Systematic Botany (2004), 29(1): pp. 4–14
© Copyright 2004 by the American Society of Plant Taxonomists

Using RAPDs to Identify DNA Sequence Loci for Species Level Phylogeny Reconstruction: an Example from *Leucaena* (Fabaceae)

C. DONOVAN BAILEY,[1,2] COLIN E. HUGHES, and STEPHEN A. HARRIS

Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, United Kingdom

Communicating Editor: Lawrence A. Alice

ABSTRACT. Resolving phylogenies among closely related species remains a perplexing problem in plant systematics. All too often cpDNA and nrDNA ITS are insufficiently variable to provide desired resolution or support. In the present study, we have adopted a RAPD-based approach to develop sequence-characterized regions, from the nuclear genome, with levels of DNA sequence variation appropriate for resolving relationships within subclades of *Leucaena* (Fabaceae). RAPDs were used to amplify a set of seven *Leucaena* accessions. Equal length fragments amplified in two or more species were sequenced and compared. Specific primers were designed for aligned product sets displaying similar or greater levels of variation than have been found in previous ITS studies. Two regions whose DNA sequences provided greater resolution and bootstrap support than ITS or cpDNA RFLP restriction site data sets for the same sample of *Leucaena* species are discussed in detail.

DNA sequence data have played a central role in phylogeny reconstruction of plant taxa for more than 15 years. Characters for DNA-based analyses have been extracted from all three plant genomes, and these data have helped to resolve angiosperm phylogenies across higher taxa, among genera, and to a certain degree, at lower taxonomic levels (reviewed by Soltis and Soltis 1998). By far the most commonly used DNA sequences include those from the chloroplast genome (cpDNA) and multicopy nuclear-encoded ribosomal DNA (nrDNA: e.g., Taberlet et al. 1991; Olmstead et al. 1993; Gielly and Taberlet 1994; Olmstead and Palmer 1994; Baldwin et al. 1995; Kallersjo et al. 1998; Soltis and Soltis 1998; Hershkovitz et al. 1999; Soltis et al. 2000). Such data have been extensively applied to phylogenetic questions because of their relatively well-characterized nature, ease of amplification as a result of both conserved primer sites and high copy number, and diverse levels of variation found within sub-regions of each of these classes of DNA. While nrDNA and cpDNA sequences routinely provide sufficient characters to resolve higher level and generic relationships, low variation and resolution have been common problems encountered when attempting to resolve relationships among closely related plant species and even among morphologically diverse species representing large genera (e.g., Soltis and Soltis 1998; Richardson et al. 2001; Sang 2002). This lack of resolution has hindered studies of infrageneric classification, hybridization and polyploidy, character evolution, biogeography, and origins of domestication (e.g., Emshwiller and Doyle 1998; Wojciechowski et al. 1999; Richardson et al. 2001; Hughes et al. 2002; Mitchell and Heenan 2002).

Furthermore, cpDNA and nrDNA are of reduced utility when reconstructing relationships among potentially reticulating species. Although a number of angiosperms can transmit paternal plastids, the more common uniparental maternal inheritance of the chloroplast genome reduces its utility for hypothesizing hybrid origins or testing specific hypotheses of hybridization when used as a sole source of data. Similarly, nrDNA multicopy sequences that have undergone complete concerted evolution do not maintain direct evidence of biparental inheritance and therefore reticulation (e.g., Baldwin 1992; Baldwin et al. 1995; Wendel et al. 1995). Conversely, a lack of concerted evolution between multicopy nrDNA repeats can create serious paralogy-related problems for phylogeny reconstruction (Doyle 1992; Baldwin et al. 1995; Álvarez and Wendel 2003).

Several molecular strategies have been attempted to resolve relationships among closely related species where conventional cpDNA and nrDNA approaches have failed, including increasing the amount of cpDNA sequence data, attempting to use the potentially complex external transcribed spacer (e.g., Linder et al. 2000), or resorting to fragment length characters derived from RFLPs, AFLPs, or RAPDs (for reviews see: Soltis and Soltis 1998; Wolfe and Liston 1998; Harris 1999). At the same time, increased interest in reconstructing phylogenies from multiple independent data sources and disentangling divergent from reticulate relationships (e.g., Wendel and Doyle 1998) has prompted increasing use of nuclear encoded low-copy number or specific members of multi-gene families to resolve phylogenetic relationships (for general reviews see Soltis and Soltis 1998; Doyle and Doyle 1999; Sang 2002). Frequently these experiments include the use of intron sequences to identify regions with high variability flanked by conserved priming sites in the exons to generate universally amplifiable loci. While some of the loci examined have considerable potential within specific plant groups (e.g., Mason et al. 1998; Em-

[1] Present Address: Department of Biology, New Mexico State University, P.O. Box 30001 Dept 3AF, Las Cruces, New Mexico 88003-8001.

[2] Author for correspondence (e-mail: dbailey@nmsu.edu)

shwiller and Doyle 1999; Sang and Zhang 1999; Evans et al. 2000), none has shown promise as a low copy number highly variable region across the angiosperms (Sang 2002). The lack of universally applicable nuclear-encoded regions, aside from nrDNA, is likely to be the result of changes in gene copy number leading to potential paralogy-related problems, high rates of change in the nuclear genome reducing the possibility of identifying universal primer sites, and frequent nuclear genome rearrangement (e.g., Rieseberg and Soltis 1988; Lagercrantz 1998; Rieseberg and Buerkle 2002).

Thus, it has become clear that nuclear DNA sequences useful for resolving relationships among closely related species generally need to be developed specifically for a group of interest. In fact, this has been the primary purpose of many of the studies that have also discussed the potential broad utility of particular low copy number nuclear genes (e.g., Mathews et al. 1995; Denton et al. 1996; Doyle et al. 1996; Galloway et al. 1998; Mason et al. 1998; Small et al. 1998; Bailey and Doyle 1999; Emshwiller and Doyle 1999; Cronn et al. 2002).

In addition to the lack of universally applicable low copy number nuclear genes, there are also questions about the levels of variation that are likely to be found between such sequences. The results of Cronn et al. (2002) and the comparative data compiled from numerous studies by Álvarez and Wendel (2003) indicate that levels of intron variation are extremely labile and that nuclear-encoded introns may, more often than not, be less variable on a site per site basis than the nrDNA ITS region. The summary of data gathered by Álvarez and Wendel (2003) suggests that only one (*pistillata*) out of the 15 nuclear genes directly compared to nrDNA ITS provided greater percent variation than the ITS region, and this exception is potentially complicated by paralogy issues (J. Beck, pers. comm.; Bailey et al. 2002). At the very least these results suggest that screening of a range of possible low copy number nuclear genes will be necessary in order to identify amplifiable regions with sufficient variation (Cronn et al. 2002).

In this study, we have used commercially available RAPD primers to develop sequence characterized amplified regions (SCARs: Melotto et al. 1996) for *Leucaena* Benth. with the purpose of building gene trees from SCAR sequences. The strategy is a modification of the AFLP-based method of McLenachan et al. (2000) for characterizing and developing population level gel-based markers. RAPDs were selected over AFLPs because of their low cost, simplicity, and the availability of a large number of commercially available primers. While the use of RAPDs in phylogenetic analyses has been questioned because of potentially co-migrating non-homologous fragments (e.g., Rieseberg 1996; Adams and Rieseberg 1998; Harris 1999), this does not represent a serious issue for the present study. RAPD fragments were sequenced, aligned, and specific SCAR primers designed for potentially useful regions (an overview of the strategy is presented in Fig. 1). The results for two SCAR-based sequence matrices are presented and compared to cpDNA RFLP and nrDNA ITS variation for the same set of *Leucaena* species.

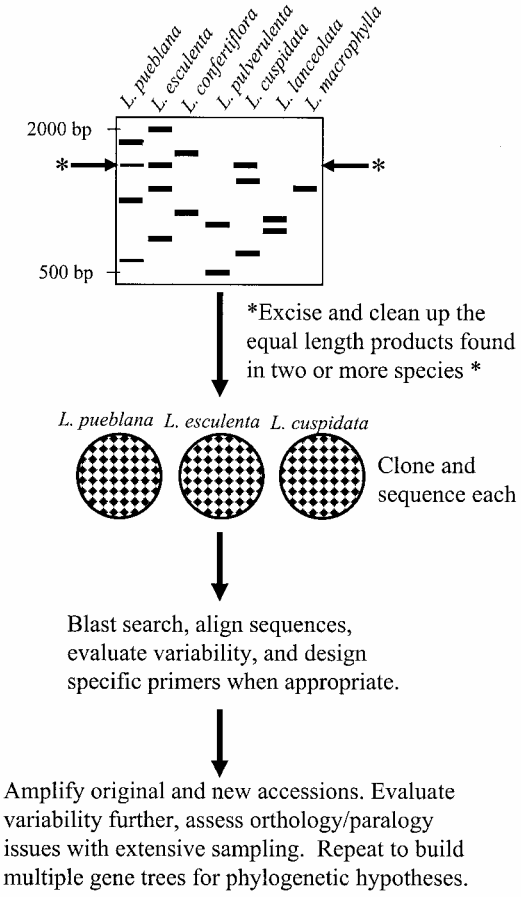

FIG. 1. An overview of the strategy used to develop highly variable loci for phylogeny reconstruction among closely related *Leucaena* species.

The examples presented here focus on resolving interspecific relationships within the genus *Leucaena*, which comprises 24 species native to the New World from Texas to northern South America (Hughes 1998) including five polyploids as well as two named and numerous putative hybrids. There is considerable evidence to suggest that hybridization and polyploidy, some of which has been precipitated by cultivation and incipient domestication, have played a major role in *Leucaena* evolution (Brewbaker and Sorensson 1990; Harris et al. 1994a; 1994b; Hughes and Harris 1994; Sorensson and Brewbaker 1994; Harris 1995; Hawkins and Harris 1998; Hughes 1998; Hughes and Harris 1998; Hughes et al. 2002). *Leucaena* phylogeny has been

investigated using data from morphology, RAPD, cpDNA RFLP, and nrDNA ITS sequences (e.g., Harris et al. 1994b; Hughes 1998; Hughes et al. 2002). While results from these studies strongly support the monophyly of *Leucaena* and the recognition of three clades within the genus, here referred to as the *L. lanceolata*, *L. esculenta*, and *L. pulverulenta* groups, relatively little or no resolution has been established within these clades. This lack of resolution has hindered our understanding of species relationships that bear heavily on the origins of the cultivated species, allopolyploids, and putative hybrids. Prior to embarking on the SCAR-based studies, variation was assessed from a number of non-coding cpDNA sequences and low copy number nuclear gene introns for a subset of *Leucaena* species. Potential problems were encountered for each of these (see below), prompting us to investigate the RAPD and SCAR based strategy for identifying potentially useful loci that we report here.

Material and Methods

An outline of the general approach described in detail below is presented in Fig. 1.

DNA Extraction. DNAs were extracted from fresh leaves of seed grown plants, herbarium specimens, or silica-gel-dried samples of field-collected leaf material (Table 1). DNA isolation followed a modified reduced volume CTAB technique (Doyle and Doyle 1987) and some samples were further purified using caesium chloride gradients (Maniatis et al. 1982). DNAs were resuspended in TE or water and stored at −20°C.

RAPD Amplification and Screening. A total of 53 RAPD primers (Operon Technology Inc.; B1–4,10; D-1,3–7,9–19; E 1, 3–11, 13–19; H 1–14) were screened using a single primer per reaction. The taxa sampled included accessions from each of the three previously characterized clades of *Leucaena* (Harris et al. 1994b; Hughes et al. 2002; Hughes et al. 2003) with one accession of each of the following species: *L. confertiflora* S. Zárate (*Hughes 1653*), *L. cuspidata* (*Hughes 1583*), *L. esculenta* (*Hughes 894*), *L. lanceolata* (*Hughes 872*), *L. macrophylla* (*Hughes 1179*), *L. pulverulenta* (*Hughes 1058*), and *L. pueblana* (*Hughes 1648*). Of the 53 original primers, 26 produced clear banding patterns with fragment classes shared between species. These were rerun in higher volume reactions for DNA recovery (below). An additional 16 reactions, each using a higher annealing and two RAPD primers simultaneously (D-13/A-10, B-10/D-05, A-10/D-05, D-13/D-055, B-01/B-03, B-04/B-08, B-08/B-10, E-06/B-10, E-06/B-08, E-04/B-04, E-01/E-03, E-04/E-06, E-06/E-01, E-04/E-01, E-01/E-15, E-15/E-16), were run to assess whether single or multiple RAPD primer reactions were likely to be more efficient for amplifying shared homologous fragment classes between ca. 500–2000 bp.

RAPD polymerase chain reactions (PCR) included ca. 1.5 units *Taq* (Qiagen, Crawley, West Sussex, UK), 165 μM of each dNTP, 1× PCR buffer, 1× Q solution, and 1.0 μM of a single primer or 0.5 μM of each primer in the double primer reactions per 25 μl of PCR mix. RAPD amplifications began with a three minute 94°C denaturation step, followed by 45 cycles of: 1) 15 sec. at 94°C; 2) 30 sec. annealing at 35°C for single primer reactions and 38°C for double primer reactions, and 3) a 1 min. 30 sec. 72°C extension. Following the cycling reactions each received a final 7 min. 72°C extension. Products were screened on 1.75% agarose gels to resolve fragments ranging from 500 to 2000 bp. Fragments of equal size generated using the same primer or primer combination in two or more species were cut from the gel and the DNA recovered using a Qiagen Gel Purification Kit. Recovered products were cloned using the Promega pGEM system (Promega, Madison, Wisconsin, USA) and sequenced using Big Dye V3 (Applied Biosystems, Inc., Warrington, UK) in both directions with M13F and M13R primers. Sequencing overlap was not necessarily achieved for products over ca. 1000 bp.

TABLE 1. Taxon sampling, voucher information, and GenBank accession numbers. Herbarium vouchers are all deposited in FHO, with duplicates variously at CAS, EAP, K, MEXU, NY, US and MO. [a] Collector of seed material, no herbarium voucher. * *L. multicapitula* accession for ITS is Hughes 1025 81/87 AF 418037.

Desmanthus fruticosus Rose—*Hughes 1532,* seedlot 109/92; nrDNA ITS AF418018, A9 N.A., 23L N.A.

Leucaena collinsii Britton & Rose subsp. *collinsii*—*Hughes 527,* seedlot 52/88; nrDNA ITS AF418020, A9 AY274286, 23L AY274272.

Leucaena collinsii Britton & Rose subsp. *zacapana* C.E. Hughes—*Hughes 1096,* seedlot 57/88; nrDNA ITS AF418023, A9 AY274287, 23L AY274273.

Leucaena cuspidata Standley—*Hughes 1583,* seedlot 89/92; nrDNA ITS AF418024, A9 AY274285, 23L AY274271.

Leucaena esculenta (Sessé & Mociño ex. DC.) Benth.—*Hughes 894,* seedlot 47/87; nrDNA ITS AF418096, A9 AY274288, 23L AY274274.

Leucaena greggii S. Watson—*Hughes 1057,* seedlot 82/87; nrDNA ITS AF418066, A9 AY274289, 23L AY274275.

Leucaena lanceolata S. Watson var. *sousae* (S. Zárate) C.E. Hughes—*Hughes 872,* seedlot 50/87; nrDNA ITS AF418051, A9 AY274290, 23L AY274276.

Leucaena lempirana C.E. Hughes—*Hughes 1447,* seedlot 5/91; nrDNA ITS AF418032, A9 AY274291, 23L AY274277.

Leucaena macrophylla Benth. subsp. *macrophylla*—*Hughes 1179,* seedlot 55/88; nrDNA ITS AF418034, A9 AY274292, 23L AY274278.

Leucaena matudae (S. Zárate) C.E. Hughes—*Hughes 879,* seedlot 49/87; nrDNA ITS AF418100, A9 AY274293, 23L AY274279.

**Leucaena multicapitula* Schery—*Hughes 1024,* seedlot 86/87; nrDNA ITS *AF418037, A9 AY274294, 23L AY274280.

Leucaena pueblana Britton & Rose—*Hughes 1648,* seedlot 125/92; nrDNA ITS AF418099, A9 AY274295 & AY274296, 23L AY274281.

Leucaena pulverulenta (Schltdl.) Benth.—*Hughes 1593*; nrDNA ITS AF418062, A9 AY274297, 23L AY274282.

Leucaena retusa Benth.—*XXX,* seedlot Bendeck[a] 23/86; nrDNA ITS AF418065, A9 AY274298, 23L AY274283.

Leucaena salvadorensis Standley ex Britton & Rose—*Hughes 1407,* seedlot 7/91; nrDNA ITS AF418038, A9 AY274299, 23L AY274284.

Sequence Alignment. Sequence fragments were edited and joined into contigs using Sequencher (Gene Codes, Ann Arbor, Michigan, USA). Complete sequences were provisionally aligned using ClustalX ver. 1.8 (Thompson et al. 1997) using the default parameters and adjusted by eye in WinClada (Nixon 1999).

Primer Design and Further Analysis. Pairwise comparisons of sequence variation were made following the sequencing of at least two, but preferably three or more products from two or more species per primer or primer pair. Aligned product sets displaying similar or greater levels of variation than identified in previous ITS studies (Hughes et al. 2002; Hughes et al. 2003) were selected for further consideration. For a subset of these, specific primers were designed from putatively ''conserved'' sites at the ends of the sequences found in the initial screen of *Leucaena* species (5′ and 3′ have little relevance in this context, the primers simply referred to as F and R, designated arbitrarily).

Sampling for the Comparative Matrices and Gene Trees Developed. Taxonomic sampling for the study included all three diploid species of both the *L. pulverulenta* and *L. esculenta* groups, and seven of the 12 diploids from the larger *L. lanceolata* group. The matrices also included *L. cuspidata*, bringing the total sampling

to 14 of the 17 diploid species of *Leucaena* and both subspecies of *L. collinsii* (Table 1). In addition, *Desmanthus fruticosus* was selected as an outgroup based on previous studies (Luckow et al. 2000; Hughes et al. 2002; Hughes et al. 2003).

The cpDNA and ITS matrices are subsets of previously published data sets (Harris et al. 1994b; Hughes et al. 2002). For each matrix, characters that were not potentially informative with the current subset taxon sampling were deactivated. The matrix and gene tree comparisons presented here include data taken from the same accessions with two exceptions (Table 1). In the cpDNA matrix, *L. pulverulenta* (Hughes 1593) was replaced by a composite terminal representing three *L. pulverulenta* accessions (Bendeck 22/86, Hughes 1051, and Hughes 1058), because the former was unavailable at the time of cpDNA analysis (Harris et al. 1994b). The three *L. pulverulenta* accessions in the cpDNA composite terminal all share identical RFLP profiles, suggesting little or no cpDNA RFLP variation within *L. pulverulenta*. The ITS matrix included *L. multicapitula* (Hughes 1025) rather than *L. multicapitula* (Hughes 1024) due to difficulties in amplifying and sequencing the latter for ITS.

PCR with the Newly Developed Primers. PCR with the newly developed primers contained ca. 1.5 units *Taq*, 100 μM of each dNTP, 1× PCR buffer, in some cases 1× Q solution (see below), and 0.5 μM of each primer per 25 μl of PCR mix. All amplifications began with a three minute 94°C denaturation step, followed by 35 rounds of: 1) 30 sec. 94°C denaturation; 2) 60 sec. annealing (temperatures dependent on the primers developed); and 3) 60–90 sec. 72°C extension. Products were cleaned using Qiagen Gel Extraction Kits for direct sequencing or cloning. Products were sequenced in both directions using the PCR primers and 'Big Dye V3' termination chemistry (Applied Biosystems Inc, Warrington UK). PCR band polymorphism or 'dirty' sequence traces for a few templates identified the potential for heterogeneous copy types. These products were cloned (pGEM; Promega Corporation, Madison, WI USA) using one half of the reaction volume described by the manufacturer. Individual clone strains were screened for the presence of an insert using the M13F and M13R primers prior to selecting clones for sequencing.

Phylogenetic Analysis. All characters were scored as unordered and equally weighted. Parsimony-based analyses were conducted with NONA (Goloboff 2000) spawned from WinClada (Nixon 1999) using 1000 random addition sequences, tree bisection and reconnection (TBR), holding 100 trees per replication, and attempting to swap to completion (h/100; mult*1000; max*). Contiguous gaps were scored as characters using the ''simple gap coding'' method formalized by Simmons and Ochoterena (2000). Sequences are available in GenBank (see Table 1) and matrices are deposited in TreeBASE (study accession number: S926; matrix accession numbers: M1529-M1532). Missing data in the A9, 23L, ITS, and cpDNA matrices constituted 6.0%, 4.7%, 8.0%, and 6.5% of the potentially informative characters respectively.

RESULTS

Twenty groups of fragments were cloned from the 42 sets of large volume RAPD reactions (26 from the single primer reactions and 16 from the double primer reactions). Seventeen of the 20 were successfully cloned from at least two species in order to make direct comparisons of divergence. Twelve of the 17 sets produced alignable sequences between species, while the other five showed no obvious sequence similarity between species.

BLAST searches (Altschul et al. 1990) were performed on all products. Of the 22 product types (12 represented by two or more clones and 10 by single clones for non-aligning fragment types), approximately half matched known plant nuclear encoded DNA.

A9 primers
A9F1 5'-GAG CTT GAA TCA GAT TCT GC-3'
A9F2 5'-TAA GAT CAC TGT GGA GCT TG-3'
A9F3 5'-ACA TTC TTA TTG GTT GAA TGT GT-3'
A9F4 5'-T(AG)T GGA GCT TGA ATA AGT AAA C-3'
A9R1 5'-AGG AAT CTG ATG AGC TCC AA-3'
A9R2 5'-CAA TTA TTG GAG ATT CCT AGA TC-3'
A9R3 5'- TGG AGG CAA CCC ATT GAT ATA G-3'
A9R4 5'-CAC TTC T(AG)C ATC ATT C(AC)A CGA TG-3'

23L primers
23LF1 5'-CTC CAT CAA TTA AGG GCA AAC-3'
23LR1 5'-GAC ATG CCT TAT AAC ACA AAT C-3'

FIG. 2. Schematic of A9 primer positions and primers sequences for A9 and 23L. The thick line region identifies the F4/R4 region used in the A9 matrix.

Characterized gene matches included, glyceraldehyde 3-phosphate dehydrogenase, chalcone synthase, alcohol dehydrogenase, helicase, glycerophosphodiesterase, isocytrate dehydrogenase, and starch synthase. The other products either had no matches or corresponded to uncharacterized portions of plant nuclear DNA. No matches longer than 22 bp were noted for non-plant DNA and no matches were noted for known mitochondrial or chloroplast DNA sequences.

Pairwise comparisons of divergence were performed on the 12 sets of alignable sequences. Five of the 12 regions showed little or no divergence between species while the other seven had levels of divergence similar to, or greater than that found with ITS (Hughes et al. 2002). Primers were subsequently designed at the ends of the seven alignments and two of these regions (A9 and 23L) were pursued in detail. The two regions pursued in detail were selected because of their ease of amplification and sequencing with the newly designed specific primers. Further work with the remaining regions would require additional primer development and/or significant cloning to reliably obtain data.

Region 23L. RAPD reactions with the primer combination OPE-03 and OPE-04 produced a size product of ca. 1000 bp in *L. pueblana* and *L. esculenta* that did not amplify in the other accessions. Sequences from the recovered products were aligned and primers 23LF1 and 23LR1 designed (Fig. 2). Following optimization to a 55°C annealing temperature with no Q solution or other PCR additives, the 23LF/23LR primer combination readily amplified single products across *Leucaena*. BLAST searches (Altschul et al. 1990) were performed using the full length *L. cuspidata* sequence. No significant matches resulted, although several 22 bp sections of *Arabidopsis thaliana* (L.) Heynh., *Homo sapiens* L., *Mus musculus* L., and *Plasmodium falciparum* DNA matched portions of the 23L fragment.

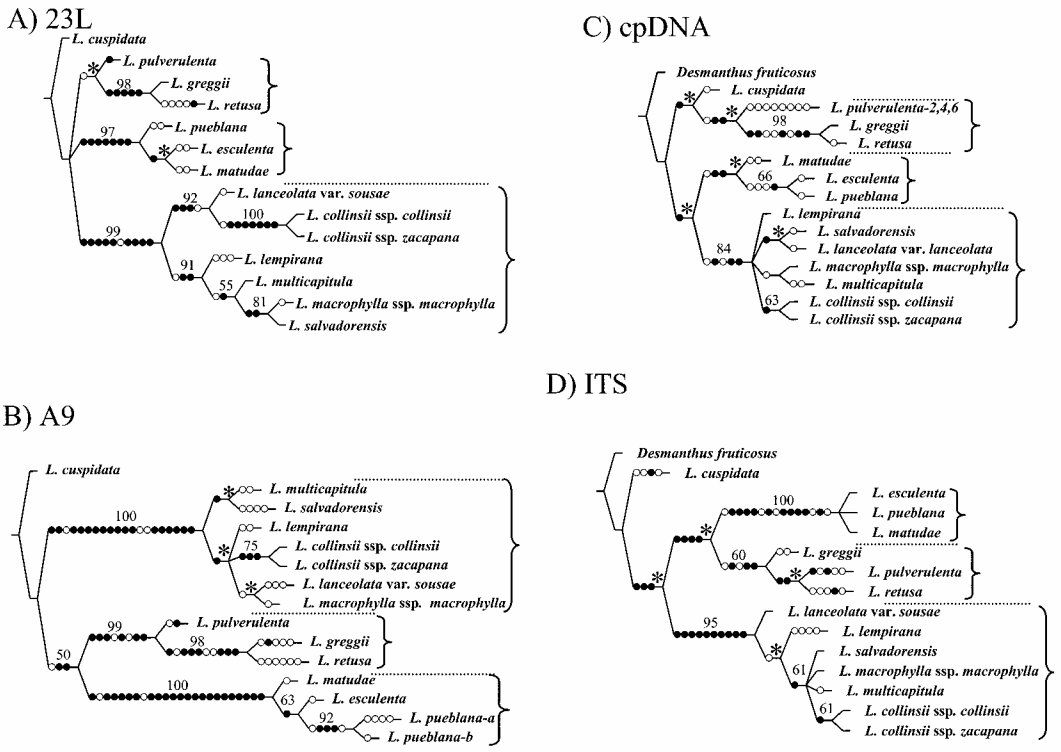

FIG. 3. Single most parsimonious gene trees. Solid and open circles along branches represent unique and homoplastic unambiguous character state transformations respectively. Values above branches are strict consensus bootstrap support values and "*" identify nodes that were not resolved in the strict consensus. a) 23L—one of four equally most parsimonious trees (EMPT); L=71; CI=0.80; RI=0.88. b) A9—one of six EMPT: L=132; CI=0.75; RI=0.88. c) cpDNA RFLP—one of 24 EMPT: L=63; CI=0.58; RI=0.69. d) ITS—one of six EMPT: L=100; CI=0.70; RI=0.81.

None of these minimal matches corresponded to a characterized gene sequence.

Sequenced products in *Leucaena* ranged from 723–798 bp long with GC contents ranging from 33.5–35.0%. The resulting 824 bp alignment included 40 potentially informative substitutions and 13 gap characters. Amplification and sequencing of *Desmanthus* outgroups, using the same primers, generated sequences ca. 630 bp long that could be completely aligned with each other, but not with the *Leucaena* sequences.

Phylogenetic analysis of the 23L *Leucaena* matrix, using the *L. cuspidata* sequence as the outgroup, resulted in four equally most parsimonious trees, one of which is presented in Fig. 3a (L=71; CI=0.80; RI=0.88). With the exception of *L. pulverulenta*, which was unresolved relative to the three main clades, resolution in the strict consensus included all three of the previously defined clades of *Leucaena* species (Harris et al. 1994b; Hughes et al. 2002; Hughes et al. 2003). Each of the three main clades had at least 97% bootstrap support and the strict consensus contains more resolution and support for nodes than the A9, ITS, or cpDNA trees despite the matrix containing fewer potentially informative characters than either the ITS or A9 matrices (Table 2).

A9 Region. RAPD reactions with the OPA-09 primer amplified a ca. 2000 bp fragment from *L. esculenta* and *L. pueblana* that was not amplified in the other accessions. Following cloning and sequencing of these products in both directions, which did not result in overlap due to the large size of the fragments, primers F1 and R1 (Fig. 2) were generated from the ends of the aligned sequences. PCR optimization to 59°C annealing without PCR additives amplified single band products in all *Leucaena* accessions except samples of the *L. lanceolata* clade, which failed to amplify under a variety of PCR conditions. Without sequence data from the *L. lanceolata* group, the assessment of possible paralogy/orthology issues within *Leucaena* would have been difficult. Therefore, additional primers were designed from the preliminary alignments of accessions representing *L. cuspidata*, the *L. esculenta* clade, and the *L. pulverulenta* clade. The new primers were designed from within the larger alignment of A9 so that subsequent amplifications included a portion of the full length A9 region. Primer combination F4/R4 (Fig. 2) optimized to a 55°C annealing temperature and 1% (w/v) Q solution produced single band amplifications from accessions representing all three clades and this is the region used in the present analysis (Fig. 2). *Leucaena esculenta* was amplified and sequenced with both

TABLE 2. Data comparison. [1] for cpDNA and ITS, the values in the lower portion of the cell were calculated without *D. fruticosus*. [2] calculations for the percentage of potentially informative sites in A9 excluded the autapomorphic *L. cuspidata* insertion from the total number of bases (i.e., the aligned length was treated as 665 bp rather than 1229 bp).

Region	Length (bp)	Gap Chars	% potentially informative sites	L; CI; RI	Aver. char. per branch	Nodes resolved in single MP trees	Nodes resolved in the consensus	Consensus nodes with ≥50% support
cpDNA	NA	NA	NA	63; 0.58; 0.69 [1]58; 0.63; 0.73	1.9 ± 2.3	9.6	5	4
ITS	638	11	8.6	100; 0.70; 0.81 [1]78; 0.79; 0.88	2.8 ± 3.6	8.3	5	5
23L	823	13	6.4	71; 0.80; 0.88	2.6 ± 2.7	9.75	8	8
A9	665 [2](1229)	11	13.2	132; 0.75; 0.88	4.4 ± 5.9	11	8	8

the A9-F1/R1 and A9-F4/R4 primer combinations. Sequences from the F4/R4 region of A9 were identical irrespective of the set of primers used to amplify the region. With the exception of *L. pueblana*, all sequences of the diploid species presented here were obtained by PCR and direct sequencing. Direct sequencing of *L. pueblana* resulted in overlapping traces. Clones of *L. pueblana* differed by two substitutions and a 6-bp indel in a 15-bp homopolymer cytosine repeat. BLAST searches (Altschul et al. 1990) were performed using the full-length *L. cuspidata* sequence with no significant matches.

The F4/R4 sequences of A9 ranged from 560–635 bp in length in all taxa except *L. cuspidata*, which has a 564 bp insertion not shared with any other taxa (overall length 1147 bp). The *L. cuspidata* indel was also noted in amplifications using alternative primer combinations (F1/R1, F3/R1, and F4/R4). GC content was 46% in each accession except *L. cuspidata*, which has 40.6% GC (46% in the overlapping region and 35% in the autapomorphic insertion). Attempts to amplify *Desmanthus* outgroups using the primers developed were unsuccessful.

The 1227 bp alignment (664 bp if the autapomorphic insertion in *L. cuspidata* is excluded) included 78 potentially informative substitutions and 11 gap characters. One of the six most parsimonious trees for the A9 region analysis is presented in Fig. 3b (L=132; CI=0.75; RI=0.88). Using *L. cuspidata* as the root, all three *Leucaena* clades are resolved and supported in the strict consensus. Considerable variation was found both within and between clades of *Leucaena* species. Furthermore, there were more nodes resolved and supported in the A9 strict consensus than in either the cpDNA RFLP or ITS-based trees (Table 2).

cpDNA and nrDNA ITS Phylogenies. Both the cpDNA and ITS analyses included a subset of terminals from previously published, more extensively sampled, matrices (Harris et al. 1994b; Hughes et al. 2002). The cpDNA matrix was originally developed from 84 probe-enzyme combinations representing roughly 85% of the chloroplast genome with 96 potentially informative sites identified. With the present sampling, the potentially informative variation was reduced to 37 sites. For the ITS region, 55 characters were potentially informative with the current sampling.

The strict consensus trees resulting from the analyses of these two data sets were entirely congruent with the previous results. However, less resolution was recovered with respect to the three previously described and well-supported clades of *Leucaena* species. Single cpDNA and ITS gene trees are presented in Figs. 3c and 3d for comparison of phylogenetic utility relative to the 23L and A9 regions. Comparative data generated for all four data sets and resulting trees are presented in Table 2.

Data Set Congruence. All three major clades of *Leucaena* species, or major components of all three clades, are resolved and supported in the four strict consensus trees. In the three instances where all three clades were not fully resolved (the cpDNA [2 cases] and band 23L [1 case] matrices), single individuals were unresolved relative to their respective clade. Thus these were not incongruent with the previous result (Hughes et al. 2002) and with the other data sets presented here. As a result of this high level of congruence at supported nodes between data sets and previous results, there was no justification for incorporating controversial incongruence tests (e.g., Davis et al. 1998; Dolphin et al. 2000).

DISCUSSION

The A9 and 23L matrices resolved more nodes with greater support than either cpDNA RFLP or ITS matrices. Furthermore, A9 provided greater percent variation than any of the other data sources. Although the 23L region contained slightly less percent variation than ITS, the added length in 23L resulted in more potentially informative characters and 23L resolved more nodes with greater support than the other matrices. The exclusion of the outgroup terminals from the ITS and cpDNA matrices facilitated a direct comparison of observed character conflict as measured by the ensemble consistency index and retention index (Kluge and Farris 1969; Farris 1989). The ITS, 23L, and

A9 all had similar levels of character conflict, while the cpDNA matrix had greater levels of conflict (Table 2).

The available data suggest that 23L and A9 contain more potentially informative characters and resulting resolution in phylogenetic trees than either the cpDNA RFLP or ITS matrices without increased character conflict. Furthermore, our interest in identifying resolution within clades of *Leucaena* species is borne out in the current results. The A9 region includes many potentially informative characters with noteworthy variation found among *L. esculenta* group sequences. In addition, unpublished results with a total of 72 *Leucaena* accessions (primers F1/R4) continue to support both between and within clade resolution (Bailey unpubl. data). Furthermore, the 23L matrix resolved more *L. lanceolata* group nodes with greater support than have been previously identified based on individual data sets or the combination cpDNA, ITS, and morphology (Hughes et al. 2002).

GC Content. The 33–35% GC content observed in sequences of the 23L region are consistent with base compositions for non-coding DNA (Initiative 2000). In contrast, the 46% GC content observed in A9 is more consistent with coding regions (Initiative 2000).

Single vs. Double Primer RAPD Amplifications. The three cloned fragment sets that showed no sequence similarity between species were all amplified using the single primer approach. The proportion of co-migrating unalignable and presumably non-homologous products is similar to those found in previous studies (e.g., Williams et al. 1993; Rieseberg 1996; Adams and Rieseberg 1998). Although our sample size is very small, it appears that equal length sequences sharing little or no sequence similarity were more commonly encountered when using a lower annealing temperature with a single RAPD primer per reaction. Thus, two or more RAPD primers in a single reaction with higher annealing temperatures is likely to provide a greater number of comparable fragments per RAPD reaction than the single primer lower annealing approach.

Sampling Versus Southern Blotting for Assessing Orthology. Broad taxonomic sampling and the development of multiple gene trees were used here as key strategies to understanding gene copy number issues and the potential impact on inferred species trees. While Southern blots are frequently used to estimate gene copy number, there are several concerns regarding the precision expected from these analyses. Sang (2002) noted that low specificity probes can result in an overestimation of gene copy number. In addition, Sang (2002) discussed difficulties in screening for low copy number genes via Southern blots because of the large quantities of high quality DNA needed. Presumably the difficulties of maximizing stringency and sensitivity simultaneously are what misled many studies of *A. thaliana* genes to hypothesize single copy number loci that were subsequently discovered to be multicopy when the *A. thaliana* genome sequence was completed (e.g., Initiative 2000).

Comparison to Low Copy Number Nuclear Sequences. Prior to embarking on the RAPD-based study, attempts were also made to amplify and characterize several low copy number nuclear genes from *Leucaena*. These included: 1) G3PDH (Strand et al. 1997); 2) glycerol-3-phosphate acyltransferase (GPAT: Tank and Sang 2001) using primer sites developed from the alignment of sequences from *Arabidopsis*, *Carthamus*, *Cucurbita*, *Paeonia*, and *Spinacea* (Bailey unpubl. data); 3) Histone H3D (Doyle et al. 1996); 4) nitrate reductase (Howarth and Baum 2002); and 5) *pistillata* intron (Bailey et al. 1997). In each case, difficulties in amplification (*pistillata* and Histone) or obvious paralogy-related problems were encountered (G3PDH, GPAT, and nitrate reductase). Paralogs were noted through the amplification of multiple fragments in diploid accessions (GPAT and nitrate reductase), and/or in the form of paralogous DNA sequences isolated from the same diploid individual (G3PDH, GPAT and nitrate reductase). Thus, considerable primer development (for *pistillata* and Histone H3) and/or characterization of paralogs (G3PDH, GPAT, and nitrate reductase) would have been necessary prior to inferring species trees from gene trees using any of these nuclear genes in *Leucaena* (Bailey unpubl. data).

Undoubtedly the multitude of genomic sequence information becoming available will streamline the potential use of predetermined sequence types (e.g., low copy number nuclear sequences) in phylogeny reconstruction. Nevertheless, the approach presented here remains a viable option until sufficient applicable genomic sequence information is available for a group of interest.

Comparison to cpDNA Sequences. In addition to comparing the A9 and 23L data to the previous cpDNA RFLP matrix and phylogeny, non-coding cpDNA sequence data were generated from 3–4 accessions representing each of the three *Leucaena* subclades. Evaluations of cpDNA variability were made for the: 1) *trnL* intron, *trnL* 3′ exon, and adjacent *trnL-F* spacer region (Taberlet et al. 1991); 2) the *trnK* intron region (Johnson and Soltis 1994); 3) the *psbA—trnH* region (Sang et al. 1997) and 4) the *rpl16* region using primers (Jordan et al. 1996) modified to match *Medicago* and *Lotus* sequences (Bailey unpubl. data). On average ca. 1% variable sites were identified (Bailey unpubl. data). If variable sites shared a 1:1 correlation with potentially informative sites, which they almost certainly underestimate (e.g., Cronn et al. 2002), ca. 8000 bp of non-coding cpDNA sequence would be required to build a matrix with the equivalent variation to A9. These results are consistent with previous direct com-

parisons of cpDNA sequence to ITS and some nuclear-encoded regions (e.g., Small et al. 1998; Bailey et al. 2002; Cronn et al. 2002; Olsen 2002).

A SCAR-Based Strategy as a Source of nDNA Sequences. The SCAR-based strategy presented here is comparable to the use of any other DNA sequence region in phylogeny reconstruction. However, the biparental inheritance patterns of nuclear loci make nuclear DNA a more powerful tool for reconstructing phylogenies (e.g., Hare 2001). Several factors suggest that A9 and 23L are both nuclear-encoded. First, several polyploid taxa and a few presumed heterozygous diploids required cloning to obtain clean sequences (Bailey unpubl. data). Second, BLAST searches using both A9 and 23L did not match the published complete angiosperm cpDNA or mtDNA sequences available in GenBank. Third, *L. mixtec*, a know sterile hybrid was recently cloned and sequenced for A9 and both parental types recovered (Bailey unpubl. data). Furthermore, none of the additional 20 RAPD fragments screened matched cpDNA or mtDNA regions. These results, along with the probabilities based on the respective size of the organellar genomes in comparison to the nuclear, suggest that the majority of these randomly amplified regions are nuclear encoded and therefore biparentally inherited (a point also supported by Li and Quiros 2001).

An additional benefit of this RAPD-based system is that no prior sequence information is necessary for potentially developing many orthologous loci in a group of interest. Universal RAPD primers can be used for the initial amplification and screening. Furthermore, the random amplification of regions across plant genomes eliminates the restriction of working within known genic regions (exons and introns), which are likely to represent the less variable half of the nuclear genome (e.g., in *A. thaliana* 53–57% of the DNA has been classified as non-coding DNA: Initiative, 2000). In addition, introns, which are generally more variable than exons, may often be both too short (nuclear introns average 168 bp in *A. thaliana*: [Initiative 2000]) and/or lack sufficient variation to provide resolution within and among closely related species. Future studies specifically evaluating levels of variation found in nuclear non-coding non-genic regions may provide important information on the levels of variation these types of non-coding regions can provide in phylogenetic studies.

Disadvantages of the SCAR-Based Strategy. Along with the potential strengths of the strategy discussed above, several limitations need to be considered. Most importantly, these include an increased potential for DNA contamination-related problems, difficulties amplifying more divergent taxa, and the lack of universally applicable primers. DNA contamination by endophytes (e.g., Liston and Alvarez-Buylla 1995; Camacho et al. 1997; Zhang et al. 1997) is a more serious issue for a RAPD-based strategy than with sequence-characterized nuclear genes, particularly relative to nuclear genes that are thought to be plant-specific. For the 23L and A9 regions BLAST searches provided no significant matches to published sequences. As a result, we cannot be certain that the DNA amplified is from *Leucaena* and not an endophyte (e.g., fungus or algae) or other potential contaminants. Although these contamination problems are an issue with any PCR-based method, they are perhaps more difficult to recognize with random DNA amplification (e.g., RAPD and AFLP). However, the low probability of developing multiple loci that produce congruent phylogenies for symbionts and parasites rather than the taxa of interest would presumably mitigate against this as a general issue (Álvarez and Wendel 2003). Nonetheless, these potential problems serve as a reminder of the critical nature of developing phylogenies from multiple independent sources (e.g., Doyle 1992; Page and Charleston 1997). For *Leucaena*, we have phylogenies generated from previous sources of data that are undoubtedly from the taxa of interest (cpDNA RFLP data generated from angiosperm-derived probes and ITS data) as well as multiple phylogenies being developed from SCARs. Futhermore, none of the 22 product types cloned in the RAPD screening matched non-plant DNA, suggesting that the RAPD fragments are from the taxa of interest.

Our reason for generating SCAR sequence data was to identify regions that provide as much, or more, variation as ITS for resolving relationships within previously characterized clades of *Leucaena* species. To accomplish this, RAPD fragments that amplified in a subset, but not all, screened accessions, were sampled. The use of this approach was based on the assumption that more variable regions would be represented by equal length fragments found in a subset of *Leucaena*. Thus, it was also assumed that less variable regions would be obtained from fragments that amplified an equal length product across all screened accessions. Although no direct comparison has been performed, six regions were isolated and two of these were sufficiently characterized to suggest that these were reasonable starting assumptions.

However, the limitation of this sampling strategy became evident when trying to amplify more divergent ingroup taxa with A9 primers, as well as outgroup taxa for both A9 and 23L. The lack of highly conserved motifs in these sequences greatly reduced the potential to amplify or align sequences from more divergent taxa (although alignment among *Leucaena* accessions was less problematic than with ITS). Similar difficulties were noted by Chase and colleagues (M. Chase, pers. comm.) while using the sequence-related amplified polymorphism strategy of Li and Quiros (2001) to ran-

domly amplify nuclear open reading frames. For the purpose of finding resolution within previously characterized clades, the sampling strategy and RAPD approach have worked. Nevertheless, had the objective been to test *Leucaena* monophyly relative to *Desmanthus*, or other mimosoid legumes, the exercise would have provided little information.

Finally, the SCAR-based data presented here are currently only of utility for interspecific *Leucaena* phylogeny reconstruction. They are of little or no use for studies of related groups of species or for broader scale insights into the molecular evolution of the SCARs amplified. However, future sequencing studies in other groups, particularly complete genome sequences of model legume species, may reveal more information about the regions used in this study. At that point, these data may lend themselves toward more universal utility in mimosoid legumes and provide information on the evolution of these sequence types across a broader range of taxa.

ACKNOWLEDGEMENTS. We would like to thank Jeff Doyle, an anonymous reviewer, and the editors for constructive comments on the manuscript. This study was supported by grants from the Leverhulme Trust and the Royal Society, as well as earlier funding from the United Kingdom Department for International Development.

LITERATURE CITED

ADAMS, R. P. and L. H. RIESEBERG. 1998. The effects of non-homology in RAPD bands on similarity and multivariate statistical ordination in *Brassica* and *Helianthus*. *Theoretical and Applied Genetics* 97: 323–326.

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

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

BAILEY, C. D. and J. J. DOYLE. 1999. Potential phylogenetic utility of the low-copy nuclear gene *pistillata* in dicotyledonous plants: comparison to nrDNA ITS and *trnL* intron in *Sphaerocardamum* and other Brassicaceae. *Molecular Phylogenetics and Evolution* 13: 20–30.

———, R. A. PRICE, and J. J. DOYLE. 2002. Systematics of the Halimolobine Brassicaceae: evidence from three loci and morphology. *Systematic Botany* 27: 318–332.

———, J. J. DOYLE, T. KAJITA, T. NEMOTO, and H. OHASHI. 1997. The chloroplast *rpl2* intron and ORF184 as phylogenetic markers in the legume tribe Desmodieae. *Systematic Botany* 22: 133–138.

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.

———, 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.

BREWBAKER, J. L. and C. T. SORENSSON. 1990. *Leucaena*. New tree crops from interspecific hybrids. Pp. 283–289 in *Advances in new crops*, eds. J. Janick and J. Simon. Portland: Timber Press.

CAMACHO, F. J., D. S. GERNANDT, A. LISTON, J. K. STONE, and A. S. KLEIN. 1997. Endophytic fungal DNA, the source of contamination in spruce needle DNA. *Molecular Ecology* 6: 983–987.

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.

DAVIS, J., M. P. SIMMONS, D. W. STEVENSON, and J. F. WENDEL. 1998. Data decisiveness, data quality and incongruence in phylogenetic analysis: an example from the monocotyledons using mitochondrial *atpA* sequences. *Systematic Biology* 47: 282–310.

DENTON, A. L., B. D. HALL, and B. L. MCCONAUGHY. 1996. *rpb2*, a nuclear gene for tracing angiosperm phylogeny. *American Journal of Botany* 83S: 150.

DOLPHIN, K., R. BELSHAW, C. D. L. ORME, and D. L. J. QUICKE. 2000. Noise and incongruence interpreting results of the incongruence length difference test. *Molecular Phylogenetics and Evolution* 17: 401–406.

DOYLE, J. J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany* 17: 144–163.

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

——— and ———. 1999. Nuclear protein-coding genes in phylogeny reconstruction and homology assessment: some examples from Leguminosae. Pp. 229–255 in *Molecular systematics and plant evolution*, eds. P. M. Hollingsworth, R. M. Bateman, and R. J. Gornall. London: The Systematics Association.

———, V. KANAZIN, and R. C. SHOEMAKER. 1996. Phylogenetic utility of histone H3 intron sequences in the perennial relatives of soybean (*Glycine*: Leguminosae). *Molecular Phylogenetics and Evolution* 6: 438–447.

EMSHWILLER, E. and J. J. DOYLE. 1998. Origins of domestication and polyploidy in oca (*Oxalis tuberosa* Oxalidaceae) nrDNA ITS data. *American Journal of Botany* 85: 975–985.

——— and ———. 1999. Chloroplast-expressed glutamine synthetase (*ncpGS*): potential utility for phylogenetic analysis with an example from *Oxalis* (Oxalidaceae). *Molecular Phylogenetics and Evolution* 12: 310–319.

EVANS, R. C., L. A. ALICE, C. S. CAMPBELL, E. A. KELLOGG, and T. A. DICKINSON. 2000. The granule-bound starch synthase (GBSSI) gene in the Rosaceae: multiple loci and phylogenetic utility. *Molecular Phylogenetics and Evolution* 17: 388–400.

FARRIS, J. S. 1989. The retention index and the rescaled consistency index. *Cladistics* 5: 417–419.

GALLOWAY, G. L., R. L. MALMBERG, and R. A. PRICE. 1998. Phylogenetic utility of the nuclear gene arginine decarboxylase: an example from Brassicaceae. *Molecular Biology and Evolution* 15: 1312–1320.

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

HARE, M. P. 2001. Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution* 16: 700–706.

HARRIS, S. A. 1995. Systematics and randomly amplified polymorphic DNA in the genus *Leucaena* (Leguminosae, Mimosoideae). *Plant Systematics and Evolution* 197: 195–208.

———. 1999. RAPDs in systematics—A useful methodology? Pp. 211–228 in *Molecular systematics and plant evolution*, eds. P. M. Hollingsworth, R. M. Bateman, and R. J. Gornall. London: The Systematics Association.

HARRIS, S. A., C. E. HUGHES, R. J. ABBOTT, and R. INGRAM. 1994a. Genetic variation in *Leucaena leucocephala* (Lam.) de Wit. (Leguminosae: Mimosoideae). *Silvae Genetica* 43: 159–167.

———, ———, R. INGRAM, and R. J. ABBOTT. 1994b. A phyloge-

netic analysis of *Leucaena* (Leguminosae Mimosoideae). *Plant Systematics and Evolution* 191: 1–26.

HAWKINS, J. A. and S. A. HARRIS. 1998. RAPD characterisation of two neotropical hybrid legumes. *Plant Systematics and Evolution* 213: 43–55.

HERSHKOVITZ, M. A., E. A. ZIMMER, and W. J. HAHN. 1999. Ribosomal DNA sequences and angiosperm systematics. Pp. 268–326 in *Molecular systematics and plant evolution,* eds. P. M. Hollingsworth, R. M. Bateman, and R. J. Gornall. London: Taylor & Francis.

HOWARTH, D. G. and D. A. BAUM. 2002. Phylogenetic utility of a nuclear intron from nitrate reductase for the study of closely related plant species. *Molecular Phylogenetics and Evolution* 23: 525–528.

HUGHES, C. E. 1998. Monograph of *Leucaena* (Leguminosae-Mimosoideae). *Systematic Botany Monographs* 55: 1–244.

——— and S. A. HARRIS. 1994. The characterization and identification of a naturally occurring hybrid in *Leucaena* Benth (Leguminosae Mimosoideae). *Plant Systematics and Evolution* 192: 177–197.

——— and ———. 1998. A second spontaneous hybrid in *Leucaena* Benth (Leguminosae Mimosoideae). *Plant Systematics and Evolution* 212: 53–77.

———, C. D. BAILEY, and S. A. HARRIS. 2002. Divergent and reticulate species relationships in *Leucaena* (Fabaceae) inferred from multiple data sources: insights into polyploid origins and nrDNA polymorphism. *American Journal of Botany* 89: 1057–1073.

———, ———, S. KROSNICK, and M. LUCKOW. 2003. Relationships among genera of the informal Dichrostachys and Leucaena groups (Mimosoideae) inferred from nuclear ribosomal ITS sequences. Pp. 221–238 in *Advances in legume systematics part 10,* eds. B. Klitgaard and A. Bruneau. Kew: Royal Botanic Gardens.

INITIATIVE, THE ARABIDOPSIS GENOME. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature London* 408: 796–815.

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

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

KALLERSJO, M., J. S. FARRIS, M. W. CHASE, B. BREMER, M. F. FAY, C. J. HUMPHRIES, G. PETERSEN, O. SEBERG, and K. BREMER. 1998. Simultaneous parsimony jackknife analysis of 2538 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants. *Plant Systematics and Evolution* 213: 259–287.

KLUGE, A. G. and J. S. FARRIS. 1969. Quantitative phyletics and the evolution of Anurans. *Systematic Zoology* 18: 1–32.

LAGERCRANTZ, U. 1998. Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* 150: 1217–1228.

LI, G. and C. QUIROS. 2001. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. *Theoretical and Applied Genetics* 103: 455–461.

LINDER, C. R., L. R. GOERTZEN, B. V. HEUVEL, J. FRANCISCO-ORTEGA, and R. K. JANSEN. 2000. The complete external transcribed spacer of 18S–26S rDNA: amplification and phylogenetic utility at low taxonomic levels in Asteraceae and closely allied families. *Molecular Phylogenetics and Evolution* 14: 285–303.

LISTON, A. and E. ALVAREZ-BUYLLA. 1995. Internal transcribed spacer sequences of conifers: ''there is a fungus among us''. *Inoculum* 46: 26.

LUCKOW, M., P. J. WHITE, and A. BRUNEAU. 2000. Relationships among the basal genera of mimosoid legumes. Pp. 165–180 in *Advances in legume systematics, Part 9,* eds. P. S. Herendeen and A. Bruneau. Kew: Royal Botanic Gardens.

MANIATIS, T., E. F. FRITSCH, and J. SAMBROOK. 1982. *Molecular cloning a laboratory manual.* New York: Cold Spring Harbor Laboratory.

MASON, G. R. J., C. F. WEIL, and E. A. KELLOGG. 1998. Granule-bound starch synthase: Structure, function, and phylogenetic utility. *Molecular Biology and Evolution* 15: 1658–1673.

MATHEWS, S., M. LAVIN, and R. A. SHARROCK. 1995. Evolution of the phytochrome gene family and its utility for phylogenetic analyses of angiosperms. *Annals of the Missouri Botanical Garden* 82: 296–321.

MCLENACHAN, P. A., K. STÖCKLER, R. C. WINKWORTH, K. MCBREEN, S. ZAUNER, and P. J. LOCKHART. 2000. Markers derived from amplified fragment length polymorphism gels for plant ecology and evolution studies. *Molecular Ecology* 9: 1899–1903.

MELOTTO, M., L. AFANADOR, and J. D. KELLY. 1996. Development of a SCAR marker linked to the I gene in common bean. *Genome* 39: 1216–1219.

MITCHELL, A. D. and P. B. HEENAN. 2002. *Sophora* sect. *Edwardsia* (Fabaceae): further evidence from nrDNA sequence data of a recent and rapid radiation around the Southern Oceans. *Botanical Journal of the Linnean Society* 140: 435–441.

NIXON, K. C. 1999. WinClada (Beta) version 09 Published by author. Ithaca, New York.

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.

———, B. BREMER, K. M. SCOTT, and J. D. PALMER. 1993. A parsimony analysis of the Asteridae sensu lato based on *rbcL* sequences. *Annals of the Missouri Botanical Garden* 80: 700–722.

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

PAGE, R. D. M. and M. A. CHARLESTON. 1997. From gene to organismal phylogeny: reconciled trees and the gene-tree species-tree problem. *Molecular Phylogenetics and Evolution* 7: 231–240.

RICHARDSON, J. E., R. T. PENNINGTON, T. D. PENNINGTON, and P. M. HOLLINGSWORTH. 2001. Rapid diversification of a species-rich genus of Neotropical rainforest trees. *Science* 293: 2242–2245.

RIESEBERG, L. H. 1996. Homology among RAPD fragments in interspecific comparisons. *Molecular Ecology* 5: 99–105.

——— and D. E. SOLTIS. 1988. A molecular reexamination of introgression between *Helianthus annuus* and *Helianthus bolanderi* Compositae. *Evolution* 42: 227–238.

——— and C. A. BUERKLE. 2002. Genetic mapping in hybrid zones. *American Naturalist* 159: S36–S50.

SANG, T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* 37: 121–147.

——— and D. ZHANG. 1999. Reconstructing hybrid speciation using sequences of low copy number genes: hybrid origins of five *Paeonia* species based on *Adh* gene phylogenies. *Systematic Botany* 24: 148–163.

———, 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.

SIMMONS, M. P. and H. OCHOTERENA. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.

SMALL, R. L., J. A. RYBURN, R. C. CRONN, T. SEELANAN, and J. F.

WENDEL. 1998. The tortoise and the hare: Choosing between non-coding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany* 85: 1301–1315.

SOLTIS, D. E. and P. S. SOLTIS. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. Pp. 1–42 in *Molecular systematics of plants II: DNA sequencing*, eds. D. E. Soltis, P. S. Soltis, and J. J. Doyle. Boston: Kluwer Academic.

———, ———, M. W. CHASE, M. E. MORT, D. C. ALBACH, M. ZANIS, V. SAVOLAINEN, W. H. HAHN, S. B. HOOT, M. F. FAY, M. AXTELL, S. M. SWENSEN, L. M. PRINCE, W. J. KRESS, K. C. NIXON, and J. S. FARRIS. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.

SORENSSON, C. T. and J. L. BREWBAKER. 1994. Interspecific compatibility among 15 *Leucaena* (Leguminosae Mimosoideae) species via artificial hybridization. *American Journal of Botany* 81: 240–247.

STRAND, A. E., M. J. LEEBENS, and B. G. MILLIGAN. 1997. Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology* 6: 113–118.

TABERLET, P., L. GIELLY, G. PAUTOU, and J. BOUVET. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1110.

TANK, D. C. and T. SANG. 2001. Phylogenetic utility of the glycerol-3-phosphate acyltransferase gene: Evolution and implications in *Paeonia* (Paeoniaceae). *Molecular Phylogenetics and Evolution* 19: 421–429.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D. G. HIGGINS. 1997. The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876–4882.

WENDEL, J. F., A. SCHNABEL, and T. SEELANAN. 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences, USA* 92: 280–284.

——— and J. J. DOYLE. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. Pp. 265–296 in *Molecular systematics of plants II. DNA sequencing*, eds. D. E. Soltis, P. S. Soltis, and J. J. Doyle. Boston: Kluwer Academic.

WILLIAMS, J. G. K., M. K. HANAFEY, J. A. RAFALSKI, and S. V. TINGEY. 1993. Genetic analysis using random amplified polymorphic DNA markers. *Methods in Enzymology* 218: 704–740.

WOJCIECHOWSKI, M. F., M. J. SANDERSON, and J. M. HU. 1999. Evidence on the monophyly of *Astragalus* and its major subgroups based on nrDNA ITS and cpDNA *trnL* intron data. *Systematic Botany* 24: 408–437.

WOLFE, A. D. and A. LISTON. 1998. Contributions of PCR-based methods to plant systematics and evolutionary biology. Pp. 43–86 in *Molecular systematics of plants II: DNA sequencing*, eds. D. E. Soltis, P. S. Soltis, and J. J. Doyle. Boston: Kluwer Academic.

ZHANG, W., J. F. WENDEL, and L. G. CLARK. 1997. Bamboozled again: inadvertent isolation of fungal rDNA sequences from Bamboo (Poaceae: Bambusoideae). *Molecular Phylogenetics and Evolution* 8: 205–217.