American Journal of Botany 85(12): 1753–1765. 1998.

CLADISTIC BIOGEOGRAPHY OF *GLEDITSIA* (LEGUMINOSAE) BASED ON *NDHF* AND *RPL16* CHLOROPLAST GENE SEQUENCES[1]

ANDREW SCHNABEL[2,3] AND JONATHAN F. WENDEL[4]

[2]Department of Biological Sciences, Indiana University South Bend, South Bend, Indiana 46634; and
[4]Department of Botany, Iowa State University, Ames, Iowa 50011

We used cladistic analysis of chloroplast gene sequences (*ndhF* and *rpl16*) to test biogeographic hypotheses in the woody genus *Gleditsia*. Previous morphological comparisons suggested the presence of two eastern Asian-eastern North American species pairs among the 13 known species, as well as other intra- and inter-continental disjunctions. Results from phylogenetic analyses, interpreted in light of the amount of sequence divergence observed, led to the following conclusions. First, there is a fundamental division of the genus into three clades, only one of which contains both Asian and North American species. Second, the widespread and polymorphic Asian species, *G. japonica*, is sister to the two North American species, *G. triacanthos* and *G. aquatica*, which themselves are closely related inter se, but are both polymorphic and paraphyletic. Third, the lone South American *Gleditsia* species, *G. amorphoides*, forms a clade with two eastern Asian species. *Gleditsia* thus appears to have only one Asian-North American disjunction and no intercontinental species pairs. Low sequence divergence between *G. amorphoides* and its closest Asian relatives implicates long-distance dispersal in the origin of this unusual disjunction. Sequence divergence between Asian and North American *Gleditsia* is much lower than between Asian and North American species of its closest relative, *Gymnocladus*. Estimates of Asian-North American divergence times for *Gymnocladus* are in general accordance with fossil data, but estimates for *Gleditsia* suggest recent divergences that conflict with ages of known North American *Gleditsia* fossils.

Key words: biogeography; chloroplast DNA; *Gleditsia*; *Gymnocladus*; Leguminosae; phylogeny.

The remarkable floristic similarities between temperate eastern Asia and eastern North America have been widely recognized for over a century (Graham, 1972; Boufford and Spongberg, 1983). Approximately 120 genera, greater than one-third of which are woody, exhibit eastern Asian-eastern North American disjunctions (Wu, 1983). Although Thomas Nuttall and Asa Gray are credited with the original idea that these disjunctions represent separate refugia of a once more widespread biota (Boufford and Spongberg, 1983), this notion was formalized later in Chaney's (1947) hypothesis of an ancient, largely unchanging ''Arcto-Tertiary Geoflora'' that was fragmented as it was forced southward by progressive cooling during the Tertiary. This paleofloristic scenario, therefore, postulates that a single vicariant event was responsible for the genesis of numerous eastern Asian-eastern North American disjunctions.

Over the past three decades, however, the concept of an Arcto-Tertiary Geoflora has been challenged by a wealth of woody plant macrofossils that support a hypothesis of much more complex and dynamic climatic and floristic changes in the northern latitudes during the Tertiary (Wolfe, 1978, 1985; Tiffney, 1985a, b). Many details of Northern hemisphere paleofloristics remain unclear, but it is known that deciduous woody taxa first appear in northern latitudes as part of a mostly broad-leaved evergreen, tropical forest in the late Eocene (Wolfe, 1969, 1972), and that cooling climates during the Oligocene and Miocene saw a diversification and expansion of broad-leaved, deciduous taxa throughout the northern latitudes of Eurasia and North America (Wolfe, 1978, 1985). Migrations between Eurasia and North America were facilitated by the presence of the Bering land bridge that connected eastern Asia with western North America throughout the mid-Tertiary and by a North Atlantic land bridge during the late Eocene (Hamilton, 1983; Hsü, 1983; McKenna, 1983; Tiffney, 1985b). These deciduous taxa, along with a smaller number of broad-leaved evergreens that were able to adapt to the cooler climate of the Neogene, formed the antecedents of the taxa currently comprising the disjunct, mixed mesophytic forests of eastern Asia and eastern North America. Continued climatic cooling in the Pliocene (2–5 million years before present [mybp]) produced a contraction of the mixed mesophytic forest and greatly reduced the possibility of migration between Eurasian and North American populations (Wolfe, 1978, 1985; Tiffney, 1985b). Further climatic changes during the Quaternary, culminating in repeated Pleistocene glaciations, effectively eliminated the mixed mesophytic forests of northern Europe, central Asia, and western North America, leaving eastern North America, eastern Asia, and to a much lesser extent, the Balkans and Caucasus, as the main refugia of many genera (Graham, 1972; Tiffney, 1985b).

[1] Manuscript received 19 February 1998; revision accepted 7 May 1998.

The authors thank K. W. Cheung, M.-G. Chung, E. Hedborn, N. T. Hiep, T. C. Khánh, G.-C. Ling, R. McDonald, B. Middleton, M. Moore, K.-W. Park, R. M. Polhill, J. Quigley, F. Santamour, and R. Vanni for sending plant material; S. Bandyopadhyay, U. C. Bhattacharyya, D. L. Dilcher, C. B. Heiser, P. S. Herendeen, K. V. Krutovskii, Z.-Y. Li, R. M. Polhill, C. Niyomdham, J. Wen, and D. Zhang for discussions and correspondences; W. C. Jordan for supplying initial aliquots of the *rpl16* PCR primers, designed by M. W. Courtney, and sharing his protocol with us; and J. J. Doyle, P. S. Herendeen, and J. Wen whose comments on an earlier version of this manuscript were very helpful. This work was supported by a grant from the National Science Foundation (DEB-9203214).

[3] Author for correspondence.

As emphasized by several authors, the diversity of taxa involved in these Asian-North American disjunctions, the spatial and temporal fluidity of plant communities, and the dynamic nature of climatic changes throughout the Oligocene and Miocene, combine to suggest that current distribution patterns may have several origins that involve both vicariance and long-distance dispersal (Wolfe, 1981; Tiffney, 1985a; Qiu, Parks, and Chase, 1995; Wen, Jansen, and Kilgore, 1996). During the past decade, two novel approaches have been taken to distinguish between the competing biogeographic hypotheses of vicariance and dispersal. First, because a vicariance hypothesis predicts concordance between the age of a species pair and the age of the barriers between them, whereas a dispersal hypothesis predicts a younger age for such species pairs relative to the barrier (Platnick and Nelson, 1978; Humphries and Parenti, 1986), several studies have attempted to use levels of genetic differentiation between well-established disjunct species pairs to estimate time since divergence of the two lineages (Liston, Rieseberg, and Elias, 1989; Parks and Wendel, 1990; Crawford, Lee, and Stuessy, 1992; Wen and Jansen, 1995; Lee et al., 1996). Second, a growing number of studies have combined phylogenetic analysis with genetic divergence or fossil data to piece together the biogeographic histories of entire genera or sections of genera (Hoey and Parks, 1991; Wen and Stuessy, 1993; Qiu, Chase, and Parks, 1995; Qiu, Parks, and Chase, 1995; Wen and Zimmer, 1996; Wen, Jansen, and Kilgore, 1996; Wen, Jansen, and Zimmer, 1996; Sang, Crawford, and Stuessy, 1997).

Biogeographic interpretations appear to be especially complicated when disjunctions involve more than two species or when more than a single vicariad species pair has been hypothesized within a given genus (Wolfe, 1981; Qiu, Chase, and Parks, 1995; Wen, Jansen, and Zimmer, 1996). An example of such a case is described herein for the genus *Gleditsia* (Leguminosae: Caesalpinioideae). Previous taxonomic work suggests the presence of two eastern Asian-eastern North American vicariad species pairs in *Gleditsia*, in addition to other intra- and intercontinental disjunctions (Gordon, 1966). To test these biogeographic hypotheses, we have used cladistic analysis of DNA sequence data from two chloroplast genes. The results support only a single Asian-North American disjunction within the genus, no Asian-American species pairs, and an unusual Asian-South American disjunction.

MATERIALS AND METHODS

The genus Gleditsia—Throughout this study, we follow the taxonomy of Gordon (1966), who determined that *Gleditsia* comprises 13 species: eight species inhabiting various regions of eastern Asia (*G. australis, G. delavayi, G. fera, G. japonica, G. macracantha, G. microphylla, G. rolfei*, and *G. sinensis*), one narrowly distributed species found in southern Azerbaijan and northern Iran (*G. caspica*), one species from the northern area of northeastern India (*G. assamica*), two species occurring in the eastern United States (*G. aquatica, G. triacanthos*), and a final species growing in south-central South America (*G. amorphoides*). *Gleditsia japonica* is recognized as a complex taxon with up to four subspecies, two of which have been created by reducing *G. delavayi* and *G. caspica* to subspecific status (Li, 1982; Paclt, 1982a, 1984).

Other studies since Gordon (1966) have recognized two narrowly endemic species, *G. pachycarpa* in Vietnam (Larsen, Larsen, and Vidal, 1980) and *G. medogensis* (Ni, 1987) in Tibet. We consider *G. medogensis* to be particularly suspicious, because there is only a single description of it in the literature (Ni, 1987), it is not mentioned in Flora Reipublicae Popularis Sinicae (Li, 1988), and it was not mentioned in any of our personal communications with Chinese legume taxonomists. *Gleditsia pachycarpa* similarly appears to be a dubiously independent evolutionary unit, in that Larsen, Larsen, and Vidal (1980) recognized the species based on examination of two specimens, both of which lacked flowers. Gordon (1966) considered *G. pachycarpa* to be conspecific with *G. fera*.

All but one of the *Gleditsia* species are thorny, deciduous trees (the Chinese *G. microphylla* is a thorny shrub up to 3 m in height) with a breeding system that is characterized by polygamodioecy and insect pollination. Chromosome counts are available for seven of the species, all of which indicate that *Gleditsia* species are diploids with a base chromosome number of $x = 14$ (Atchison, 1949; Goldblatt, 1981). Several authors have noted a high degree of morphological similarity among *Gleditsia* species (Gordon, 1966; Isely, 1975; Tucker, 1991). Data on crossing ability between species is limited to evidence of successful crosses between *G. triacanthos* and four other species: *G. aquatica* (Gordon, 1966), *G. amorphoides* (Gordon, 1966), *G. japonica* (Santamour, 1976), and *G. caspica* (A. Schnabel and K. V. Krutovskii, unpublished data).

Despite the fact that no cladistic analysis of *Gleditsia* has been published, the discussion of Gordon (1966) provides a ''scenario'' that leads to several specific phylogenetic hypotheses. Of particular importance to the question of intercontinental disjunctions are two putative clades involving eastern Asian and eastern North American species. First, Gordon (1966) proposed that *G. microphylla* (China) and *G. aquatica* (eastern United States) are closest relatives, because the fruits of these species lack the sweet pulp found in the fruits of all other *Gleditsia* species and because they have many fewer ovules per carpel (1–4 vs. >10 for all other species). Second, Gordon (1966) viewed *G. japonica, G. delavayi,* and *G. caspica* as close relatives with affinities to the second eastern North American species, *G. triacanthos,* a hypothesis that was later reiterated by Isely (1975). *Gleditsia triacanthos* (honeylocust) is the best known and most widely studied species in the genus. It grows naturally throughout the eastern United States and has been promoted and used as a fodder tree, a wind-row tree, and an ornamental and shade tree in the United States and other countries, such as Australia, Azerbaijan, New Zealand, Russia, and Ukraine (Gold and Hanover, 1993; Csurhes and Kriticos, 1994; A. Schnabel, personal observations).

Taxon sampling—We included 11 of the 13 *Gleditsia* species described by Gordon (1966). For those 11 species, we obtained accessions from a variety of sources, and we included multiple accessions whenever possible. For example, to begin to understand the evolutionary history of *G. japonica,* we sampled accessions of this species from China, Korea, and Japan (Table 1). To minimize the possibility of species misidentification, we used accessions collected by us or for us directly from natural populations, or we tried to use only those samples from botanical gardens and arboreta that had voucher information clearly indicating that the specimen was collected from the wild. Nonetheless, although we obtained two putative *G. macracantha* accessions from two botanical gardens, our DNA sequencing suggested that both of them were probably *G. triacanthos.* Several authors suggest that *G. macracantha* may not be distinct from *G. sinensis* or at least should be very similar to *G. sinensis* (Gordon, 1966; Isely, 1975; Paclt, 1982b; Li, 1988). Because our sequences showed clear differences between *G. triacanthos* and *G. sinensis* but no differences between *G. triacanthos* and putative *G. macracantha,* we concluded that the two *G. macracantha* accessions had been misidentified, and we excluded them from further analysis. In addition, we were unable to obtain tissue samples of the narrowly distributed species, *G. assamica,* from northeastern India. Published reports indicate that *G. assamica* may be extinct or highly

TABLE 1. *Gleditsia* and *Gymnocladus* accessions used to obtain *ndhF* and *rpl16* sequences. All voucher specimens are deposited in ISC, unless otherwise noted in parentheses.

Taxon	Voucher or arboretum accession no.	Origin of material	Collection location
Genus *Gleditsia* L.			
G. triacanthos L.	Small 159	Cultivated	Ames, Iowa, USA
G. triacanthos	M. Moore 2602	Wild collected	Athens, Georgia, USA
G. triacanthos	B. Middleton s.n.	Wild collected	Carbondale, Illinois, USA
G. aquatica Marsh	Schnabel 9	Wild collected	Felsenthal National Wildlife Refuge, Arkansas, USA
G. aquatica	No voucher available	Wild collected	Four Holes Swamp, South Carolina, USA
G. aquatica	No voucher available	Wild collected	Edisto River, South Carolina, USA
G. aquatica	B. Middleton s.n.	Wild collected	Perks, Illinois, USA
G. amorphoides (Gris.) Taub.	R. Vanni & G. Lopez 3291	Wild collected	Corrientes, Argentina
G. caspica Desf.	Arnold Arboretum #461-78	Wild collected	Ariamehr Botanical Garden, Tehran, Iran
G. caspica	Schnabel 10	Wild collected	Astara, Azerbaijan
G. japonica Miq.	Arnold Arboretum #13-38	Wild collected	China
G. japonica	Schnabel 11	Wild collected	Mt. Myonggi, South Korea
G. japonica	Schnabel 4	Cultivated	Gyeongsang National University, South Korea
G. japonica	S. Tsugaru & M. Sawada 14667 (Missouri Bot. Gard.)	Wild collected	Honshu, Japan
G. japonica	Arnold Arboretum #357-81	Wild collected	Hangzhou Botanical Garden, China
G. delavayi Franch.	Schnabel 1	Wild collected	Kunming, China
G. sinensis Lam.	Morton Arboretum #294-81	Wild collected	China
G. microphylla Gordon	Schnabel 7	Unknown	Tashkent, Uzbekistan
G. microphylla	Schnabel 8	Unknown	Tashkent, Uzbekistan
G. rolfei Vid.	No voucher available	Wild collected	Kenting National Park, Taiwan
G. fera (Lour.) Merr.	Schnabel 2	Unknown	Kowloon, Hong Kong
G. australis Hemsl.	Schnabel 3	Unknown	Hanoi, Vietnam
G. australis	Schnabel 6	Unknown	Hanoi, Vietnam
Genus *Gymnocladus* Lam.			
G. dioica (L.) Koch	Small 160	Cultivated	Ames, Iowa, USA
G. chinensis Baill.	Schnabel 5	Wild collected	Kunming, China

endangered (Sanjappa, 1990), and it apparently has neither been collected for many years nor been preserved in local arboreta (U. C. Bhattacharayya, personal communication).

Seed accessions were germinated and grown in the Iowa State University greenhouse, and voucher specimens are deposited in the Ada Hayden Herbarium (ISC) at Iowa State University. The clear outgroup for this study was *Gymnocladus*, which has long been considered to be sister to *Gleditsia*, the two genera forming a recognizable ''*Gleditsia* group'' within Tribe Caesalpineae (Gordon, 1966; Polhill and Vidal, 1981; Polhill, 1994; Doyle et al., 1997). *Gymnocladus* itself has an eastern North American-eastern Asian distribution, and in the phylogenetic analyses, we included one accession each of the North American species, *Gymnocladus dioica* (L.) Koch, and an Asian species, *Gymnocladus chinensis* Baill.

PCR amplification and DNA sequencing—We extracted total cellular DNA from fresh, frozen, or silica-gel desiccated leaves using methods detailed in Paterson, Brubaker, and Wendel (1993). The polymerase chain reaction (PCR) was employed to amplify large portions of the *ndhF* and *rpl16* chloroplast genes for DNA sequencing. The amplification primers and internal sequencing primers for *ndhF* are described in Olmstead and Sweere (1994). Because of low-yield amplifications using their 5′F primer, we designed an additional primer that begins 16 base pairs (bp) downstream from the 5′ end of the gene (5′-GAATATGCATGGATCATACC-3′). We initially amplified the gene in two overlapping segments using primer pairs 16F-1318R and 803F-2110R. These double-stranded amplifications were performed in a 50-μL volume containing 10 mmol/L Tris-HCl pH 9.0, 50 mmol/L KCl, 0.1% Triton X-100, 1.5 mmol/L $MgCl_2$, 0.25 mmol/L of each dNTP, and 0.2 μmol/L of each primer. Aliquots (5 μL) of the double-stranded PCR products were used directly as templates in asymmetric amplifications to generate single-stranded *ndhF* DNA for sequencing (Kaltenboeck et al., 1992). Reaction components for the asymmetric amplifications were identical to those for symmetric amplifications, except that we used only one of the two primers and we doubled the total volume of each reaction mixture. For the symmetric amplifications, the PCR cycling parameters were 1.5 min at 94°C for denaturation, 2 min at 42°C for primer annealing, and 3 min at 72°C for extension. A final extension period of 10 min at 72°C was added after the completion of 30 cycles. The asymmetric amplification conditions were nearly identical, except that the annealing temperature was raised to 48°C, and the number of cycles was decreased to 20.

The *rpl16* gene contains two exons separated by an intron that varies in length from ~1000 bp to 1500 bp in the species studied to date (Posno, van Vliet, and Groot, 1986; Tanaka et al., 1986; Jordan, Courtney, and Neigel, 1996; Kelchner and Wendel, 1996). From this gene, we amplified nearly the entire intron and a large portion of the 3′ exon using the F71 and R1661 primers of Jordan, Courtney, and Neigel (1996). The F71 primer begins 3 bp upstream of the short, three-codon exon at the 5′ end of the gene, and the R1661 primer begins at base 167 of the 3′ exon. Reaction components for these amplifications were generally the same as for the *ndhF* amplifications, except that we used 0.4 μmol/L of each primer in both symmetric and asymmetric amplifications. Cycling parameters, designed to avoid problems of mismatching and premature denaturation in the highly AT-rich intron, were 1 min at 95°C for denaturation, 1 min at 50°C for primer annealing, a slow increase in temperature (8 s/°C) to 65°C, and 4 min at 65°C for extension. A 5-min denaturation period at 95°C was included prior to the first cycle, and a final 10-min extension period at 65°C was added after the final cycle. Symmetric and asymmetric amplifications were run for 35 and 25 cycles, respectively.

Single-stranded DNAs were purified using Microcon 100 ultrafilters (Amicon, Beverly, Massachusetts) prior to sequencing. We sequenced both *ndhF* and *rpl16* using standard methods of dideoxy sequencing

with Sequenase 2.0 (United States Biochemical, Cleveland, Ohio) and α-35S]dATP, electrophoresis in Long Ranger (AT Biochem, Malvern, Pennsylvania) sequencing gels, and autoradiography. All internal sequencing primers for *ndhF* are specified in Olmstead and Sweere (1994). Initial sequencing of *rpl16* was accomplished using the PCR primers and a single internal primer (F435; Jordan, Courtney, and Neigel, 1996). We subsequently designed four internal primers; coding strand (forward) primers are F220 (5′-CTGATTATGAGTTGTGAAGC-3′), F590 (5′-GCGGAAGGAACCAAGAAC-3′), and F1034 (5′-TCCAATAGACCAATAGAT-3′), whereas R270 (5′-TCACCCTTT-CATTTATCC-3′) is a complementary strand (reverse) primer. The resulting 17 unique *ndhF* sequences have been deposited in GenBank under the accession numbers GBANAF020469–GBANAF020485, and 19 unique *rpl16* sequences have been deposited under the accession numbers GBANAF040726–GBANAF040744.

Phylogenetic analyses—We conducted phylogenetic analyses on matrices of aligned DNA sequences for each gene separately and then combined the two data sets for a final analysis. Alignment of the *ndhF* sequences was trivial, as only two insertion/deletion events (indels) needed to be postulated, each of which involved a single AAA codon. For the *rpl16* data, an initial alignment produced by CLUSTALW version 1.6 (Thompson, Higgins, and Gibson, 1994) was adjusted manually to produce a final alignment that postulated 19 indels, the lengths of which ranged from one to 32 nucleotides. All maximum parsimony analyses were performed using PAUP version 3.1.1 (Swofford, 1993). For the analyses of individual *ndhF* and *rpl16* data sets, we performed heuristic searches, with 1000 random stepwise additions, tree bisection-reconnection branch swapping, and the MULPARS and furthest addition sequence options on. The analysis of the combined data involved fewer accessions, so we were able to search for the shortest trees using the branch-and-bound algorithm. In all analyses, nucleotide transformations within characters were equally weighted, and postulated indels were either treated as missing data or were additionally reintroduced into the data set as presence-absence characters. For the results based on combined data, the strength of individual clades was evaluated using decay analysis (Bremer, 1988; Donoghue et al., 1992), in which we constructed strict consensus trees for all topologies that were up to five steps longer than the most parsimonious trees. For comparisons of divergence levels, we used the program MEGA (Kumar, Tamura, and Nei, 1993) to calculate Kimura two-parameter distances (Kimura, 1980) between all pairs of combined *ndhF-rpl16* sequences. In the calculation of distances, nucleotide positions that involved gaps and missing data in one or more accessions were removed only in pairwise comparisons.

RESULTS

Variability at ndhF—We obtained nearly complete *ndhF* nucleotide sequences for two accessions of *Gymnocladus, G. dioica* and *G. chinensis,* and 25 accessions of *Gleditsia,* representing 11 of the 13 species recognized by Gordon (1966). Five of the *Gleditsia* accessions that we received either as seeds or silica-gel dried leaves turned out to have *ndhF* and *rpl16* gene sequences that were identical to accessions of other species. Based on those data, we decided that those five accessions had been incorrectly identified, and we excluded them from all further analyses. In particular, as discussed above, both of our *G. macracantha* accessions had sequences identical to *G. triacanthos,* so *G. macracantha* is not represented in our study.

The portion of *ndhF* sequenced in *Gleditsia* and *Gymnocladus* represents 2095–2101 nucleotides of the gene, beginning at a position equivalent to nucleotide 36 of the tobacco *ndhF* gene. Length differences result from an apparent insertion or deletion of one or two phenylalanine codons (AAA) within a region of the gene (beginning at nucleotide position 1454 of the *Gleditsia* sequence) that has a continuous string of ten adenine nucleotides in *Gymnocladus dioica,* 17 adenines in *Gymnocladus chinensis,* and 14 or 15 adenines in all *Gleditsia* species.

In *Gleditsia* and *Gymnocladus,* the *ndhF* gene has an average GC content of 31%. Excluding indels, 84 of the 2095 nucleotide sites (4.2%) were variable within the entire data set, but the *Gleditsia* sequences varied at only 44 sites (2.1%), and only 18 (0.9%) of those were potentially phylogenetically informative. The frequency of variable sites was not random along the length of the gene, but tended to be higher in the latter third of the sequences (41 of 84 variable sites). We also found an increased proportion of nonsynonymous substitutions in the 3′ half of the sequences, where 29% of 34 substitutions were nonsynonymous in the first 1050 nucleotides and 62% of the 50 substitutions were nonsynonymous in the second 1050 nucleotides.

Multiple accessions were sequenced for *G. aquatica* (3), *G. australis* (2), *G. japonica* (4), *G. microphylla* (2), and *G. triacanthos* (3). From these sequences, we observed two cases of intraspecific variation. The two *G. australis* accessions differed by a single nucleotide substitution, and, although the sequences of the Korean and Japanese accessions of *G. japonica* were identical to one another, they differed from the Chinese *G. japonica* accessions by three substitutions. Most importantly from a phylogenetic perspective, all accessions of the North American species, *G. triacanthos* and *G. aquatica,* were identical to one another.

Overall, the *Gleditsia* accessions showed low levels of *ndhF* sequence divergence, as estimated by Kimura two-parameter distances with pairwise deletion. Excluding *G. microphylla,* which showed sequence divergence of 0.6–1.0% from all other members of the genus, the Asian *Gleditsia* species all differed from one another by 0.5% or less. Similarly, estimates of sequence differences between Asian and North American accessions varied from 0.3 to 0.5%. On the other hand, Asian and North American species of *Gymnocladus,* with a sequence divergence estimate of 1.5%, were considerably more differentiated from one another than were any Asian/North American pair of *Gleditsia.*

Phylogenetic analyses of ndhF—The heuristic search identified two most parsimonious trees, each with a length of 95, a consistency index of 0.86, excluding uninformative characters, and a retention index of 0.92. In both trees, *Gleditsia* and *Gymnocladus* are well differentiated, and *Gleditsia* has three major clades, each of which is supported by a single synapomorphy (Fig. 1). The largest clade posits that the North American species, *G. triacanthos* and *G. aquatica,* are sister to the species of the Asian *G. japonica* group (*G. japonica, G. caspica,* and *G. delavayi*). The second clade pairs the South American species, *G. amorphoides,* with a clade containing two Asian species, *G. sinensis* and *G. rolfei.* The only difference between the two shortest trees was in the relationships among species of the third major clade. In one arrangement, identical to the strict consensus (Fig. 1), the southeast Asian species, *G. australis* and *G. fera,* form a

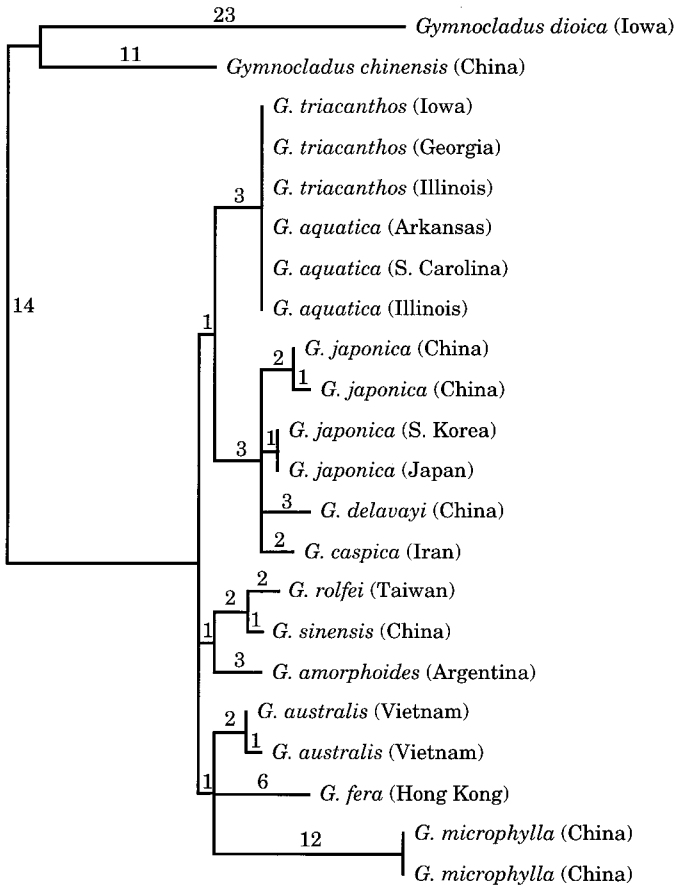

Fig. 1. Relationships among 11 species of *Gleditsia* based on cladistic analysis of the *ndhF* chloroplast gene sequences. The strict consensus of the two most parsimonious trees uncovered by a heuristic search is shown. Numbers above branches indicate minimum character support.

trichotomy with the northern Chinese species, *G. microphylla,* whereas in the second possible arrangement, *G. australis* and *G. microphylla* appear as a sister clade to *G. fera.*

Variability in rpl16—For the analysis of *rpl16* sequences, we used the same 22 accessions analyzed for *ndhF* and added one new accession each of *G. aquatica, G. japonica,* and *G. caspica.* The aligned sequences of these 25 accessions represented all but the first 13 bp of the *rpl16* intron (which overlap with the primer used for PCR amplification) and 125 of the 399 nucleotides in the 3′ exon. Alignment required the hypothesis of 19 insertion/deletion events ranging from 1 to 32 bp in length. Fourteen of those gaps involve exact or nearly exact duplications of immediately adjacent or nearby sequences, whereas five of the hypothesized gaps are most likely short deletions. The aligned length of the *rpl16* intron alone is 1245 nucleotides, with actual lengths varying from 1078 nucleotides in the *Gymnocladus dioica* accession to 1167 nucleotides in one of the *Gleditsia triacanthos* accessions.

Like the *ndhF* gene, the *rpl16* gene (exons + intron) has a GC content of ~31% in *Gleditsia* and *Gymnocladus.* Excluding the 19 potential indels, 5.9% (64/1078) of the nucleotide positions were variable within the data set as a whole. All but two of those variable sites were found within the intron region. For *Gleditsia* alone, 2.9% (31/1078) of the sites were polymorphic, and 2.0% (22/1078) were potentially phylogenetically informative, one of which was within the 3′ exon region. The *rpl16* intron thus provided slightly more nucleotide substitution data than the *ndhF* gene.

As with the *ndhF* sequences, levels of intraspecific polymorphism in *rpl16* were low, but the pattern of variation differed from *ndhF* in two respects. First, no sequence polymorphism was found within *G. japonica,* despite the addition of a fifth accession. Second, sequence variation was found within both *G. triacanthos* and *G. aquatica.* Two of the *G. triacanthos* sequences (from Illinois and Georgia) differed at a single nucleotide position from the Iowa accession, whereas one of the *G. aquatica* accessions from South Carolina differed by two substitutions from the other South Carolina accession and from the Illinois and Arkansas accessions. A complex pattern of variation in putative indels was also seen within each of the North American species (Fig. 2).

Levels of sequence divergence between species were slightly higher for *rpl16* than for *ndhF.* The Kimura two-parameter estimate for the two *Gymnocladus* accessions was 2.1%, whereas average sequence divergence between the two genera was 2.6% (range of 2.1–3.1%). The two North American *Gleditsia* species exhibited divergence of only 0.1%, and average divergence among the Asian species (0.7%) was approximately equal to divergence between American and Asian species (0.8%). The greatest level of divergence within *Gleditsia* involved the South American species, *G. amorphoides,* which differed by 0.8–1.5% from all species except *G. rolfei* and *G. sinensis,* to which its *rpl16* sequence is very similar (0.5% divergence).

Phylogenetic analyses of rpl16 sequences—We performed two heuristic searches for the most parsimonious relationships among the 25 *rpl16* sequences. In both analyses, hypothesized indels were treated as missing data, but in one of the analyses, all 19 indels, regardless of length, were reintroduced into the data set as presence/absence characters. The first analysis recovered six most-parsimonious trees, each with a length of 72, consistency index of 0.87, excluding uninformative characters, and retention index of 0.95. The second analysis, using phylogenetic information available in gapped positions, resulted in 96 shortest trees, each with a length of 94, a consistency index of 0.83, excluding uninformative characters, and a retention index of 0.94. The increase in number of shortest trees in the second analysis was due entirely to an increase in the number of most parsimonious relationships among the *G. triacanthos* and *G. aquatica* sequences. In the first analysis, the same relationship among the seven North American accessions was found in all six shortest trees, whereas the second analysis uncovered 16 most parsimonious explanations of the substitution and indel data for the North American accessions. In none of the trees from either analysis do *G. triacanthos* and *G. aquatica* accessions form separate clades.

Other than the *G. triacanthos* and *G. aquatica* se-

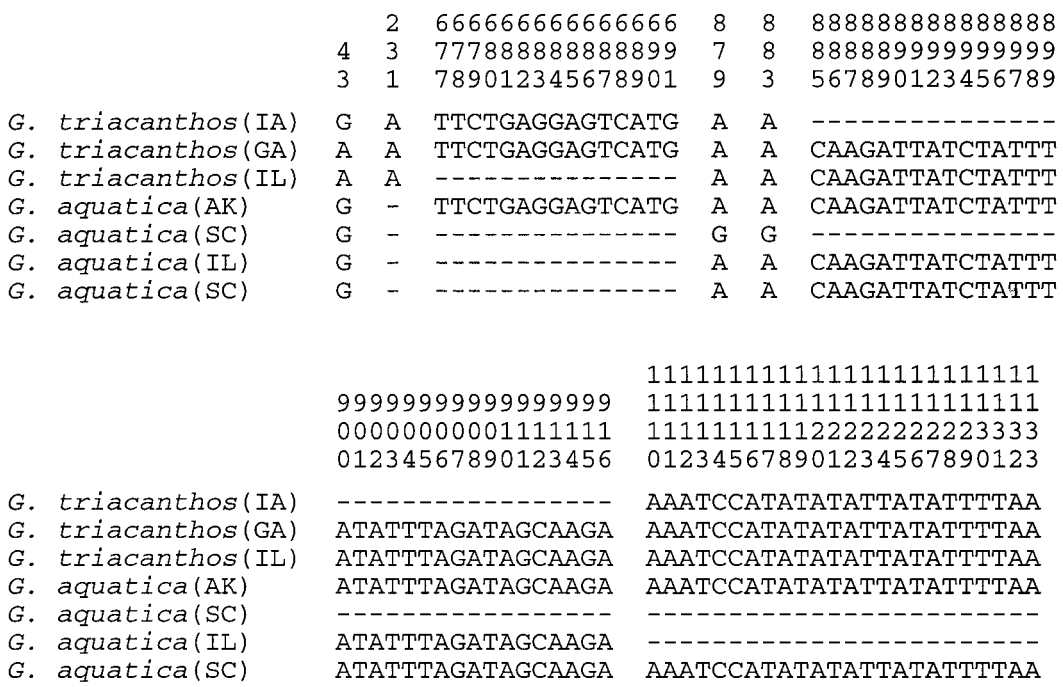

Fig. 2. Patterns of three nucleotide substitution and five indel polymorphisms for the *rpl16* intron in *Gleditsia triacanthos* and *G. aquatica.* Numbers above each column indicate position in the aligned data set of all 25 accessions analyzed. Abbreviations in parentheses indicate the states within the eastern United States where accessions were collected.

quences, the main sources of ambiguity in the *rpl16* data are the sequences of the three Asian species, *G. australis, G. microphylla,* and *G. fera,* which show the same four possible arrangements in both analyses. In one of these topologies (Fig. 3a), which is nearly identical to that obtained using the *ndhF* sequences (Fig. 1), the three species form a separate clade that is part of a larger trichotomy. In the other three topologies (Fig. 3b–d), and in strong contrast to the *ndhF* trees, either two or three of these Asian species appear as sister to the *G. japonica*/*G. triacanthos*/*G. aquatica* clade. The strict consensus trees from the two analyses were identical except for the relationships between the North American accessions, and we therefore present only the consensus from the analysis with gaps included as data (Fig. 4).

Despite the lack of resolution with respect to some relationships, all *rpl16* trees show several similarities to the topologies found in analyses of the *ndhF* sequences (cf. Figs. 1 and 4). First, *Gymnocladus* and *Gleditsia* can be rooted so as to appear as separate, well-differentiated clades. Second, the *G. japonica* group of species forms a major clade with the two North American species, *G. triacanthos* and *G. aquatica.* Third, *G. triacanthos* and *G. aquatica* show a very close relationship, although in the case of *rpl16,* there are both substitution and indel differences within and between the species, and one of the indel differences is species specific (Fig. 2). Fourth, *G. amorphoides,* from Argentina, forms a well-supported clade with the two Asian species, *G. sinensis* and *G. rolfei.*

Combined analysis of ndhF and rpl16 sequences—Given the many similarities between the *ndhF* and *rpl16* trees, and the fact that the two chloroplast genes are part of the same historically nonrecombinant molecule, we felt justified in combining data sets and performing a joint analysis. For this analysis, we used only those accessions for which we had both *ndhF* and *rpl16* sequences, and we merged all identical sequences of the same species. Moreover, because no new insights into the relationships between *G. triacanthos* and *G. aquatica* could be gained by joint analysis of the two genes (all *G. triacanthos* and *G. aquatica* sequences were identical at *ndhF,* and all phylogenetic information was found solely in the *rpl16* data), we eliminated all but one each of the *G. triacanthos* and *G. aquatica* sequences. The final combined data set, therefore, contained 15 *Gleditsia* and two *Gymnocladus* sequences, which were analyzed using a branch-and-bound algorithm, with gaps treated as presence-absence characters. That analysis uncovered four most parsimonious trees, each of which has a length of 184, a consistency index of 0.86, excluding uninformative characters, and a retention index of 0.92. The four trees differ only in the placement of the *G. australis*/*G. microphylla*/*G. fera* clade and in the relationships among those three species. In three of the topologies, the *G. australis*/*G. microphylla*/*G. fera* clade is placed sister to the *G. triacanthos*/*G. aquatica*/*G. japonica* clade, but that grouping is supported by only one synapomorphy. Those three trees vary only in the hypothesized relationships among *G. australis*/*G. microphylla*/*G. fera,* in that *G. microphylla* is placed sister to either *G. australis* or *G. fera,* or the three species form an unresolved trichotomy. The fourth most parsimonious topology shows the three major clades within the genus in an unresolved trichotomy, which is the same arrangement shown in the strict consensus of all four trees (Fig. 5).

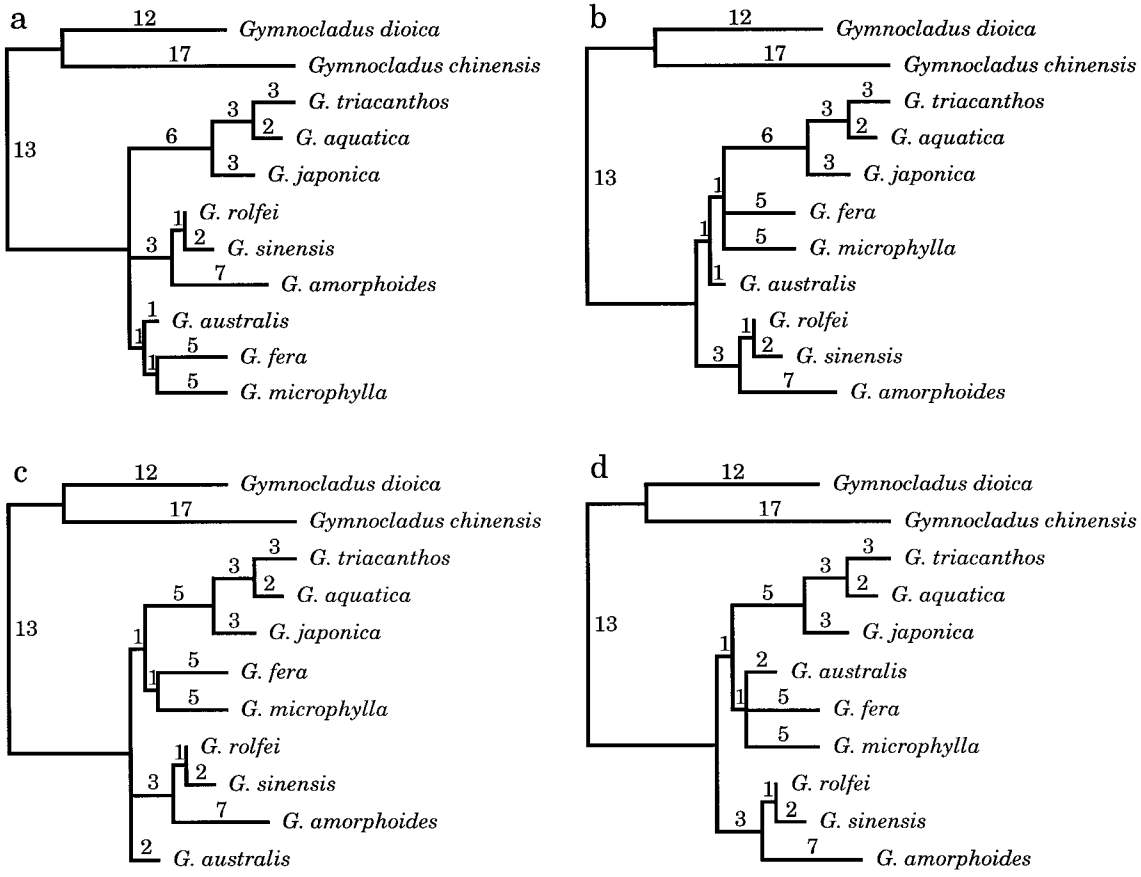

Fig. 3. Cladograms showing four equally parsimonious arrangements of *Gleditsia australis, G. fera,* and *G. microphylla* resulting from cladistic analysis of *rpl16* gene sequence data. For the sake of clarity, *Gleditsia delavayi* and *G. caspica* were not included in this analysis, and only one accession was used from each of the remaining nine species. Numbers of supporting characters are given above branches.

DISCUSSION

Our *ndhF* and *rpl16* chloroplast sequence data both support and conflict with previous phylogenetic hypotheses of intra- and intercontinental disjunctions in *Gleditsia,* and they also raise new hypotheses not previously considered. In particular, the cladistic analyses provide insights into (1) the origin of the genus as a whole; (2) biogeographic origins of the single South American species, *G. amorphoides*; and (3) intracontinental biogeography of the *G. japonica* group. Most importantly from the standpoint of eastern Asian-eastern North American disjunctions, our results suggest only a single Asian-North American disjunction and no intercontinental sister-species relationships, thereby conflicting with Gordon (1966), whose morphological comparisons suggested the presence of two Asian-North American species pairs within the genus.

Timing of biogeographic events—As a supplement to considerations of the tree topologies, we can use estimates of molecular divergence and a molecular clock hypothesis to place some bounds on the timing of important evolutionary and biogeographic events in *Gleditsia* and *Gymnocladus.* As emphasized by several authors (e.g., Li, 1993; Hillis, Mable, and Moritz, 1996), the variance associated with estimates of divergence times based on molecular data can be very large for a number of reasons, such as a lack of proper calibration of nucleotide substitution rates, taxon- and gene-specific substitution rates, and heterogeneity of substitution rates among lineages within a taxon. For example, nearly 15-fold variation in divergence levels based on cpDNA restriction sites and fourfold variation in genetic identities based on allozyme data have been reported for eastern Asian-eastern North American disjunct taxa (Qiu, Parks, and Chase, 1995; Lee et al., 1996; Wen, Jansen, and Zimmer, 1996), and it is not clear how much of this rate heterogeneity is due to factors other than the timing of divergence between the disjuncts.

Time since divergence is calculated by dividing the Kimura distances by twice the rate of nucleotide substitution. Because the rates of substitution have not been calibrated for either *ndhF* or *rpl16,* however, it is unclear what rates are appropriate for estimating divergence times in this case. Olmstead and Sweere (1994) hypothesized that *ndhF* evolves at about twice the rate of *rbcL,* which has an estimated average substitution rate of 2×10^{-10} substitutions per site per year (Albert et al., 1994), so we have chosen to use a rate of 4×10^{-10} substitutions per site per year for calculations based on that gene. The *rpl16* intron is more problematic, because little is known about its evolution. Based on estimates presented in

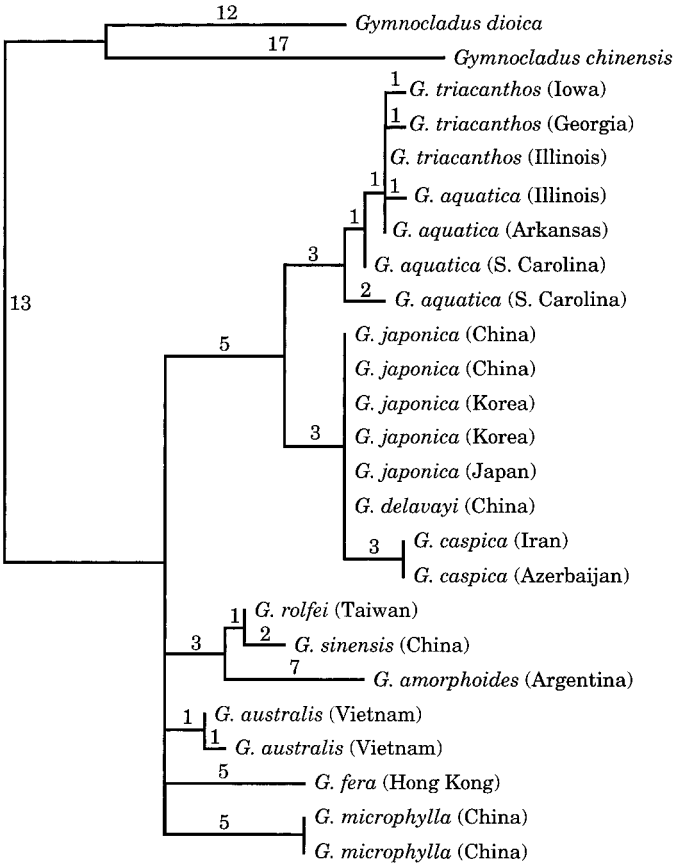

Fig. 4. Relationships among 11 species of *Gleditsia* suggested by cladistic analysis of *rpl16* chloroplast gene sequences. The strict consensus of the 96 most parsimonious trees uncovered by a heuristic search is shown. Numbers above branches indicate minimum character support.

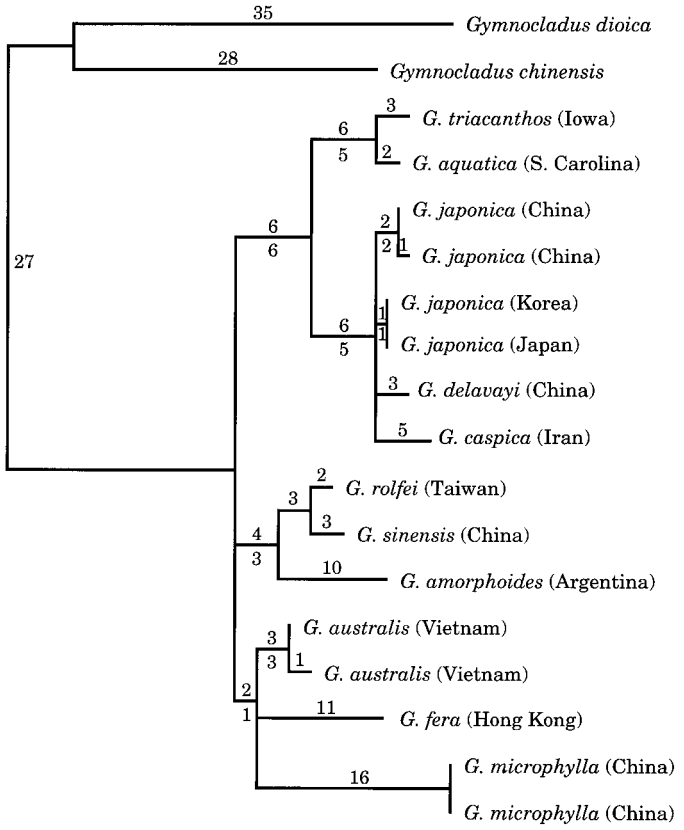

Fig. 5. Relationships among 11 species of *Gleditsia* based on cladistic analysis of combined *ndhF* and *rpl16* gene sequence data. The strict consensus of the four most parsimonious trees uncovered by a branch-and-bound search is shown. Numbers above branches are minimum number of supporting characters, and numbers below branches are decay indices.

Palmer (1991), the average rate of substitutions across a large number of chloroplast genes is $\sim 5 \times 10^{-10}$ substitutions per site per year. Because this average includes both synonymous and nonsynonymous substitutions, it probably underestimates substitution rates for strictly noncoding regions, like the *rpl16* intron. This hypothesis is supported by estimates of percentage divergence between *Gleditsia* taxa, which are 30–100% higher for the *rpl16* intron than for *ndhF*. Eyre-Walker and Gaut (1997) report a synonymous substitution rate at *rbc*L for grasses of 12×10^{-10} substitutions per site per year, whereas palms have a much lower substitution rate of $\sim 3.5 \times 10^{-10}$ substitutions per site per year. An average of these two rates is $\sim 8 \times 10^{-10}$ subtitutions per site per year, but we might expect synonymous substitution rates in long-lived, woody genera, like *Gleditsia* and *Gymnocladus*, to be slightly less than that, because species with long generation times often have lower substitution rates than species with shorter generation times (Gaut et al., 1992; Li, 1993; Qiu, Parks, and Chase, 1995; Li et al., 1996; Gaut et al., 1997). We therefore chose to use a rate of 6×10^{-10} substitutions per site per year in all calculations of divergence times based on *rpl16* sequences.

Given the many caveats listed above, we can proceed with caution in calculating divergence times for some of the major bifurcations shown in Fig. 4, beginning with the divergence of *Gleditsia* and *Gymnocladus.* Based largely on floral morphology and ontogeny, these genera were for many years considered to be possible examples of the most primitive genera within the Leguminosae (Gordon, 1966; Polhill, Raven, and Stirton, 1981; Tucker, 1991). Recent morphological and molecular phylogenetic analyses, however, negate that hypothesis, indicating instead that these genera were not involved in the earliest dichotomies within the Leguminosae (summarized in Doyle et al., 1997). In agreement with this conclusion are reevaluations of legume fossils, which demonstrate that although most major lineages within the family were present by the Eocene (35–56 mybp), the oldest reliable fossil *Gymnocladus* and *Gleditsia* are from Oligocene sediments in North America (23–35 mybp; Herendeen, Crepet, and Dilcher, 1992). The 1.7 and 2.6% sequence divergences between *Gleditsia* and *Gymnocladus* for *ndhF* and *rpl16,* respectively, suggest that the two genera diverged 21–22 mybp, in reasonably close agreement with the fossil data.

If *Gleditsia* and *Gymnocladus* diverged from a common ancestor between 23 and 35 mybp, there would have been ample opportunity for the genera to become distributed in both North America and Asia, because a land bridge across the Bering Sea was in continuous existence

from the middle Eocene through the late Miocene (Tiffney, 1985a). Up until the very latest part of the Miocene, climates in Alaska were apparently warm enough to support broad-leaved, deciduous taxa (Wolfe, 1972, 1978; Tiffney, 1985a), and underlying temperature fluctuations may have produced three especially favorable periods for dispersal (Wolfe, 1978). Thus, Asian-North American distributions within *Gleditsia* and *Gymnocladus* could have been attained at very different times. This possibility is suggested by the sequence data from both *ndhF* and *rpl16,* which show that sequence differences between *Gymnocladus dioica* and *G. chinensis* (1.5% for *ndhF* and 2.1% for *rpl16*) are 2–5 times as great as differences between Asian and North American *Gleditsia* (0.3–0.8% for *ndhF* and 0.4–1.2% for *rpl16*). These distance estimates for *Gymnocladus* translate into a minimum range of divergence times of 17.5–18.8 mybp, or only shortly after the initial split between the two genera. On the other hand, estimates of divergence times for even the most genetically distant Asian and North American *Gleditsia* species (*G. triacanthos* vs. *G. microphylla*) are only 10.2 mybp based on *ndhF* data and 9.7 mybp based on *rpl16* data and are much less for *G. japonica* and *G. triacanthos*/*G. aquatica* (4.7 and 3.9 mybp, based on *ndhF* and *rpl16,* respectively).

Although these recent divergence times in the late Miocene and early Pliocene conveniently span the final of three proposed warm periods (5–6 mybp) in the Northern Hemisphere during the Oligocene and Miocene (Wolfe, 1978), they are inconsistent with present fossil data. The oldest fossil *Gleditsia* are Oligocene in origin and are found in North America (Herendeen, Crepet, and Dilcher, 1992), but North American *Gleditsia* species are not basal in any of the trees recovered by cladistic analyses (Figs. 1, 4, 5). The phylogenetic results instead support three main lineages within the genus, only one of which includes both Asian and North American species. Our results thus imply that the most basal divergence within the genus probably involved only Asian species. In support of this hypothesis, the legume fossils from China include two distinct *Gleditsia* species from Miocene sediments of eastern China, *G. parajaponica* Guo & Zhou and *G. miosinensis* Hu & Chaney (Guo and Zhou, 1992). The names of these fossil species reflect their similarities to the extant *G. japonica* and *G. sinensis,* both of which are prominent members of the two most strongly supported clades within the genus (Fig. 5). Moreover, fossil *Gleditsia* is known from Miocene deposits in Japan (Tanai, 1972), and a fossil species, *G. allemanica* Heer, which shows close affinities to *G. caspica,* is found in Miocene deposits of Georgia, Armenia, and the northern Caucasus (Shakryl, 1992). The *G. japonica* lineage thus had been in existence long enough in the Miocene to become distributed from Japan to the northern Caucasus. The fossil record from Asia and North America, therefore, disagrees with our estimates of intercontinental divergence times and indicates instead that *Gleditsia* was probably widespread in both Asia and North America throughout the Miocene.

Either of two hypotheses could begin to resolve the conflict between the fossil and molecular data. First, the earliest North American lineages of *Gleditsia* from the Oligocene may have become extinct, with the genus reestablishing itself in North America at a later point in the Tertiary, perhaps as recently as 4–5 mybp, as suggested by the molecular divergence data. *Gleditsia* is known from several Miocene deposits in North America (e.g., McCartan et al., 1990), but as for nearly all taxa, its fossil record is far from complete. Second, the rate of sequence evolution in *Gleditsia* may be significantly slower than in *Gymnocladus,* which would result in underestimates of divergence times. Significant differences in substitution rates among angiosperm taxa are well documented for chloroplast genes (e.g., Bousquet et al., 1992; Gaut et al., 1992; Barraclough, Harvey, and Nee, 1996), although we note that the two genera under consideration share most of the life-history traits that are thought to influence substitution rates. Testing this hypothesis would require a more complete fossil record, especially for the Oligocene in China and the Miocene and Pliocene in North America.

Origin of G. amorphoides in South America—One of the most puzzling biogeographic questions in *Gleditsia* is the presence of a single South American species, *G. amorphoides,* in Paraguay, Uraguay, southernmost Brazil, and northern Argentina. Raven and Polhill (1981), who viewed this phenomenon as a disjunction between North and South America, noted that such disjunctions involving woody genera are rare and their origins are poorly understood. Gordon (1966), however, hinted that *G. amorphoides* may be closely related either to *G. microphylla* (based on presence of persistent sepaloid bracts subtending the flowers) or to *G. rolfei* (based on leaf characters). Our data support this latter hypothesis in that *G. amorphoides* falls sister to *G. sinensis* and *G. rolfei,* which show little divergence from one another. The presence of *G. amorphoides* in South America therefore appears to represent a temperate South American-Asian disjunction, which is considered to be even rarer than those between North and South America (Thorne, 1972).

Three possible hypotheses could explain the origin of an eastern Asian-eastern South American disjunction within the genus. First, Raven and Polhill (1981) proposed that *Gleditsia* reached South America by way of Africa in the early Tertiary. Thus, if *Gleditsia* arose in Asia, it may have dispersed to Africa (where there are neither extant nor fossil *Gleditsia* species) and subsequently to South America. This hypothesis, however, rests on the assumption that *Gleditsia* is a very old genus (evolving in the early Tertiary) and possibly is among the most basal lineages within the entire family (Polhill, Raven, and Stirton, 1981). Contradicting this proposal are (1) recent cladistic analyses that place *Gleditsia* and *Gymnocladus* among the more recently diverged lineages of the Caesalpinioideae (Tucker and Douglas, 1994; Chappill, 1995; Doyle et al., 1997); (2) reevaluations of the fossil evidence, which indicate that *Gleditsia* and *Gymnocladus* may have arisen as late as 35 mybp in the Oligocene (Herendeen, Crepet, and Dilcher, 1992); and (3) low sequence divergence between *G. amorphoides* and all Asian species, which suggests a much more recent divergence, possibly in the upper Miocene. Second, *G. amorphoides* may have originated from a North American ancestor that in turn was closely related to the Asian lineage from which *G. sinensis* and *G. rolfei* arose. Al-

though this scenario would allow for a more recent divergence consistent with both fossil and molecular data, it suffers from an incomplete fossil record for the Miocene and Pliocene of North and South America and from the lack of a close relationship between *G. amorphoides* and the extant North American species. Third, this disjunction may have arisen through long-distance dispersal across the Pacific Ocean from Asia to South America. The fruits of *G. rolfei* and *G. sinensis* have no obvious adaptations for water dispersal, and the viability of *Gleditsia* seeds after long exposure to seawater is not known, but the presence of *G. rolfei* on the Chinese mainland as well as on the islands of the Philippines and Celebes perhaps implicates water as a repeatedly successful means of long-distance dispersal within this lineage. Moreover, if the substitution rates we are using to calculate divergence times are close to actual rates, then the divergence time between these highly disjunct lineages is estimated to be 5.6–7.8 mybp. Because at that time, the continents were approximately in their present positions, a scenario of long-distance dispersal would fit the present data.

Relationships within the G. japonica group—Considerable controversy has arisen over the classification of *G. japonica* and its close relatives in Asia. Gordon (1966) recognized three separate species (*G. japonica, G. delavayi,* and *G. caspica*) and considered *G. japonica* to comprise two subspecies, *G. japonica* ssp. *japonica* and *G. japonica* ssp. *stenocarpa,* where the latter subspecies is confined to the Korean Peninsula and the former inhabits China, Korea, and Japan. Li (1982) subsequently reclassified *G. delavayi* as one of three Chinese subspecies of *G. japonica* (*G. japonica* ssp. *japonica, G. japonica* ssp. *velutina,* and *G. japonica* ssp. *delavayi*). Paclt (1982a, 1984) also reduced *G. caspica* to a fourth subspecies of *G. japonica.* In doing so, Paclt (1982a) argued that (1) the original range of *G. japonica* probably stretched from Japan and Korea to Transcaucasia and possibly Himalaya; (2) *G. caspica* represents a remnant of that once broader distribution, which has contracted markedly as a result both of several Pleistocene glaciations and of agricultural activities in China; and (3) the natural postglacial range of *G. japonica* var. *japonica* was restricted to Korea and Japan with subsequent reintroduction to China by humans.

Although our sampling of this diverse group was admittedly limited, our results generally agree with Paclt's hypotheses (Fig. 5). First, all accessions within the group form a single, well-supported clade, but weaker resolution was possible among the six accessions within that clade for which we have both *ndhF* and *rpl16* data. Second, *G. caspica,* which according to Paclt's scenario would have been isolated from *G. japonica* for much of the Pleistocene and which is now found only in a small area near the southern and southwestern edges of the Caspian Sea, was the most divergent taxon within the group, differing by 5–8 substitutions and indels from the other species. Third, the *ndhF* data distinguish a Chinese clade and a Korean/Japanese clade among the four *G. japonica* accessions (Fig. 1), as would be expected if Korea and Japan served as glacial refugia one or more times during the Pleistocene. More extensive sampling of this group needs to be conducted to obtain a clearer picture of its evolutionary history.

Relationship between G. triacanthos and G. aquatica—It has been suggested that *G. triacanthos* is sister to *G. japonica* (Isely, 1975), or that it forms a sister relationship with the clade composed of *G. japonica, G. caspica,* and *G. delavayi* (Gordon, 1966). In contrast, although our phylogenetic analyses place *G. triacanthos* and the *G. japonica* group within one of three major clades in the genus, *G. triacanthos* appears to have its closest relationship with the only other North American species, *G. aquatica.* Thus, this result also conflicts with the hypothesis that *G. microphylla* and *G. aquatica* form a second Asian-North American species pair (Gordon, 1966).

Either of two independent, although not mutually exclusive, hypotheses could explain the relationship between *G. triacanthos* and *G. aquatica* suggested by the *rpl16* data. The first explanation assumes that the cpDNA data are recovering the correct relationship between these two species in that *G. triacanthos* and *G. aquatica* are sister to the Asian *G. japonica.* If this is the case, the identity of the two North American species at *ndhF* and the low sequence divergence (0.1%) at *rpl16* would indicate a recent separation from a common ancestor, and the paraphyly suggested by the *rpl16* data (with only one species-specific indel between *G. triacanthos* and *G. aquatica*) would then be ascribed to incomplete lineage sorting following that separation. More specifically, given that (1) *G. aquatica* is the only species within the genus to have adapted to permanently flooded habitats such as swamps and sloughs, and (2) *G. triacanthos* more closely resembles *G. japonica* morphologically than does *G. aquatica* (Gordon, 1966; Isely, 1975), it seems reasonable under this scenario to assume that *G. aquatica* evolved from an ancestor closely resembling *G. triacanthos.* Although the range of *G. triacanthos* has greatly expanded since the time of European settlement, it was originally reported mainly as a member of streamside and bottomland forest communities (Blair, 1989), and Gordon (1966) reported that ''this species attains its greatest size along the Wabash River Valley in southern Indiana and Illinois.'' Moreover, small-scale experiments by Heiser (1985) suggest that *G. triacanthos* seedlings have short-term tolerance to flooding or at least to fully saturated soils. An evolutionary shift to tolerance of permanent flooding, therefore, might not have required large genetic changes and may have happened relatively rapidly.

If such an adaptive shift did take place, it was accompanied by rapid evolution in fruit morphology. The elongate fruits of *G. triacanthos* resemble those of most species within the genus in having many seeds (>10 seeds), a thick, tough wall, and a sugary pulp, whereas the fruits of *G. aquatica* are 1–2 seeded and dry with thinner, papery walls. Much of the evolution in fruit morphology in *G. aquatica* has been assumed to result from adaptation for water dispersal (Gordon, 1966), suggesting a selective regime for the evolution of novel fruit traits. Heiser (1985) reported, however, that the fruits of both species can remain afloat for up to 2 wk.

A second explanation for the close relationship between *G. triacanthos* and *G. aquatica* assumes that the

two species are in fact more distantly related, as suggested by Gordon (1966) and that the cpDNA data do not recover the correct relationship due to interspecific hybridization followed by introgression of cytoplasmic genes from one species into the other. Such asymmetric introgression is well documented in plants and is often the source of phylogenetic incongruence between data sets (Rieseberg, Whitton, and Linder, 1996; Wendel and Doyle, 1998). Several authors have suggested the sporadic occurrence of natural hybrids between *G. triacanthos* and *G. aquatica* (Sargent, 1922; Gordon, 1966; Isely, 1975; Heiser, 1985). Formal designation of these putative hybrids as *G.* × *texana* Sarg. emphasizes the location of the original description in southeastern Texas (Sargent, 1922), but Gordon (1966) noted that rare hybrids have also been identified in Arkansas, Mississippi, Louisiana, and southern Indiana, which are areas where the native ranges of the two species overlap (Vines, 1960; Isely, 1975; Blair, 1989). Both Gordon (1966) and Isely (1975) state that *G.* × *texana* is intermediate between *G. triacanthos* and *G. aquatica* for many morphological characters, which suggests that *G.* × *texana* populations consist mostly of first-generation hybrids and that backcrossing and introgression of genes in either direction are rare. Heiser (1985) alone hypothesizes that introgression of genes from *G. aquatica* may partially account for the great morphological diversity seen across the range of *G. triacanthos.*

Of these two major scenarios, recent divergence with subsequent incomplete lineage sorting appears to be the simplest explanation of all current morphological, molecular, and ecological data. Because we sampled both species within the area of range overlap (southern Illinois and Arkansas), as well as in areas where only one or the other species is found (Iowa, South Carolina coastal plain, Georgia piedmont), the alternative hybridization scenario would imply either much more extensive hybridization than suggested by previous morphological identification of hybrids or rapid spread of cytoplasmic genes of one species across the entire range of the other despite the rarity of hybridization events to initiate this spread. Moreover, if *G. aquatica* is truly sister to *G. microphylla,* hybridization between *G. aquatica* and *G. triacanthos* would imply successful mating between two of the most genetically distant lineages within the genus (Fig. 5). We know of no crossing experiments that have tested the viability of hybrids between *G. microphylla* and either *G. aquatica* or *G. triacanthos.*

We therefore propose that only one eastern Asian-eastern North American disjunction exists within *Gleditsia* and that the morphological similarities suggested by Gordon (1966) to be indicative of a close relationship between *G. aquatica* and *G. microphylla* (small, few-seeded pods that lack pulp) instead imply convergent evolution under very different selective regimes. Verification of this proposal, as well as the other biogeographic conclusions reached here, will emerge from additional studies employing appropriate nuclear genetic markers.

LITERATURE CITED

ALBERT, V., A. BACKLUND, K. BREMER, M. W. CHASE, J. R. MANHART, B. D. MISCHLER, AND K. NIXON. 1994. Functional constraints and *rbc*L evidence for land plant phylogeny. *Annals of the Missouri Botanical Garden* 81: 534–567.

ATCHISON, E. 1949. Studies in the Leguminosae. IV. Chromosome numbers and geographical relationships of miscellaneous Leguminosae. *Journal of the Elisha Mitchell Society* 65: 118–122.

BARRACLOUGH, T. G., P. H. HARVEY, AND S. NEE. 1996. Rate of *rbc*L gene sequence evolution and species diversification in flowering plants. *Proceedings of the Royal Society of London B* 263: 589–591.

BLAIR, R. M. 1989. *Gleditsia triacanthos* L. (honeylocust). *In* R. M. Burns and B. H. Honkala (tech. coords.), Silvics of North America 2. Hardwoods. Agricultural Handbook 654, 358–364. U.S. Department of Agriculture, Forest Service, Washington, DC.

BOUFFORD, D. E., AND S. A. SPONGBERG. 1983. Eastern Asian-eastern North American phytogeographical relationships—a history from the time of Linnaeus to the twentieth century. *Annals of the Missouri Botanical Garden* 70: 423–439.

BOUSQUET, J., S. H. STRAUSS, A. H. DOERKSEN, AND R. A. PRICE. 1992. Extensive variation in evolutionary rate of *rbc*L gene sequences among seed plants. *Proceedings of the National Academy of Sciences, USA* 89: 7844–7848.

BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* 42: 795–803.

CHANEY, R. W. 1947. Tertiary centers and migration routes. *Ecological Monographs* 17: 139–148.

CHAPPILL, J. A. 1995. Cladistic analysis of the Leguminosae: the development of an explicit phylogenetic hypothesis. *In* M. Crisp and J. J. Doyle [eds.], Advances in legume systematics, part 7, phylogeny, 1–9. Royal Botanic Gardens, Kew.

CRAWFORD, D. J., N. S. LEE, AND T. F. STUESSY. 1992. Plant species disjunctions: perspectives from molecular data. *Aliso* 13: 395–409.

CSURHES, S. M., AND D. KRITICOS. 1994. *Gleditsia triacanthos* L. (Caesalpiniaceae), another thorny, exotic fodder tree gone wild. *Plant Protection Quarterly* 9: 101–105.

DONOGHUE, M. J., R. G. OLMSTEAD, J. F. SMITH, AND J. D. PALMER. 1992. Phylogenetic relationships of Dipsicales based on *rbc*L sequences. *Annals of the Missouri Botanical Garden* 79: 333–345.

DOYLE, J. J., J. L. DOYLE, J. A. BALLENGER, E. E. DICKSON, T. KAJITA, AND H. OHASHI. 1997. A phylogeny of the chloroplast gene *rbc*L in the Leguminosae: taxonomic correlations and insights into the evolution of nodulation. *American Journal of Botany* 84: 541–554.

EYRE-WALKER, A., AND B. S. GAUT. 1997. Correlated rates of synonymous site evolution across plant genomes. *Molecular Biology and Evolution* 14: 455–460.

GAUT, B. S., L. G. CLARK, J. F. WENDEL, AND S. V. MUSE. 1997. Comparisons of the molecular evolutionary process at *rbc*L and *ndhF* in the grass family (Poaceae). *Molecular Biology and Evolution* 14: 769–777.

———, S. V. MUSE, W. D. CLARK, AND M. T. CLEGG. 1992. Relative rates of nucleotide substitution at the *rbc*L locus of monocotyledonous plants. *Journal of Molecular Evolution* 35: 292–303.

GOLD, M. A., AND J. W. HANOVER. 1993. Honeylocust (*Gleditsia triacanthos*), a multipurpose tree for the temperate zone. *International Tree Crops Journal* 7: 189–207.

GOLDBLATT, P. 1981. Cytology and phylogeny of Leguminosae. *In* R. M. Polhill and P. H. Raven [eds.], Advances in legume systematics, part 1, 427–463. Royal Botanic Gardens, Kew.

GUO, S.-X., AND Z.-K. ZHOU. 1992. The megafossil legumes from China. *In* P. S. Herendeen and D. L. Dilcher [eds.], Advances in legume systematics, part 4, the fossil record, 207–223. Royal Botanic Gardens, Kew.

GORDON, D. 1966. A revision of the genus *Gleditsia* (Leguminosae). Ph.D. dissertation, Indiana University, Bloomington, IN.

GRAHAM, A. 1972. Outline of the origin and historical recognition of floristic affinities between Asia and eastern North America. *In* A. Graham [ed.], Floristics and paleofloristics of Asia and eastern North America, 1–18. Elsevier, Amsterdam.

HAMILTON, W. 1983. Cretaceous and Cenozoic history of the northern continents. *Annals of the Missouri Botanical Garden* 70: 440–458.

HERENDEEN, P. S., W. L. CREPET, AND D. L. DILCHER. 1992. The fossil history of the leguminosae: phylogenetic and biogeographic implications. *In* P. S. Herendeen and D. L. Dilcher [eds.], Advances in

legume systematics, part 4, the fossil record. Royal Botanic Gardens, Kew.

HEISER, C. B., JR. 1985. Of plants and people. University of Oklahoma Press, Norman, OK.

HILLIS, D. M., B. K. MABLE, AND C. MORITZ. 1996. Applications of molecular systematics. *In* D. M. Hillis, C. Moritz, and B. K. Mable [eds.], Molecular systematics, 515–543. Sinauer, Sunderland, MA.

HOEY, M. T., AND C. R. PARKS. 1991. Isozyme divergence between eastern Asian, North American, and Turkish species of *Liquidambar* (Hamamelidaceae). *American Journal of Botany* 78: 938–947.

HSÜ, J. 1983. Late Cretaceous and Cenozoic vegetation in China, emphasizing their connections with North America. *Annals of the Missouri Botanical Garden* 70: 490–508.

HUMPHRIES, C. J., AND L. R. PARENTI. 1986. Cladistic biogeography. Clarendon, Oxford.

ISELY, D. 1975. Leguminosae of the United States: II. Subfamily Caesalpinioideae. *Memoirs of the New York Botanical Garden* 25: 1–228.

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

KALTENBOECK, B., J. W. SPATAFORA, X. ZHANG, K. G. KOUSOULAS, M. BLACKWELL, AND J. STORZ. 1992. Efficient production of single-stranded DNA as long as 2 kb for sequencing of PCR-amplified DNA. *BioTechniques* 12: 164–171.

KELCHNER, S. A., AND J. F. WENDEL. 1996. Hairpins create minute inversions in non-coding regions of chloroplast DNA. *Current Genetics* 30: 259–262.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111–120.

KUMAR, S., K. TAMURA, AND M. NEI. 1993. MEGA: molecular evolutionary genetics analysis, version 1.01. Pennsylvania State University, University Park, PA.

LARSEN, K., S. S., LARSEN, AND J. E. VIDAL. 1980. Flore du Cambodge, du Laos, et du Viêt-Nam. 18. Légumineuses—Césalpinioïdées. Muséum National D'Histoire Naturelle, Paris.

LEE, N. S., T. SANG, D. J. CRAWFORD, S. H. YEAU, AND S.-C. KIM. 1996. Molecular divergence between disjunct taxa in eastern Asia and eastern North America. *American Journal of Botany* 83: 1373–1378.

LI, L.-C. 1982. Two new varieties of *Gleditsia japonica* Miq. *Acta Phytotaxonomica Sinica* 20: 228–229.

———. 1988. *Gleditsia*. *In* Flora Reipublicae Popularis Sinicae, Tomus 39 (Leguminosae, Caesalpinioideae), 80–90. Science Press, Beijing.

LI, W.-H. 1993. So, what about the molecular clock hypothesis? *Current Opinion in Genetics and Development* 3: 896–901.

———, D. L. ELLSWORTH, J. KRUSHKAL, B. H.-J. CHANG, AND D. H. EMMET. 1996. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Molecular Phylogenetics and Evolution* 5: 182–187.

LISTON, A., L. H. RIESEBERG, AND T. S. ELIAS. 1989. Morphological stasis and molecular divergence in the intercontinental disjunct genus *Datisca* (Datiscaceae). *Aliso* 12: 525–542.

MCCARTAN, L., B. H. TIFFNEY, J. A. WOLFE, S. L. WING, L. A. SIRKIN, L. W. WARD, AND J. BROOKS. 1990. Late Tertiary floral assemblage from upland gravel deposits of the southern Maryland Coastal Plain. *Geology* 18: 311–314.

MCKENNA, M. C. 1983. Holarctic landmass rearrangement, cosmic events and Cenozoic terrestrial organisms. *Annals of the Missouri Botanical Garden* 70: 459–489.

NI, Z.-C. 1987. New taxa of the Leguminosae from Xizang (Tibet). *Acta Phytotaxonomica Sinica* 25: 231–234.

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

PACLT, J. 1982a. *Gleditsia caspia*, not a distinct species (Leguminosae). *Taxon* 31: 336–339.

———. 1982b. On the repeatedly confused nomenclature of Chinese species of *Gleditsia* (Caesalpiniaceae). *Taxon* 31: 551–553.

——— 1984. A note on *Triaenodendron*, with new combinations in *Gleditsia* (Caesalpiniaceae). *Taxon* 33: 100–101.

PALMER, J. D. 1991. Plastid chromosome: structure and evolution. *In* L. Bogorad and I. K. Vasil [eds.], The molecular biology of plastids, 5–53. Academic Press, San Diego, CA.

PARKS, C. R., AND J. F. WENDEL. 1990. Molecular divergence between Asian and North American species of *Liriodendron* (Magnoliaceae) with implications for interpretation of fossil floras. *American Journal of Botany* 77: 1243–1256.

PATERSON, A. H., C. L. BRUBAKER, AND J. F. WENDEL. 1993. A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. *Plant Molecular Biology Reporter* 11: 112–127.

PLATNICK, N. I., AND G. NELSON. 1978. A method of analysis for historical biogeography. *Systematic Zoology* 27: 1–16.

POLHILL, R. M. 1994. Classification of the Leguminosae. *In* F. A. Bisby, J. Buckingham, and J. B. Harborne [eds.], Phytochemical dictionary of the Leguminosae, xxxv–lvii. Chapman and Hall, New York, NY.

———, P. H. RAVEN, AND C. H. STIRTON. 1981. Evolution and systematics of the Leguminosae. *In* R. M. Polhill and P. H. Raven [eds.], Advances in legume systematics, part 1, 1–26. Royal Botanic Gardens, Kew.

———, AND J. E. VIDAL. 1981. Caesalpinieae. *In* R. M. Polhill and P. H. Raven [eds.], Advances in legume systematics, part 1, 81–96. Royal Botanic Gardens, Kew.

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

QIU, Y.-L., M. W. CHASE, AND C. R. PARKS. 1995. A chloroplast DNA phylogenetic study of the eastern Asian-eastern North American disjunct section *Rytidospermum* of *Magnolia* (Magnoliaceae). *American Journal of Botany* 82: 1582–1588.

———, C. R. PARKS, AND M. W. CHASE. 1995. Molecular divergence in the eastern Asia-eastern North America disjunct section *Rytidospermum* of *Magnolia* (Magnoliaceae). *American Journal of Botany* 82: 1589–1598.

RAVEN, P. H., AND R. M. POLHILL. 1981. Biogeography of the Leguminosae. *In* R. M. Polhill and P. H. Raven [eds.], Advances in legume systematics, part 1, 27–34. Royal Botanic Gardens, Kew.

RIESEBERG, L. H., J. WHITTON, AND C. R. LINDER. 1996. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Botanica Neerlandica* 45: 243–262.

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

SANJAPPA, M. 1990. *Gleditsia assamica* Bor. *In* M. P. Nayar and A. R. K. Sastry [eds.], Red data book of Indian plants, 139–140. Botanical Survey of India, Calcutta.

SANTAMOUR, F. S. 1976. Metaxenia in interspecific honeylocust crosses. *The Journal of Heredity*. 67: 185–186.

SARGENT, C. S. 1922. Notes on North American trees. X. New species and varieties of *Crataegus*. *Journal of the Arnold Arboretum* 3: 182–207.

SHAKRYL, A. K. 1992. Leguminosae species from the Tertiary of Abkhasia. *In* P. S. Herendeen and D. L. Dilcher [eds.], Advances in legume systematics, part 4, the fossil record, 189–206. Royal Botanic Gardens, Kew.

SWOFFORD, D. L. 1993. PAUP: phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign, IL.

TANAI, T. 1972. Tertiary history of vegetation in Japan. *In* A. Graham [ed.], Floristics and paleofloristics of Asia and eastern North America, 235–255. Elsevier, Amsterdam.

TANAKA, M., T. WAKASUGI, M. SUGITA, K. SHINOZAKI, AND M. SUGIURA. 1986. Genes for eight ribosomal proteins are clustered on the chloroplast genome of tobacco (*Nicotiana tabacum*): similarity to the S10 and *spc* operons of *Escherichia coli*. *Proceedings of the National Academy of Sciences, USA* 83: 6030–6034.

THOMPSON, J. D., D. G. HIGGINS, AND T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap pen-

alties and weight matrix choice. *Nucleic Acids Research* 22: 4673–4680.

THORNE, R. F. 1972. Major disjunctions in the geographic ranges of seed plants. *The Quarterly Review of Biology* 47: 365–410.

TIFFNEY, B. H. 1985a. Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. *Journal of the Arnold Arboretium* 66: 73–94.

———. 1985b. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. *Journal of the Arnold Arboretium* 66: 243–273.

TUCKER, S. C. 1991. Helical floral organogenesis in *Gleditsia*, a primitive caesalpinioid legume. *American Journal of Botany* 78: 1130–1149.

———, AND A. W. DOUGLAS. 1994. Ontogenetic evidence and phylogenetic relationships among basal taxa of legumes. *In* I. K. Ferguson and S. C. Tucker [eds.], Advances in legume systematics, part 6, structural botany, 11–32. Royal Botanic Gardens, Kew.

VINES, R. A. 1960. Trees, shrubs, and woody vines of the Southwest. University of Texas Press, Austin, TX.

WEN, J., AND R. K. JANSEN. 1995. Morphological and molecular comparisons of *Campsis grandiflora* and *C. radicans* (Bignoniaceae), an eastern Asian and eastern North American vicariad species pair. *Plant Systematics and Evolution* 196: 173–183.

———, ———, AND K. KILGORE. 1996. Evolution of the eastern Asian and eastern North American genus *Symplocarpus* (Araceae): insights from chloroplast DNA restriction site data. *Biochemical Systematics and Ecology* 24: 735–748.

———, ———, AND E. A. ZIMMER. 1996. Phylogenetic relationships and DNA sequence divergence of Eastern Asian and Eastern North American disjunct plants. *In* M. Nei and N. Takahata [eds.], Current topics on molecular evolution, 37–44. Pennsylvania State University and The Graduate University for Advanced Studies, Hayama, Japan.

———, AND T. F. STUESSY. 1993. The phylogeny and biogeography of *Nyssa* (Cornaceae). *Systematic Botany* 18: 68–79.

———, AND E. A. ZIMMER. 1996. Phylogeny and biogeography of *Panax* L. (the ginseng genus, Araliaceae): inferences from ITS sequences of nuclear ribosomal DNA. *Molecular Phylogenetics and Evolution* 6: 167–177.

WENDEL, J. F., AND J. J. DOYLE. 1998. Phylogenetic incongruence: window into genome history and speciation. *In* D. E. Soltis, P. S. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants II. Chapman and Hall, New York, NY.

WOLFE, J. A. 1969. Neogene floristic and vegetational history of the Pacific northwest. *Madroño* 20: 83–110.

———. 1972. An interpretation of Alaskan Tertiary floras. *In* A. Graham [ed.], Floristics and paleofloristics of Asia and eastern North America, 201–233. Elsevier, Amsterdam.

———. 1978. A paleobotanical intrepretation of Tertiary climates in the Northern Hemisphere. *American Scientist* 66: 694–703.

———. 1981. Vicariance biogeography of angiosperms in relation to paleobotanical data. *In* G. J. Nelson and D. E. Rosen [eds.], Vicariance biogeography, 413–427. Columbia University Press, New York, NY.

———. 1985. Distribution of major vegetational types during the Tertiary. *In* E. T. Sundquist and W. S. Broecker [eds.], The carbon cycle and atmosphere CO_2: natural variations Archean to Present, 357–375. American Geophysical Union Geophysical Monograph 32.

WU, Z. 1983. On the significance of Pacific intercontinental discontinuity. *Annals of the Missouri Botanical Garden* 70: 577–590.