388
Euasterid II (Asteraceae)											
Eupatorium rotundifolium L.	AY727090	—	AY727503	AY727260	AY736007	AY736007	AY727210	AY727210	AY727311	AY727375	AY727393
Eupatorium hyssopifolium L.	AY727089	—	AY727501	AY727261	AY736008	AY736008	AY727211	AY727211	AY727315	AY727374	AY727392
O.G. = *Eupatorium capillifolium* (Lamarck) Small	AY727088	—	AY727502	AY727259	AY736009	AY736009	AY727212	AY727212	AY727304	AY727373	AY727394
Carphephorus corymbosus (Nutt.) Torr. & A. Gray	—	—	AY727500	AY727257	AY736010	AY736010	AY727214	AY727214	AY727307	AY727376	AY727390
Trilisa paniculata (Willd.) Cass.	—	—	AY727499	AY727258	AY736011	AY736011	AY727213	AY727213	AY727318	AY727377	AY727391
O.G. = *Eupatorium capillifolium* (Lamarck) Small	AY727088	—	AY727502	AY727259	AY736009	AY736009	AY727212	AY727212	AY727304	AY727373	AY727394

matK5PSIF (CTA TGG CTC CAA TTC TGG T), **matK5PSIR** (CCG CAT CAG GCA CTA ATC TA).

Hibiscus and *Minuartia* protocol: Amplification of the *matK*-5′*trnK* spacer used the matK6 and matK5′R primers with the PCR parameters 80°C, 5 min; 35× (95°C, 1 min; 50°C, 1 min with a ramp of 0.3°C/s; 65°C, 5 min); 65°C, 5 min. This spacer was sequenced with the matK6 primer. The *psbA-trnK-matK* spacers were amplified using the matKM (*Hibiscus*) or ALS-11F (*Minuartia*) and psbA5′R primers using the parameters 80°C, 5 min; 30× (94°C, 30 s; 50°C, 30 s; 72°C, 2 min); 72°C, 5 min. This region was sequenced using the psbA5′R primer.

Magnolia, *Prunus*, and *Gratiola* protocol: Amplification of the *matK*-5′*trnK* spacer used the matK6 and matK5 primers with the parameters 80°C, 5 min; 30–35× (94°C, 1 min; 50°C, 1 min; 72°C, 1.5 min); 72°C, 5 min. Amplification of the *psbA*-3′*trnK-matK* spacers was done using the matK8F and psbA5′R primers with the same PCR protocol.

Trillium-Pseudotrillium protocol: Amplification of the *matK*-5′*trnK* spacer used the matK6 and matKAR primers with the parameters 80°C, 5 min; 30–35× (94°C, 1 min; 50°C, 1 min; 72°C, 2 min); 72°C, 5 min. Amplification of the *psbA*-3′*trnK-matK* spacers used the matK8F and psbA5′R primers with the same PCR parameters. Because of two poly-A/T runs, matK5PSIF and matK5PSIR were used for internal sequencing.

Solanum, *Carphephorus-Trilisa*, *Eupatorium* protocol: Amplification of the *matK*-5′*trnK* spacer used the matK6 and matK5 primers with the parameters 80°C, 5 min; 35× (95°C, 1 min; 50°C, 1 min; 65°C, 5 min); 65°C, 5 min. Both primers were also used for sequencing reactions. The *psbA*-3′*trnK-matK* spacers were amplified with the psbA5′R and ALS-11F for *Solanum americanum* and *S. physalifolium*, matKM for *S. ptychanthum*, and matK8F for *Eupatorium* and *Carphephorus-Trilisa* with the above parameters. All were sequenced using only the psbA5′R primer.

rpS16—This region was amplified using the parameters 80°C, 5 min; 35× (94°C, 30 s; 50–55°C, 30 s; 72°C, 1 min); 72°C, 5 min, with primers **rpS16F** (AAA CGA TGT GGT ARA AAG CAA C) and **rpS16R** (AAC ATC WAT TGC AAS GAT TCG ATA), which are modified from Oxelman et al. (1997). Both primers were also used in sequencing reactions. This region amplified and sequenced easily for all angiosperm taxa and two of the three gymnosperm representatives with minimal troubleshooting. Despite trying several different PCR programs, annealing temperatures, and $MgCl_2$ concentrations, we were unable to amplify this region for *Cryptomeria japonica*.

trnS^{GCU}-*trnG*^{UUC}-*trnG*^{UUC}—For this region, three different protocols were used and in most cases the *trnS-trnG* spacer and the *trnG* intron were amplified as one fragment. For most taxa protocol 1was successful. Both protocols 1 and 2 used the primers **trnS**^{GCU} (AGA TAG GGA TTC GAA CCC TCG GT) and **3′trnG**^{UUC} (GTA GCG GGA ATC GAA CCC GCA TC). Additional primers **5′trnG2G** (GCG GGT ATA GTT TAG TGG TAA AA) (toward *trnG*) and **5′trnG2S** (TTT TAC CAC TAA ACT ATA CCC GC) (toward *trnS*) were sometimes used to amplify only the *trnG* intron, and for sequencing longer fragments and templates with a difficult poly-A repeat.

Protocol 1: This is a two-step PCR protocol with primer annealing and chain extension occurring at the same temperature, using the parameters 80°C, 5 min; 30× (95°C, 1 min; 66°C, 4 min); 66°C, 10 min. A final $MgCl_2$ concentration of 1.5 mmol/L (rather than 3.0 mmol/L) was used.

Protocol 2: This protocol was used when amplification with protocol 1 was problematic. The parameters are 80°C, 5 min; 35× (95°C, 1 min; 50°C, 1 min with a ramp of 0.3°C/s; 65°C, 5 min); 65°C, 10 min. This protocol always coamplifies the *trnS*^{UGA} and *trnG*^{GGC} part of the *trnS*^{UGA}-*trnfM*^{CAU} spacer. The result of this protocol yields two equal-intensity, but well-separated bands in a test gel, the larger of which was always the target *trnS*^{GCU}-*trnG*^{UUC}. The desired fragment was excised from the gel and cleaned with a QIAquick Gel Extraction Kit. Because of the sequence similarity of these two different *trnS* and *trnG* genes, primer design was difficult and the protocols needed to be very specific to amplify only the correct region.

Protocol 3: Independent inversions in monocots (Hiratsuka et al., 1989) and Asteraceae (Jansen and Palmer, 1987) interrupt the *trnS*^{UGA}-*trnG*^{GGC} spacer preventing amplification. However, using the 3′trnG and 5′trnG2G primers, we successfully amplified and sequenced the *trnG* intron for *Trillium-Pseudotrillium*, *Carphephorus-Trilisa*, and *Eupatorium*. The amplification parameters for the *trnG* intron are 80°C, 5 min; 35× (95°C, 1 min; 50°C, 1 min with a ramp of 0.3°C/s; 65°C, 5 min); 65°C, 5 min.

rpoB-trnC^{GCA}—This region amplified easily for most angiosperm taxa using primers **trnC**^{GCA}**R** (CAC CCR GAT TYG AAC TGG GG) and **rpoB** (CKA CAA AAY CCY TCR AAT TG), modified from Ohsako and Ohnishi (2000). The PCR parameters for this region are 80°C, 5 min; 30–35× (96°C, 1 min; 50–57°C, 2 min; 72°C, 3 min); 72°C, 5 min. For unknown reasons, we were unable to amplify this region for *Taxodium*, *Glyptostrobus*, or *Cryptomeria*.

trnC^{GCA}-*ycf6-psbM-trnD*^{GUC}—Two different, but equally successful, protocols were used to amplify this region. For *Gratiola*, *Hibiscus*, *Magnolia*, *Minuartia*, *Prunus*, and *Taxodium*, we amplified the entire approximately 3-kb *trnC* to *trnD* fragment. For *Carphephorus-Trilisa*, *Eupatorium*, *Solanum*, and *Trillium-Pseudotrillium*, we amplified the fragments *trnC-psbM* and *ycf6-trnD*. Both protocols used the same PCR parameters, which were 80°C, 5 min; 35× (94°C, 1 min; 50–55°C, 1 min; 72°C, 3.5 min); 72°C, 5 min. PCR and sequencing primers included **trnC**^{GCA}**F** (CCA GTT CRA ATC YGG GTG) (modified from Demesure et al., 1995), **ycf6R** (GCC CAA GCR AGA CTT ACT ATA TCC AT), **ycf6F** (ATG GAT ATA GTA AGT CTY GCT TGG GC), **psbMR** (ATG GAA GTA AAT ATT CTY GCA TTT ATT GCT), **psbMF** (AGC AAT AAA TGC RAG AAT ATT TAC TTC CAT), **Taxodium-psbMF2** (CTT TTG TTC GGG TGA GAA AGG), and **trnD**^{GUC}**R** (GGG ATT GTA GYT CAA TTG GT) (modified from Demesure et al., 1995). This region required only moderate troubleshooting. After trying several different PCR modifications, we were unable to obtain the *psbM-trnD* segment for *Carphephorus-Trilisa*. In nearly all surveyed lineages, a poly-A/T run exists between *psbM* and *trnD*, but created sequencing difficulties in only a few cases.

trnD^{GUC}-*trnT*^{GGU}—This spacer amplified easily for most taxa using Demesure et al. (1995) primers **trnD**^{GUC}**F** (ACC AAT TGA ACT ACA ATC CC) and **trnT**^{GGU} (CTA CCA CTG AGT TAA AAG GG). The PCR parameters for this region are 80°C, 5 min; 30× (94°C, 45 s; 52-58°C, 30 s; 72°C, 1 min); 72°C, 5 min. Internal sequencing primers **trnE**^{UUC} (AGG ACA TCT CTC TTT CAA GGA G) and **trnY**^{GUA} (CCG AGC TGG ATT TGA ACC A) were created because of poly-A/T repeats that were difficult to sequence and the atypically large size of the region in a few taxa. A large inversion in the Asteraceae, excluding the Barnadesieae (Jansen and Palmer, 1987), interrupts the *trnD-trnT* spacer precluding its use. This region also appears to be absent in the *Pinus* chloroplast genome (Wakasugi et al., 1994), which may explain why we were unable to amplify this region for *Taxodium*, *Glyptostrobus*, or *Cryptomeria*.

trnS^{UGA}-*trnfM*^{CAU}—The amplification parameters for this region are 80°C, 5 min; 30× (94°C, 30 s; 55°C, 30 s; 72°C, 2 min); 72°C, 5 min, using Demesure et al. (1995) primers **trnS**^{UGA} (GAG AGA GAG GGA TTC GAA CC) and **trnfM**^{CAU} (CAT AAC CTT GAG GTC ACG GG). This region amplified and sequenced easily for most taxa with minimal troubleshooting.

As explained in the *trnS*^{GCU}-*trnG*^{UUC}-*trnG*^{UUC} region above, *trnG*^{GCC} occurs between *trnS*^{UGA}-*trnfM*^{CAU}. Because there is so little difference between the sequences of these *trnS* and *trnG* genes, the two independent *trnS-trnG* regions will coamplify under certain amplification parameters. However, a seemingly counterintuitive advantage to such sequence similarity is that primer **3′trnG**^{UUC} (and possibly primers **5′trnG2G** and **5′trnG2S**) can be used as an internal sequencing primer for the *trnS*^{UGA}-*trnfM*^{CAU} region.

trnS^{GGA}-*rpS4-trnT*^{UGU}-*trnL*^{UAA}-*trnL*^{UAA}-*trnF*^{GAA}—Because of an initial lack of communication, we PCR amplified several of the taxa using different primer combinations, all of which worked well. However, for all of the lineages of angiosperm taxa, this region was easily amplified in two fragments. The first, *trnS*-5′*trnL*, was amplified using primers **trnS**^{GGA} (TTA CCG AGG GTT CGA ATC CCT C) and **5′trnL**^{UAA}**R (TabB)** (TCT ACC GAT TTC GCC ATA TC) (Taberlet et al., 1991) with the parameters 96°C, 5 min; 35× (96°C, 1 min; 50–55°C, 2 min; 72°C, 2.5 min); 72°C, 5 min. The second fragment, *trnL*5′-

trnF, was amplified using primers **trnL5′^{UAA}F (TabC)** (CGA AAT CGG TAG ACG CTA CG) (Taberlet et al., 1991) and **trnF^{GAA} (TabF)** (ATT TGA ACT GGT GAC ACG AG) (Taberlet et al., 1991) with the parameters 80°C, 5 min; 35× (94°C, 1 min; 50°C, 1 min; 72°C, 2 min); 72°C, 5 min. Several internal sequencing primers were used and included **rpS4R2** (CTG TNA GWC CRT AAT GAA AAC G), **trnT^{UGU}R** (AGG TTA GAG CAT CGC ATT TG), **trnT^{UGU}F (TabA)** (CAT TAC AAA TGC GAT GCT CT) (Taberlet et al., 1991), **trnT^{UGU}2F** (CAA ATG CGA TGC TCT AAC CT) (trnA2 of Cronn et al., 2002), **3′trnL^{UAA}R (TabD)** (GGG GAT AGA GGG ACT TGA AC) (Taberlet et al., 1991), and **3′trnL^{UAA}R (TabE)** (GGT TCA AGT CCC TCT ATC CC) (Taberlet et al., 1991).

5′rpS12-rpL20—This region amplified and sequenced easily for almost all taxa using primers **5′rpS12** (ATT AGA AAN RCA AGA CAG CCA AT) and **rpL20** (CGY YAY CGA GCT ATA TAT CC), both modified from Hamilton (1999a). Amplification parameters were 96°C, 5 min; 35× (96°C, 1 min; 50–55°C, 1 min; 72°C, 1 min); 72°C, 5 min. Although amplification of this region was successful for *Trillium ovatum*, sequencing reactions using either primer failed repeatedly, even for several different accessions of this species.

psbB-psbH—This region amplified and sequenced easily for all taxa using primers **psbB** (TCC AAA AAN KKG GAG ATC CAA C) and **psbH** (TCA AYR GTY TGT GTA GCC AT), both modified from Hamilton (1999a). Amplification parameters were 80°C, 5 min; 35× (94°C, 30 s; 57–60°C, 30 s; 72°C, 1 min); 72°C, 5 min.

rpL16—This region amplified and sequenced easily for all taxa with minimal troubleshooting using primers **rpL16F71** (GCT ATG CTT AGT GTG TGA CTC GTT G) and **rpL16R1516** (CCC TTC ATT CTT CCT CTA TGT TG) (Small et al., 1998). Amplification parameters were 80°C, 5 min; 35× (95°C, 1 min; 50°C, 1 min with a ramp of 0.3°C/s; 65°C, 5 min); 65°C, 4 min.

cpDNA compilation and analysis—Sequencher 3.0 (Gene Codes Corp., 1998) was used to compile contiguous sequences (contigs) of each accession from electropherograms generated on the automated sequencer. Positions of coding and noncoding (gene, exon, and intron) borders were determined by comparison with either *Arabidopsis* (NC 000932), *Lotus* (NC 001874), or *Nicotiana* (NC 002694) entire cpDNA sequences in GenBank. Terminal coding regions and, in a few rare cases, unreadable ends of the PCR amplicons were excluded from the contigs. Small coding regions within some of the noncoding regions (e.g., *trnE*^{UUC} and *trnY*^{GUA} within the *trnD*^{GUC}-*trnT*^{GGU} spacer) were not excluded from the contigs. Sequences of each of the three-species groups were aligned using Clustal X (Thompson et al., 2001) and manually corrected using McClade v. 4.0 to produce an alignment with the fewest number of changes (indels or nucleotide substitutions). All polymorphic sites found in the three-species groups were rechecked against the original electropherograms. Alignments are available upon request from J. Shaw, E. B. Lickey, or R. L. Small.

The number of nucleotide substitutions, indels, and inversions (hereafter referred to collectively as Potentially Informative Characters or PICs) between the two ingroup species and between either ingroup species and the outgroup species were tallied for each noncoding cpDNA region in each of the lineages. Because indels have been shown to be prevalent and often phylogenetically informative (Golenberg et al., 1993; Morton and Clegg, 1993; Gielly and Taberlet, 1994), they were scored in this study, as were inversions. Indels, any nucleotide substitutions within the indels, and inversions were scored as independent, single characters. We then estimated the proportion of observed mutational events for each noncoding cpDNA region using a modified version of the formula used in O'Donnell (1992) and Gielly and Taberlet (1994). The proportion of mutational events (or % variability) = [(NS + ID + IV) / L] × 100, where NS = the number of nucleotide substitutions, ID = the number of indels, IV = the number of inversions, and L = the total sequence length.

Assessment of a correlation between variability and length—To assess whether or not the length of the different noncoding cpDNA regions accounts for the number of PICs observed within a particular region, we used a simple regression analysis. Because of the variation in phylogenetic distance between species in the different lineages we could not combine all lineages in a single regression. Instead, we performed 10 separate regressions (one per lineage) and calculated r^2 for each to determine how much of the variation seen in the PIC values is explained by the length of the region.

Cost/benefit analysis of coamplifiable noncoding cpDNA regions—In the above analyses, each noncoding region was treated individually. However, several adjacent, shorter, noncoding cpDNA regions may be coamplified as a single contiguous unit. We surveyed several cpDNA region combinations to assess the potential phylogenetic utility of coamplifiable regions from a cost/benefit perspective. For example, the *trnL* intron and *trnL-trnF* spacer are often coamplified, and most of the time these two regions are sequenced with the same two primers that were used in PCR (TabC and TabF). From a cost/benefit perspective, it is beneficial to amplify and sequence both of these regions together instead of separately by maximizing the number of characters obtained per two sequencing reactions. Our sequencing reactions always yielded easily readable sequence data of 800 bp from a single-primer sequencing reaction. We therefore limited what we categorize as ''coamplifiable'' regions to those whose total length average is < approximately 1500 bp and can be sequenced entirely with two sequencing reactions. These coamplifiable regions include *psbA-3′trnK-matK*, *trnS-trnG-trnG*, *trnC-ycf6-psbM*, *ycf6-psbM-trnD*, *rps4-trnT-trnL*, and *trnL-trnL-trnF*.

Assessment of the predictive value of a three-species sample study—Our inferences from these data rely on the assumption that a sample of three species is predictive of the overall levels of variation that will be found in an entire data set. To test the predictive power of a three-species survey we compared the number of PICs among the three species with the respective complete data sets of 18 taxa of *Prunus* sect. *Prunocerasus* (Shaw and Small, 2004) and nine taxa of *Hibiscus* sect. *Furcaria* (R. L. Small et al., unpublished data), each with a single outgroup. The comparison of the *Prunus* data sets was made with introns *trnL*, *trnG*, *rpS16*, and *rpL16* and intergenic spacers *trnL-trnF*, *trnH-psbA*, and *trnS-trnG*, and the comparison of the *Hibiscus* data sets was made with introns *rpS16*, *rpL16*, and *trnG* and intergenic spacers *trnD-trnT*, *rpoB-trnC*, *trnH-psbA*, and *trnS-trnG*. Regression lines were calculated and their slopes were compared on a scatterplot for each data set comparison.

RESULTS

Assessment of the noncoding cpDNA regions surveyed across phanerogam lineages—A few regions were excluded from analysis because they were missing in some taxa or their lineages due to inversions such as those observed in the *trnS-trnG* spacer in *Trillium-Pseudotrillium*, *Eupatorium*, and *Carphephorus-Trilisa*, the *rpoB-trnC* spacer in *Eupatorium* and *Carphephorus-Trilisa*, and the *trnD-trnT* spacer in *Taxodium*, *Eupatorium*, and *Carphephorus-Trilisa*. Some regions were excluded because of their inconsistency or inability to amplify such as the *rpoB-trnC* region in *Taxodium* and the *rpS16* intron in *Cryptomeria*. Others had to be excluded because, for unknown reasons, they could not be sequenced, such as the *psbM-trnD* region in *Carphephorus-Trilisa* and the *5′rpS12-rpL20* region in *Trillium ovatum*. After the exclusion of the problematic regions, taxa, and most coding regions, 133 504 bp from 21 noncoding cpDNA regions from 10 phanerogamic lineages were sequenced. Of that, we observed 2968 nucleotide substitutions, 1260 indels, and six inversions for a total of 4234 PICs. Nucleotide substitutions account for 70.1% of the variable characters, while indels and inversions account for 29.8% and 0.14%, respectively. No obvious differences were observed in the amount of variability or number of PICs between intergenic spacers and introns.

We did not apply statistical analyses to these data because of potentially different rates of evolution among the different lineages, the incongruent phylogenetic distances between the species in each lineage, and the exclusion of some regions because of structural rearrangement of the cpDNA molecule or PCR amplification or sequencing difficulties. Thus, the following discussion is based on our qualitative interpretation of the results, which are compiled in Table 2 and simplified in Fig. 4.

Figure 4 illustrates the disparity in the PICs offered by different noncoding cpDNA regions and that this trend exists across all phylogenetic lineages. Of the 21 separate noncoding cpDNA regions surveyed in this investigation, several consistently provided more PICs per our cost/benefit criterion of two sequencing reactions than other regions across all lineages. Specifically, five intergenic spacers (*trnD-trnT, trnS-trnG, rpoB-trnC, trnT-trnL*, and *trnS-trnfM*) provided more PICs than the other surveyed regions. The PIC value for each region was averaged across the six lineages for which there are complete data sets to illustrate more clearly the general trend (Fig. 5).

Assessment of a correlation between PICs and length—A scatterplot showing the relationship of the PIC values with respect to the length of the region is shown in Fig. 6. For each lineage, a regression line was drawn (not shown) and coefficients of determination were calculated. Coefficients of determination ranged from 22% in *Eupatorium* to 83% in *Taxodium*, which means that length of the region explains 83% of the variation in PIC value observed in *Taxodium*, while it only explains 22% of the variation in PIC value observed in *Eupatorium*. However, there is an apparent outlier within the *Taxodium* data set (circled in the upper right corner of Fig. 6), the removal of which drops the r^2 value from 0.83 to 0.54. Although Fig. 6 reveals the intuitively obvious conclusion that the length of a region accounts for a proportion of the PICs, this figure shows that length does not explain all, or even a majority in many cases, of the variability within a particular region. Many examples can be found where very different PIC values are found within regions that are nearly the same length.

Cost/benefit analysis of coamplifiable noncoding cpDNA regions—Several regions may be coamplified, sequenced, and successfully contiged with the same two PCR primers, and from a cost perspective, they are equal to amplifying and sequencing a portion of each alone. These combined regions include *psbA*-3′*trnK-matK, trnS-trnG-trnG, trnC-ycf6-psbM, ycf6-psbM-trnD, rps4-trnT-trnL*, and *trnL-trnL-trnF*. The results of the comparison between the coamplifiable regions are shown in Fig. 7. This figure shows that the *trnS-trnG* spacer combined with the *trnG* intron (*trnS-trnG-trnG*) potentially provides the greatest number of PICs compared to all other regions that could be amplified and sequenced with two primers.

Assessment of the predictive value of a three-species sample study—Analysis of the predictive value of a three-species survey of a particular cpDNA region indicates that as the number of PICs in a three-species survey increases, so too will the actual number of variable characters in a complete data set generated by that region (Fig. 8). A regression analysis of each of the *Prunus* and *Hibiscus* data sets reveals remarkably similar slopes. It is apparent from the linear relationship that a preliminary survey of three species is highly predictive of the amount of information that a noncoding cpDNA region might offer to a phylogenetic investigation.

DISCUSSION

Since 1995, the number of molecular systematic investigations that employ noncoding cpDNA sequence data has increased every year (Fig. 1). However, most of these studies (about 77% from 1995 to 2002) have used some portion of the *trnL-trnL-trnF* or *trnK-matK-trnK* regions and very few investigators have sampled from the myriad other noncoding regions of the cpDNA molecule. Because of this, little is known about the relative rates of evolution among the different noncoding cpDNA regions, and most investigators continue to rely on these two very popular regions.

The initial goal of this investigation was to provide a comparison of noncoding cpDNA regions to see if there are any that reliably yield a greater number of variable characters (PICs) at low taxonomic levels, and thus would be of greater value to systematic studies than the often used *trnL-trnL-trnF* or *trnK-matK-trnK* regions. To do so we used three-species surveys representing most of the major phylogenetic lineages of phanerogams (sensu APG II, 2003). To test the predictive power of a three-species survey we compared the surveys of seven regions in *Prunus* and eight regions in *Hibiscus* with their respective complete data sets (Fig. 8). Figure 8 shows that as the number of PICs in survey of three species increases, so will the actual number of variable characters in a complete data set generated from those regions. Therefore, a survey of three species is highly predictive of the amount of information that a noncoding cpDNA region might offer to a phylogenetic investigation and is an effective means of comparison between different noncoding cpDNA regions.

Most investigators, when comparing different DNA regions, have used either of two metrics that are not wholly separate. One tallies the number of variable characters including nucleotide substitutions, indels, and inversions (PICs), while the other calculates the percent variability, or percent divergence of a region, by dividing the total number of variable characters by the total length of the region. It is necessary to emphasize that, from the viewpoint of systematists, the total number of variable characters offered by a region is more important than the percent variability. A highly variable but extremely short region may not provide a sufficient number of variable characters with which to generate a resolved phylogeny. As systematists, we are interested in obtaining the greatest number of variable characters per sequencing reaction, arguably the costliest portion of sequence acquisition, where current techniques and equipment allow for 600–800 bp of easily readable nucleotides per reaction. Therefore, it would be ideal to use cpDNA regions that combine high variability in fragments of approximately 700–1500 bp that can easily be sequenced with one or two primers, ideally the original amplification primers. To show that the number of PICs offered to systematic studies is not due solely to total length of a region, we regressed PICs on length of the region for all of the regions surveyed in this study (Fig. 6). It is apparent that while the length of the region accounts for some proportion of the PIC value, there is a large amount of unaccountable variation in this trend. Within *Prunus*, for example, regions that are between 261 and 307 bp contain between 2 and 14 PICs, while regions that are from

TABLE 2. Quantitative data collected in this study. Each cell (cpDNA region/three-species survey), contains data regarding: aligned length of the three-species survey; the number of indels (between the ingroup taxa/between the ingroup and the outgroup taxon); average indel length; number of nucleotide substitutions (between the ingroup taxa/between the ingroup and the outgroup taxon); PICs = total indels + nucleotide substitutions + inversions (inversions are indicated by underlined PIC values). The percent variability was calculated by dividing the PIC value by the three-species aligned length. Table abbreviations: L. = length, avg. = average, subst = nucleotide substitutions, PICs = potentially informative characters.

	Metrics	*trnH-psbA*	*psbA-3'trnK*	*3'trnK-matK*	*matK-5'trnK*	*rpS16*	*trnS-trnG*
GYMNOSPERM: ***Cryptomeria / Taxodium / Glyptostrobus***	Aligned L. (bp)	519	430	-	-	784	800
	Indels: in / out	2 3	2 1	- -	- -	5 -	3 3
	Avg. indel L. (bp)	2.6	1.7	-	-	5	1.8
	Subst: in / out	7 6	12 10	- -	- -	14 -	10 21
	Total PICs	18	25	-	-	19	37
	% variability	3.47	5.81	-	-	2.42	4.63
MAGNOLIID: ***Liriodendron / Magnolia***	Aligned L. (bp)	416	282	257	704	838	687
	Indels: in / out	4 1	1 2	0 2	1 2	1 4	1 1
	Avg. indel L. (bp)	5	6.7	2	5.3	2.75	9
	Subst: in / out	16 3	0 9	2 10	3 11	5 8	7 14
	Total PICs	24	12	14	17	18	23
	% variability	5.77	4.26	5.45	2.41	2.15	3.35
MONOCOT: ***Pseudotrillium / Trillium***	Aligned L. (bp)	1077	248	257	860	795	-
	Indels: in / out	8 3	0 1	1 1	5 3	5 3	- -
	Avg. indel L. (bp)	6.1	11	5.5	8.6	6.6	-
	Subst: in / out	20 10	4 3	3 4	11 10	4 4	- -
	Total PICs	41	8	9	29	16	-
	% variability	3.81	3.23	3.50	3.37	2.01	-
CARYOPHYLLID: ***Minuartia***	Aligned L. (bp)	198	256	306	719	883	652
	Indels: in / out	0 1	1 3	2 2	3 7	6 9	4 8
	Avg. indel L. (bp)	10	16	5.75	5	4.1	8.5
	Subst: in / out	1 0	0 10	4 13	6 25	13 43	7 24
	Total PICs	2	14	22	41	71	44
	% variability	1.01	5.47	7.19	5.70	8.04	6.75
EUROSID I: ***Prunus***	Aligned L. (bp)	296	261	306	742	882	709
	Indels: in / out	3 4	1 0	1 2	0 0	2 5	4 13
	Avg. indel L. (bp)	7	6	12	0	4.6	7.9
	Subst: in / out	2 5	0 1	0 3	2 0	3 7	2 15
	Total PICs	14	2	6	2	17	34
	% variability	4.73	0.77	1.96	0.27	1.93	4.80
EUROSID II: ***Hibiscus***	Aligned L. (bp)	518	322	336	758	893	1035
	Indels: in / out	4 6	2 3	2 3	3 3	7 3	8 16
	Avg. indel L. (bp)	21.6	8.2	3.8	5.5	2.6	9.8
	Subst: in / out	5 5	1 2	2 0	9 2	2 10	5 9
	Total PICs	20	8	7	17	23	38
	% variability	3.86	2.48	2.08	2.24	2.58	3.67
EUASTERID I: ***Gratiola***	Aligned L. (bp)	373	244	277	707	946	836
	Indels: in / out	7 9	0 1	7 1	3 2	5 4	12 11
	Avg. indel L. (bp)	15.4	8	3.9	1.4	6.5	3.3
	Subst: in / out	8 25	10 6	5 6	15 14	24 17	29 30
	Total PICs	49	17	19	34	51	82
	% variability	13.14	6.97	6.86	4.81	5.39	9.81
EUASTERID I: ***Solanum***	Aligned L. (bp)	487	214	533	709	855	619
	Indels: in / out	0 2	0 0	0 0	1 1	0 0	0 0
	Avg. indel L. (bp)	7	0	0	1.5	0	0
	Subst: in / out	1 2	0 0	0 1	0 1	3	1 0
	Total PICs	5	0	1	3	3	1
	% variability	1.03	0.00	0.19	0.42	0.35	0.16
EUASTERID II: ***Eupatorium***	Aligned L. (bp)	386	212	273	773	789	-
	Indels: in / out	1 2	1 1	0 1	1 0	1 0	- -
	Avg. indel L. (bp)	2	1	1	5	2	-
	Subst: in / out	2 5	0 0	3 1	2 4	0 6	- -
	Total PICs	10	2	5	7	7	-
	% variability	2.59	0.94	1.83	0.91	0.89	-
EUASTERID II: ***Carphephorus / Eupatorium / Trilisa***	Aligned L. (bp)	387	212	285	749	794	-
	Indels: in / out	0 4	0 0	1 3	0 2	1 3	- -
	Avg. indel L. (bp)	9.5	0	3.5	3	3.25	-
	Subst: in / out	1 14	0 3	3 4	0 12	4 21	- -
	Total PICs	19	3	11	14	29	-
	% variability	4.91	1.42	3.86	1.87	3.65	-

TABLE 2. Continued.

trnG	*rpoB-trnC*	*trnC-ycf6*	*ycf6-psbM*	*psbM-trnD*	*trnD-trnT*	*trnS-trnfM*	*trnS-rpS4*
785	-	311	758	1801	-	891	218
3 4	- -	0 0	7 6	23 13	- -	10 5	2 1
2	-	0	3.3	15.3	-	2.2	1
8 5	- -	8 10	9 18	40 37	- -	13 18	2 6
20	-	18	40	113	-	46	11
2.55	-	5.79	5.28	6.27	-	5.16	5.05
783	1240	1071	764	1202	1180	1059	270
3 4	3 7	0 3	2 3	5 8	7 5	0 2	1 2
2.8	6.3	2.6	4.4	4.2	4	2	3
4 10	7 18	7 22	5 11	7 23	1 26	7 19	0 4
21	35	32	21	43	40	28	7
2.68	2.82	2.99	2.75	3.58	3.39	2.64	2.59
722	914	840	706	1055	1113	1017	314
1 0	8 5	2 2	6 5	3 6	6 10	4 5	1 0
6	6.6	62.8	13.9	39	8.2	5.6	18
4 2	8 14	4 9	11 6	11 10	11 10	9 8	4 3
7	35	17	28	30	37	26	8
0.97	3.83	2.02	3.97	2.84	3.32	2.56	2.55
719	938	246	406	751	1185	1064	209
0 4	1 16	0 2	5 6	1 8	6 17	4 11	2 2
5	5.1		6	49	17	4.66	4.8
4 28	7 40	0 10	5 17	3 43	12 52	10 33	0 9
36	64	13	33	55	87	58	13
5.01	6.82	5.28	8.13	7.32	7.34	5.45	6.22
725	1239	982	1283	1195	1000	1199	294
1 8	1 9	1 4	3 15	0 10	6 11	1 7	0 1
4.2	8.5	17.6	6.05	3.6	6.4	27	5
3 12	7 13	5 11	3 19	2 14	1 13	4 6	1 6
24	30	21	40	26	31	18	8
3.31	2.42	2.14	3.12	2.18	3.10	1.50	2.72
1008	1285	594	1218	506	1403	1804	303
11 7	6 8	1 4	6 8	3 2	11 9	14 20	0 1
13.3	9.8	6.4	5.6	3.4	10.3		12
9 6	6 7	4 4	4 2	2 4	8 10	8 10	0 2
33	27	13	20	11	38	52	3
3.27	2.10	2.19	1.64	2.17	2.71	2.88	0.99
726	1296	767	1029	517	578	1076	255
8 5	12 8	5 5	2 6	4 3	5 2	5 4	3 1
1.8	10.8	18.7	50.9	4.57	6.6	3.4	5.5
17 17	30 33	27 13	33 21	14 7	32 16	18 20	3 4
47	83	50	62	28	55	47	11
6.47	6.40	6.52	6.03	5.42	9.52	4.37	4.31
697	1309	871	1107	1025	1004	856	300
0 1	0 0	0 0	1 0	1 1	3 1	1 0	0 0
1	0	0	4	2	2	1	0
1 0	2 1	2 1	1 1	0 0	2 2	0 0	0 0
2	3	3	3	2	8	1	0
0.29	0.23	0.34	0.27	0.20	0.80	0.12	0.00
729	-	599	490	637	-	1118	307
5 0	- -	1 0	0 2	1 1	- -	4 0	2 0
2.8	-	4	59.5	72	-	7.5	2
5 3	- -	4 9	5 5	2 4	- -	1 4	1 2
13	-	14	12	8	-	9	5
1.78	-	2.34	2.45	1.26	-	0.81	1.63
732	-	620	490	-	-	1104	268
0 2	- -	2 8	0 2	- -	- -	1 3	1 0
15	-	5.3	59.9	-	-	4	5
2 5	- -	3 21	2 18	- -	- -	8 17	1 2
9	-	34	22	-	-	29	4
1.23	-	5.48	4.49	-	-	2.63	1.49

709 to 783 bp contain between 2 and 34 PICs, with the largest region not accounting for the greatest PIC value.

Our results clearly show that a disparity exists in the information offered to phylogenetic investigations by different noncoding cpDNA regions. Additionally, we show that the most widely used noncoding cpDNA regions in infrageneric systematic investigations, namely the *trnL-trnL-trnF* and *trnK-matK* intron regions, consistently provide fewer PICs than several other choices, such as *trnS-trnG-trnG, trnC-ycf6-psbM, trnD-trnT, trnT-trnL*, and *rpoB-trnC*.

TABLE 2. Continued.

rpS4-trnT	*trnT-trnL*	*trnL*	*trnL-trnF*	*5'rpS12-rpL20*	*psbB-psbH*	*rpL16*
351	*-*	*489*	*275*	*741*	*250*	*811*
1 1	*- -*	*3 3*	*6 2*	*8 5*	*0 3*	*1 3*
9	*-*	*3.5*	*4.4*	*3.4*	*2*	*10.3*
1 8	*- -*	*1 6*	*2 8*	*8 13*	*4 6*	*7 12*
11	*-*	*13*	*18*	*34*	*13*	*23*
3.13	*-*	*2.66*	*6.55*	*4.59*	*5.20*	*2.84*
364	*794*	*485*	*362*	*783*	*591*	*999*
1 3	*2 9*	*0 1*	*1 2*	*0 0*	*1 1*	*1 4*
2.5	*4.8*	*1*	*3*	*0*	*2*	*1.6*
2 4	*5 13*	*3 7*	*4 7*	*4 7*	*3 5*	*6 10*
10	*29*	*11*	*14*	*11*	*10*	*21*
2.75	*3.65*	*2.27*	*3.87*	*1.40*	*1.69*	*2.10*
328	*777*	*566*	*384*	*-*	*579*	*1055*
1 1	*2 1*	*6 2*	*3 4*	*- -*	*2 1*	*9 4*
4	*27.6*	*4.8*	*2.6*	*-*	*2.3*	*8.9*
6 2	*10 5*	*1 4*	*7 2*	*- -*	*2 10*	*14 10*
10	*18*	*13*	*16*	*-*	*15*	*37*
3.05	*2.32*	*2.30*	*4.17*	*-*	*2.59*	*3.51*
382	*644*	*572*	*413*	*846*	*595*	*939*
4 7	*3 8*	*1 9*	*1 5*	*2 5*	*0 7*	*1 8*
9.5	*18.8*	*4.8*	*3.8*	*3.71*	*4.6*	*6*
6 13	*6 26*	*1 14*	*3 29*	*7 27*	*2 14*	*5 35*
30	*43*	*25*	*38*	*41*	*23*	*49*
7.85	*6.68*	*4.37*	*9.20*	*4.85*	*3.87*	*5.22*
401	*997*	*523*	*387*	*791*	*604*	*1165*
1 2	*8 8*	*0 1*	*0 5*	*1 3*	*0 2*	*3 10*
7.3	*8*	*7*	*7.6*	*1*	*7*	*6.6*
1 6	*6 17*	*1 4*	*1 8*	*1 8*	*2 7*	*3 13*
10	*39*	*6*	*14*	*13*	*11*	*29*
2.49	*3.91*	*1.15*	*3.62*	*1.64*	*1.82*	*2.49*
785	*1023*	*602*	*439*	*760*	*584*	*1208*
15 14	*6 6*	*3 3*	*6 6*	*2 2*	*0 1*	*9 5*
11.4	*14.1*	*8*	*6.4*	*4.3*	*5*	*9*
1 1	*7 3*	*2 7*	*0 1*	*3 4*	*1 0*	*4 10*
31	*22*	*15*	*13*	*11*	*2*	*28*
3.95	*2.15*	*2.49*	*2.96*	*1.45*	*0.34*	*2.32*
353	*803*	*487*	*207*	*866*	*579*	*852*
2 5	*14 9*	*0 3*	*3 1*	*5 6*	*3 2*	*4 5*
5.3	*7.4*	*3.3*	*2.5*	*6.1*	*5.2*	*3.33*
5 5	*23 19*	*5 9*	*3 3*	*14 14*	*7 6*	*19 20*
17	*65*	*17*	*10*	*39*	*18*	*48*
4.82	*8.09*	*3.49*	*4.83*	*4.50*	*3.11*	*5.63*
367	*674*	*475*	*474*	*816*	*321*	*983*
0 0	*0 0*	*1 0*	*1 0*	*2 2*	*0 0*	*1 0*
0	*0*	*1*	*0*	*1.25*	*0*	*1*
2 0	*0 0*	*1 0*	*0 0*	*0 1*	*0 1*	*1 3*
2	*0*	*2*	*1*	*5*	*1*	*5*
0.54	*0.00*	*0.42*	*0.21*	*0.61*	*0.31*	*0.51*
345	*528*	*395*	*344*	*715*	*579*	*1005*
1 1	*1 3*	*0 0*	*0 0*	*0 0*	*1 0*	*0 1*
6	*2*	*0*	*0*	*0*	*1*	*6*
2 3	*2 0*	*1 1*	*4 2*	*1 3*	*2 2*	*6 5*
7	*6*	*2*	*6*	*4*	*5*	*12*
2.03	*1.14*	*0.51*	*1.74*	*0.56*	*0.86*	*1.19*
345	*527*	*395*	*344*	*728*	*585*	*1005*
1 2	*0 3*	*0 1*	*0 0*	*0 3*	*0 0*	*0 5*
4.3	*2*	*1*	*0*	*5*	*0*	*5.6*
1 8	*0 12*	*0 6*	*0 7*	*4 14*	*0 7*	*1 22*
12	*15*	*7*	*7*	*21*	*7*	*28*
3.48	*2.85*	*1.77*	*2.03*	*2.88*	*1.20*	*2.79*

Discussion of each of the regions—Below is a summary of each of the 21 different noncoding cpDNA regions that we have surveyed in this study including a brief history of their utility in previous studies and an assessment of their utility based on the results of this study. Because there is no intuitively straightforward way to rank each of the regions, we have divided the regions into three tiers based on their overall qualitative usefulness (Fig. 5). Tier 1 contains five regions that on average consistently provide the greatest number of PICs across all phylogenetic lineages. Tier 2 includes the next five

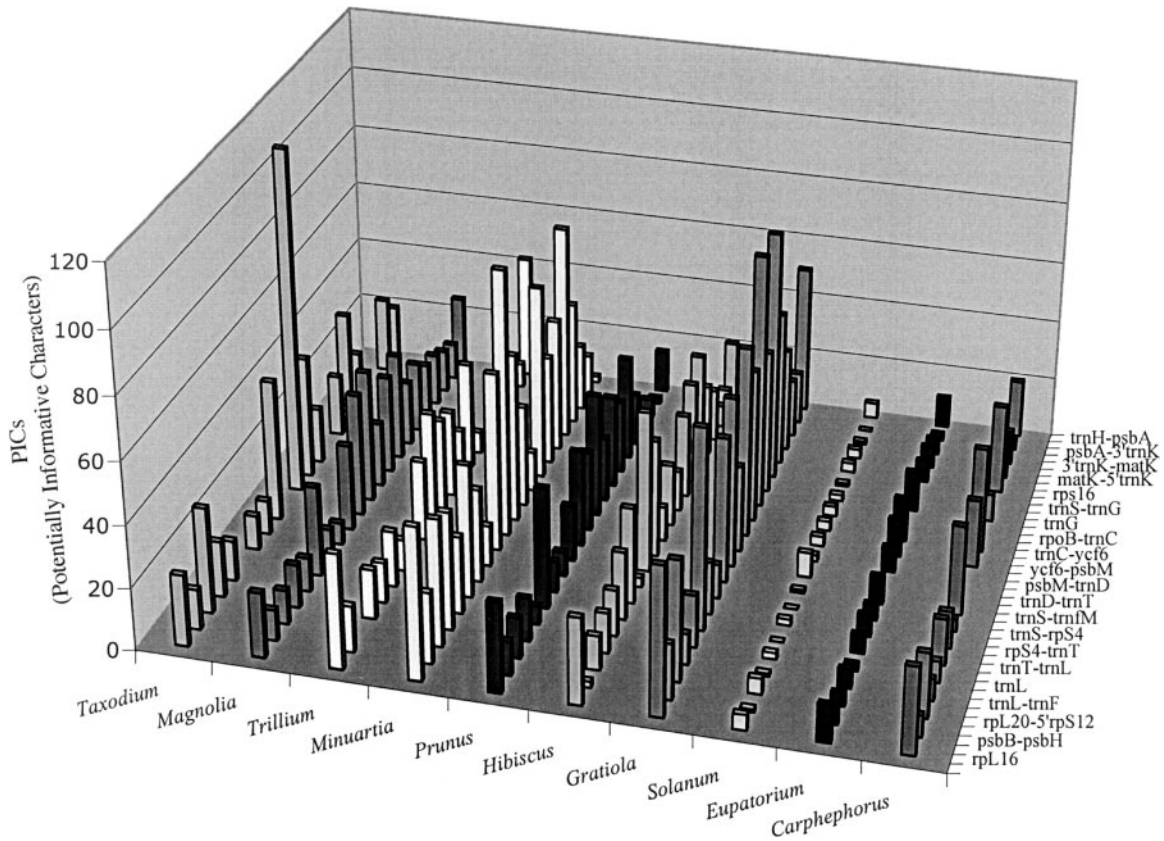

Fig. 4. Representation of the PIC (potentially informative character) values among the noncoding regions (*z*-axis) across taxonomic groups (*x*-axis). These data, summarized from Table 2, indicate that a similar trend exists among the taxonomic groups in the number of PICs provided by each region.

regions that may provide some useful information, but they may be less than optimal in providing the number of characters needed for a well-resolved phylogenetic study. Tier 3 comprises those regions that consistently provide the fewest PICs across all lineages and are therefore not recommended for low-level studies because better noncoding cpDNA choices exist. Ranking these regions in three tiers offers information relevant to studies focused on very low taxonomic levels where researchers might opt to choose one or more regions that likely contain the highest number of PICs. In addition, this ranking scheme is also useful in providing information to researchers who may wish to couple quickly evolving regions with more slowly evolving Tier 2 or Tier 3 regions, which might allow for resolution within the clade of interest in addition to confidence alignment with an outgroup (Asmussen and Chase, 2001).

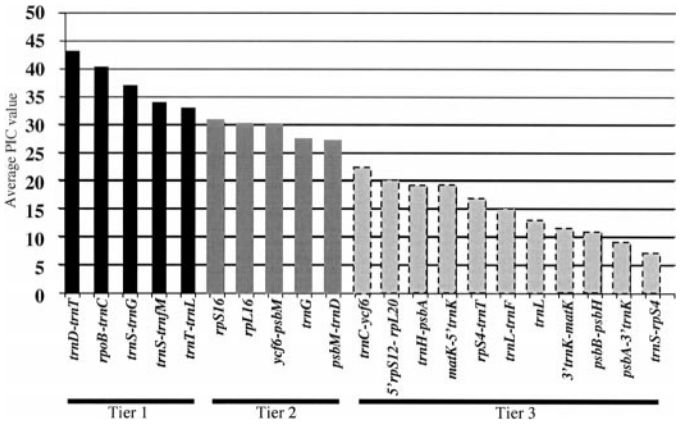

Fig. 5. The average PIC (potentially informative character) value of the six lineages with complete data sets for each region. The 21 regions, oriented in order of most to least number of PICs, are grouped into three tiers based on their qualitative value. The five Tier 1 regions are shown with black bars, the five Tier 2 regions with dark-gray bars, and the Tier 3 regions with light-gray bars with dashed outlines.

trnH^{GUG}*-psbA (Tier 3)*—Inquiry into the *trnH-psbA* intergenic spacer began with Aldrich et al. (1988) who showed that indels were prevalent in this region, even between closely related species. An early study that showed this region to be of value to systematics is Sang et al. (1997) who noted that it was highly variable compared to *matK* and *trnL-trnF*. The utility of *trnH-psbA* was also shown by Hamilton (1999b) who used it for an intraspecific study within *Corythophora* (Lecythidaceae). Subsequent to these two studies, several investigators have used this region to study closely related genera and species (Azuma et al., 1999; Chandler et al., 2001; Mast and Givnish, 2002; Fukuda et al., 2003; Miller et al., 2003; Tate and Simpson, 2003). It has also been used in an intraspecific investigation (Holdregger and Abbott, 2003). At higher levels, *trnH-psbA* has proven to be largely unalignable (Laurales: Renner, 1999; Saxifragaceae: Soltis et al., 2001; Lecythidaceae: Hamilton et al., 2003). In a study of the relative rates of nucleotide and indel evolution, Hamilton et al. (2003) showed *trnH-psbA* to be more divergent, based on percent

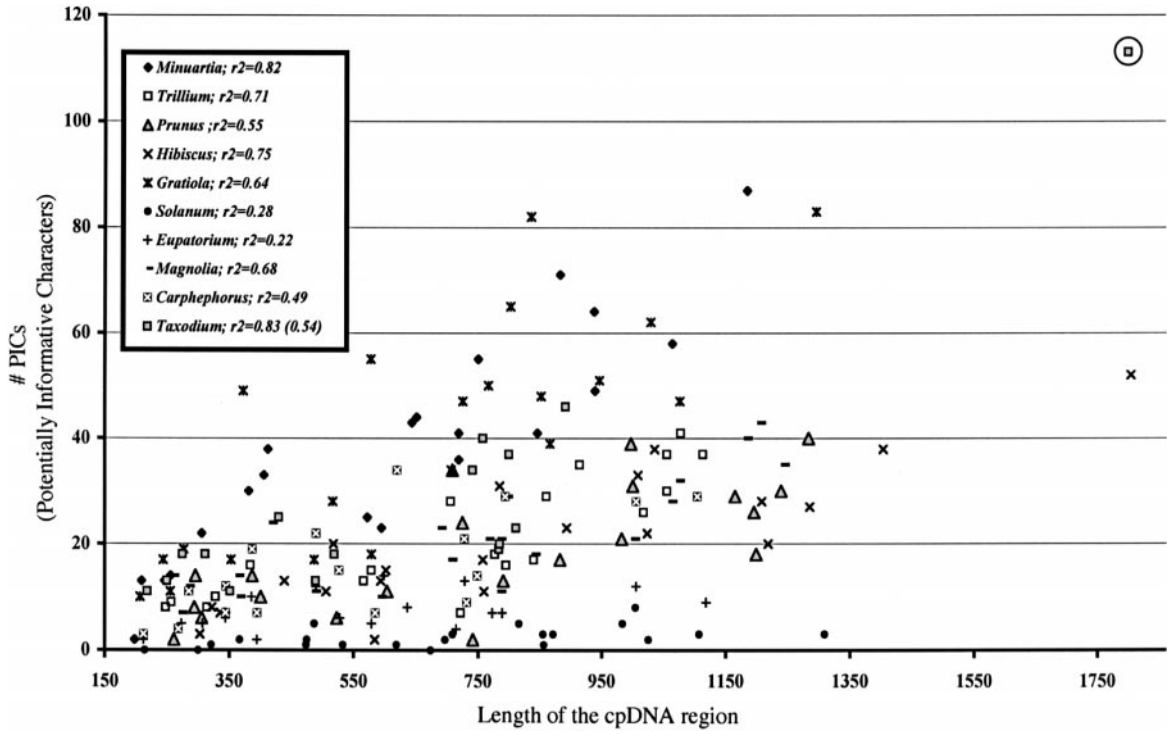

Fig. 6. Scatterplot showing the relationship of region length and its PIC value for each lineage. A key to the symbols representing each lineage is provided, including the r^2 value for each lineage. Because of space constraints, regression lines are not included. An apparent outlier in the *Taxodium* data set is indicated by a circle (upper right). The exclusion of this outlier decreases its r^2 value as shown in parentheses.

variability, than *trnS-trnG*, *psbB-psbH*, *atpB-rbcL*, *trnL-trnF*, and 5′*rpS12-rpL20*. Although studies have shown that *trnH-psbA* contains a very high percentage of variable characters (Azuma et al., 2001; Hamilton et al., 2003), this spacer is usually coupled with other regions because it is comparatively short and may not yield enough characters with which to build a well-resolved phylogeny.

The average length of *trnH-psbA* is 465 bp, and it ranges from 198 to 1077 bp. Based on our data, and data of the previous workers listed above, the 1077-bp length found in *Trillium-Pseudotrillium* is atypical. Although this spacer is the second-most variable on a percent basis, we include it in Tier 3 because its relatively short length provides few overall characters. However, it amplified and sequenced easily across all lineages and can be sequenced with only one primer in most taxa. It is also worth noting that the ends of this spacer, roughly 75 bp from either gene, are relatively conserved compared to the middle portion of this spacer, which is highly indel prone (Aldrich et al., 1988), and contains several poly-A/T runs. Most of the numerous observed indels were relatively short, but a 132-bp indel was observed among the *Hibiscus*

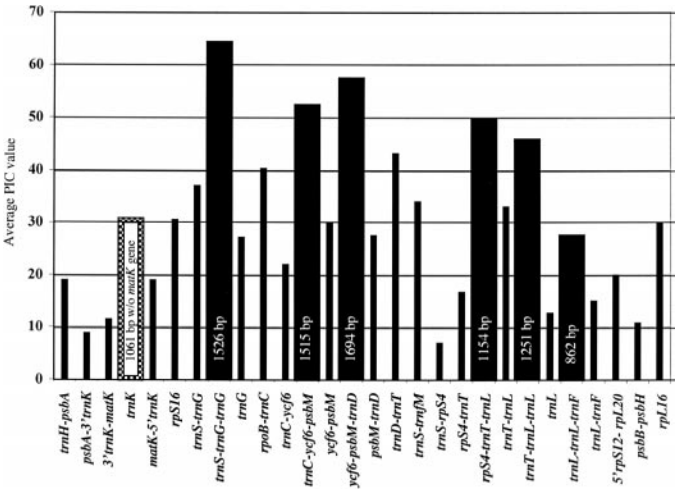

Fig. 7. The average PIC (potentially informative character) value of six lineages with complete data sets for each region compared to easily coamplifiable regions. The 21 single (narrow bars) and seven combined (thick bars) regions are oriented left to right in order as they appear in *Nicotiana* (Wakasugi et al., 1998). The size of the combined regions is included. Additionally, both halves of the *trnK* intron are shown combined (thick checkered bar) because they can be amplified as a single fragment, and each end of the intron can be sequenced completely with one primer each.

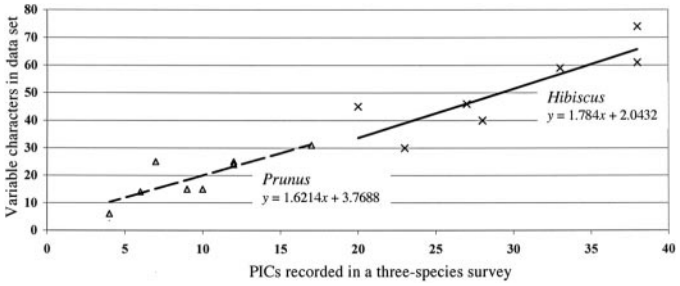

Fig. 8. Comparison of the predictability of a three-species survey with complete data sets used to generate phylogenetic hypotheses for *Prunus* (Shaw and Small, 2004) and *Hibiscus*. Regression lines are shown with their respective equations.

accessions. Among more distantly related taxa, this indel-prone middle region may generate a relatively high amount of homoplasy due to apparent indel ''hot spots'' with numerous, repeating, and overlapping indels.

psbA-3′trnK^{UUU}*-[matK]-5′trnK*^{UUU} *(Tier 3 + Tier 3 + Tier 3)*—The *matK* gene region (*trnK-matK-trnK*) or some portion of it was first employed in intrafamilial phylogenetic studies by Steele and Vilgalys (1994) and Johnson and Soltis (1994). Since then, this region has been a primary tool in phylogenetic investigations below the family level, but it has also been suggested as an effective tool above the familial level (Hilu and Liang, 1997; Hilu et al., 2003). The frequency of infrageneric phylogenetic use of this region is second only to *trnL-trnL-trnF*, representing 22 vs. 55%, respectively, of studies in 2002 (Fig. 1). Several studies have used the entire *trnK-matK-trnK* region (e.g., Johnson and Soltis, 1994; Sang et al., 1997; Hardig et al., 2000; Miller and Bayer, 2001), while most have carved out various portions depending on variable primer success and availability. Additionally, some investigators have used the intergenic spacer between *psbA* and 3′*trnK* (Winkworth et al., 2002; Pedersen and Hadenäs, 2003). In some studies the 3′*trnK* intron to some 3′ portion of *matK* was used (Wang et al., 1999; Schultheis, 2001; Winkworth et al., 2002; Hufford et al., 2003; Salazar et al., 2003). Others have used some 5′ portion of *matK* to 5′*trnK* (Plunkett et al., 1996; Ohsako and Ohnishi, 2000, 2001; Chandler et al., 2001), and still others have used part of the *matK* gene only (Kajita et al., 1998; Bayer et al., 2002; Cuénoud et al., 2002; Ge et al., 2002; Samuel et al., 2003). In many of the abovementioned investigations, several sequencing primers were required in addition to the PCR primers to piece together sequences for the entire desired region. Also, truly universal primers cannot be designed due to the variability of the gene across broad phylogenetic lineages, and often primers have to be made that are specific to different groups (e.g., Wang et al., 1999; Hardig et al., 2000; Hu et al., 2000; Miller and Bayer, 2001; Mort et al., 2001; Pridgeon et al., 2001; Bayer et al., 2002; Hilu et al., 2003). Therefore, in terms of cost, the *matK* region is relatively expensive because it often involves several sequencing reactions from multiple unique primers. Although *matK* is putatively the most variable coding region found within cpDNA (Neuhaus and Link, 1987; Olmstead and Palmer, 1994), it was excluded from this study primarily because it is a coding region and not part of our focus. Furthermore, the gene's large size would require the development of several internal sequencing primers, and with few strategically placed conserved regions, the number of primers for specific lineages becomes too cumbersome for the scope of this investigation. Therefore, we only included both ends of the *trnK* intron in addition to the *psbA*-3′*trnK* intergenic spacer.

Although the above discussion may read as to denigrate the *psbA*-3′*trnK*-[*matK*]-5′*trnK* region, this was not our intent. Because of the plethora of data already available, this region is valuable with respect to its potential in comparative studies (e.g., the placement of taxa whose phylogenetic positions are ambiguous).

The *psbA*-3′*trnK* intergenic spacer is usually shorter than either portion of the *trnK* intron, averaging 268 bp in length and ranging from 212 to 430 bp. The 5′ end of the *trnK* intron is consistently larger, with an average of 747 bp and a range of 704–860 bp, than the 3′ end of the *trnK* intron, which averages 314 bp and ranges from 257 to 533 bp. While the 3′*trnK* portion of the intron is more variable on a percent basis than the 5′ portion, the 5′*trnK* portion consistently provides more PICs, as was reported in *Acacia* by Miller and Bayer (2001). Compared to other noncoding cpDNA regions surveyed in this study, both the *trnK* intron and *psbA*-3′*trnK* spacer provide relatively few variable characters and are ranked in Tier 3. Even combining the two halves of the intron yields an average PIC value below that of several other regions (Fig. 7). It is our opinion that the entire *psbA-trnK-matK-trnK* region is less suitable for infrageneric phylogenetic investigation than several other choices because it is very large, less informative than other regions, lacks sufficiently conserved coding regions where ''universal'' primers can be anchored, and was inexplicably problematic during sequencing.

rpS16 (Tier 2)—The ribosomal protein 16 small subunit gene (*rpS16*) contains a group II intron that was first used in a phylogenetic context by Oxelman et al. (1997). Since this initial investigation the *rpS16* intron has been used to successfully resolve relationships among genera in Rubiaceae subfamily Rubioideae (Andersson and Rova, 1999), Arecaceae subfamily Calamoideae (Baker et al., 2000), Arecaceae (Asmussen and Chase, 2001), Fabaceae tribe Glycininae (Lee and Hymowitz, 2001), Marantaceae (Andersson and Chase, 2001), Apiaceae subfamily Apioideae (Downie and Katz-Downie, 1999), and Colchicaceae (Vinnersten and Reeves, 2003). However, within each of these studies, infrageneric resolution was weak. Other studies have also shown that *rpS16* is usually not variable enough to resolve infrageneric relationships (Baker et al., 2000; Edwards and Gadek, 2001; Wanntorp et al., 2001; Popp and Oxelman, 2001; Aagesen and Sanso, 2003; Ingram and Doyle, 2003).

Originally, the *rps16* intron was suggested to be a valuable tool for investigation at the family level and below (Oxelman et al., 1997), but the accumulated literature in addition to our data suggests it will often not provide enough characters to resolve relationships below generic levels. The intron averages 846 bp in length and ranges from 784 to 946 bp. The *rpS16* intron is typically more informative than the *trnL-trnL-trnF* region (Fig. 7), but it frequently contains fewer PICs than other choices and is therefore included in Tier 2. A poly-A/T run in most lineages (especially *Prunus* and *Hibiscus*) at the 3′ end of the intron may be problematic in sequencing from that direction. This region cannot be used in some taxa because all or some part of the *rpS16* gene is absent from some members of Linaceae, Malpighiaceae, Passifloraceae, Salicaceae, Polygalaceae, Turneraceae, Violaceae (see Downie and Palmer, 1992), Connaraceae, Eucommiaceae, Fagaceae, Krameriaceae, Fabaceae (see Doyle et al., 1995), *Marchantia polymorpha* (Ohyama et al., 1986), *Pinus thunbergii* (Tsudzuki et al., 1992), *Pisum sativum* (Nagano et al., 1991) and *Epifagus virginiana* (Wolfe et al., 1992).

trnS^{GCU}*-trnG*^{UUC}*-trnG*^{UUC} *(Tier 1 + Tier 2)*—Hamilton (1999a, b) designed primers for the intergenic spacer between *trnS* and *trnG* (*trnS-trnG*) to study population dynamics within a tropical tree species in *Corythophora* (Lecythidaceae) and subsequently published them along with the suggested amplification protocol. Xu et al. (2000) designed nearly the same primers for this spacer for use in *Glycine* (Fabaceae). Subsequent studies have shown this region to be highly variable. Olson (2002a) showed that the *trnS-trnG* spacer sequences are largely unalignable between genera in the Caricaceae-Morin-

gaceae clade, and Xu et al. (2000) showed the *trnS-trnG* spacer to be among the most informative of nine noncoding cpDNA regions within two closely related subgenera of *Glycine*. In another study within *Glycine*, Sakai et al. (2003) showed the *trnS-trnG* spacer to contain many more PICs than *atpB-rbcL, rpS11-rpL36*, and *rpS3-rpL16*. The *trnS-trnG* spacer was reported to show intraspecific variation in *Moringa* (Moringaceae) by Olson (2002b) and *Corythophora* (Lecythidaceae) by Hamilton (1999a, b). Perret et al. (2003) showed the *trnS-trnG* spacer to provide more PICs than *rpL16, trnL* intron, *trnL-trnF* spacer, *trnT-trnL* spacer, and *atpB-rbcL* spacer in the tribe Sinningieae (Gesneriaceae) and Hamilton et al. (2003) showed it to be more informative than *psbB-psbH, atpB-rbcL, trnL-trnF*, and 5′*rpS12-rpL20* in *Corythophora* (Lecythidaceae). However, Shönenberger and Conti (2003) showed the *trnS-trnG* spacer to contain fewer PICs than the *rpS16* and *rpL16* introns, but having more PICs than the *trnH-psbA* spacer, *atpB-rbcL* spacer, and part of the *matK* exon. Gaskin and Schaal (2003) showed that *trnS-trnG* is five times more variable than the *trnL-trnF* spacer and contained more variable characters than nuclear ribosomal ITS in *Tamarix*. Lastly, Pacak and Szweykowska-Kulinska (2000) designed primers to study the group II *trnG* intron which was not included in the abovementioned studies. They found that this intron provided several nucleotide substitutions in a group where the *trnL* intron was invariant. Pedersen and Hedenäs (2003) also used the *trnG* intron and showed that, although it did not contain as many variable characters as *rpL16*, it provided nearly twice as many as *trnL-trnF*.

Because Hamilton's (1999a) ''G'' primer was designed in the 5′ *trnG* exon and the *trnG* intron was shown to be relatively variable, we designed a 3′ exon *trnG* primer that would allow the *trnS-trnG* intergenic spacer and the *trnG* intron to be coamplified. Additionally, we created internal sequencing primers that are located in the internal 5′ *trnG* exon near the position of Hamilton's (1999a) *trnG* primer.

The *trnG* intron averages 763 bp in length and ranges from 697 to 1008 bp, while the *trnS-trnG* spacer averages 763 bp and ranges from 619 to 1035 bp. Alone the *trnS-trnG* spacer ranks in Tier 1, while the *trnG* intron ranks in Tier 2. Additionally, the *trnS-trnG* spacer not only provides the greatest number of PICs, but also has the highest percent variability with an average of 4.74%. However, when combined as a coamplifiable unit, *trnS-trnG-trnG* averages approximately 1500 bp and provides the greatest number of PICs per two (very rarely three) sequencing reactions compared to any other regions, single or combined, surveyed in this study (Figs. 5 and 7). All lineage representatives included in this study have a poly-A/T run in the *trnG* intron near the 3′*trnG* end, which is usually not long enough to affect PCR or sequencing. Sometimes, as is the case in *Taxodium*, this poly-A/T run is over 30 bp and prohibits sequencing from that direction. Sequencing with internal primers usually alleviates this problem. Approximately the same number of indels were found in both the *trnS-trnG* spacer and the *trnG* intron, and large indels were noted in *Hibiscus* (73 bp) and in the complete data set of *Prunus* (358 bp) (Shaw and Small, 2004).

Because of independent structural rearrangements in both monocots (Hiratsuka et al., 1989) and Asteraceae, excluding Barnadesieae (Jansen and Palmer, 1987), the *trnS-trnG* spacer does not exist in these taxa. However, the *trnG* intron can be used in both of these groups, but there may be better choices for such studies (Table 2, Fig. 5).

rpoB-trnC^{GCA} (Tier 1)—The *rpoB-trnC*^{GCA} region was first used by Ohsako and Ohnishi (2000, 2001) in their study of intra- and interspecific relationships in *Fagopyrum* (Polygonaceae). They showed this spacer to contain enough variable characters to distinguish between closely related species in addition to showing some intraspecific variation. Studying only intraspecific relationships in *Fagopyrum cymosum*, Yamane et al. (2003) found 15 informative characters. The *rpoB-trnC* spacer has also been used in an intergeneric study of subtribe Clematidinae (Ranunculaceae) where O. Miikeda et al. (Tokyo Metropolitan Mizuho-nogei High School, personal communication) found 66 potentially informative characters.

The average length of this region is 1174 bp with a range of 914–1309 bp, comparable with that found in the studies listed above. The center region of this intergenic spacer contains several relatively small (<8 bp) poly-A/T strings, and the *rpoB-trnC* spacer contains several relatively large indels (47 bp in *Prunus*, 58 bp in *Hibiscus*, 81 bp in *Gratiola*). This spacer, ranked in Tier 1, is among the most informative regions with respect to the number of PICs offered for infrageneric investigations (Figs. 5 and 7).

trnC^{GCA}-ycf6-psbM-trnD^{GUC} (Tier 3 + Tier 2 + Tier 2)—The *trnC*^{GCA}*-trnD*^{GUC} region is approximately 3200 bp long in *Nicotiana* (Wakasugi et al., 1998) and includes the genes *ycf6* and *psbM* which are 90 and 105 bp long, respectively. This region was first identified as a potential region for phylogenetic study by Demesure et al. (1995) who reported a length of 3000 bp in *Quercus* (Fagaceae). Demesure et al. (1996) subsequently used the region in a PCR-RFLP phylogeographic study in *Fagus* (Fagaceae). In their PCR-RFLP study of the interspecific relationships in *Allium* (Alliaceae), Mes et al. (1997) also used the *trnC-trnD* region. Hartmann et al. (2002) compared sequences of a 1149-bp portion of this region among eleven cactus species across five genera to determine the origin of *Lophocereus* (Cactaceae), but only found nine informative characters. Sequences of this region have been used to elucidate infrageneric relationships in *Humulus* (Cannabaceae) (A. Murakami, Kirin Brewery Company, Ltd., unpublished data) and *Panax* (Araliaceae) (Lee and Wen, 2004). In both studies, internal primers were designed to completely sequence this region which is about 2600 bp in *Humulus* and up to 3000 bp in *Panax*. Within the aligned data set for *Panax*, Lee and Wen (2004) report 71 informative characters plus 20 informative indels.

Because this entire region is large enough to be cumbersome and each of the three intergenic spacers greatly varies in length, we have analyzed the three intergenic spacers separately. The average length of the entire region is 2480 bp with a range of 1726–3460 bp. The *trnC-ycf6* intergenic spacer averages 690 bp with a range of 246–1071 bp, the *ycf6-psbM* spacer averages 825 bp with a range of 406–1283 bp, and the *psbM-trnD* spacer averages 965 bp and ranges from 506 to 1801 bp. All three of these regions appear to be prone to large indels. We observed indels of 232 bp in *Trillium-Pseudotrillium*, and 64 and 89 bp in *Gratiola* in the *trnC-ycf6* spacer, indels of 107 bp in *Carphephorus-Trilisa*/*Eupatorium*, 86 bp in *Trillium-Pseudotrillium*, and 371 bp in *Gratiola* in the *ycf6-psbM* spacer, and indels of 40 bp in *Minuartia*, 142 bp in *Eupatorium*, 360 and 45 bp in *Taxodium-Glyptostrobus-Cryptomeria* in the *psbM-trnD* spacer. Of the three regions, *ycf6-psbM* and *psbM-trnD* rank in Tier 2, while *trnC-ycf6* ranks at the top of Tier 3. Of these three regions, the *ycf6-psbM* inter-

genic spacer provides the greatest number of PICs. The value of *ycf6-psbM* may actually be an underestimate because many potential characters may have been hidden in the large portions of missing data caused by large indels among the three-species sequence data sets. When combined with either *trnC-ycf6* or *psbM-trnD*, the *ycf6-psbM* spacer is the second-most variable coamplifiable region behind *trnS-trnG-trnG*. When the entire *trnC-ycf6-psbM-trnD* region is compared to *trnS-trnG-trnG*, it provides a greater number of PICs on average, 78 vs. 64 respectively (Fig. 7), but this much larger fragment requires at least two additional sequencing reactions. Therefore, there are likely better combinations of regions to obtain a greater number of characters.

trnD^{GUC}*-trnT*^{GGU} *(Tier 1)*—Using the aligned sequences of *Oryza*, *Nicotiana*, and *Marchantia*, Demesure et al. (1995) developed a primer pair anchored within the *trnD* and *trnT* genes to amplify the noncoding intergenic spacer and the embedded *trnY*^{GUA} and *trnE*^{UUC} genes, which are 84 and 73 bp, respectively. Friesen et al. (2000) used sequence data from this region to investigate the phylogenetic relationships among some *Allium* species and the monotypic *Milula* (Alliaceae), in which the resulting parsimony trees were well resolved and comparable to those generated with nuclear ITS sequences. The *trnD-trnT* region provided sufficient characters to separate populations in an intraspecific study of *Cunninghamia konishii* (Cupressaceae) and the very closely related *Cunninghamia lanceolata* (Lu et al., 2001). In a phylogenetic study of the arecoid lineage of palms (Arecaceae), Hahn (2002) showed that this region provided more variable characters than either the *trnQ-rpS16* intergenic spacer or the *atpB* and *rbcL* genes combined. To trace wild parentage and potential hybridization among cultivated rootstocks of *Juglans* (Juglandaceae), Potter et al. (2002) sequenced the *trnD-trnT* region, along with *trnT-trnL* and *trnL-trnF*, and ITS. Although they could not obtain the middle portion of the *trnD-trnT* spacer because of a large poly-A/T, this region still provided more variable characters than *trnT-trnL* and *trnL-trnF* and provided the same number of variable characters as those two regions put together, while providing only one fewer character than ITS. That the *trnD-trnT* spacer might provide an equivalent number of variable characters as ITS was also suggested by Feliner et al. (2002). In a phylogenetic study of *Brassica* and *Raphanus* (Brassicaceae), Yang et al. (2002) showed that the ~1150 bp *trnD-trnT* region evolves 1.1 times faster than the ~1641-bp *trnT-trnL-trnL-trnF* and that both regions provided nearly the same number of nucleotide substitutions, 345 and 346, respectively.

The *trnD-trnT* intergenic spacer averages 1066 bp, ranging from 578 to 1403 bp, and provides the greatest number of PICs compared to all of the uncombined regions we surveyed (Fig. 5). This region amplified and sequenced easily for all of the lineages except for the two representatives of the Asteraceae clade where this spacer is interrupted by the same inversion involving the *trnS-trnG* spacer (Jansen and Palmer, 1987). As an aside, we attempted to amplify the resulting *trnD*^{GUC}*-trnS*^{GCU} and *trnT*^{GGU}*-trnG*^{UUC} using the primers above for our Asteraceae representatives, but these attempts were not entirely successful and provided no useful information (data not shown). Within the *trnD-trnT* region, several relatively large indels were noted (246 bp in *Minuartia*, 42 bp in *Trillium*, 87 bp in *Hibiscus*) as well as poly-A/T runs and poly-AT repeats. Because of these runs and repeats, we designed (but never used beyond making sure that they work), universal internal sequencing primers embedded in the *trnE* and *trnY* genes.

trnS^{UGA}*-trnfM*^{CAU} *(Tier 1)*—A universal primer pair for the amplification of the *trnS-trnfM* intergenic spacer was developed by Demesure et al. (1995), using aligned sequences from *Oryza*, *Nicotiana*, and *Marchantia*. Subsequent studies used the PCR-RFLP method to investigate geographically structured intraspecific variation (e.g., El Mousadik et al., 1996; Stehlik, 2002; Stehlik et al., 2002). Zuber and Widmer (2000) used sequences of the *trnS-trnfM* region to assess genetic variation within and among host-specific subspecies of *Viscum album* (Viscaceae). They showed that the *trnS-trnfM* spacer provided more nucleotide substitutions than *trnL-trnL-trnF*, *trnH-trnK*, or ITS, 8 vs. 2, 7, 5, respectively. Chassot et al. (2001) showed that *trnS-ycf9*, approximately one-third of the *trnS-trnfM* region (Fig. 3), provided 76 informative characters, whereas *trnL-trnF* provided 83 and the *trnL* intron provided 59. Although Hartmann et al. (2002) investigated the phylogenetic origins of *Lophocereus* (Cactaceae) using *trnS-trnfM* (and others), they only sequenced approximately 375 bp of the ~1.5 kb spacer; therefore a comparison of the relative utility within this study could not be made.

The average length of the *trnS-trnfM* region is 1119 bp with a range of 856-1804 bp. Embedded in this region are coding regions *trnG*^{GCC}, *ycf9*, and *psbZ*. Although it contains three genes, our data, in addition to the previously mentioned studies, indicate this region exhibits a relatively high PIC value. Within *Minuartia* and *Prunus*, two large indels of 43 and 141 bp, respectively, were observed. Several of the taxa surveyed here showed poly-A/T runs, most of which were not long enough to affect sequencing. However, an approximately 250-bp portion of the *Hibiscus* data set was excluded from our analyses because it consisted of nearly all A's and T's and could not be confidently aligned.

trnS^{GGA}*-rpS4 (Tier 3)*—Cranfill (2001) suggested that the *trnS-rpS4* spacer might be of phylogenetic utility and described it as useful for investigations below the family level because it evolves at a rate similar to ITS. Subsequently, Smith and Cranfill (2002) used this region in intrafamilial reconstruction of the thelypteroid ferns, although it was combined with the *rpS4* gene and the *trnL-trnF* spacer and comparative analyses among these regions were not discussed. Hennequin et al. (2002) coupled this spacer with *rbcL* and the *rpS4* genes in an investigation of *Hymenophyllum* (Hymenophyllaceae) and showed that it provided more variable characters than either. However, the utility of this spacer with respect to Hennequin et al. (2002) should be taken with caution because it was compared only to coding regions.

The *trnS-rpS4* intergenic spacer averages 273 bp and ranges from 209 to 314 bp. This Tier 3 intergenic spacer yielded the lowest PIC value of any region surveyed in this study and is therefore not recommended as a systematic tool for infrageneric studies.

rpS4-trnT^{UGU} *(Tier 3)*—Saltonstall (2001) amplified the *rpS4-trnT* intergenic spacer along with several other intergenic spacers to test the ''universality'' of the primers in one member of each of the six major subfamilies of Poaceae. She then used sequence data from *rpS4-trnT*, along with several other regions, to assess the amount of intraspecific polymorphism within each of these regions in *Phragmites australis*. Because

only numbers of haplotypes were reported, it is difficult to make inferences as to the comparative utility of regions within her study. However, she did report that this spacer provided more haplotypes than *trnH-psbA, trnT-trnE, rpoB-trnC, trnL, trnL-trnF* and some lesser known intergenic spacers. Before the start of this study, we (J. Shaw and R. L. Small) failed in several attempts to obtain the *trnT-trnL* intergenic spacer in *Prunus* and *Hibiscus* because of the problematic nature of the TabA primer (see *trnT*^{UGU}-*trnL*^{UAA}-*trnL*^{UAA}-*trnF*^{GAA} discussion below). Therefore, we designed a primer embedded in the *rpS4* gene that allowed us to obtain the *trnT-trnL* spacer in addition to the *rpS4-trnT* spacer.

Although Saltonstall (2001) reported that the *rpS4-trnT* intergenic spacer is 750–950 bp in grasses, we found it averages 402 bp with a range of 345 to 785 bp. This region shows a relatively low PIC value (see Fig. 5), and is ranked in Tier 3.

trnT^{UGU}*-trnL*^{UAA}*-trnL*^{UAA}*-trnF*^{GAA} *(Tier 1 + Tier 3 + Tier 3)*—One of the first sets of universal PCR primers for noncoding cpDNA was published by Taberlet et al. (1991). These primer sets span a region comprising three tRNA genes—*trnT*^{UGU}, *trnL*^{UAA}, and *trnF*^{GAA}. The noncoding portions of the region include a Group I intron that interrupts the *trnL* gene, as well as the intergenic spacers between *trnT-trnL* and *trnL-trnF*. Taberlet et al. (1991) described primer sequences situated in conserved regions of the tRNA genes for amplifying each of these regions and demonstrated amplification in land plants ranging from bryophytes to pteridophytes, gymnosperms, and angiosperms. Because of the near-universal nature of the primers and their early publication, these regions have become the most widely used noncoding cpDNA sequences in plant systematics. As of December 2003, Web of Science lists 579 citations of the Taberlet et al. (1991) paper. Usually these regions are employed in studies of closely related species or genera, but a recent study by Borsch et al. (2003) used the entire region to evaluate relationships among basal angiosperms. Renner (1999) used the *trnT-trnL* and *trnL-trnF* spacers, along with other coding and noncoding regions, in an analysis of the order Laurales. Bremer et al. (2002) also employed the entire region, along with other coding and noncoding regions, in a phylogenetic analysis of Asterids, as did Stech et al. (2003) who analyzed *trnL* intron and *trnL-trnF* spacer sequences in a broad survey of land plants and algae.

The *trnT-trnL* intergenic spacer has been the least used of the Taberlet et al. (1991) regions due to difficulties with PCR amplification in many plant groups (personal observation and personal communication from colleagues). This difficulty apparently stems from the *trnT* primer, Taberlet et al. (1991) primer ''A,'' but a new PCR-amplification primer designed by Cronn et al. (2002) works in all of the taxa surveyed in this study (data not shown). Studies that have used the *trnT-trnL* spacer often report that it provides greater variation than other surveyed regions, including the *trnL* intron and the *trnL-trnF* spacer (e.g., Böhle et al., 1994; Small et al., 1998; Cronn et al., 2002; Downie et al., 2002). This spacer exhibits a wide range of sizes in different plant groups from ~400–1500 bp and often includes large A/T rich regions that may be difficult to align among divergent sequences.

The *trnL* intron is the only chloroplast Group I intron (Palmer, 1991). It has a specific secondary structure and several highly conserved regions that are found among all Group I introns (Westhof and Michel, 1996; Stech et al., 2003). This intron ranges in size from as small as ~250 bp in pteridophytes and bryophytes (Stech et al., 2003) to over 1400 bp in some angiosperms (e.g., *Disa* [Orchidaceae], Bellstedt et al. [2001]). The *trnL-trnF* spacer is generally shorter than the *trnL* intron, ranging from less than 100 bp in mosses and liverworts (Stech et al., 2003) up to ~500 bp in seed plants.

Sequences of the *trnL* intron and *trnL-trnF* spacer have been employed in numerous studies, oftentimes together because they can be coamplified using the ''C'' and ''F'' primers of Taberlet et al. (1991). In studies where both the *trnL* intron and the *trnL-trnF* spacer have been sequenced for a common set of taxa, the number of parsimony-informative characters in the *trnL-trnF* spacer is often greater than or equal to the *trnL* intron, despite the fact that the *trnL* intron is usually larger than the *trnL-trnF* spacer. For example, in *Lepidium* (Brassicaceae) the aligned length of the *trnL* intron was 519 bp vs. 350 bp in the *trnL-trnF* spacer, yet only 46 parsimony-informative sites were detected in the *trnL* intron vs. 52 in the *trnL-trnF* spacer (Mummenhoff et al., 2001). Similarly, in *Disa* (Orchidaceae) the *trnL* intron was 1412 bp aligned vs. 359 bp for the *trnL-trnF* spacer, yet 72 parsimony-informative substitutions were found in each region, despite the fact that the *trnL* intron is almost four times longer (Bellstedt et al., 2001). Lastly, Yang et al. (2002) showed that the rate of nucleotide substitution in the *trnL* intron is about 33% of that within the *trnT-trnL* or *trnL-trnF* spacers. Presumably, this observation is due to greater functional constraints on the *trnL* intron that must assume a correct secondary structure for proper removal.

A comprehensive list of studies using the *trnT-trnL-trnL-trnF* regions is spatially impossible and the majority of the papers cited in this work contain at least a portion of this region. However, the following list provides some representative papers and emphasizes those studies that have employed combinations of *trnT-trnL* and *trnL-trnF* spacer and/or *trnL* intron sequences or comparisons of these sequences with other coding or noncoding regions. These studies include: Gielly and Taberlet (1994); Sang et al. (1997); Small et al. (1998); Mort et al. (2001); Bremer et al. (2002); Cronn et al. (2002); Goldblatt et al. (2002); Hartmann et al. (2002); Mast and Givnish (2002); Borsch et al. (2003); Fukuda et al. (2003); Jobson et al. (2003); Miller et al. (2003); Salazar et al. (2003); Simpson et al. (2003); and Stech et al. (2003).

Our data support the previous findings that the *trnT-trnL* spacer is much more variable than the *trnL-trnF* spacer and that the *trnL* intron is the least variable of these three regions. The *trnT-trnL* spacer averages 752 bp, ranging from 527 to 1023 bp, and exhibits a few large indels (174 bp in *Minuartia*, 46 bp in *Hibiscus*, 61 bp in *Gratiola*, 63 bp in *Trillium*). The *trnL* intron averages 499 bp, ranging from 395 to 602 bp, whereas the *trnL-trnF* spacer averages 362 bp and ranges from 207 to 474 bp. Separately, the *trnL* intron and the *trnL-trnF* spacer are Tier 3 regions, while the *trnT-trnL* spacer is a Tier 1 region. When the *trnL* intron and the *trnL-trnF* spacer (*trnL-trnL-trnF*) are combined and compared to the other combined regions, this region still ranks behind several others (Fig. 7). However, when the *trnT-trnL* spacer is coamplified with either the *rps4-trnT* spacer or the *trnL* intron (Fig. 7), this combined region is among the most variable of the combined regions, behind only *trnS-trnG-trnG, trnC-ycf6-psbM*, or *ycf6-psbM-trnD*.

5′rpS12-rpL20 (Tier 3)—Hamilton (1999a, b) designed primers for the intergenic spacer between 5′*rps12* and *rpL20* to study population dynamics within a tropical tree species in the

genus *Corythophora* (Lecythidaceae) and subsequently published them along with the suggested amplification protocol.

The 5′*rps12-rpL20* intergenic spacer averages 783 bp and ranges from 715 to 866 bp. This is comparable to the length of approximately 880 bp in *Corythophora alta* reported by Hamilton (1999b). Contrary to Hamilton (1999a), this Tier 3 region consistently shows a relatively low PIC value and does not appear suitable for low level investigations.

psbB-psbH (Tier 3)—Hamilton (1999a, b) designed primers for the intergenic spacer between *psbB* and *psbH* to study population dynamics within a tropical tree species in *Corythophora* (Lecythidaceae) and subsequently published them along with the suggested amplification protocol. Xu et al. (2000) showed this region to be less informative than *trnH-psbA*, *trnS-trnG*, and *trnT-trnL*, but more informative than *atpB-rbcL* and *ndhD-ndhE*. Schütze et al. (2003) also used this region in addition to the *atpB-rbcL* spacer in an investigation of the Suaedoideae (Chenopodiaceae) and showed that it was only slightly more than half as informative as *atpB-rbcL*. One reason for the lack of variable characters within *psbB-psbH* is that two genes, *psbT* (~100 bp) and *psbN* (~130 bp), comprise nearly half of the approximately 527-bp spacer between *psbB* and *psbH*.

The *psbB-psbH* intergenic region averages 527 bp, ranging from 250 to 604 bp, and contains approximately 230 bp of coding sequence, which is relatively invariable across all lineages. In reference to the number of PICs offered to investigators, this Tier 3 region ranks toward the bottom and is therefore not a suitable region for low-level investigation, contrary to Hamilton (1999a).

rpL16 (Tier 2)—Posno et al. (1986) first demonstrated homology between the chloroplast *rpL*16 region from *Spirodela* (Lemnaceae) and the ribosomal protein L16 of *E. coli*, and observed the presence of a Group II intron that split the coding region into shorter and longer exons. Comparisons of whole chloroplast sequences confirmed its presence and pointed to the *rpL16* region as having high sequence divergence in flowering plants (Wolfe et al., 1987). Jordan et al. (1996) published the first attempt to use *rpL16* intron sequence data for phylogenetic studies, but reported relatively little variation in Lemnaceae. Application of data from the *rpL16* region has been primarily for phylogenetic analysis at the infrageneric and familial levels (Kelchner and Clark, 1997; Baum et al., 1998; Schnabel and Wendel, 1998; Small et al., 1998; Seelanan et al., 1999; Applequist and Wallace, 2000; Downie et al., 2000; Zhang, 2000; Shaw, 2000; Baumel et al., 2001; Mast et al., 2001; Butterworth et al., 2002; Cronn et al., 2002; Les et al., 2002; Mast and Givnish, 2002; Pfeil et al., 2002; Pires and Sytsma, 2002; Kimball et al., 2003; Perret et al., 2003). There have been some reports of variability within species (e.g., Seelanan et al., 1999; Les et al., 2002), but only a few studies have reported using it specifically for examination of intraspecific variation (Xu et al., 2000; Kimura et al., 2003).

The *rpL16* intron averages 1002 bp, ranging from 811 to 1208 bp, and is especially indel prone in the D3 bulge region (Baum et al., 1998; Kelchner, 2000; Kelchner, 2002; Pfeil et al., 2002). Relative to the other regions surveyed here, this intron ranks in Tier 2. Finally, the *rpL16* intron is absent in at least some Geraniaceae, Goodeniaceae, and Plumbaginaceae (Campagna and Downie, 1998), precluding its use in some groups.

Indels vs. nucleotide substitutions—A number of authors have addressed the issue of the relative frequencies of nucleotide substitutions and indels in noncoding cpDNA sequences. Clegg et al. (1994) noted that indels may occur more frequently than nucleotide substitutions. Golenberg et al. (1993) and Gielly and Taberlet (1994) suggested that indels occur with nearly the same frequency as nucleotide substitutions. On the contrary, our results agree more with Small et al. (1998). We found that nucleotide substitutions account for 70.1% of the PIC value, while indels account for only 29.8% and inversions only 0.14% of the 35% of the noncoding LSC region we surveyed. As systematic investigations move more toward lower levels and any variable characters become important, indels, which may be homoplasious at deeper levels (Golenberg et al., 1993), are of great utility for infrageneric studies.

Several of the regions were rich in strings of mononucleotide repeats and/or small tandem repeat units that are likely the result of slipped-strand mispairing (Levinson and Gutman, 1987). Polynucleotide (A/T) repeats and/or small tandem repeats (AT) were especially noted in the *trnH-psbA*, *psbA-3′trnK*, *matK-5′trnK*, *trnS-trnfM*, *trnS-trnG*, *trnD-trnT*, *trnT-trnL* spacers and in the *rpS16* and *trnG* introns. Length variation, because of relatively large indels, was noted in several regions, described above.

Implications of this study—The results of our study do not point to a ''holy grail'' of noncoding cpDNA regions that can be universally used for low-level systematic studies. However, our results do highlight several noncoding cpDNA regions that are better suited for low-level investigation than many commonly employed choices. The Tier 1 intergenic spacers that provided the greatest numbers of PICs in order of most to least include *trnD-trnT*, *rpoB-trnC*, *trnS-trnG*, *trnS-trnfM*, and *trnT-trnL*. Several other regions, designated as Tier 2, may also be useful to investigators if one or more of the Tier 1 regions cannot be obtained or simply to add more data, including *rpS16*, *rpL16*, *ycf6-psbM*, *trnG*, and *psbM-trnD*. Regions that consistently provided the fewest PICs include *trnC-ycf6*, 5′*rpS12-rpL20*, *trnH-psbA*, *matK-5′trnK*, *rpS4-trnT*, *trnL-trnF*, *trnL*, 3′*trnK-matK*, *psbB-psbH*, *psbA-3′trnK*, and *trnS-rpS4*. Because several of the more variable regions are often adjacent to other surveyed regions, they are easily coamplifiable and may be amplified and sequenced with little to no additional cost. As we have shown, we can amplify a combined *trnS-trnG-trnG* fragment (Tier 1 + Tier 2 region) to yield a fragment that exhibits the greatest number of PICs per two sequencing reactions as compared to other noncoding cpDNA choices. Other combinable regions that yield relatively high PIC values include *ycf6-psbM-trnD*, *trnC-ycf6-psbM*, and *rpS4-trnT-trnL* (Fig. 7). Because these combined regions include small internal coding regions, internal primers already exist and may be used in cases where sequencing is difficult. The five shorter, uncombined Tier 1 regions may provide a more cost-efficient alternative if your sequencing reactions yield confidently useable reads of <700 bp.

An additional important finding of this work is that a preliminary three-species survey can be used to determine the relative utility of a given region prior to implementing a full scale sequencing project. Such a full-scale assault, blindly choosing a reportedly useful region which may yield little information, may be a risky and costly venture in terms of time and resources. While the regions we identify as Tier 1 are consistently among the most variable among all lineages we

tested, there is still variation within and among lineages with respect to phylogenetic utility of the regions. A pilot study employing a small number (e.g., three in this study) of taxa can quickly identify which particular Tier 1 region may likely be the most informative within a particular species group.

In summary, the data we present indicate that there is indeed phylogenetically significant and predictable rate heterogeneity among noncoding cpDNA regions. While considerable variation exists among lineages, several noncoding cpDNA regions are identified that consistently provide greater levels of sequence variation compared to other regions that consistently yield low levels of variation, such as the commonly employed *trnL-trnL-trnF* and *trnK*/*matK*. More phylogenetically informative variation appears to be present in the chloroplast genome than previously thought, based on the accumulated evidence from a small number of apparently more slowly evolving noncoding regions. In addition, we show the importance and applicability of performing pilot studies to identify appropriate regions for further study. A small survey with as few as three species can be predictive of the overall levels of variation likely to be found in a larger scale study. The application of the top tier regions we have identified for future infrageneric studies and the continued exploration of noncoding regions of the chloroplast genome for variable markers are warranted.

Relative to the cpDNA genome, comparatively few noncoding cpDNA regions were surveyed in this study, and unsampled cpDNA regions may be found that yield a greater number of PICs than any of the top tier regions of this study. Therefore, we are currently in the midst of a companion study where we are adding data from other noncoding cpDNA regions to this data set in continuation of our search for the ''holy grail.''

LITERATURE CITED

AAGESEN, L., AND A. M. SANSO. 2003. The phylogeny of the Alstroemeriaceae, based on morphology, *rps16* intron, and *rbcL* sequence data. *Systematic Botany* 28: 47–69.

ALDRICH, J., B. W. CHERNEY, E. MERLIN, AND L. CHRISTOPHERSON. 1988. The role of insertions/deletions in the evolution of the intergenic region between *psbA* and *trnH* in the chloroplast genome. *Current Genetics* 14: 137–146.

ÁLVAREZ, I., AND J. F. WENDEL. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417–434.

ANDERSSON, C. L., AND J. H. E. ROVA. 1999. The *rpS16* intron and the phylogeny of the Rubioideae (Rubiaceae). *Plant Systematics and Evolution* 214: 161–186.

ANDERSSON, C. L., AND M. W. CHASE. 2001. Phylogeny and classification of Marantaceae. *Botanical Journal of the Linnean Society* 135: 275–287.

AOKI, K., T. SUZUKI, AND N. MURAKAMI. 2003. Intraspecific sequence variation of chloroplast DNA among the component species of evergreen broad-leaved forests in Japan. *Journal of Plant Research* 116: 337–344.

APG II. 2003. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141: 399–436.

APPLEQUIST, W. L., AND R. S. WALLACE. 2000. Phylogeny of the Madagascan endemic family Didiereaceae. *Plant Systematics and Evolution* 221: 157–166.

ASMUSSEN, C. B., AND M. W. CHASE. 2001. Coding and noncoding plastid DNA in palm systematics. *American Journal of Botany* 88: 1103–1117.

AZUMA, H., J. G. GARCÍA-FRANCO, V. RICO-GRAY, AND L. B. THIEN. 2001. Molecular phylogeny of the Magnoliaceae: the biogeography of tropical and temperate disjunctions. *American Journal of Botany* 88: 2275–2285.

AZUMA, H., L. B. THIEN, AND S. KAWANO. 1999. Molecular phylogeny of *Magnolia* (Magnoliaceae) inferred from cpDNA sequences and evolutionary divergence of the floral scents. *Journal of Plant Research* 112: 291–306.

BAKER, W. J., T. A. HEDDERSON, AND J. DRANSFIELD. 2000. Molecular phylogenetics of subfamily Calamoideae (Palmae) based on nrDNA ITS and cpDNA *rpS16* intron sequence data. *Molecular Phylogenetics and Evolution* 14: 195–217.

BALDWIN, B. G. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Molecular Phylogenetics and Evolution* 1: 3–16.

BALDWIN, B. G., M. J. SANDERSON, J. M. PORTER, M. F. WOJCIECHOWSKI, C. S. CAMPBELL, AND M. J. DONOGHUE. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247–277.

BAUM, D. A., R. L. SMALL, AND J. F. WENDEL. 1998. Biogeography and floral evolution of baobabs (*Adansonia*, Bombacaceae) as inferred from multiple data sets. *Systematic Biology* 47: 181–207.

BAUMEL, A., M. L. AINOUCHE, AND J. E. LEVASSEUR. 2001. Molecular investigations in populations of *Spartina anglica* C. E. Hubbard (Poaceae) invading coastal Brittany (France). *Molecular Ecology* 10: 1689–1701.

BAYER, R. J., D. G. GREBER, AND N. H. BAGNALL. 2002. Phylogeny of Australian Gnaphalieae (Asteraceae) based on chloroplast and nuclear sequences, the *trnL* intron, *trnL-trnF* intergenic spacer, *matK*, and ETS. *Systematic Botany* 27: 801–814.

BELL, C. D., AND R. W. PATTERSON. 2000. Molecular phylogeny and biogeography of *Linanthus* (Polemoniaceae). *American Journal of Botany* 87: 1857–1870.

BELLSTEDT, D. U., H. P. LINDER, AND E. H. HARLEY. 2001. Phylogenetic relationships in *Disa* based on noncoding *trnL-trnF* chloroplast sequences: evidence of numerous repeat regions. *American Journal of Botany* 88: 2088–2100.

BÖHLE, U.-R., H. HILGER, R. CERFF, AND W. F. MARTIN. 1994. Noncoding chloroplast DNA for plant molecular systematics at the infrageneric level. *In* B. Schierwater, B. Streit, G. P. Wagner, and R. DeSalle [eds.], Molecular ecology and evolution: approaches and applications, 391–403. Birkhäuser Verlag, Basel, Germany.

BORSCH, T., K. W. HILU, D. QUANDT, V. WILDE, C. NEINHUIS, AND W. BARTHLOTT. 2003. Noncoding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology* 16: 558–576.

BORTIRI, E., S. H. OH, J. G. JIANG, S. BAGGETT, A. GRANGER, C. WEEKS, M. BUCKINGHAM, D. POTTER, AND D. E. PARFITT. 2001. Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast *trnL-trnF* spacer DNA. *Systematic Botany* 26: 797–807.

BREMER, B., K. BREMER, N. HEIDARI, P. ERIXON, R. G. OLMSTEAD, A. A. ANDERBERG, M. KÄLLERSJÖ, AND E. BARKHORDARIAN. 2002. Phylogenetics of asterids based on 3 coding and 3 noncoding chloroplast DNA markers and the utility of noncoding DNA at higher taxonomic levels. *Molecular Phylogenetics and Evolution* 24: 274–301.

BUTTERWORTH, C. A., J. H. COTA-SANCHEZ, AND R. S. WALLACE. 2002. Molecular systematics of the tribe Cacteae (Cactaceae: Cactoideae): a phylogeny based on *rpl16* intron sequence variation. *Systematic Botany* 27: 257–270.

CAMPAGNA, M. L., AND S. R. DOWNIE. 1998. The intron in chloroplast gene *rpL16* is missing from the flowering plant families Geraniaceae, Goodeniaceae, and Plumbaginaceae. *Transactions of the Illinois State Academy of Science* 91: 1–11.

CHANDLER, G. T., R. J. BAYER, AND M. D. CRISP. 2001. A molecular phylogeny of the endemic Australian genus *Gastrolobium* (Fabaceae: Mirbelieae) and allied genera using chloroplast and nuclear markers. *American Journal of Botany* 88: 1675–1687.

CHASE, M. W., A. Y. DE BRUIJN, A. V. COX, G. REEVES, P. J. RUDALL, M. A. T. JOHNSON, AND L. E. EGUIARTE. 2000. Phylogenetics of Asphodelaceae (Asparagales): an analysis of plastid *rbcL* and *trnL-F* DNA sequences. *Annals of Botany* 86: 935–951.

CHASE, M. W., ET AL. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528–580.

CHASSOT, P., S. NEMOMISSA, Y.-M. YUAN, AND P. KÜPFER. 2001. High paraphyly of *Swertia* L. (Gentianaceae) in the *Gentianella*-lineage as revealed by nuclear and chloroplast DNA sequence variation. *Plant Systematics and Evolution* 229: 1–21.

CLARK, L. G., W. ZHANG, AND J. F. WENDEL. 1995. A phylogeny of the grass

family (Poaceae) based on *ndhF* sequence data. *Systematic Botany* 20: 436–460.

CLEGG, M. T., B. S. GAUT, G. H. LEARN, JR., AND B. R. MORTON. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences, USA* 91: 6795–6801.

CRANFILL, R. B. 2001. Phylogenetic utility of plastid ribosomal protein S4 (*rpS4*) in land plants. *American Journal of Botany* 87: 121 (Abstract).

CRONN, R. C., R. L. SMALL, T. HASELKORN, AND J. F. WENDEL. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *American Journal of Botany* 89: 707–725.

CUÉNOUD, P., M. A. DEL PERO MARTINEZ, P. A. LOIZEAU, R. SPICHIGER, S. ANDREWS, AND J. F. MANEN. 2000. Molecular phylogeny and biogeography of the genus *Ilex* L. (Aquifoliaceae). *Annals of Botany* 85: 111–122.

CUÉNOUD, P., V. SAVOLAINEN, L. W. CHATROU, M. POWELL, R. J. GRAYER, AND M. W. CHASE. 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL, atpB,* and *matK* DNA sequences. *American Journal of Botany* 89: 132–144.

DEMESURE, B., B. COMPS, AND R. PETIT. 1996. Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution* 50: 2515–2520.

DEMESURE, B., N. SODZI, AND R. J. PETIT. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* 4: 129–131.

DOWNIE, S. R., R. L. HARTMAN, S.-F. JIE, AND D. S. KATZ-DOWNIE. 2002. Polyphyly of the spring-parsleys (*Cymopterus*): molecular and morphological evidence suggests complex relationships among the perennial endemic genera of western North American Apiaceae. *Canadian Journal of Botany* 80: 1295–1324.

DOWNIE, S. R., AND D. S. KATZ-DOWNIE. 1999. Phylogenetic analysis of chloroplast *rps16* intron sequences reveals relationships within the woody southern African Apiaceae subfamily Apioideae. *Canadian Journal of Botany* 77: 1120–1135.

DOWNIE, S. R., D. S. KATZ-DOWNIE, AND K. J. CHO. 1996. Phylogenetic analysis of Apiaceae subfamily Apioideae using nucleotide sequences from the chloroplast *rpoC1* intron. *Molecular phylogenetics and Evolution* 6: 1–18.

DOWNIE, S. R., D. S. KATZ-DOWNIE, AND M. F. WATSON. 2000. A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA *rpl16* and *rpoC1* intron sequences: towards a suprageneric classification of subfamily Apioideae. *American Journal of Botany* 87: 273–292.

DOWNIE, S. R., AND J. D. PALMER. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. *In* Soltis et al. [eds.], Molecular systematics of plants, 14–35. Chapman and Hall, New York, New York, USA.

DOWNIE, S. R., G. M. PLUNKETT, M. F. WATSON, K. SPALIK, D. S. KATZ-DOWNIE, C. M. VALIEJO-ROMAN, E. I. TERENTIEVA, A. V. TROITSKY, B.-Y. LEE, AND A. EL-OQLAH. 2001. Tribes and clades within Apiaceae subfamily Apioideae: the contribution of molecular data. *Edinburg Journal of Botany* 58: 301–330.

DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

DOYLE, J. J., J. L. DOYLE, AND J. D. PALMER. 1995. Multiple independent losses of two genes and one intron from legume chloroplast genomes. *Systematic Botany* 20: 272–294.

DUMOLIN-LAPEGUE, S., M. H. PEMONGE, AND R. J. PETIT. 1997. An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology* 6: 393–397.

EDWARDS, K. J., AND P. A. GADEK. 2001. Evolution and biogeography of *Alectryon* (Sapindaceae). *Molecular Phylogenetics and Evolution* 20: 14–26.

EHRENDORFER, F., J. F. MANEN, AND A. NATALI. 1994. cpDNA intergene sequences corroborate restriction site data for reconstructing Rubiaceae phylogeny. *Plant Systematics and Evolution* 190: 245–248.

EL MOUSADIK, A., AND R. J. PETIT. 1996. Chloroplast DNA phylogeography of the Argan tree of Morocco. *Molecular Ecology* 5: 547–555.

FELINER, G. N., J. F. AGUILAR, AND J. A. ROSSELLO. 2002. Reticulation or divergence: the origin of a rare serpentine endemic assessed with chloroplast, nuclear, and RAPD markers. *Plant Systematics and Evolution* 231: 19–38.

FRIESEN, N., R. M. FRITSCH, S. POLLNER, AND F. R. BLATTNER. 2000. Molecular and morphological evidence for an origin of the aberrant genus *Milula* within the Himalayan species of *Allium* (Alliaceae). *Molecular Phylogenetics and Evolution* 17: 209–218.

FUKUDA, T., J. YOKOYAMA, AND H. TSUKAYA. 2003. Phylogenetic relationships among species in the genera *Chisocheton* and *Guarea* that have unique indeterminate leaves as inferred from sequences of chloroplast DNA. *International Journal of Plant Sciences* 164: 13–24.

GASKIN, J. F., AND B. A. SCHAAL. 2003. Molecular phylogenetic investigation of U.S. Invasive *Tamarix*. *Systematic Botany* 28: 86–95.

GE, S., A. LI, B. R. LU, S. Z. ZHANG, AND D. Y. HONG. 2002. A phylogeny of the rice tribe Oryzeae (Poaceae) based on *matK* sequence data. *American Journal of Botany* 89: 1967–1972.

GENE CODES CORP. 1998. Sequencher, version 3.1.1. Gene Codes, Ann Arbor, Michigan, USA.

GIELLY, L., AND P. TABERLET. 1994. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. *Molecular Biology and Evolution* 11: 769–777.

GOLDBLATT, P., V. SAVOLAINEN, O. PORTEOUS, I. SOSTARIC, M. POWELL, G. REEVES, J. C. MANNING, T. G. BARRACLOUGH, AND M. W. CHASE. 2002. Radiation on the Cape flora and the phylogeny of peacock irises *Moraea* (Iridaceae) based on four plastid DNA regions. *Molecular Phylogenetics and Evolution* 25: 341–360.

GOLENBERG, E. M., M. T. CLEGG, M. L. DURBIN, J. DOEBLEY, AND D. P. MA. 1993. Evolution of a noncoding region of the chloroplast genome. *Molecular Phylogenetics and Evolution* 2: 52–64.

HAHN, W. J. 2002. A phylogenetic analysis of the Arecoid line of palms based on plastid DNA sequence data. *Molecular Phylogenetics and Evolution* 23: 189–204.

HAMILTON, M. B. 1999a. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* 8: 521–523.

HAMILTON, M. B. 1999b. Tropical tree gene flow and seed dispersal: deforestation affects the genetic structure of the surviving forest fragments. *Nature* 401: 129–130.

HAMILTON, M. B., J. M. BRAVERMAN, AND D. F. SORIA-HERNANZ. 2003. Patterns and relative rates of nucleotide and insertion/deletion evolution at six chloroplast intergenic regions in New World species of the Lecythidaceae. *Molecular Biology and Evolution* 20: 1710–1721.

HARDIG, T. M., P. S. SOLTIS, AND D. E. SOLTIS. 2000. Diversification of the North American shrub genus *Ceanothus* (Rhamnaceae): conflicting phylogenies from nuclear ribosomal DNA and chloroplast DNA. *American Journal of Botany* 87: 108–123.

HARTMANN, S., J. D. NASON, AND D. BHATTACHARYA. 2002. Phylogenetic origins of *Lophocereus* (Cactaceae) and the senita cactus-senita moth pollination mutualism. *American Journal of Botany* 89: 1085–1092.

HENNEQUIN, S., A. EBIHARA, M. ITO, K. IWATSUKI, AND J. Y. DUBUISSON. 2002. Molecular systematics of the fern genus *Hymenophyllum s.l.* (Hymenophyllaceae) based on chloroplastic coding and noncoding regions. *Molecular Phylogenetics and Evolution* 27: 283–301.

HILU, K. W., AND H. LIANG. 1997. The *matK* gene: sequence variation and application in plant systematics. *American Journal of Botany* 84: 830–839.

HILU, K. W., T. BORSCH, K. MÜLLER, D. E. SOLTIS, P. S. SOLTIS, V. SAVOLAINEN, M. W. CHASE, M. P. POWELL, L. A. ALICE, R. EVANS, H. SAUQUET, C. NEINHUIS, T. A. B. SLOTTA, J. G. ROHWER, C. S. CAMPBELL, AND L. W. CHATROU. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90: 1758–1776.

HIRATSUKA, J., H. SHIMADA, R. WHITTIER, T. ISHIBASHI, M. SAKAMOTO, M. MORI, C. KONDO, Y. HONJI, C. R. SUN, B. Y. MENG, Y. Q. LI, A. KANNO, Y. NISHIZAWA, A. HIRAI, K. SHINOZAKI, AND M. SIGIURA. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Molecular General Genetics* 217: 185–194.

HODGES, S. A., AND M. L. ARNOLD. 1994. Columbines: a geographically widespread species flock. *Proceedings of the National Academy of Sciences, USA* 91: 5129–5132.

HOLDREGGER, R., AND R. J. ABBOTT. 2003. Phylogeography of the Arctic-Alpine *Saxifraga oppositifolia* (Saxifragaceae) and some related taxa based on cpDNA and ITS sequence variation. *American Journal of Botany* 90: 931–936.

HOOT, S. B., A. CULHAM, AND P. R. CRANE. 1995. The utility of *atpB* gene sequences in resolving phylogenetic relationships: comparison with *rbcL*

and 18S ribosomal DNA sequences in the Lardizabalaceae. *Annals of the Missouri Botanical Garden* 82: 194–208.

HU, J.-M., M. LAVIN, M. F. WOJCIECHOWSKI, AND M. J. SANDERSON. 2000. Phylogenetic systematics of the tribe Millettieae (Leguminosae) based on chloroplast *trnK/matK* sequences and its implications for evolutionary patterns in Papilionoideae. *American Journal of Botany* 87: 418–430.

HUFFORD, L., M. M. MCMAHON, A. M. SHERWOOD, G. REEVES, AND M. W. CHASE. 2003. The major clades of Loasaceae: phylogenetic analysis using the plastid *matK* and *trnL-F* regions. *American Journal of Botany* 90: 1215–1228.

INGRAM, A. L., AND J. J. DOYLE. 2003. The origin and evolution of *Eragrostis tef* (Poaceae) and related polyploids: evidence from nuclear *waxy* and plastid *rps16*. *American Journal of Botany* 90: 116–122.

JANSEN, R. K., AND J. D. PALMER. 1987. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proceedings of the National Academy of Sciences, USA* 84: 5818–5822.

JENSEN, U., S. B. HOOT, J. T. JOHANSSON, AND K. KOSUGE. 1995. Systematics and phylogeny of the Ranunculaceae—a revised family concept on the basis of molecular data. *Plant Systematics and Evolution* Suppl. 9: 273–280.

JOBSON, R. W., J. PLAYFORD, K. M. CAMERON, AND V. A. ALBERT. 2003. Molecular phylogenetics of Lentibulariaceae inferred from plastid *rpS16* intron and *trnL-F* DNA sequences: implications for character evolution and biogeography. *Systematic Botany* 28: 157–171.

JOHNSON, L. A., AND D. E. SOLTIS. 1994. *matK* DNA sequence and phylogenetic reconstruction in Saxifragaceae *s.s. Systematic Botany* 19: 143–156.

JORDAN, W. C., M. W. COURTNEY, AND J. E. NEIGEL. 1996. Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeeds (Lemnaceae). *American Journal of Botany* 83: 430–439.

KAJITA, T., K. KAMIYA, K. NAKAMURA, H. TACHIDA, R. WICKNESWARI, Y. TSUMURA, H. YOSHIMARU, AND T. YAMAZAKI. 1998. Molecular phylogeny of Dipterocarpaceae in southeast Asia based on nucleotide sequences of *matK, trnL* intron, and the *trnL-trnF* intergenic spacer region in chloroplast DNA. *Molecular Phylogenetics and Evolution* 10: 202–209.

KAZEMPOUR OSALOO, S., F. H. UTECH, M. OHARA, AND S. KAWANO. 1999. Molecular systematics of Trilliaceae I. Phylogenetic analyses of *Trillium* using *matK* gene sequences. *Journal of Plant Research* 112: 35–49.

KELCHNER, S. A. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* 87: 482–498.

KELCHNER, S. A. 2002. Group II introns as phylogenetic tools: structure, function, and evolutionary contraints. *American Journal of Botany* 89: 1651–1669.

KELCHNER, S. A., AND L. G. CLARK. 1997. Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). *Molecular Phylogenetics and Evolution* 8: 385–397.

KIM, K. J., AND R. K. JANSEN. 1995. *ndhF* sequence evolution and the major clades in the sunflower family. *Proceedings of the National Academy of Sciences, USA* 92: 10379–10383.

KIMBALL, R. T., D. J. CRAWFORD, AND E. B. SMITH. 2003. Evolutionary processes in the genus *Coreocarpus*: insights from molecular phylogenetics. *Evolution* 57: 52–61.

KIMURA, T., H. IKETANI, K. KOTOBUKI, N. MATSUTA, Y. BAN, T. HAYASHI, AND T. YAMAMOTO. 2003. Genetic characterization of pear varieties revealed by chloroplast DNA sequences. *Journal of Horticultural Science and Biotechnology* 78: 241–247.

KLAK, C., T. A. HEDDERSON, AND P. LINDER. 2003. A molecular systematic study of the *Lampranthus* group (Aizoaceae) based on the chloroplast *trnT-trnF* and nuclear ITS and 5S NTS sequence data. *Systematic Botany* 28: 70–85.

KUSUMI, J., Y. TSUMURA, H. YOSHIMARU, AND H. TACHIDA. 2000. Phylogenetic relationships in Taxodiaceae and Cupressaceae sensu stricto based on *matK* gene, *chlL* gene, *trnL-trnF* IGS region, and *trnL* intron sequences. *American Journal of Botany* 87: 1480–1488.

LEE, C., AND J. WEN. 2004. Phylogeny of *Panax* using chloroplast *trnC-trnD* intergenic region and the utility of *trnC-trnD* in interspecific studies of plants. *Molecular Phylogenetics and Evolution* 31: 894–903.

LEE, J., AND T. HYMOWITZ. 2001. A molecular phylogenetic study of the subtribe Glycininae (Leguminosae) derived from the chloroplast DNA *rps16* intron sequences. *American Journal of Botany* 88: 2064–2073.

LES, D. H., D. J. CRAWFORD, E. LANDOLT, J. D. GABEL, AND R. T. KEMBALL. 2002. Phylogeny and systematics of Lemnaceae, the duckweed family. *Systematic Botany* 27: 221–240.

LEVINSON, G., AND G. A. GUTMAN. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4: 203–221.

LU, S.-Y., C. PENG, Y.-P. CHENG, K.-H. HONG, AND T.-Y. CHIANG. 2001. Chloroplast DNA phylogeography of *Cunninghamia konishii* (Cupressaceae), an endemic conifer of Taiwan. *Genome* 44: 797–807.

MANEN, J. F., A. NATALI, AND F. EHRENDORFER. 1994. Phylogeny of Rubiaceae-Rubieae inferred from the sequence of a cpDNA intergene region. *Plant Systematics and Evolution* 190: 195–211.

MAST, A. R., AND T. J. GIVNISH. 2002. Historical biogeography and the origin of stomatal distributions in *Banksia* and *Dryandra* (Proteaceae) based on their cpDNA phylogeny. *American Journal of Botany* 89: 1311–1323.

MAST, A. R., S. KELSO, A. J. RICHARDS, D. J. LANG, D. M. S. FELLER, AND E. CONTI. 2001. Phylogenetic relationships in *Primula* L. and related genera (Primulaceae) based on noncoding chloroplast DNA. *International Journal of Plant Sciences* 162: 1381–1400.

MAYER, V., M. MOLLER, M. PERRET, AND A. WEBER. 2003. Phylogenetic position and generic differentiation of Epithemateae (Gesneriaceae) inferred from plastid DNA sequence data. *American Journal of Botany* 90: 321–329.

MES, T. H. M., N. FRIESEN, R. M. FRITSCH, M. KLAAS, AND K. BACHMANN. 1997. Criteria for sampling in *Allium* based on chloroplast DNA PCR-RFLP's. *Systematic Botany* 22: 701–712.

MILLER, J. T., AND R. J. BAYER. 2001. Molecular phylogenetics of *Acacia* (Fabaceae: Mimosoideae) based on the chloroplast *matK* coding sequence and flanking *trnK* intron spacer regions. *American Journal of Botany* 88: 697–705.

MILLER, J. T., J. W. GRIMES, D. J. MURPHY, R. J. BAYER, AND P. Y. LADIGES. 2003. A phylogenetic analysis of the Acacieae and Ingeae (Mimosoideae: Fabaceae) based on *trnK, matK, psbA-trnH,* and *trnL/trnF* sequence data. *Systematic Botany* 28: 558–566.

MORT, M. E., D. E. SOLTIS, P. S. SOLTIS, O. J. FRANCISCO, AND A. SANTOS-GUERRA. 2001. Phylogenetics and evolution of the Macaronesian clade of Crassulaceae inferred from nuclear and chloroplast sequence data. *Systematic Botany* 27: 271–288.

MORTON, B. R., AND M. T. CLEGG. 1993. A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbcL* in the grass family (Poaceae). *Current Genetics* 24: 357–365.

MUELLNER, A. N., R. SAMUEL, S. A. JOHNSON, M. CHEEK, T. D. PENNINGTON, AND M. W. CHASE. 2003. Molecular phylogenetics of Meliaceae (Sapindales) based on nuclear and plastid DNA sequences. *American Journal of Botany* 90: 471–480.

MUMMENHOFF, K., H. BRÜGGEMANN, AND J. L. BOWMAN. 2001. Chloroplast DNA phylogeny and biogeography of *Lepidium* (Brassicaceae). *American Journal of Botany* 88: 2051–2063.

NAGANO, Y., R. MATSUNO, AND S. SASAKI. 1991. Sequence and transcriptional analysis of the gene cluster *trnQ-zfpA-psaI-ORF231-petA* in pea chloroplasts. *Current Genetics* 20: 431–436.

NEUHAUS, H., AND G. LINK. 1987. The chloroplast tRNA Lys (UUU) gene from mustard (*Sinapis alba*) contains a class II intron potentially coding for a maturase-related polypeptide. *Current Genetics* 11: 251–257.

NYFFELER, R. 2002. Phylogenetic relationships in the cactus family (Cactaceae) based on evidence from *trnK/matK* and *trnL-trnF* sequences. *American Journal of Botany* 89: 312–326.

O'DONNELL, K. 1992. Ribosomal DNA internal transcribed spacers are highly divergent in the phytopathogenic ascomycete *Fusarium sambucinum* (*Gibberella pulicaris*). *Current Genetics* 22: 213–220.

OHSAKO, T., AND O. OHNISHI. 2000. Intra- and interspecific phylogeny of wild *Fagopyrum* (Polygonaceae) species based on nucleotide sequences of noncoding regions in chloroplast DNA. *American Journal of Botany* 87: 573–582.

OHSAKO, T., AND O. OHNISHI. 2001. Nucleotide sequence variation of the chloroplast *trnK/matK* region in two wild *Fagopyrum* (Polygonaceae) species, *F. leptopodum* and *F. statice*. *Genes and Genetic Systems* 76: 39–46.

OHYAMA, K., H. FUKUZAWA, T. KOHCHI, H. SHIRAI, T. SANO, S. SANO, K. UMESONO, Y. SHIKI, M. TAKEUCHI, Z. CHANG, S. AOTA, H. INOKUCHI, AND H. OZEKI. 1986. Chloroplast gene organization deduced from com-

plete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322: 522–574.

OLMSTEAD, R. G., AND J. D. PALMER. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany* 81: 1205–1224.

OLMSTEAD, R. G., AND P. A. REEVES. 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbcL* and *ndhF* sequences. *Annals of the Missouri Botanical Garden* 82: 176–193.

OLMSTEAD, R. G., AND J. A. SWEERE. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Systematic Biology* 43: 467–481.

OLSON, M. E. 2002a. Intergeneric relationships within the Caricaceae-Moringaceae clade (Brassicales) and potential morphological synapomorphies of the clade and its families. *International Journal of Plant Sciences* 163: 51–65.

OLSON, M. E. 2002b. Combining data from DNA sequences and morphology for a phylogeny of Moringaceae (Brassicales). *Systematic Botany* 27: 55–73.

OXELMAN, B., M. LIDÉN, AND D. BERGLUND. 1997. Chloroplast *rps16* intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Systematics and Evolution* 206: 393–410.

PACAK, A., AND Z. SZWEYKOWSKA-KULINSKA. 2000. Molecular data concerning alloploid character and the origin of chloroplast and mitochondrial genomes in the liverwort species *Pellia borealis*. *Journal of Plant Biotechnology* 2: 101–108.

PALMER, J. D. 1991. Plastid chromosomes: structure and evolution. *Cell Culture and Somatic Cell Genetics of Plants* 7A: 5–53.

PANERO, J. L., AND B. S. CROZIER. 2003. Primers for PCR amplification of Asteraceae chloroplast DNA. *Lundellia* 6: 1–9.

PEDERSEN, N., AND L. HEDENÄS. 2003. Phylogenetic investigations of a well supported clade within the acrocarpous moss family Bryaceae: evidence from seven chloroplast DNA sequences and morphology. *Plant Systematics and Evolution* 240: 115–132.

PERRET, M., A. CHAUTEMS, R. SPICHIGER, G. KITE, AND V. SAVOLAINEN. 2003. Systematics and evolution of tribe Sinningieae (Gesneriaceae): evidence from phylogenetic analysis of six plastid DNA regions and nuclear *ncpGS*. *American Journal of Botany* 90: 445–460.

PFEIL, B. E., C. L. BRUBAKER, L. A. CRAVEN, AND M. D. CRISP. 2002. Phylogeny of *Hibiscus* and the tribe Hibisceae (Malvaceae) using chloroplast DNA sequences of *ndhF* and the *rpl16* intron. *Systematic Botany* 27: 333–350.

PIRES, J. C., AND K. J. SYTSMA. 2002. A phylogenetic evaluation of a biosystematic framework: *Brodiaea* and related petaloid monocots (Themidaceae). *American Journal of Botany* 89: 1342–1359.

PLUNKETT, G. M., D. E. SOLTIS, AND P. S. SOLTIS. 1996. Evolutionary patterns in Apiaceae: inferences based on *matK* sequence data. *Systematic Botany* 21: 477–495.

POPP, M., AND B. OXELMAN. 2001. Inferring the history of the polyploid *Silene aegaea* (Caryophyllaceae) using plastid and homoeologous nuclear DNA sequences. *Molecular Systematics and Evolution* 20: 474–481.

POSNO, M., A. VAN VLIET, AND G. S. GROOT. 1986. The gene for *Spirodela oligorhiza* chloroplast ribosomal protein homologous to *E. coli* ribosomal protein *L16* is split by a large intron near its 5′ end: structure and expression. *Nucleic Acids Research* 14: 3181–3195.

POTTER, D., F. GAO, S. BAGGETT, J. R. MCKENNA, AND G. H. MCGRANAHAN. 2002. Defining the sources of paradox: DNA sequence markers for North American walnut (*Juglans* L.) species and hybrids. *Scientia Horticulturae* 94: 157–170.

PRIDGEON, A. M., R. SOLANO, AND M. W. CHASE. 2001. Phylogenetic relationships in Pleurothallidinae (Orchidaceae): combined evidence from nuclear and plastid DNA sequences. *American Journal of Botany* 88: 2286–2308.

RENNER, S. S. 1999. Circumscription and phylogeny of the Laurales: evidence from molecular and morphological data. *American Journal of Botany* 86: 1301–1315.

RICHARDSON, J. E., M. F. FAY, Q. C. B. CRONK, D. BOWMAN, AND M. W. CHASE. 2000. A phylogenetic analysis of Rhamnaceae using *rbcL* and *trnL-F* sequences. *American Journal of Botany* 87: 1309–1324.

RITLAND, K., AND M. T. CLEGG. 1987. Evolutionary analysis of plant DNA sequences. *American Naturalist* 130: S74–S100.

SAKAI, M., A. KANAZAWA, A. FUJII, F. S. THSENG, J. ABE, AND Y. SHIMAMOTO. 2003. Phylogenetic relationships of the chloroplast genomes in the genus *Glycine* inferred from four intergenic spacer sequences. *Plant Systematics and Evolution* 239: 29–54.

SALAZAR, G. A., M. W. CHASE, M. A. S. ARENAS, AND M. INGROUILLE. 2003. Phylogenetics of Cranichideae with emphasis on Spiranthinae (Orchidaceae, Orchidoideae): evidence from plastid and nuclear DNA sequences. *American Journal of Botany* 90: 777–795.

SALTONSTALL, K. 2001. A set of primers for amplification of noncoding regions of chloroplast DNA in the grasses. *Molecular Ecology Notes* 1: 76–78.

SAMUEL, R., T. F. STUESSY, K. TREMETSBERGER, C. M. BAEZA, AND S. SILJAK-YAKOVLEV. 2003. Phylogenetic relationships among species of *Hypochaeris* (Asteraceae, Cichorieae) based on ITS, plastid *trnL* intron, *trnL-F* spacer, and *matK* sequences. *American Journal of Botany* 90: 496–507.

SANG, T., D. J. CRAWFORD, AND T. F. STUESSY. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84: 1120–1136.

SCHNABEL, A., AND J. F. WENDEL. 1998. Cladistic biogeography of *Gleditsia* (Leguminosae) based on *ndhF* and *rpl16* chloroplast gene sequences. *American Journal of Botany* 85: 1753–1765.

SCHÖNENBERGER, J., AND E. CONTI. 2003. Molecular phylogeny and floral evolution of the Penaeaceae, Oliniaceae, Rhynchocalycaceae, and Alzateaceae (Myrtales). *American Journal of Botany* 90: 293–309.

SCHULTHEIS, L. M. 2001. Systematics of *Downingia* (Campanulaceae) based on molecular sequence data: implications for floral and chromosome evolution. *Systematic Botany* 26: 603–621.

SCHÜTZE, P., H. FREITAG, AND K. WEISING. 2003. An integrated molecular and morphological study of the subfamily Suaedoideae Ulbr. (Chenopodiaceae). *Plant Systematics and Evolution* 239: 257–286.

SEELANAN, T., C. L. BRUBAKER, J. M. STEWART, L. A. CRAVEN, AND J. F. WENDEL. 1999. Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). *Systematic Botany* 24: 183–208.

SHAW, A. J. 2000. Phylogeny of the Sphagnopsida based on chloroplast and nuclear DNA sequences. *Bryologist* 103: 277–306.

SHAW, J., AND R. L. SMALL. 2004. Addressing the ''hardest puzzle in American pomology:'' phylogeny of *Prunus* sect. *Prunocerasus* (Rosaceae) based on seven noncoding chloroplast DNA regions. *American Journal of Botany* 91: 985–996.

SIMPSON, D. A., C. A. FURNESS, T. R. HODKINSON, A. M. MUASYA, AND M. W. CHASE. 2003. Phylogenetic relationships in Cyperaceae subfamily Mapanioideae inferred from pollen and plastid DNA sequence data. *American Journal of Botany* 90: 1071–1086.

SMALL, R. L., J. A. RYBURN, R. C. CRONN, T. SEELANAN, AND J. F. WENDEL. 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogenetic reconstruction in a recently diverged plant group. *American Journal of Botany* 85: 1301–1315.

SMITH, A. R., AND R. B. CRANFILL. 2002. Intrafamilial relationships of the thelypteroid ferns (Thelypteridaceae). *American Fern Journal* 92: 131–149.

SOLTIS, D. E., R. K. KUZOFF, M. E. MORT, M. ZANIS, M. FISHBEIN, L. HUFFORD, J. KOONTZ, AND M. K. ORROYO. 2001. Elucidating deep-level phylogenetic relationships in Saxifragaceae using sequences for six chloroplastic and nuclear DNA regions. *Annals of the Missouri Botanical Garden* 88: 669–693.

STECH, M., D. QUANDT, AND W. FREY. 2003. Molecular circumscription of the hornworts (Anthocerotophyta) based on the chloroplast DNA *trnL-trnF* region. *Journal of Plant Research* 116: 389–398.

STEELE, K. P., AND R. VILGALYS. 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. *Systematic Botany* 19: 126–142.

STEFANOVIC, S., L. KRUEGER, AND R. G. OLMSTEAD. 2002. Monophyly of the Convolvulaceae and circumscription of their major lineages based on DNA sequences of multiple chloroplast loci. *American Journal of Botany* 89: 1510–1522.

STEHLIK, I. 2002. Glacial history of the Alpine herb *Rumex nivalis* (Polygonaceae): a comparison of common phylogeographic methods with nested clade analysis. *American Journal of Botany* 89: 2007–2016.

STEHLIK, I., F. R. BLATTNER, R. HOLDREGGER, AND K. BACHMANN. 2002. Nunatak survival of the high Alpine plant *Eritrichium nanum* (L.) Gaudin in the central Alps during the ice ages. *Molecular Ecology* 11: 2027–2036.

TABERLET, P., L. GIELLY, G. PAUTOU, AND J. BOUVET. 1991. Universal prim-

ers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1109.

TATE, J. A., AND B. B. SIMPSON. 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany* 28: 723–737.

THOMPSON, J. D., D. G. HIGGINS, AND T. J. GIBSON. 2001. ClustalX. Program available at ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/.

TSUDZUKI, J., K. NAKASHIMA, T. TSUDZUKI, J. HIRATSUKA, M. SHIBATA, T. WAKASUGI, AND M. SUGIURA. 1992. Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ, trnK, psbA, trnI,* and *trnH* and the absence of *rpS16*. *Molecular and General Genetics* 232: 206–214.

VINNERSTEN, A., AND G. REEVES. 2003. Phylogenetic relationships within Colchicaceae. *American Journal of Botany* 90: 1455–1462.

WAKASUGI, T., J. TSUDZUKI, S. ITO, K. NAKASHIMA, T. TSUDZUKI, AND M. SUGIURA. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proceedings of the National Academy of Sciences, USA* 91: 9794–9798.

WAKASUGI, T., M. SUGITA, T. TSUDZUKI, AND M. SUGIURA. 1998. Updated gene map of tobacco chloroplast DNA. *Plant Molecular Biology Reporter* 16: 231–241.

WANG, X. R., Y. TSUMURA, H. YOSHIMARU, K. NAGASAKA, AND A. E. SZMIDT. 1999. Phylogenetic relationships of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL, matK, rpL20-rpS18* spacer, and *trnV* intron sequences. *American Journal of Botany* 86: 1742–1753.

WANNTORP, L., H. E. WANNTORP, B. OXELMAN, AND M. KALLERSJÖ. 2001. Phylogeny of *Gunnera*. *Plant Systematics and Evolution* 226: 85–107.

WESTHOF, E., AND F. MICHEL. 1996. The structural domains of group I introns. *In* R. Green and R. Schroeder [eds.], Ribosomal RNA and group I introns. Chapman and Hall, New York, New York, USA.

WINKWORTH, R. C., J. GRAU, A. W. ROBERTSON, AND P. J. LOCKHART. 2002. The origins and evolution of the genus *Myosotis* L. (Boraginaceae). *Molecular Phylogenetics and Evolution* 24: 180–193.

WOLF, P. G. 1997. Evaluation of *atpB* nucleotide sequences for phylogenetic studies of ferns and other pteridophytes. *American Journal of Botany* 84: 1429–1440.

WOLFE, K. H., W.-H. LI, AND P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054–9058.

WOLFE, K. H., C. W. MORDEN, AND J. D. PALMER. 1992. Function and evolution of a minimal plastid genome from a non-photosynthetic parasitic plant. *Proceedings of the National Academy of Sciences, USA* 89: 10648–10652.

XU, D. H., A. J. SAKAI, M. KANAZAWA, A. SHIMAMOTO, AND Y. SHIMAMOTO. 2000. Sequence variation of non-coding regions of chloroplast DNA of soybean and related wild species and its implications for the evolution of different chloroplast haplotypes. *Theoretical and Applied Genetics* 101: 724–732.

YAMANE, K., Y. YASUI, AND O. OHNISHI. 2003. Intraspecific cpDNA variations of diploid and tetraploid perennial buckwheat, *Fagopyrum cymosum* (Polygonaceae). *American Journal of Botany* 90: 339–346.

YANG, Y.-W., P.-Y. TAI, Y. CHEN, AND W.-H. LI. 2002. A study of the phylogeny of *Brassica rapa*, *B. nigra*, *Raphanus sativus*, and their related genera using noncoding regions of chloroplast DNA. *Molecular Phylogenetics and Evolution* 23: 268–275.

ZHANG, W. P. 2000. Phylogeny of the grass family (Poaceae) from *rpL16* intron sequence data. *Molecular Phylogenetics and Evolution* 15: 135–146.

ZUBER, D., AND A. WIDMER. 2000. Genetic evidence for host specificity in the hemi-parasitic *Viscum album* L. (Viscaceae). *Molecular Ecology* 9: 1069–1073.

ZURAWSKI, G., AND M. T. CLEGG. 1987. Evolution of higher plant chloroplast DNA-coded genes: implications for structure-function and phylogenetic studies. *Annual Review of Plant Physiology* 38: 391–418.