Pl. Syst. Evol. 252: 199–214 (2005)
DOI 10.1007/s00606-004-0294-0

Plant Systematics and Evolution
Printed in Austria

Phylogeny of the New World diploid cottons (*Gossypium* L., Malvaceae) based on sequences of three low-copy nuclear genes

I. Álvarez[1], **R. Cronn**[2], and **J. F. Wendel**[3]

[1]Real Jardín Botánico de Madrid, CSIC, Madrid, Spain
[2]Pacific Northwest Research Station, USDA Forest Service, Corvallis, Oregon, USA
[3]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, USA

Received October 4, 2004; accepted December 15, 2004
Published online: May 9, 2005
© Springer-Verlag 2005

Abstract. American diploid cottons (*Gossypium* L., subgenus *Houzingenia* Fryxell) form a monophyletic group of 13 species distributed mainly in western Mexico, extending into Arizona, Baja California, and with one disjunct species each in the Galapagos Islands and Peru. Prior phylogenetic analyses based on an alcohol dehydrogenase gene (*Adh*A) and nuclear ribosomal DNA indicated the need for additional data from other molecular markers to resolve phylogenetic relationships within this subgenus. Toward this end, we sequenced three nuclear genes, the anonymous locus *A1341*, an alcohol dehydrogenase gene (*Adh*C), and a cellulose synthase gene (*Ces*A1b). Independent and combined analyses resolved clades that are congruent with current taxonomy and previous phylogenies. Our analyses diagnose at least two long distance dispersal events from the Mexican mainland to Baja California, following a rapid radiation of the primary lineages early in the diversification of the subgenus. Molecular data support the proposed recognition of a new species closely related to *Gossypium laxum* that was recently collected in Mexico.

Key words: *Gossypium*, *Houzingenia*, cotton, phylogeny, low-copy nuclear genes, alcohol dehydrogenase, cellulose synthase.

Introduction

New World, diploid *Gossypium* species comprise a morphological and cytogenetic (D-genome) assemblage (Cronn et al. 2002, Endrizzi et al. 1985, Wendel 1995, Wendel and Cronn 2003) that taxonomically is recognized as subgenus *Houzingenia* (Fryxell 1969, 1979, 1992). This group of plants includes 11 species distributed primarily in SW Mexico and extending northward into Arizona, in addition to two other species with disjunct distributions in Peru and the Galapagos Islands (Fryxell 1992). Although none of these species produces commercially important cotton fiber, the fact that one of the parental lineages of allotetraploid cultivated cotton (*G. hirsutum* L. and *G. barbadense* L.) belongs to this group (Cronn et al. 1999, Endrizzi et al. 1985, Small et al. 1998, Small and Wendel 2000a) gives special relevance to the understanding of their systematics and evolutionary relationships.

Numerous molecular phylogenetic analyses have demonstrated that the subgenus is monophyletic (reviewed in Wendel and Cronn 2003). However, while the circumscription of the subgenus and species boundaries within this

clade are reasonably well-understood (Fryxell 1979, 1992), phylogenetic relationships among species remain unclear, despite numerous studies (Cronn et al. 1996, DeJoode 1992, Fryxell 1971, Liu et al. 2001, Seelanan et al. 1997, Small and Wendel 2000b, Wendel 1995, Wendel and Albert 1992). Phylogenetic analysis based on chloroplast restriction site analysis and chloroplast DNA sequences (Cronn et al. 2003, DeJoode 1992, Seelanan et al. 1997, Wendel and Albert 1992) have led to a number of phylogenetic conclusions that are at odds with numerous, unlinked nuclear markers and morphological trends in the genus. Results from nuclear ribosomal sequences (Cronn et al. 1996, Seelanan et al. 1997) have provided equally controversial resolutions (e.g. Wendel et al. 1995a,b), and resolution of these closely related species appears hampered by the presence of intraindividual polymorphism, some of which appears trans-specific.

The use of low-copy nuclear genes to infer plant phylogenies is rapidly increasing, due in part to the recent accessibility of many nuclear genes (characterized in gene discovery and genome sequencing projects) and their higher resolution, as has been demonstrated for various groups (reviewed in Sang 2002, Small et al. 2004). In *Gossypium*, the characterization of numerous low-copy nuclear genes (Cronn et al. 2002, Cronn et al. 1999, Senchina et al. 2003, Small and Wendel 2000a) has yielded a wealth of nuclear gene markers for studying *Gossypium* (Cronn et al. 2002, Small et al. 2004, Small et al. 1998, Small and Wendel 2000b) and related genera (Wendel et al. 2002). To date, these markers have been applied primarily to just a few representatives of subgenus *Houzingenia*, as studies to date have focused on higher-level phylogenetic relationships (Cronn et al. 2002, 2003; Wendel et al. 2002).

The most comprehensive molecular systematic study of New World cottons to date was performed by Small and Wendel (2000b) using a member of the *Adh* gene family in *Gossypium* (Cronn et al. 1999, Small and Wendel 2000a). While two to three allozyme loci can be resolved for *Adh* in *Gossypium* (Wendel, unpublished), the gene family is much larger, including up to seven discrete loci in some diploid cotton species (Small and Wendel 2000a). One of these loci, *Adh*A, was selected for phylogenetic applications (Small and Wendel 2000b) based on its homologous chromosome location in three genetic maps including diploid and allotetraploid cottons (Brubaker et al. 1999), and on the results of Southern hybridization analyses (Cronn et al. 1999) that indicate the existence of only one copy per genome in two species of the subgenus (*G. raimondii* Ulbrich and *G. trilobum* (DC.) Skovsted). The phylogeny based on *Adh*A (Small and Wendel 2000b) supports the taxonomically recognized subsections and is generally congruent with previous analyses in the subgenus (Cronn et al. 1996, DeJoode 1992, Seelanan et al. 1997, Wendel and Albert 1992), although relationships among sections and subsections remained unresolved.

More recently, Cronn et al. (2003) used four chloroplast genes and eight low-copy nuclear genes to reevaluate the evolutionary history of *G. gossypioides* (Ulbrich) Standley. Among the nuclear genes were two alcohol dehydrogenase genes (*Adh*A, *Adh*C), two cellulose synthase genes (*Ces*A1, *Ces*A1b), a fatty acid desaturase intron (*FAD2-1 intron*), and the anonymous genes *A1341*, *G1121*, and *G1262*. In this study, six of the 13 D-genome species were included. The individual molecular markers used in this study revealed levels of variation that provided modest resolution of New World species; however, results from this study indicated that combining the most informative genes might have the potential to resolve phylogenetic relationships among all 13 species in the subgenus. With this objective in mind, we sequenced the three most informative nuclear genes (*A1341*, *Ces*A1b, and *Adh*C) that showed variation at different levels (Cronn et al. 2002, Seelanan et al. 1999, Senchina et al. 2003, Small et al. 1998) from representatives of all species. To these data we added previously generated sequences of the *Adh*A gene (Small and Wendel 2000b).

Insights from these four nuclear genes are compared to results obtained with nuclear ribosomal DNA, and the most recent taxonomic treatment of Fryxell (1992) is evaluated in light of these combined results.

Materials and methods

Plant materials. Sampling included the 13 American diploid cottons: *Gossypium aridum* (Rose & Standley) Skovsted, *G. armourianum* Kearny, *G. davidsonii* Kellogg, *G. gossypioides* (Ulbrich) Standley, *G. harknessii* Brandegee, *G. klotzschianum* Andersson, *G. laxum* Phillips, *G. lobatum* Gentry, *G. raimondii* Ulbrich, *G. schwendimanii* Fryxell & Koch, *G. thurberi* Todaro, *G. trilobum*, and *G. turneri* Fryxell. Besides, we included one specimen (*Gossypium* sp.) that is suggested to be a new species related to *G. aridum* (Ulloa et al., unpublished). In some cases, more than one accession per species was sampled, based on our assessments of variability within each species (some species are narrowly distributed and relatively invariable morphologically and with respect to allozyme markers, whereas others are more widespread and/or exhibit greater variation). Six species that belong to different cytogenetic and taxonomic groups: *G. anomalum* Wawra ex Wawra & Peyritch, *G. bickii* Prokhanov, *G. longicalyx* J. B. Hutchinson & Lee, *G. robinsonii* F. von Mueller, and *Gossypioides kirkii* (Mast.) J. B. Hutchinson and *Kokia drynarioides* (Seemann) Lewton were included as an outgroup (Table 1). At least one accession per species is identical to those used by Small and Wendel (2000b), so that direct evaluation of the *Adh*A data could be made. For the genes *A1341*, *Adh*C, and *Ces*A1b, we included sequences already published for six species and the outgroup (Cronn et al. 2002), and for the remaining we used DNAs available from previous studies (Cronn et al. 2002, Cronn et al. 1996, DeJoode 1992, Seelanan et al. 1997, Wendel and Albert 1992). In a few cases, we newly isolated total DNA from plants grown in the greenhouse (at Iowa State University), using fresh leaf tissue and the Plant DNeasy kit (Qiagen) following the manufacturer's instructions. Vouchers for these plants were deposited at the Ada Hyden Herbarium (ISC) at Iowa State University, Ames.

Molecular markers. We used sequences of three independent nuclear genes (*A1341*, *Adh*C, and *Ces*A1b) as molecular markers. This selection was based on copy number of each gene (inferred to be single-copy from earlier work) and on the knowledge of their orthology across different genomes in cotton (Brubaker et al. 1999, Cronn and Wendel 1998, Small and Wendel 2000a). Additionally, we selected genes that in previous analyses (Cronn et al. 2002) showed a relatively high ratio of phylogenetically informative sites (PI) compared to other low-copy nuclear genes. The *A1341* locus (0.7 kb) is an anonymous gene that corresponds to a *Pst*I mapping probe (Brubaker and Wendel 1994, Cronn and Wendel 1998, Cronn et al. 1999); this gene has a PI ratio slightly higher than the *Adh*A gene used in the previous phylogenetic analysis of the subgenus (Small and Wendel 2000b). A region of the *Ces*A1b gene (Cronn et al. 2002, Cronn et al. 1999) has a similar PI ratio to plastid genes, although this low ratio is compensated by its length (1.15 kb). From the *Adh* gene family, we sequenced a portion (0.94 kb) of the *Adh*C gene that has a high PI ratio similar to some chloroplast spacers, and higher than other nuclear genes in a previous analysis (Cronn et al. 2002).

Amplification primers were those used previously (Cronn et al. 2002), namely A1341F and A1341R for the *A1341* locus, CelAF and CelAR for the *Ces*A1b partial gene, and ADHx4-3 and ADH-P2 for the *Adh*C partial gene. For accessions of *G. lobatum*, a new set of internal forward and reverse primers (GTG AGG CTT CTA GGA TCA TTG G and CCA ATG ATC CTA GAA GCC TCA C, respectively), were used in a second PCR in order to obtain enough amplification product to sequence the *Adh*C partial gene. To amplify the three loci we followed protocols already described (Cronn et al. 1999, Small et al. 1998). The DNA Sequencing Facility of Iowa State University carried out direct sequencing of all amplification products.

Data analysis. Sequence alignment was performed manually using BioEdit v.5.0.9 (Hall 1999). Genomic sequences were aligned to previously published (Cronn et al. 1999) exon sequences for the corresponding gene, which aided determination of intron/exon boundaries. Alignment was straightforward in all cases, as indels were rare and uncomplicated. Data matrices are available at http://www.eeob.iastate.edu/faculty/WendelJ/datasets.htm

Table 1. Plant materials used, indicating geographic origin, voucher, and GenBank accession numbers for the three genes sequenced (*A1341*, *Adh*C, and *Ces*A1b respectively). Sequences in italics are from previous work (Cronn et al. 2002)

Species	Geographic origin	Voucher ID	GenBank accession numbers
G. anomalum	Africa	JFW &TDC 305	*AF403074, AF419966, AF419974*
G. aridum	Mexico, Jalisco	DRD 185	AY699077; AY699104; AY699084
G. aridum	Mexico, Colima	DRD 168	AY699105; AY699085
G. aridum	Mexico, Colima	IA 1-4	AY699106; AY699086
G. aridum	Mexico, Guerrero	IA 14-1	AY699107; AY699087
G. aridum	Mexico, Sinaloa	IA 36-1	AY699109; AY699089
G. armourianum	Mexico, Baja California	D2-1-7	AY699078; AY699110; AY699090
G. bickii	Australia	JFW &TDC 557	*AF403077, AF419968, AF419977*
G. davidsonii	Mexico, Baja California	32	*AF520737, AY125059, AY125071*
G. gossypioides	Mexico, Oaxaca	D6-2	*AF520736, AY125058, AY125070*
G. harknessii	Mexico, Baja California	D2-2	AY699079; AY699111; AY699091
G. klotzschianum	Galapagos Islands	D3k-3	AY699080; AY699112; AY699092
G. klotzschianum	Galapagos Islands	IA 54	AY699113; AY699093
G. laxum	Mexico, Guerrero	L. Phillips	AY699081; AY699119; AY699094
G. laxum	Mexico, Guerrero	IA 26-2	AY699120; AY699095
G. laxum	Mexico, Guerrero	IA 25-2	AY699121; AY699096
G. laxum	Mexico, Guerrero	IA 40-4	AY699122; AY699097
G. lobatum	Mexico, Michoacan	DRD 157	AY699123; AY699100
G. lobatum	Mexico, Michoacan	DRD 161	AY699082; AY699115; AY699099
G. lobatum	Mexico, Michoacan	IA 57-6	AY699114; AY699098
G. longicalyx	Tanzania	TS 8	*AF403076, AF419967, AF419976*
G. raimondii	Peru	No accession no.	*AF136815, AF036568, AF139449*
G. robinsonii	Australia	AZ-50	*AF136817, AF036567, AF139451*
G. schwendimanii	Mexico, Michoacan	No accession no.	*AF520738, AY125060, AY125072*
G. schwendimanii	Mexico, Michoacan	IA 56-3	AY699116; AY699101
G. thurberi	Arizona	D1-17	AY699083; AY699117; AY699102
G. thurberi	Arizona	D1-8	AY699118; AY699103
G. trilobum	Mexico, Sinaloa	No accession no.	*AF520739, AY125061, AY125073*
G. turneri	Mexico, Sonora	D10-3	*AF520740, AY125062, AY125074*
Gossypium sp.	Mexico, Guerrero	IA 64-1 (US72)	AY699108; AY699088
Gossypioides kirkii	Madagascar	TS 3	*AF201877, AF169254, AF201887*
Kokia drynarioides	Hawaiian Islands	TS 6	*AF403078, AF419969, AF419978*

Phylogenetic analyses were conducted using maximum-parsimony as implemented in PAUP*4.0b10 (Swofford 1999) and a heuristic search with the TBR and ACCTRAN option for character optimization. Gaps were treated as missing data. To obtain the most parsimonious trees (m.p.t.), 100 random addition sequences were performed, saving 1000 trees per replicate. Relative support for tree branches was assessed by using decay and bootstrap analyses. Decay values were obtained by the converse constraints approach (Bremer 1994) with the aid of the program AutoDecay (Eriksson 1998) for PAUP*. Bootstrap analyses were performed with a fast-heuristic search of 1000 replicates.

To assess congruence among datasets, all possible combinations of the three datasets obtained plus the *Adh*A dataset (Small and Wendel

2000b) were analyzed and the results compared. For the independent analyses, we included all samples, and for the combined analysis, we used reduced matrices (only one individual per species) that contained a selection of the accessions that also were present in the *Adh*A matrix (Small and Wendel 2000b). To examine topological congruence between the consensus trees of all pairwise combinations, we computed the partition metric (PM) by Robinson and Foulds (1981), and the greatest agreement subtree metric (D_1) by Kubicka et al. (1995) indices. The PM measures the rearrangements needed to transform one of the two trees in a comparison into the other, and D_1 measures the number of taxa that have to be pruned in two trees to get a minimum topology in which the two trees agree. Congruence between datasets was estimated calculating both Miyamoto (I_M) by Miyamoto in Swofford (1991), and Mickevich and Farris (I_{MF}) by Mickevich and Farris (1981) incongruence metric indices. I_M calculates the extra homoplasy needed to explain each data set on the topology recovered from the alternative data set, and I_{MF} indicates the number of homoplasies required by each individual data set to explain the shortest tree recovered from the combined matrix. In addition, we applied a significance test for heterogeneity, the partition homogeneity test (HT_F) by Johnson and Soltis (1998), to assess when a congruence index is indicating a serious conflict between datasets. The HT_F test measures I_{MF} for a number of random partitions of the combined data set, each partition being formed from two subsets of the same size as the two data sets. When 95% or more of those random partitions have an I_{MF} smaller than the original, the null hypothesis of homogeneity is rejected, and the data sets are inferred to be significantly heterogeneous. All indices and tests were conducted using PAUP*4.0b10 (Swofford 1999).

Results

***Adh*C sequences.** A total length of 768 nucleotides (nt) of aligned sequences were analyzed for the partial *Adh*C gene in 26 ingroup plus 6 outgroup specimens, including partial sequences of exon 1 and 5, and complete sequences of exons 2 through 4 and of introns 1 through 4 (Small et al. 1998). Absolute sequence lengths ranged from 749 to 760nt, and the total number of indels was 17, all within the introns. Indel size varied from 1–3 nt in the ingroup to 1—6 nt including the outgroup. In the aligned matrix, 369 nt were in exons and 399 were in introns. Within the ingroup, 102 sites were variable (13.3%), of which 70 (9.1%) were parsimony-informative. Of the 102 variable sites, 36 were in exons and 66 occurred within introns. Of the 36 substitutions in exons, 20 were synonymous and 16 were replacement changes.

For all four accessions of *G. laxum* and one accession of *G. lobatum* (DRD 161), we obtained sequences with an unusual terminal (3′) intron splice site dinucleotide (i.e. the conventional AG splice signal mutated to AT) for the first intron at position 167 of the alignment. In addition, the single accession of *G. trilobum* displayed an unconventional intron splice site dinucleotide in the second intron (position 252 of the alignment), as the typical GT splicing signal mutated to GA. One stop codon was present in positions 227–229 of the alignment for the accession IA 26–2 of *G. laxum*. These features indicate that *Adh*C from at least some D-genome species may be pseudogenes, as has been convincingly demonstrated for A-genome African cotton species (Small et al. 1998, Small and Wendel 2000a).

***Ces*A1b sequences.** Cellulose synthase gene sequences were obtained for the same 26 ingroup and 6 outgroup individuals as for *Adh*C. This gene was formerly named *Cel*A2 or *Ces*A2 (Cronn et al. 2002, Cronn et al. 1999, Senchina et al. 2003) but is now called *Ces*A1b gene due to its high sequence similarity to *Ces*A1 (Cronn et al. 2002). The aligned matrix has a total length of 1177nt, and includes partial exons 1 and 5 and complete exons 2 through 4, and introns 1 through 4. Absolute sequence lengths ranged from 1166 to 1175nt, and the total number of indels was 7, all occurring within introns. Indel sizes varied from 1–13nt in the ingroup to 1–49nt including the outgroup. In the aligned matrix 520nt were in exons and 657 were in introns. Within the ingroup, 62 sites were variable (5.3%), of which 32 (2.7%) were parsimony informative.

Of the 62 variable sites, exactly half (31) occurred in exons, and of these, 8 were synonymous and 23 were replacement changes. The unusually large number of replacements in this data set raises the possibility that this locus also encodes a pseudogene. *Ces*A1b orthologs from A-genome cotton species *G. herbaceum* and *G. arboreum* show the hallmarks of pseudogenes, such as non-consensus intron splice dinucleotides, a premature stop codon in exon 4 (Cronn et al. 2002, Cronn et al. 1999), and a lack of detection via RT-PCR (Cronn, unpublished). Nonetheless, these hallmarks are lacking from D-genome species, and preliminary evidence from RT-PCR indicates that *Ces*A1b is expressed at the level of mRNA in *G. raimondii* and the D-genome from *G. hirsutum*.

***A1341* sequences.** In a preliminary alignment of the first fourteen accessions sequenced, we found no intra-specific variation. Therefore, we selected one accession per species (those used in the *Adh*A matrix; Small and Wendel 2000b) to sequence the *A1341* locus. To make feasible the comparison between this and the *Adh*A matrix, a total of 13 ingroup sequences and 1 outgroup sequence were included. The aligned matrix had a total length of 661nt, including positions 37 through 681 of the previously published data (Cronn et al. 2002, Cronn et al. 1999, Senchina et al. 2003). Absolute sequence lengths varied from 627 to 649nt, and the total number of indels was 42. Sequences were treated as non-coding, as BLAST searches against databases revealed no significant homologies with known genes. Within the ingroup, 24 sites were variable (3.6%), of which 7 (1.1%) were parsimony informative. All indels within the ingroup are of one or two nucleotides in length, excepting one deletion of 18nt in position 445 of the alignment, present in both *G. davidsonii* and *G. klotzschianum*.

Independent analyses. Parsimony analyses were performed for each independent dataset (*Adh*C, *Ces*A1b, and *A1341*). The *Adh*C matrix resulted in 210 most-parsimonious trees (m.p.t.) (length = 248; consistency index (CI), excluding uninformative characters = 0.74; retention index (RI) = 0.87). The strict consensus tree of all m.p.t. is shown in Fig. 1. All ingroup taxa belong to a unique clade, with decay support (d = 3). As indicated in a previous analysis (Cronn et al. 2003), *G. gossypioides* (subsection *Selera* (Ulbrich) Fryxell), appears sister to the remaining members of the subgenus, with a low decay value (d = 1) but with reasonable bootstrap support (b = 87%). Within the large ingroup clade three clades are resolved, although relationships among them are not. One of these three groups includes all the species of subsection *Caducibracteata* Mauer (*G. armourianum*, *G. harknessii*, and *G. turneri*) with high support (d = 5, b = 97%). A second group includes all species of subsection *Erioxylum* (Rose & Standley) Prokhanov (*G. aridum*, *G. laxum*, *G. lobatum*, and *G. schwendimanii*) plus *Gossypium* sp., although not all their accessions (exceptions are two of *G. lobatum* and one of *G. aridum*), with moderate branch support (d = 3, b = 94%). The third clade include all accessions of subsections *Austroamericana* Fryxell (*G. raimondii*), *Integrifolia* (Todaro) Todaro (*G. davidsonii*, and *G. klotzschianum*), and *Houzingenia* (*G. thurberi*, and *G. trilobum*), as well as the two aforementioned accessions of *G. lobatum* (DRD 157, IA 57-6).

Within the '*Erioxylum*' clade relationships are essentially unresolved. The *G. schwendimanii* accessions were united with high support (d = 6, b = 100%). Also, all *G. laxum* plus one *G. lobatum* (DRD 161) accessions form a strongly supported clade (d = 6, b = 100%). *Gossypium aridum* accessions resolve into two different clades, and the accession IA 64-1 (US72) of *Gossypium* sp. appears sister to the *G. laxum*/*G. lobatum* clade, albeit with minimal support (d = 1, b = 58%). Within the '*Caducibracteata*' clade, *G. armourianum* is sister to *G. harknessi* and *G. turneri*, the latter comprising a clade with strong branch support (d = 3, b = 98%). In the third clade, subsection *Houzingenia* appears monophyletic (d = 3, b = 88%), as does subsection *Integrifolia* (d = 1, b = 63%). The single species in subsection *Austroamericana*,

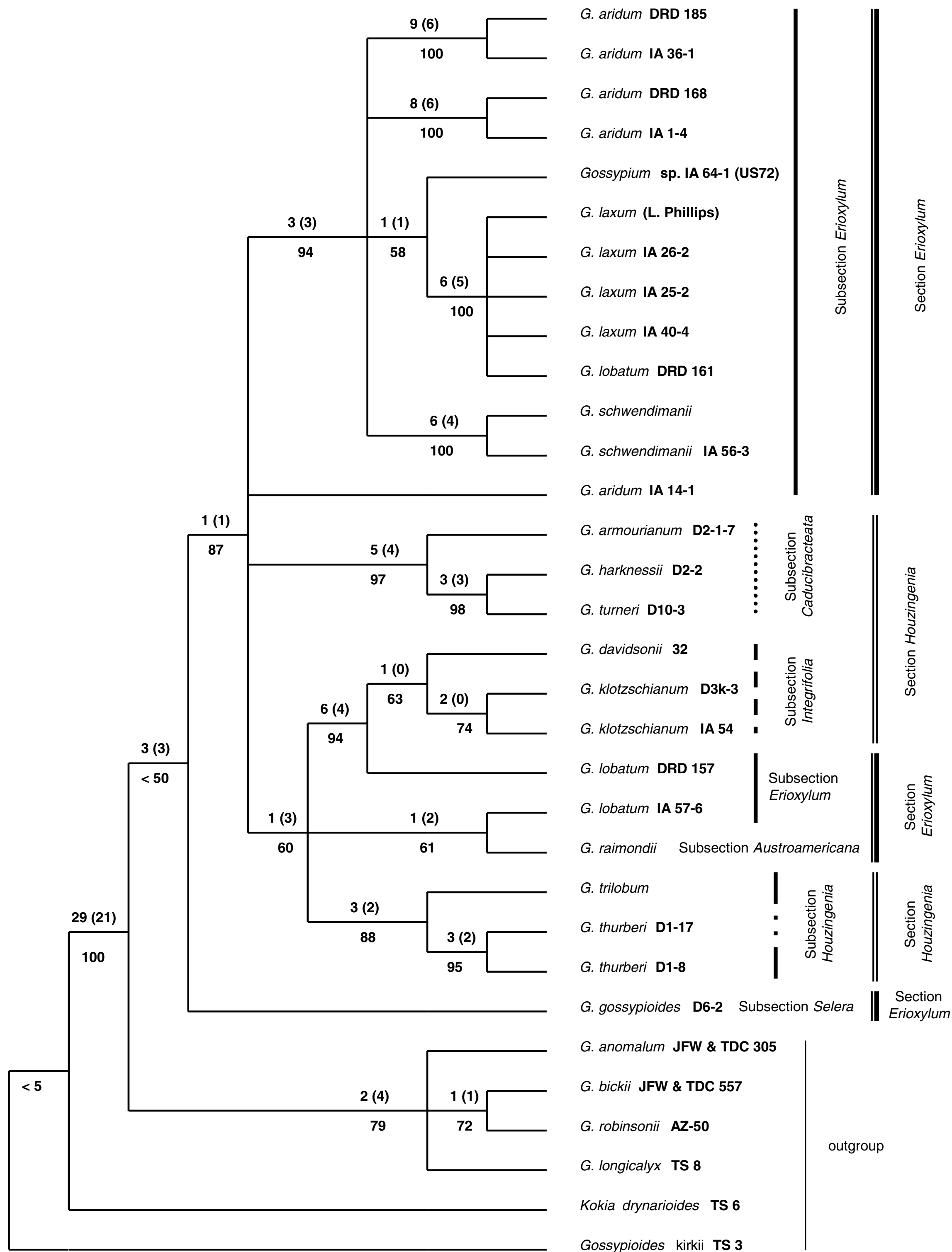

Fig. 1. Strict consensus of the 210 most parsimonious trees obtained with the analysis of the *Adh*C matrix (length of 248; CI = 0.74; RI = 0.87). Numbers above branches indicate decay values, followed by the number of synapomorphic characters for that branch (within parentheses). Numbers below branches indicate bootstrap values

G. raimondii, is sister to *G. lobatum* accession IA 57-6, although support for this relationship is modest (d = 1, b = 61%).

The *CesA1b* matrix resulted in 4 m.p.t. (length = 188; CI = 0.81; RI = 0.90). The strict consensus tree of the m.p.t. is shown in Fig. 2. Ingroup taxa appear monophyletic (d = 2, b = 100%), and *G. gossypioides* is sister to the remaining members of the ingroup, but with low support (d = 1, b = 64%). The remainder of the ingroup is divided into two clades. The largest includes all members of subsections *Caducibracteata* and *Erioxylum*, although with virtually no branch support (d = 1, b = 10%). The second group includes all accessions of the subsections *Austroamericana*, *Houzingenia*, and *Integrifolia*, also with low branch support (d = 1, b = 29%). Three of the six subsections are resolved with branch support values ranging from low (i.e. *Erioxylum* d = 1, b = 51%) to high (*Houzingenia* d = 4, b = 84%; *Integrifolia* d = 4, b = 87%). In contrast to the *Adh*C analysis, all accessions within a species coalesce to form monophyletic units. The single accession IA 64-1 (US72) of the suggested new species, appears related to *G. laxum*, as well as with the *Adh*C data, with relatively high support (d = 3, b = 82%).

The *A1341* matrix, which included only one accession per species, resulted in two m.p.t. (length = 32; CI = 0.92; RI = 0.91). The strict consensus tree (not shown) offers almost no resolution, although a few groups have moderate branch support. These are: (1) a large clade including all ingroup accessions except *G. gossypioides* (d = 4, b = 93%); (2) the monophyletic subsection *Integrifolia* (d = 3, b = 95%); (3) the sister-species relationship between *G. harknessii* and *G. turneri* (d = 2, b = 85%); and (4) the grouping of *G. aridum* and *G. lobatum,* with low branch support (d = 1, b = 57%). None of these relationships contradict results from the analyses of the other genes.

Congruence indices. Indices for incongruence obtained in the independent and combined analyses of the four datasets (*A1341*, *Adh*A, *Adh*C, and *Ces*A1b) are shown in Table 2. Both PM and D_1 topological congruence indices are mostly higher comparing trees generated from the same dataset than between trees from different datasets. The values were normalized (0 to 1, indicating least to most similar topologies). Partition metric values (PM) range from 0.77 to 0.98 between trees of the same datasets, and from 0.45 to 0.98 between trees of different datasets. Greatest agreement subtree metric values (D_1) range from 0.81 to 0.98 between trees of the same datasets, and from 0.64 to 0.98 between trees of different datasets. Although there is not a clear agreement between these indices, both suggest the most topological incongruence when comparing the *A1341* trees to others. Values for the two character congruence indices (I_M, and I_{MF}) were also normalized (0% to 100%, with the latter indicating the most incongruence). Both indices indicate that the most incongruent combination is *A1314* and *Adh*A, with 94% and 50% values respectively, although the significance test (HT_F) does not return a significant value (Table 2). Other combinations with high incongruent values are *Adh*A plus *Adh*C (I_M = 69%, I_{MF} = 20%), and *Adh*C plus *Ces*A1b including all accessions (I_M = 56%, I_{MF} = 24%). The remaining combinations show low to moderate incongruence values that range from 31 to 71% for the I_M index, and from zero to 10% for the I_{MF} index, and in no case are these values significant.

The results of the significance test for heterogeneity (HT_F) suggest that the *Adh*C dataset is incongruent with two other datasets (*Adh*A, and *Ces*A1b). In the case of *Adh*C vs. *Ces*A1b, this incongruence is not significant when the *Adh*C matrix is reduced to one accession per species. Pseudogenization (in this case, sequences with stop codons and intron splicing abnormalities) could be a possible cause of significant data set heterogeneity. However, congruence indices indicated significant differences (P = 0.02) even after when putative pseudogenes were eliminated from the data set. Another reason for the apparent conflict may be the inadvertent isolation of a previously unknown paralogous sequences

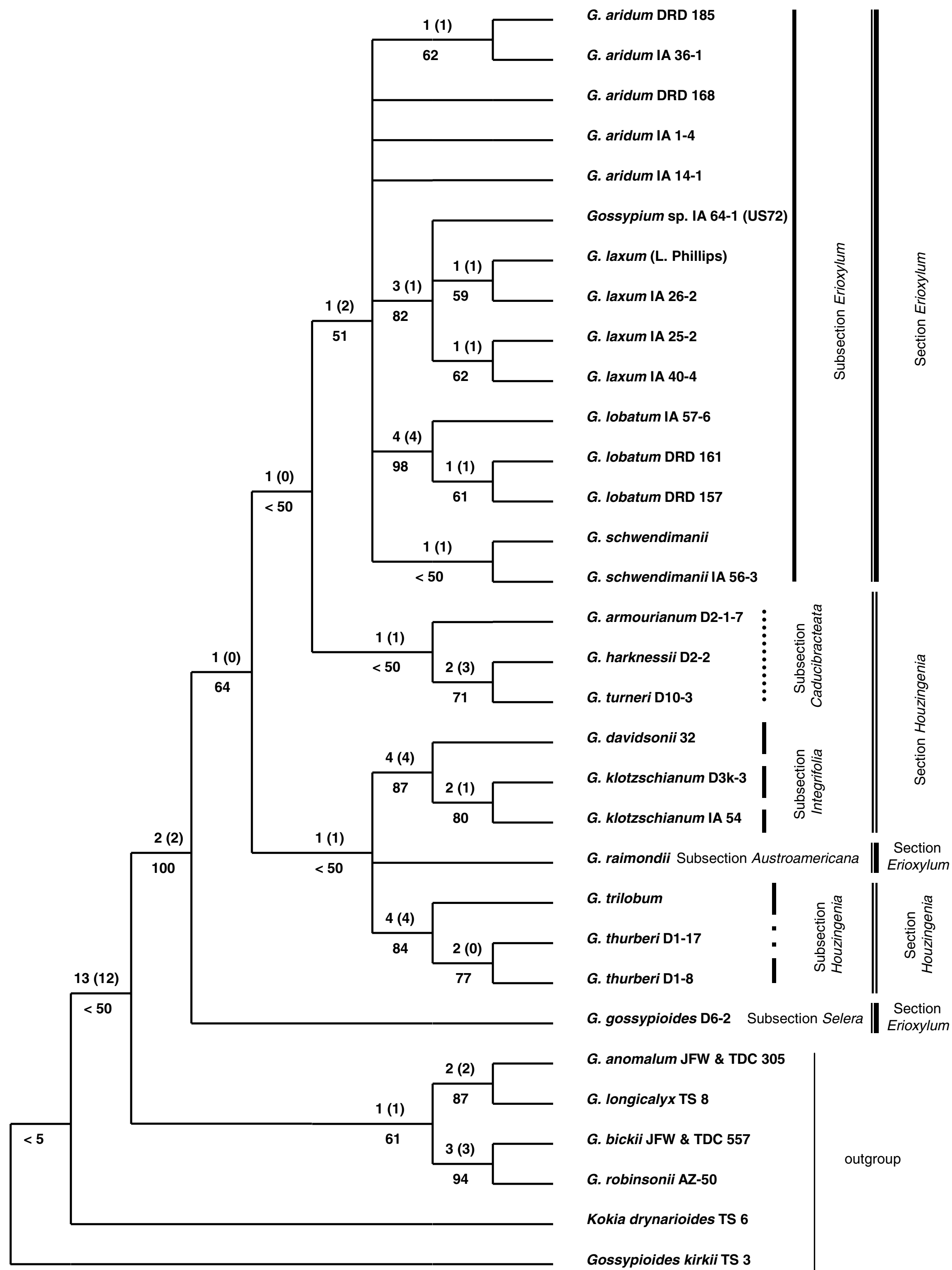

Fig. 2. Strict consensus of the four most parsimonious trees obtained with the analysis of the *Ces*A1b matrix (length of 188; CI = 0.81; RI = 0.9). Numbers above branches indicate decay values, followed by the number of synapomorphic characters for that branch (within parentheses). Numbers below branches indicate bootstrap values

Table 2. Congruence indices for independent and combined analyses between datasets. PM = Partition metric index for topological incongruence. D_1 = Greatest agreement subtree metric index for topological incongruence. Values were normalized (0 least similar, 1 most similar). I_M = Miyamoto index for incongruence between data sets. I_{MF} = Mickevich and Farris index for incongruence between data sets. Zero percentage indicates least incongruence and 100% most incongruence. H_{TF} = Farris significance test for heterogeneity. An asterisk indicates significant values. ([1]) Matrices that include complete sampling

	PM		D_1		I_M	I_{MF}	HT_F
	Mean	Range	Mean	Range			
A1341	0.96	–	0.96	–	–	–	–
*Adh*A	0.96	–	0.91	–	–	–	–
*Adh*C	0.91	0.96–0.86	0.91	0.96–0.86	–	–	–
*Ces*A1b	0.86	0.96–0.77	0.91	0.96–0.82	–	–	–
*Adh*C[1]	0.91	0.98–0.83	0.83	0.98–0.81	–	–	–
*Ces*A1b[1]	0.96	0.98–0.97	0.96	0.98–0.97	–	–	–
A1341 vs. *Adh*A	0.72	0.96–0.59	0.77	0.96–0.68	94%	50%	0.53
A1341 vs. *Adh*C	0.77	0.96–0.59	0.77	0.96–0.64	71%	6%	0.71
A1341 vs. *Ces*A1b	0.82	0.96–0.59	0.82	0.96–0.64	60%	0%	1
*Adh*A vs. *Adh*C	0.77	0.96–0.59	0.82	0.96–0.68	69%	20%	0.01*
*Adh*A vs. *Ces*A1b	0.82	0.96–0.45	0.82	0.96–0.68	61%	10%	0.31
*Adh*C vs. *Ces*A1b	0.82	0.96–0.64	0.86	0.96–0.77	31%	0%	1
*Adh*C[1] vs. *Ces*A1b[1]	0.91	0.98–0.6	0.88	0.98–0.74	56%	24%	0.01*

other than pseudogenes. This could explain the recovery of paraphyletic/polyphyletic species (cf. Fig. 1). In this regard we note that the *Adh*C locus in *Gossypium* is dynamic, and that phylogenetically local gene duplications have been reported (Small and Wendel 2000b). Finally, the unexpected resolution of *G. lobatum* alleles from accessions DRD 157 and IA 57–6 could reflect a failure of coalescence within this species, or possibly the remnants of an historical hybridization involving the ancestor(s) of subsections *Houzingenia* and *Austroamericana*.

Combined analysis. For the reasons discussed above, we excluded the *Adh*C dataset from the total combined analysis. Thus, three datasets, *A1341*, *Adh*A, and *Ces*A1b were concatenated to form a final matrix with a total length of 2827 nt for the combined analysis. To make feasible this combination, only accessions shared by the three datasets were selected. Accordingly, only 13 ingroup sequences (one per species) and one outgroup sequence (*G. robinsonii*) were included in this particular analysis. This final matrix led to the recovery of two m.p.t. (length = 158; CI = 0.8; RI = 0.92). A strict consensus tree of the two trees gives a well-resolved phylogeny, although support for relationships among groups often is low (Fig. 3). Ingroup taxa appear as a monophyletic group although with a bootstrap value less than 5%, and *G. gossypioides* sister to the remainder of the clade with high branch support (d = 8, b = 100%). Within the remaining cottons are two clades, one weakly inclusive of subsections *Caducibracteata* and *Erioxylum* and the other inclusive of subsections *Austroamericana*, *Houzingenia*, and *Integrifolia*. Both clades are only weakly supported (d = 1, b = 15% and d = 1, b = 31%, respectively), but subsections typically have high branch support (Fig. 3). Subsection *Austroamericana* (*G. raimondii*) appears sister to subsection *Integrifolia* (*G. davidsonii* and *G. klotzschianum*), although this relationship is not strongly supported (d = 1, b = 28%).

Discussion

Phylogeny and classification. At the taxonomic level of subsection, results from this analysis

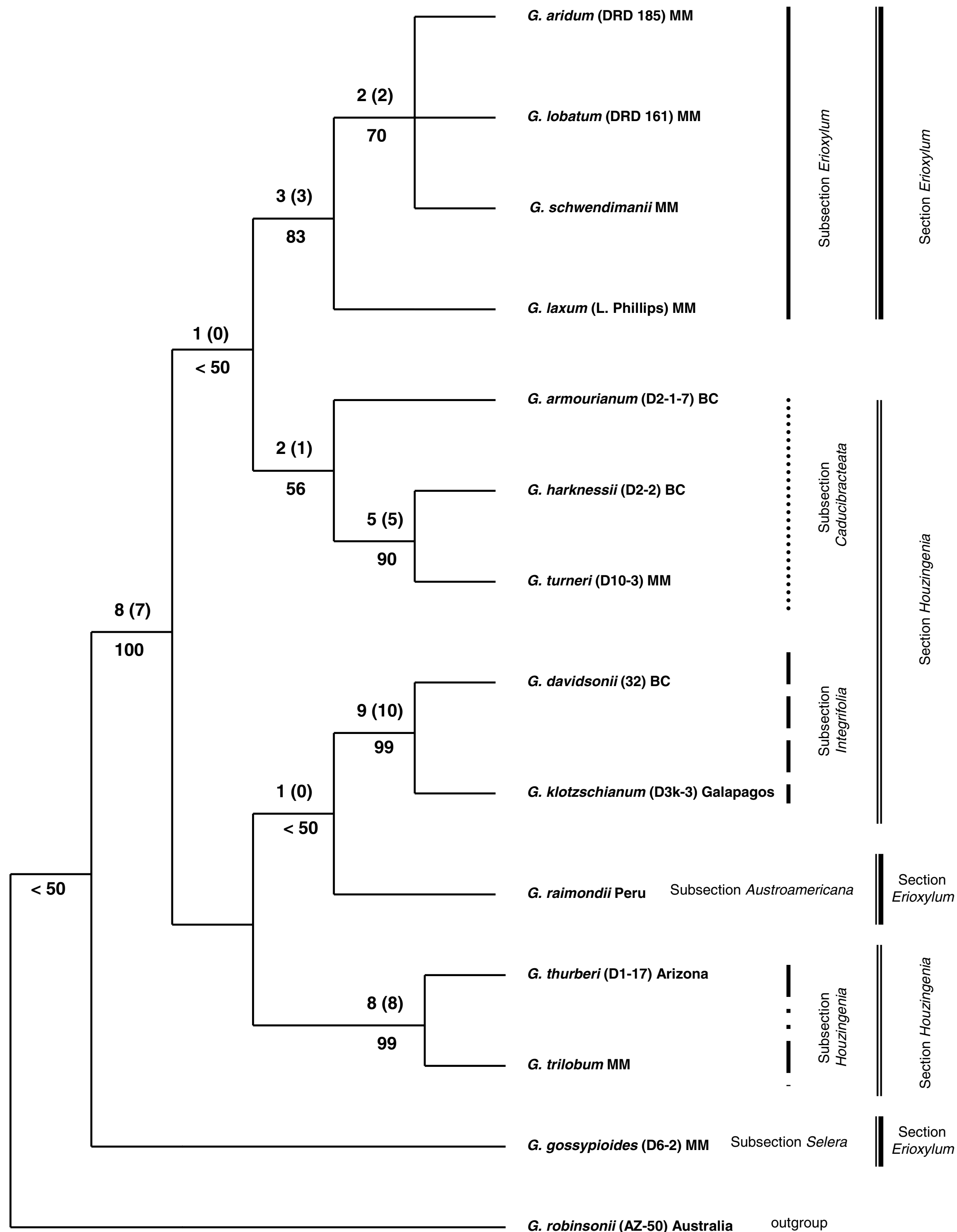

Fig. 3. Strict consensus of the two most parsimonious trees obtained with the analysis of the final matrix (*A1341* + *Adh*A + *Ces*A1b); length of 158; CI = 0.8; RI = 0.92. Numbers above branches indicate decay values, followed by the number of synapomorphic characters for that branch (within parentheses). Numbers below branches indicate bootstrap values. Accession numbers are within parenthesis. MM = Mexico mainland; BC = Baja California

are largely congruent with those obtained in previous molecular phylogenetic analyses (Cronn et al. 1996, Cronn et al. 1999, Liu et al. 2001, Small and Wendel 2000b) and morphological studies (Fryxell 1971, 1992). In particular, the data support the monophyly of the four multi-species subsections (*Caducibracteata, Erioxylum, Houzingenia* and *Integrifolia*), as well as the distinctiveness of the two subsections represented by single species (*Austroamericana, Selera*) (Fryxell 1992). Exceptions of subsectional monophyly were detected in the *Adh*C analysis, with two individuals of *G. lobatum* (DRD 157 and IA 57–6) and one of *G. aridum* (IA 14–1) resolving outside of the expected subsection (*Erioxylum*). The four species comprising subsection *Erioxylum* are among the most distinctive in the genus, as they are arid-season deciduous trees (as opposed to evergreen shrubs) that flower when leafless. Potential morphological synapomorphies for this group include sympodial to fascicled flowers (as opposed to solitary for the majority of the genus), and an epicalyx that is smaller than the true calyx (the opposite is seen in the remainder of the genus). In addition, sequences from 5S rDNA (Cronn et al. 1996), ITS (Wendel et al. 1995b), and the nuclear gene *Adh*A (Small and Wendel 2000b) support the monophyly of this subsection. For these reasons, we conclude that observations of paraphyly at *Adh*C involve comparisons of paralogous sequences (see above), or possibly supra-specific coalescence of *Adh*C alleles (i.e. the paraphyletic alleles are older than the divergence events separating extant species; Small et al. 2004).

The sectional-level taxonomy of this group remains unresolved since the sections proposed by Fryxell (1992) either demonstrate clear evidence of paraphyly (sect. *Erioxylum*) or fail to show monophyly (sect. *Houzingenia*) in these molecular phylogenetic analyses (Figs. 1–3). The unique sister position of *G. gossypioides* to the remainder of the New World clade is supported here by all analyses (Figs. 1–3), and has been discussed previously (Cronn et al. 2003). The evolutionary history of this species apparently includes nuclear introgression from an African lineage followed by a cytoplasmic capture of an American species. Given the divergence of this species from other members of section *Erioxylum*, *G. gossypioides* should be removed from this group, and possibly recognized as a distinct section *Selera*.

In these analyses, the sole representative of subsection *Austroamericana* (*G. raimondii*) shows a modest affinity to subsection *Integrifolia*, albeit with minimal support ($d = 0$, $b < 50$). This same resolution was obtained in the study of Cronn et al. (2003) that was based on a larger sampling of genes (albeit with lower taxon sampling density). Since subsections *Austroamericana* and *Integrifolia* are classified in different sections (*Erioxylum* and *Houzingenia*, respectively), these two lineages could possibly be recognized as a distinct lineage at the sectional level. It is important to note that these species are linked by two potential morphological synapomorphies, one based on leaf shape (leaves are unlobed, a trait that is unusual in the genus), and a second based on leaf pubescence (leaves are strongly pubescent, a feature shared with *G. aridum, G. laxum* and *G. lobatum* from subsection *Erioxylum*).

Other potentially important associations are evident in our analyses, even though most lack significant support. For example, subsection *Caducibracteata* appears closely related to subsection *Erioxylum* in most analyses (Figs. 2–3). If this resolution is correct, these lineages would be united by two potential morphological synapomorphies. First, this group shows solitary flowers, while fascicled flowers are evident in all other New World species (except *G. lobatum*). Second, epicalyx shape in this group tends towards ligulate-lanceolate, while cordate-ovate epicalyces predominate in New World cottons. Within subsection *Caducibracteata*, *G. harknessii* and *G. turneri* are closely related in all analyses (Figs. 1–3), including the *A1341* matrix (data not shown), with moderate to high branch support. Relationships within subsection *Erioxylum* remain unclear since there is no agreement among analyses. The analysis of the

combined data matrix gives a basal position of *G. laxum* to the polytomy formed by *G. aridum*, *G. lobatum*, and *G. schwendimanii* with moderate support (d = 2, b = 70%). It should be noted that this resolution contradicts morphology, as *G. laxum*/*G. lobatum* share morphological features (shallow leaf lobing; punctate style) that clearly distinguish them from other arborescent cottons. Since these species occupy a large geographic range and are known to exhibit substantial intraspecific variation (Wendel and Albert 1992), clarification of relationships among species from subsection *Erioxylum* will require greater sampling of the allelic variation contained within these weakly-differentiated species.

Coalescence and polymorphism. Although we did not sample deeply within species, the detection of intraspecific variation in one of the nuclear genes sequenced (*Adh*C) led to phylogenetically variable placements of individuals of *G. aridum* and *G. lobatum* (Fig. 1). As noted above, variable placement may reflect paralogy or coalescence that antedates divergence among this set of morphologically similar yet genetically variable Mexican, arborescent cottons. As in a previous work using *Adh*A (Small and Wendel 2000b), sequences of multiple clones from more individuals would be necessary to unravel the complexity and history of *Adh*C sequences in this group. Intraspecific variation is also present in the *Ces*A1b sequences of *G. aridum*, *G. laxum* and *G. lobatum* (Fig. 2), although in this case this variation does not override species boundaries (i.e. alleles coalesce within species).

Based on *Ces*A1b sequences, the accession IA 64-1 (also called US72 in Ulloa et al., unpublished) cladistically aligns with *G. laxum* instead of the other *G. aridum* accessions as expected. This is an interesting result in that this specimen, recently collected (2002) in Guerrero, is suggested to be a new species of *Gossypium* related to *G. aridum* (Ulloa et al., unpublished). The fact that the same result (sister to *G. laxum*) was obtained in a NJ distance analysis of AFLP data based on a broad survey of 28 *G. aridum* and eight *G. laxum* populations (Álvarez and Wendel, unpublished data) gives support to the hypothesis that this individual represents a new species. Like other members of section *Erioxylum*, the plant is arborescent in dry deciduous forests and defoliates during the dry season. The leaf morphology is variable in that juvenile leaves and mature leaves developed under stress are cordate (cf. *G. aridum*), but unstressed mature leaves are palmately lobed (cf. *G. laxum*).

Biogeographic considerations. Results from the combined analysis (*A1341* + *Adh*A + *Ces*A1b) do not suggest a simple biogeographic scenario. The basal divergence of New World cottons, represented by subsection *Selera*, is restricted to a narrow range in mainland Mexico (primarily, Oaxaca). Members of subsection *Erioxylum* cover much of the western and southwestern Mexican mainland (Sinaloa, Puebla, Guerrero, Michoacan, Colima, Jalisco). Included in this group is the most widespread New World species, *G. aridum*, which also is the most variable morphologically and from which geographically marginal populations have been reported (e.g. from Veracruz). The remaining species within this subsection are more locally distributed.

The Mexican arborescent clade appears to be genetically and morphologically equidistant to members of subsection *Caducibracteata*, and the monophyletic group composed of subsections *Houzingenia*, *Integrifolia* and *Austroamericana*. Of the three species placed in subsection *Caducibracteata*, one resides primarily in Sonora (*G. turneri*), and the remaining species (*G. armourianum*, *G. harknessii*) are localized across the Gulf of California to Baja California. Subsection *Integrifolia*, is only found on the lower half of Baja California (*G. davidsonii*) and the Galapagos Islands (*G. klotzschianum*), while the sole representative of subsection *Austroamericana* (G. raimondii) occurs in Peru, and subsection *Houzingenia* that includes two species have an aggregate distribution in northern Mexico and Arizona.

The phylogeny (Fig. 3) indicates that Baja California was colonized from two indepen-

dent lineages, the first from subsection *Caducibracteata*, and the second from subsection *Integrifolia*. The Galapagos Islands distribution reflects a relatively recent long-distance dispersal event across $\sim$ 800 km of ocean, giving rise to the island endemic *G. klotzschianum* (Wendel and Percival 1990). *Gossypium raimondii* (subsection *Austroamericana*), is similarly diagnosed as having arisen following a relatively recent long-distance dispersal event to Peru. Relationships between these two subsections were reported previously (Liu et al. 2001).

The geographic pattern discussed above, in conjunction with the phylogenetic evidence (Figs. 1–3) suggest that a rapid radiation of the American diploid cotton lineage took place somewhere in western Mexico, and that it was followed by differentiation (Small and Wendel 2000b). The radiation might have occurred prior to the separation of Baja California from the mainland, or it could have involved at least two long-distance dispersal events from the Mexican mainland to the Baja California peninsula. The Gulf of California is thought to have formed about 7–12 MYA (Ferrari et al. 1999, Lonsdale 1991), near the time of origin of the American diploid cottons, estimated to have arisen approximately 6.7 MYA (Senchina et al. 2003). If the Gulf had already formed at the time of basal radiation, long-distance dispersal events would be implicated, although these have turned out to be surprisingly common in the evolutionary history of the genus (Wendel and Cronn 2003). Taking all evidence into consideration, our results support the previously proposed scenario for the evolutionary history of the American diploid cottons (Small and Wendel 2000b). This scenario is one of relatively rapid, basal radiation around 7 MYA in western Mexico, giving rise to different groups that evolved into the current taxonomically recognized subsections. Subsequently, at least two lineages from different groups (subsections *Caducibracteata*, and a common ancestor of subsections *Houzingenia*, *Austroamericana*, and *Integrifolia*) colonized the Baja California peninsula, with later colonizations of the Galapagos Islands and Peru.

We thank Ed Percival and Dan DeJoode for some collections of plant material, Mauricio Ulloa and James McD. Stewart for seeds of accession IA 64-1 (US72) and worthy comments on this paper. Research support was provided by the National Science Foundation and by the Spanish Ministry of Science, Education and Sports.

References

Bremer K. (1994) Branch support and tree stability. Cladistics 10: 295–304.

Brubaker C. L., Wendel J. F. (1994) Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*, Malvaceae) using nuclear Restriction-Fragment-Length-Polymorphisms (RFLPs). Amer. J. Bot. 81: 1309–1326.

Brubaker C. L., Paterson A. H., Wendel J. F. (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42: 184–203.

Cronn R., Zhao X., Paterson A. H., Wendel J. F. (1996) Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J. Molec. Evol. 42: 685–705.

Cronn R., Wendel J. F. (1998) Simple methods for isolating homeologous loci from allopolyploid genomes. Genome 41: 756–762.

Cronn R., Small R. L., Wendel J. F. (1999) Duplicated genes evolve independently after polyploid formation in cotton. Proc. Natl. Acad. Sci. U.S.A. 96: 14406–14411.

Cronn R., Small R. L., Haselkorn T., Wendel J. F. (2002) Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. Amer. J. Bot. 89: 707–725.

Cronn R., Small R. L., Haselkorn T., Wendel J. F. (2003) Cryptic repeated genomic recombination during speciation in *Gossypium gossypioides*. Evolution 57: 2475–2489.

DeJoode D. R. (1992) Molecular insights into speciation in the genus *Gossypium* L. (Malvaceae). Iowa State University, Ames.

Endrizzi J. E., Turcotte E. L., Kohel R. J. (1985) Genetics, cytology, and evolution of *Gossypium*. Adv. Genet. 23: 271–375.

Eriksson T. (1998) AutoDecay ver. 4.0. Departament of Botany, Stockholm Universtiy. Stockholm.

Ferrari L., López-Martínez M., Aguirre-Díaz G., Carrasco-Núez G. (1999) Space-time patterns of Cenozoic arc volcanism in central Mexico: From the Sierra Madre Occidental to the Mexican Volcanic Belt. Geology 27: 303–306.

Fryxell P. A. (1969) A classification of *Gossypium* L. (Malvaceae). Taxon 18: 585–591.

Fryxell P. A. (1971) Phenetic analysis and phylogeny of diploid species of *Gossypium* L. (Malvaceae). Evolution 25: 554–562.

Fryxell P. A. (1979) Genus *Gossypium* Linnaeus. In: Fryxell P. A. (ed.) The natural history of the cotton tribe. Texas A&M University Press, College Station, Texas, pp. 37–71.

Fryxell P. A. (1992) A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 2: 108–165.

Hall T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41: 95–98.

Johnson L. A., Soltis D. E. (1998) Assessing congruence: empirical examples from molecular data. In: Doyle J. J. (ed.) Molecular systematics of plants II. DNA sequencing. Kluwer Academic Publishers, Boston, pp. 297–348.

Kubicka E., Kubicka G., McMorris F. R. (1995) An algorithm to find agreement subtrees. J. Classif. 12: 91–99.

Liu Q., Brubaker C. L., Green A. G., Marshall D. R., Sharp P. J., Singh S P. (2001) Evolution of the *FAD2-1* fatty acid desaturase 5′ UTR intron and the molecular systematics of *Gossypium* (Malvaceae). Amer. J. Bot. 88: 92–102.

Lonsdale P. (1991) Structural pattern of the pacific floor offshore peninsular California. In: Dauphin J. P., Simoneit B. R. T. (eds.) The gulf and the peninsular province of the Californias. Am. Assoc. Petr. Geol., Memoir. 47: 87–125.

Mickevich M. F., Farris J. S. (1981) The implications of congruence in Menidia. Syst. Zool. 30: 351–370.

Robinson D. F., Foulds L. R. (1981) Comparison of phylogenetic trees. Math. Biosci. 53: 131–147.

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

Seelanan T., Schnabel A., Wendel J. F. (1997) Congruence and consensus in the cotton tribe (Malvaceae). Syst. Bot. 22: 259–290.

Seelanan T., Brubaker C. L., Stewart J. M., Craven L. A., Wendel J. F. (1999) Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). Syst. Bot. 24: 183–208.

Senchina D. S., Álvarez I., Cronn R., Liu B., Rong J., Noyes R. D., Paterson A. H., Wing R. A., Wilkins T. A., Wendel J. F. (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Molec. Biol. Evol. 20: 633–643.

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

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

Small R. L., Wendel J. F. (2000b) Phylogeny, duplication, and intraspecific variation of *Adh* sequences in New World diploid cottons (*Gossypium* L., Malvaceae). Mol. Phylogenet. Evol. 16: 73–84.

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

Swofford D. L. (1991) When are phylogeny estimates from molecular and morphological data incongruent? In: Cracraft J. (ed.) Phylogenetic analysis of DNA sequences. Oxford University Press, New York, pp. 295–333.

Swofford D. L. (1999) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.02b. Sinauer, Sunderland, MA.

Wendel J. F. (1995) Cotton. In: Simmonds N. W. (ed.) Evolution of crop plants. Longman Scientific & Technical, Essex, England, pp. 358–366.

Wendel J. F., Percival A. E. (1990) Molecular divergence in the Galapagos Island-Baja California species pair, *Gossypium klotzschianum* and *G. davidsonii* (Malvaceae). Plant Syst. Evol. 171: 99–115.

Wendel J. F., Albert V. A. (1992) Phylogenetics of the cotton genus (*Gossypium*). Character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. Syst. Bot. 17: 115–143.

Wendel J. F., Schnabel A., Seelanan T. (1995a) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc. Natl. Acad. Sci. U.S.A. 92: 280–284.

Wendel J. F., Schnabel A., Seelanan T. (1995b) An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. Molec. Phylogenet. Evol. 4: 298–313.

Wendel J. F., Cronn R., Johnston J. S., Price H. J. (2002) Feast and famine in plant genomes. Genetica 115: 37–47.

Wendel J. F., Cronn R. (2003) Polyploidy and the evolutionary history of cotton. Adv. Agron. 78: 139–186.

Addresses of the authors: Inés Álvarez (e-mail: ines@ma-rjb.csic.es), Real Jardín Botánico de Madrid, CSIC, Plaza de Murillo, 2, 28014 Madrid, Spain. Richard Cronn (e-mail: rcronn@fs.fed.us), Pacific Northwest Research Station, USDA Forest Service, 3200 SW Jefferson Way, Corvallis, Oregon 97331, USA. Jonathan F. Wendel (e-mail: jfw@iastate.edu), Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, 50011, USA.