

# Phylogeny of the Cucurbitales based on DNA sequences of nine loci from three genomes: Implications for morphological and sexual system evolution

Li-Bing Zhang<sup>a,\*</sup>, Mark P. Simmons<sup>a</sup>, Alexander Kocyan<sup>b</sup>, Susanne S. Renner<sup>b</sup>

<sup>a</sup> Department of Biology, Colorado State University, Fort Collins, CO 80523, USA

<sup>b</sup> Department of Biology, Ludwig Maximilians University, Menzinger-Str. 67, D-80638 Munich, Germany

Received 1 July 2005; revised 30 September 2005; accepted 1 October 2005

Available online 15 November 2005

## Abstract

The Cucurbitales are a clade of rosids with a worldwide distribution and a striking heterogeneity in species diversity among its seven family members: the Anisophylleaceae (29–40 species), Begoniaceae (1400 spp.), Coriariaceae (15 spp.), Corynocarpaceae (6 spp.), Cucurbitaceae (800 spp.), Datisceae (2 spp.), and Tetramelaceae (2 spp.). Most Cucurbitales have unisexual flowers, and species are monoecious, dioecious, andromonoecious, or androdioecious. To resolve interfamilial relationships within the order and to polarize morphological character evolution, especially of flower sexual systems, we sequenced nine plastids (*atpB*, *matK*, *ndhF*, *rbcL*, the *trnL-F* region, and the *rpl20-rps12* spacer), nuclear (18S and 26S rDNA), and mitochondrial (*nadI* b/c intron) genes (together ~12,000 bp) of 26 representatives of the seven families plus eight outgroup taxa from six other orders of the Eurosids I. Cucurbitales are strongly supported as monophyletic and are closest to Fagales, albeit with moderate support; both together are sister to Rosales. The deepest split in the Cucurbitales is that between the Anisophylleaceae and the remaining families; next is a clade of Corynocarpaceae and Coriariaceae, followed by Cucurbitaceae, which are sister to a clade of Begoniaceae, Datisceae, and Tetramelaceae. Based on this topology, stipulate leaves, inferior ovaries, parietal placentation, and one-seeded fruits are inferred as ancestral in Cucurbitales; exstipulate leaves, superior ovaries, apical placentation, and many-seeded fruits evolved within the order. Bisexual flowers are reconstructed as ancestral, but dioecy appears to have evolved already in the common ancestor of Begoniaceae, Cucurbitaceae, Datisceae, and Tetramelaceae, and then to have been lost repeatedly in Begoniaceae and Cucurbitaceae. Both instances of androdioecy (*Datisca glomerata* and *Schizopepon bryoniifolius*) evolved from dioecious ancestors, corroborating recent hypotheses about androdioecy often evolving from dioecy.

© 2005 Elsevier Inc. All rights reserved.

**Keywords:** Dioecy; Character evolution; Cucurbitaceae; Cucurbitales; Androdioecy; Long-branch attraction; Sexual systems; Monoecy

## 1. Introduction

The core families of the order Cucurbitales—Begoniaceae, Cucurbitaceae, Datisceae, and Tetramelaceae—have long been recognized as closely related to each other (Cronquist, 1988; Dahlgren, 1983, 1988; Hutchinson,

1973; Matthews and Endress, 2004; Melchior, 1964; Takhtajan, 1969). The second-largest of these families, the Cucurbitaceae, is among the most economically important families of flowering plants and accordingly has received much attention from morphologists and, of course, plant breeders. Cucurbitaceae contain squashes, melons, gourds, and many other species that are used as food, medicine, or storage containers for dry or liquid items. Cultivars of *Citrullus lanatus* (watermelon), *Cucumis sativus* (cucumber), *Cucumis melo* (melon), and *Cucurbita pepo* (pumpkin) are widely cultivated food crops. Species of

\* Corresponding author. Present address: Department of Integrative Biology, 401 WIDB, Brigham Young University, Provo, UT 84602, USA. Fax: +1 801 422 0090.

E-mail address: [gonggashan@gmail.com](mailto:gonggashan@gmail.com) (L.-B. Zhang).

*Begonia* (Begoniaceae) are popular ornamental plants; *Datisca cannabina* (Datisceae) historically was used to produce the yellow dye for silk (Kumar and Sinha, 2004); and *Tetrameles nudiflora* (Tetramelaceae) is an important source of timber in Southeast Asia (Thin, 1997).

The order Cucurbitales, established by Dumortier (1829), is recognized in many classifications (e.g., APG, 1998, 2003; Bremer et al., 1997; Dahlgren, 1983; Hutchinson, 1973; Melchior, 1964; Takhtajan, 1969, 1997). Except for the four core families, however, the circumscription of the order has long been controversial (Appendix A); Melchior's (1964) and Takhtajan's (1969) Cucurbitales contained only the Cucurbitaceae, whereas Hutchinson's (1973) and Dahlgren's (1983) Cucurbitales also included Begoniaceae and Datisceae *sensu lato* (*s.l.*; including Tetramelaceae; Appendix A). Cronquist (1988) also put these three families together, though without giving a suprafamilial rank.

With the advent of molecular phylogenetics, it became apparent that the Cucurbitales core families are closely related to Anisophylleaceae, Coriariaceae, and Corynocarpaceae. Based on *rbcL* gene sequences, Albert et al. (1992) associated *Coriaria* with *Begonia*, and greater taxon sampling showed that both belong near Cucurbitaceae (Chase et al., 1993) as do Corynocarpaceae (Swensen, 1996). Using 16S rDNA sequences of the nitrogen-fixing and plant-symbiotic bacterium *Frankia* isolated from its hosts, Mirza et al. (1994) found a close relationship between *Coriaria* (Coriariaceae) and *Datisca* (Datisceae). Accepting these data, Bremer et al. (1997) re-circumscribed Cucurbitales to contain Begoniaceae, Coriariaceae, Corynocarpaceae, Cucurbitaceae, Datisceae, and Tetramelaceae. The APG classifications (1998, 2003; Appendix A) then added the Anisophylleaceae based on the *rbcL* studies of Setoguchi et al. (1999) and Schwarzbach and Ricklefs (2000). Subsequent analyses supported this circumscription, but either did not sample all ingroup families, provided only weak support for the order's monophyly or intra-ordinal relationships, and/or did not include closely related families as outgroups (Clement et al., 2004; Hilu et al., 2003; Savolainen et al., 2000a,b; Soltis et al., 1995, 2000, 2003; Swensen et al., 1994, 1998; Wagstaff and Dawson, 2000).

The lack of morphological cohesion of the Cucurbitales is also evident from their families having been assigned to at least 17 orders of eudicots (Cronquist, 1981, 1988; Dahlgren, 1983, 1988; Engler, 1896; Hutchinson, 1973; Melchior, 1964; Takhtajan, 1969, 1980, 1997; Thorne, 2000). The morphological heterogeneity, caused especially by the Anisophylleaceae (Matthews and Endress, 2004), parallels widely disjunct geographic ranges and a striking heterogeneity among the families in species diversity: Anisophylleaceae have 29–40 species, Begoniaceae 1400, Coriariaceae 15, Corynocarpaceae 6, Cucurbitaceae 800, Datisceae 2, and Tetramelaceae also 2.

The early fossil record of Cucurbitales is sparse, but there are seeds from the Uppermost Paleocene and Lower

Eocene London Clay [65 million years ago (mya)] that represent Cucurbitaceae (Chandler, 1964; Collinson, 1986; Collinson et al., 1993); a *Coriaria*-like flowering branchlet from the Upper Oligocene Armissan beds in France (33–34 mya; Saporta, 1865); Miocene Coriariaceae leaves from Japan (Ozaki, 1991), Miocene Coriariaceae seeds from Germany (Gregor, 1980); *Corynocarpus laevigatus*-like fruits from the Miocene of New Zealand (24 mya; Campbell, 2002); and Middle and Late Miocene pollen of Anisophylleaceae (Anderson and Muller, 1975; Morley, 1977). Woods described as *Tetrameleoxylon prenudiflora* [Datisceae *s.l.*] from the Tertiary Deccan Intertrappean beds of India (Lakhanpal and Verma, 1965; Lakhanpal, 1970) have characteristics consistent with *Tetrameles* (E. Wheeler, Department of Wood and Paper Science at North Carolina State University, personal communication to S.R., August, 2005). The North American Paleocene leaves of *Vitis lobata* (Knowlton) Brown, mentioned as possibly cucurbitaceous in Raven and Axelrod (1974), look like Vitaceae, not Cucurbitaceae (R. Burnham and S. Renner, August 2005, based on images of the type material; further study of this material is needed).

Over the past 12 years, molecular data have revealed a previously unsuspected placement of the Cucurbitales among angiosperms, namely as sister to the Fagales. This relationship was first found by Chase et al. (1993), albeit with weak support, but has since been supported with better taxon sampling (Schwarzbach and Ricklefs, 2000; Setoguchi et al., 1999; Soltis et al., 2000, 2003; Swensen et al., 1994). Molecular studies also suggested two other affinities of the Cucurbitales. Savolainen et al. (2000b), using *rbcL* sequences and nearly complete family sampling in the Eudicots, found Zygophyllales as sister to Cucurbitales, whereas Hilu et al. (2003) resolved Cucurbitales as sister to a clade of Fabales, Fagales, Rosales, and Zygophyllales. However, none of these three ordinal affinities of the Cucurbitales was well supported.

Most of the ~2250 species of Cucurbitales have unisexual flowers and are variously monoecious, dioecious, andromonoecious, or androdioecious. Compared to other ordinal-level clades of flowering plants, Cucurbitales are among the top 10 in absolute and relative frequencies of dioecious species (Renner and Ricklefs, 1995; and SSR unpublished data). Their two species-rich families, Begoniaceae (1400 spp.) and Cucurbitaceae (800 spp.), have strictly unisexual flowers (except for mutant flowers or polygamous populations). On the other hand, several of the small families have bisexual flowers, namely Anisophylleaceae, which have bisexual or unisexual flowers (Tobe and Raven, 1988), Coriariaceae, which have bisexual or unisexual flowers (Thompson and Gornall, 1995), and Corynocarpaceae, which have entirely bisexual flowers (Matthews and Endress, 2004). Within the Cucurbitaceae, most species are dioecious, a few monoecious; the opposite is the case in the Begoniaceae. The Cucurbitales also contain two of the best-studied cases of androdioecy (bisexual plants intermixed with pure males) in the angiosperms,

the Cucurbitaceae *Schizopepon bryoniifolius* (Akimoto et al., 1999; Fukuhara and Akimoto, 1999) and the Datisceae *Datisca glomerata* (Fritsch and Rieseberg, 1992; Liston et al., 1992; Rieseberg et al., 1992; Wolf et al., 2001). Androdioecy is an exceedingly rare sexual system recently shown to be favored under selective conditions involving reproductive assurance, that is, an ability for flowers to serve as pollen sources for each other when genetically different conspecific plants are rare or absent (Pannell, 2002). Reproductive assurance is greater in monoecious than dioecious individuals.

For this study, we sampled nine loci from the plastid, mitochondrial, and nuclear genomes for 26 ingroup taxa, representing all seven families, and eight outgroups to address the following questions: (1) Is the morphologically poorly defined order Cucurbitales monophyletic and what are its closest relatives among Eurosids? (2) What are the relationships among the seven families within the order? (3) What are the evolutionary transformations of key morphological characters, such as stipules, ovary position, placentation, and fruit type? (4) Are perfect or unisexual flowers more likely to be the ancestral condition in the Cucurbitales? And what is the likely ancestral sexual system from which the two androdioecious species evolved? For *Datisca glomerata*, an earlier answer to this last question was that it evolved from dioecious ancestors (Rieseberg et al., 1992), but a later study concluded that it could have evolved from monoecious or dioecious ancestors (Swensen et al., 1998) because of the unclear sister group(s) of the Datisceae.

## 2. Materials and methods

### 2.1. Sampling strategy

The 34 species included in the analysis with voucher information (for new sequences) and GenBank accession numbers are listed in Table 1. We sampled three species from two of the four genera of Anisophylleaceae; three species of both genera of Begoniaceae; four species of the monogeneric Coriariaceae; one species of the monogeneric Corynocarpaceae; 11 species from 11 of the 121 currently recognized genera of Cucurbitaceae (Jeffrey, 2005), chosen to span the root of Cucurbitaceae based on another study that sampled most of the genera (Kocyan et al., 2004); both species of Datisceae; and both species of Tetramelaceae (Table 1). Thus, except in Anisophylleaceae and Cucurbitaceae, all genera were sampled from each family. The other two genera of Anisophylleaceae have also been sequenced and were found not to affect the placement of this family (Zhang, Simmons, and Renner, unpublished data).

Combined plastid *atpB*, *rbcL*, and 18S rDNA strongly supported the monophyly of Eurosids I (APG, 1998, 2003) to which Cucurbitales belong (Soltis et al., 2000). Our outgroup sample includes representatives of six of the seven orders of Eurosids I, namely Celastrales, Fabales, Fagales, Oxalidales, Rosales, and Zygophyllales. Malpighiales were

not sampled because they are well supported as distantly related to Cucurbitales (Zhang and Simmons, 2006).

Molecular-marker sampling follows the approach that “well-supported congruent phylogenetic estimates from all three genomic compartments would result in the highest confidence of angiosperm relationships” (Barkman et al., 2000, p. 13166:) and the idea that genes that are evolving at different rates may in combination resolve different levels of a phylogenetic tree (Graybeal, 1994; Hillis, 1987; Pennington, 1996). Nine loci from the plastid, mitochondrial, and nuclear genomes that evolve at different rates (Demesure et al., 1995; Hoot et al., 1995; Kuzoff et al., 1998; Olmstead and Sweere, 1994; Soltis et al., 2000; Taberlet et al., 1991) were sampled.

### 2.2. DNA isolation, amplification, and sequencing

Total genomic DNA was isolated from silica-dried leaves or herbarium material using DNeasy Plant Mini kits (Qiagen, Valencia, CA), NucleoSpin-Plant kits (Macherey-Nagel, Düren, Germany), or the CTAB method (Doyle and Doyle, 1987), with 4% CTAB used instead of 2% CTAB. When using the kits, the time of incubation at 65 °C was extended to 20 min and 500 µl extraction buffer/100 mg material was used. EB buffer instead of AE buffer was used for elution with the DNeasy kits to avoid possible inhibition of DNA amplifications caused by EDTA. DNA amplification was performed following the protocol described in Zhang and Renner (2003).

The plastid ATP synthase  $\beta$  subunit (*atpB*) gene was amplified with forward primers S385R, S766R, S1186R, and S1494R and reverse primers rbcL-1, S2, S20, S335, S611, and S1022 (Hoot et al., 1995). The plastid maturase K (*matK*) gene was amplified with primers trnK710 (Johnson and Soltis, 1995), trnK-2R (Steele and Vilgalys, 1994), matK-AF and matK-8R (Ooi et al., 1995; matK-8R is complement to matK8 of Steele and Vilgalys, 1994), and newly designed primers matK-299F (GGRTTYKCRV TYATTKTGGAATTCC), matK-639F (TCCTATATA ATTYTYATGTRTVBRAAT), matK-880F (RCSTWTT YTRATRAATAARTGGAA), matK-441R (GGTATTM RTACAWCKRAYAYATAAT), matK-699R (AVGATK TTRAYCGTAAATGARAAK), and matK-1018R (GTA CYACYGAAKRATYBAGYCSCACM). The last six primers, designed for rosids, are modifications of Yokoyama et al.’s (2000) F1, F2, F3, R1, R2, and R3, respectively. One portion of approximately 630 bp of the plastid NADH dehydrogenase subunit F (*ndhF*) gene was amplified with primers 972F, 1318F, 1318R (complement to 1318F) and 1603R, 1955R (Olmstead and Sweere, 1994). When these primers did not work, primers 924F (AG CWMTTGCTCAAARGAYATTA) and 951F (RRRTT TRGCYTATTCTACMATGTC) as forward primers, and 1318R, 1785R (KMYGGVNTKAAMAATTTKGATA AK), and 1955R as reverse primers were used instead. The external primers 1F (Fay et al., 1997) and 1460R (Olmstead et al., 1992) were used to amplify plastid

Table 1  
Sources of plant material and GenBank accession numbers for the phylogenetic analysis of the Cucurbitales

Taxon	Plastid DNA							mt DNA <i>nad1</i> b/c and intron	Nuclear DNA		Provenience/voucher
	<i>atpB</i> gene	<i>matK</i> gene	<i>ndhF</i> gene	<i>rbcL</i> gene	<i>trnL</i> intron	<i>trnL-trnF</i> spacer	<i>rpl20/rps12</i> spacer		18S gene	26S gene	
Ingroup (Cucurbitales)											
Anisophylleaceae											
<i>Anisophyllea corneri</i> Ding Hou	AY968424 <sup>1</sup>	AY968444 <sup>1</sup>	AY968487 <sup>2</sup>	AF027109 <sup>3</sup>	AY968559 <sup>2</sup>	AY968375 <sup>2</sup>	AY968527 <sup>2</sup>	AY968461 <sup>1</sup>	AY968390 <sup>1</sup>	AY968401 <sup>1</sup>	<sup>1</sup> Malaysia: S. FRI 40360 (KEP); <sup>2</sup> Singapore: S. Lum s.n. (no voucher); <sup>3</sup> GenBank
<i>A. fallax</i> S. Elliot	AY935849 <sup>1</sup>	AY935923 <sup>1</sup>	AY968488 <sup>1</sup>	AF127696 <sup>2</sup>	AY968560 <sup>1</sup>	AY935779 <sup>1</sup>	AY968528 <sup>1</sup>	AY968462 <sup>1</sup>	AY929365 <sup>1</sup>	AY935807 <sup>1</sup>	<sup>1</sup> Madagascar: G. Schatz et al. 3808 (MO); <sup>2</sup> GenBank
<i>Combretocarpus rotundatus</i> Danser	AY968428 <sup>1</sup>	AY968447 <sup>1</sup>	AY968492 <sup>1</sup>	AF127698 <sup>2</sup>	AY968561 <sup>1</sup>	AY968376 <sup>1</sup>	AY968532 <sup>1</sup>	AY968465 <sup>1</sup>	AY968393 <sup>1</sup>	AY968405 <sup>1</sup>	<sup>1</sup> Indonesia: E. Mirmanto s.n. (BO); <sup>2</sup> GenBank
Begoniaceae											
<i>Begonia herbacea</i> Vell.	AY968425 <sup>1</sup>	AY968460 <sup>1</sup>	AY968489 <sup>1</sup>	U59816 <sup>2</sup>	AY968562 <sup>1</sup>	AY968377 <sup>1</sup>	AY968529 <sup>1</sup>	AY968463 <sup>1</sup>	AY968391 <sup>1</sup>	AY968402 <sup>1</sup>	<sup>1</sup> America: L. Forrest 163 (E); <sup>2</sup> GenBank
<i>B. oxyloba</i> Welw. ex Hook. f.	AY968426 <sup>1</sup>	AY968445 <sup>1</sup>	AY968490 <sup>1</sup>	U59815 <sup>3</sup>	AY968563 <sup>1</sup>	AY968378 <sup>2</sup>	AY968530 <sup>1</sup>	AY256883 <sup>3</sup>	AY968392 <sup>1</sup>	AY968403 <sup>1</sup>	<sup>1</sup> Africa: Hughes s.n. (L. Forrest 279) (E); <sup>2</sup> Tanzania: S. S. Renner 2716 (MO); <sup>3</sup> GenBank
<i>Hillebrandia sandwicensis</i> Oliv.	AY968437 <sup>1</sup>	AY968452 <sup>1</sup>	AY968504 <sup>1</sup>	U59822 <sup>2</sup>	AY968564 <sup>1</sup>	AY968379 <sup>1</sup>	AY968544 <sup>1</sup>	AY968472 <sup>1</sup>	AY968398 <sup>1</sup>	AY968416 <sup>1</sup>	<sup>1</sup> Hawaii (cult. Montreal Bot Gard); acc. 2960 57; <sup>2</sup> GenBank
Coriariaceae											
<i>Coriaria myrtifolia</i> L.	AJ235443 <sup>2</sup>	AB016459 <sup>2</sup>	AY968493 <sup>1</sup>	AY968521 <sup>1</sup>	AY091824 <sup>2</sup>	AY091824 <sup>2</sup>	AY968533 <sup>1</sup>	AY968466 <sup>1</sup>	AF206891 <sup>2</sup>	AY968406 <sup>1</sup>	<sup>1</sup> Mediterranean (cult. Munich Bot Gard); S. S. Renner 2810 (M); <sup>2</sup> GenBank
<i>C. nepalensis</i> Wall.	AY968429 <sup>1</sup>	AB016460 <sup>2</sup>	AY968494 <sup>1</sup>	AY968522 <sup>1</sup>	AY091825 <sup>2</sup>	AY091825 <sup>2</sup>	AY968534 <sup>1</sup>	AY968467 <sup>1</sup>	AY968394 <sup>1</sup>	AY968407 <sup>1</sup>	<sup>1</sup> China: D.F. Chen 1 (no voucher); <sup>2</sup> GenBank
<i>C. ruscifolia</i> L.	AY968430 <sup>2</sup>	AB016462 <sup>3</sup>	AY968495 <sup>1</sup>	AF148999 <sup>3</sup>	AY091827 <sup>3</sup>	AY968380 <sup>1</sup>	AY968535 <sup>1</sup>	AY968468 <sup>1</sup>	AY968395 <sup>1</sup>	AY968408 <sup>2</sup>	<sup>1</sup> Mexico: M. Olson 836 (MEXU); <sup>2</sup> Mexico: G. Flores-Franco et al. 4420 (MO); <sup>3</sup> GenBank
<i>C. sarmentosa</i> Forst. f.	AY968431 <sup>1</sup>	AB016464 <sup>2</sup>	AY968496 <sup>1</sup>	AF149000 <sup>2</sup>	AY091829 <sup>2</sup>	AY968381 <sup>1</sup>	AY968536 <sup>1</sup>	AY256884 <sup>2</sup>	AY968396 <sup>1</sup>	AY968409 <sup>1</sup>	<sup>1</sup> New Zealand: CHR 512491; <sup>2</sup> GenBank
Corynocarpaceae											
<i>Corynocarpus laevigatus</i> J.R. Forster & G. Forster	AJ235446 <sup>1</sup>	AY968448 <sup>1</sup>	AY968497 <sup>1</sup>	AF148994 <sup>2</sup>	AY968565 <sup>1</sup>	AY968382 <sup>1</sup>	AY968537 <sup>1</sup>	AY256885 <sup>2</sup>	AF206892 <sup>2</sup>	AF479110 <sup>2</sup>	<sup>1</sup> New Zealand: CHR 420527; <sup>2</sup> GenBank
Cucurbitaceae											
<i>Coccinia sessilifolia</i> (Sond.) Cogn.	AY968427	AY968446	AY968491	AY968520	AY968568	AY968385	AY968531	AY968464	AY973011	AY968404	Africa (cult. Mainz Bot Gard), S. S. Renner et al. 2763 (M)
<i>Dendrosicyos socotranus</i> Balf. f.	AY968433 <sup>1</sup>	AY973018 <sup>2</sup>	AY968500 <sup>1</sup>	AY973022 <sup>2</sup>	AY973005 <sup>2</sup>	AY973005 <sup>2</sup>	AY968540 <sup>1</sup>	AY256880 <sup>3</sup>	AY968397 <sup>1</sup>	AY968412 <sup>1</sup>	<sup>1</sup> Mexico: M. Olson s.n. (MO); <sup>2</sup> J. Lavranos s.n. (M), cult. Munich BG; <sup>3</sup> GenBank
<i>Ecballium elaterium</i> (L.) A. Rich. ssp. <i>elaterium</i>	AY968434 <sup>1</sup>	AY973019 <sup>2</sup>	AY968501 <sup>1</sup>	AY973023 <sup>2</sup>	AY973006 <sup>2</sup>	AY973006 <sup>2</sup>	AY968541 <sup>1</sup>	AY968470 <sup>1</sup>	AY973012 <sup>2</sup>	AY968413 <sup>1</sup>	<sup>1</sup> Mediterranean (cult. Kew Bot Gard): 1970-624 (M. Chase 922; K); <sup>2</sup> Mediterranean (cult. Mainz Bot Gard); S. S. Renner et al. 2768 (M); <sup>3</sup> GenBank
<i>Gurania tubulosa</i> Cogn. (syn.: <i>G. megistantha</i> J. D. Sm.)	AY968435 <sup>1</sup>	AY968450 <sup>1</sup>	AY968502 <sup>1</sup>	AY973024 <sup>2</sup>	AY968569 <sup>1</sup>	AY968386 <sup>1</sup>	AY968542 <sup>1</sup>	AY256881 <sup>3</sup>	AY973013 <sup>1</sup>	AY968414 <sup>1</sup>	<sup>1</sup> South America (cult. Missouri Bot Gard); acc. 1993-1657-4; <sup>2</sup> sequence for <i>G. makoyana</i> (S. S. Renner et al. 2771 [M]); <sup>3</sup> GenBank
<i>Gynostemma pentaphyllum</i> Makino	AY968436	AY968451	AY968503	AY968523	AY973007	AY973007	AY968543	AY968471	AY973014	AY968415	Japan: H. Takahashi 20712 (GIFU)
<i>Lagenaria breviflora</i> (Benth.) Roberty	AY968438	AY935934	AY968505	AY935747	AY968570	AY935788	AY973020	AY968473	AY929371	AY935817	Ghana: M. Merello et al. 1331 (MO)
<i>Marah macrocarpus</i> Greene	AY968439 <sup>1</sup>	AY968453 <sup>1</sup>	AY968506 <sup>1</sup>	AY968524 <sup>1</sup>	AY968571 <sup>1</sup>	AY968387 <sup>1</sup>	AY973021 <sup>3</sup>	AY968474 <sup>1</sup>	AY973015 <sup>2</sup>	AY968417 <sup>1</sup>	<sup>1</sup> USA, Sonora Desert: D. Arisa & S. Swensen 1009 (RSA); <sup>2</sup> M. Olson s.n., 26 Nov, 2001 (MO); <sup>3</sup> sequence for <i>M. fabaceus</i> (R. Ricklefs & S. S. Renner 1 [MO])

<i>Neosalsmitra sarcophylla</i> (Wall.) Hutchinson	AY968440 <sup>1</sup>	AY968454 <sup>1</sup>	AY968507 <sup>1</sup>	AY968525 <sup>1</sup>	AY968572 <sup>1</sup>	AY973008 <sup>1</sup>	AY968545 <sup>1</sup>	AY974333 <sup>1</sup>	AY968399 <sup>1</sup>	AY968418 <sup>1</sup>	<sup>1</sup> Germany (cult. Mainz Bot Gard): S. S. Renner et al. 2778 (M); <sup>2</sup> GenBank
<i>Schizopepon bryoniifolius</i> Maxim.	AY968442	AY968456	AY968509	AY973025	AY973009	AY973009	AY968547	AY968476	AY968400	AY968420	Japan: T. Fukuhara leg. seeds, cult. in St. Louis by SR
<i>Seyrigia humbertii</i> Keraudr.	AY968443 <sup>1</sup>	AY968457 <sup>1</sup>	AY968510 <sup>1</sup>	AY968526 <sup>1</sup>	AY973010 <sup>1</sup>	AY973010 <sup>1</sup>	AY968548 <sup>1</sup>	AY968477 <sup>1</sup>	AY973016 <sup>1</sup>	AY968421 <sup>1</sup>	<sup>1</sup> Madagascar (cult. Missouri Bot Gard): acc. 1996-3485
<i>Xerosicyos danguyi</i> Humbert	AJ235648 <sup>3</sup>	AY968459 <sup>1</sup>	AY968512 <sup>1</sup>	AY973026 <sup>2</sup>	AY968573 <sup>1</sup>	AY968388 <sup>1</sup>	AY968550 <sup>1</sup>	AY968479 <sup>1</sup>	AY973017 <sup>2</sup>	AY968423 <sup>1</sup>	<sup>1</sup> Madagascar (cult. Missouri Bot Gard): acc. 1984-0142; <sup>2</sup> Madagascar (cult. Mainz Bot Gard): S. S. Renner et al. 2807 (M); <sup>3</sup> GenBank
Datisceae											
<i>Datisca cannabina</i> L.	AJ235450 <sup>2</sup>	AB016467 <sup>2</sup>	AY968498 <sup>1</sup>	L21939 <sup>2</sup>	AY968566 <sup>1</sup>	AY968383 <sup>1</sup>	AY968538 <sup>1</sup>	AY968469 <sup>1</sup>	AF008952 <sup>2</sup>	AY968410 <sup>1</sup>	<sup>1</sup> West Asia (cult. Tokyo Bot Gard): acc. 87-98 (TI); <sup>2</sup> GenBank
<i>Datisca glomerata</i> (Presl) Baill.	AY968432 <sup>1</sup>	AY968449 <sup>1</sup>	AY968499 <sup>1</sup>	L21940 <sup>2</sup>	AY968567 <sup>1</sup>	AY968384 <sup>1</sup>	AY968539 <sup>1</sup>	AY256882 <sup>2</sup>	U42426 <sup>2</sup>	AY968411 <sup>1</sup>	<sup>1</sup> USA, California: H. van der Werff 14002 (MO); <sup>2</sup> GenBank
Tetramelaceae											
<i>Octomeles sumatrana</i> Miq.	AY968441 <sup>2</sup>	AY968455 <sup>1</sup>	AY968508 <sup>1</sup>	L21942 <sup>3</sup>	AY968574 <sup>1</sup>	AY968389 <sup>1</sup>	AY968546 <sup>2</sup>	AY968475 <sup>1</sup>	AF008953 <sup>3</sup>	AY968419 <sup>1</sup>	<sup>1</sup> Papua, New Guinea: W. Takeuchi & D. Ama 15674 (LAE); <sup>2</sup> Papua, New Guinea: W. Takeuchi & D. Ama 16151 (LAE); <sup>3</sup> GenBank
<i>Tetrameles nudiflora</i> R. Br.	AF209689 <sup>2</sup>	AY968458 <sup>1</sup>	AY968511 <sup>1</sup>	L21943 <sup>2</sup>	AY091831 <sup>2</sup>	AY091831 <sup>2</sup>	AY968549 <sup>1</sup>	AY968478 <sup>1</sup>	U41502 <sup>2</sup>	AY968422 <sup>1</sup>	<sup>1</sup> China: Y. H. Ji 3003 (KUN); <sup>2</sup> GenBank
Outgroups											
Celastrales											
Celastraceae											
<i>Brexia madagascariensis</i> Thou. ex Ker Gawl.	AJ235419 <sup>2</sup>	AY935899 <sup>1</sup>	AY968514 <sup>1</sup>	L11176 <sup>2</sup>	AY968576 <sup>1</sup>	AY935754 <sup>1</sup>	AY968553 <sup>1</sup>	AY968482 <sup>1</sup>	U42543 <sup>2</sup>	AF222408 <sup>2</sup>	<sup>1</sup> USA (cult.): S. D. Wikoff 1390 (BH); <sup>2</sup> GenBank
Fabales											
Fabaceae											
<i>Glycine max</i> (L.) Merr.	AY935856 <sup>1</sup>	AF142700 <sup>2</sup>	AY968515 <sup>1</sup>	Z95552 <sup>2</sup>	AY968577 <sup>1</sup>	AY935785 <sup>1</sup>	AY968554 <sup>1</sup>	AJ278415 <sup>2</sup>	X02623 <sup>2</sup>	AY935814 <sup>1</sup>	<sup>1</sup> USA (cult.): L.-B. Zhang 4001 (CS); <sup>2</sup> GenBank
Fagales											
Fagaceae											
<i>Fagus grandifolia</i> Ehrh.	AY935855 <sup>1</sup>	AY042400 <sup>2</sup> , AB046507 <sup>3</sup>	AY968513 <sup>1</sup>	AY935745 <sup>1</sup>	AB066497 <sup>2</sup>	AY935784 <sup>1</sup>	AY968551 <sup>1</sup>	AY968480 <sup>1</sup>	AF206910 <sup>2</sup>	AY935813 <sup>1</sup>	<sup>1</sup> USA (cult.): P. Renner 1 (MO); <sup>2</sup> GenBank; <sup>3</sup> GenBank for <i>F. sylvestris</i>
Juglandaceae											
<i>Juglans nigra</i> L.	AF209609 <sup>2</sup>	AF118036 <sup>2</sup>	U92851 <sup>2</sup>	U00437 <sup>2</sup>	AY968575 <sup>1</sup>	AF303783 <sup>2</sup>	AY968552 <sup>1</sup>	AY968481 <sup>1</sup>	AF206943 <sup>2</sup>	AF479105 <sup>2</sup>	<sup>1</sup> USA (cult.): S. S. Renner 2190 (MO); <sup>2</sup> GenBank
Oxalidales											
Oxalidaceae											
<i>Oxalis stricta</i> L.	AY935861 <sup>1</sup>	AY935936 <sup>1</sup>	AY968517 <sup>1</sup>	L01938 <sup>2</sup>	AY968579 <sup>1</sup>	AY935789 <sup>1</sup>	AY968556 <sup>1</sup>	AY968484 <sup>1</sup>	AF206978 <sup>2</sup>	AY935819 <sup>1</sup>	<sup>1</sup> USA, Colorado: M. P. Simmons & L.-B. Zhang 1905 (CS); <sup>2</sup> GenBank for <i>O. dillenii</i>
Rosales											
Rhamnaceae											
<i>Ziziphus obtusifolia</i> (Torrey & A.Gray) A. Gray	AY935863 <sup>1</sup>	AY935939 <sup>1</sup>	AY968519 <sup>1</sup>	U60313 <sup>2</sup>	AY968581 <sup>1</sup>	AJ225799 <sup>3</sup>	AY968558 <sup>1</sup>	AY968486 <sup>1</sup>	AY929374 <sup>1</sup>	AY935822 <sup>1</sup>	<sup>1</sup> USA (cult. Desert Bot Gard, Arizona): D. Damrel s.n. (acc. 1977047601); <sup>2</sup> GenBank for <i>Z. sp.</i> ; <sup>3</sup> GenBank for <i>Z. glabrata</i>
Rosaceae											
<i>Prunus persica</i> (L.) Batsch	AF209660 <sup>2</sup>	AF288117 <sup>2</sup>	AY968518 <sup>1</sup>	AF411492 <sup>2</sup>	AY968580 <sup>1</sup>	AF429938 <sup>2</sup>	AY968557 <sup>1</sup>	AY968485 <sup>1</sup>	L28749 <sup>2</sup>	AY935820 <sup>1</sup>	<sup>1</sup> USA (cult.): L.-B. Zhang 4003 (CS); <sup>2</sup> GenBank
Zygophyllales											
Zygophyllaceae											
<i>Larrea tridentata</i> (DC.) Coville	AY935860 <sup>1</sup>	AY935935 <sup>1</sup>	AY968516 <sup>1</sup>	AY935748 <sup>1</sup>	AY968578 <sup>1</sup>	AJ387951 <sup>2</sup>	AY968555 <sup>1</sup>	AY968483 <sup>1</sup>	AY929372 <sup>1</sup>	AY935818 <sup>1</sup>	<sup>1</sup> USA, New Mexico: M. P. Simmons & C. D. Bailey 1904 (CS); <sup>2</sup> GenBank

ribulose-1,5-bisphosphate carboxylase large subunit (*rbcL*) gene. Internal primers 636F, 724R, 724F (complement to 724R; Lledó et al., 1998), and newly designed primers 227F (TCTTGATCGTTAYAAAGG), 1094F (GCAGTTATTGATAGACAGA), 579R (AAATCAAGTCCACC GCG), and 915R (ATACCRTGATTYTTYTGCTR) were frequently used to amplify poor-quality templates. The numbers of all new primers with sequences above refer to the corresponding positions in *Lotus japonicus* (GenBank Accession No. NC002694; Kato et al., 2000). To amplify the plastid *trnL* intron and adjacent spacer before the *trnF* gene, the universal primers c, d, e, and f of Taberlet et al. (1991) were used. The plastid *rpl20-rps12* intergenic spacer of most taxa between the ribosomal protein genes S12 and L20 was generally amplified using primers 'rpl20' and 'rps12' (Hamilton, 1999). Rosid-specific internal primers rpl20-538F (TAACCTTCCC VACCACKAT) and rps12-916R (KRMAAGAACGGACTAASAG), and external primers rpl20-Cuc (CTR TCCCGATGAGCCG AAACYAAAG), rpl20-N (TTTKTYCTVCGYYTYC GMGC), and rps12-Cuc (AGCCAATCMGAAAYG TCACGAAATC) were designed for *Anisophyllea fallax*, *Combretocarpus sumatrana*, and *Larrea tridentata*. Parts of exons b and c of the mitochondrial NADH dehydrogenase gene (*nad1*) and the complete intron between them were amplified using primers 'exon B' and 'exon C' (Demesure et al., 1995), BF2, BF3 (Sanjur et al., 2002), 1289F (GCCGCAGCGGGACTACCA), 1447R (CTTTCATCAAATGATGCATG), and BR3 (CCATCACCTACAGCCCTTTC). The latter three were designed for this study. The 18S nrDNA gene was amplified with primers 25eF, 530F, 922F, 1322F, 626R, 1131R, 1433R, 1769R (Nickrent and Starr, 1994), and ITS2p (Swensen et al., 1998). A portion of the 5' end of the nuclear ribosomal 26S gene was amplified by use of primers S1, S2, S3, S4, 268R, 641R, and 950R (Kuzoff et al., 1998).

Amplified fragments were purified by running the entire product on a 1% agarose gel and then recovering the DNA with QIAquick Gel Extraction Kits (Qiagen). Cycle sequencing of the purified PCR products used the BigDye Terminator Cycle Sequencing Kit (vers. 3.0 and 3.1, Applied Biosystems [ABI], Norwalk, Connecticut) according to the manufacturer's suggested protocol. The dye was removed by 2  $\mu$ l of 3 mol/L NaOAc (pH 4.6), and 50  $\mu$ l ethanol precipitation, and samples were then run on an ABI377, ABI3100 Avant automated sequencer, or an ABI 3730XL Analyzer. A total of 234 new sequences were generated for this study and have been deposited in GenBank as Accession Nos. AY968375–AY968581 and AY973005–AY973026 (Table 1).

### 2.3. Sequence alignments

Preliminary alignments of nucleotides were obtained independently for each of the nine loci using the default alignment parameters in Clustal X ver. 1.83 (Thompson et al., 1997). Manual adjustments to the Clustal-based

alignments were performed using the procedure outlined by Simmons (2004), following Zurawski and Clegg (1987), with reading frames maintained for the protein-coding genes. To maintain the alignment of repeats created by insertions, however, additional positions were added to the Clustal-based alignments for some loci. Also, an 8-bp gap was added to positions 641–649 of the *rpl20-rps12* matrix to accommodate a long insertion in *Glycine max*. Alignment of the *trnL-F* spacer among the outgroup sequences, and between outgroup and ingroup sequences, was problematic with several long, ambiguously aligned regions. The *trnL-F* spacer was therefore scored as uncertain ("?") for the outgroup sequences in the gene-tree analyses. The *trnL* intron and the *trnL* and *trnF* genes could be aligned for all sequences and were included in the analyses.

For ambiguously aligned regions where one or more sequences had a duplicate insertion (or the others had a deletion of one of two repeats) and the character-state distribution among the characters in the ambiguously aligned region was identical for those sequences that had both repeats, such that the character-state distribution among the positions in question would be identical for either of the alternative alignments, the ambiguously aligned region was arbitrarily aligned with one of the two repeats, and the entire region was kept in the analysis. A total of 302 ambiguously-aligned positions were excluded from the analyses (18S rDNA, 2; 26S rDNA, 6; *atpB*, 0; *matK*, 0; *nad1*, 72; *ndhF*, 0; *rbcL*, 0; *rpl20-rps12*, 124; *trnL-F* region, 98). Ambiguously aligned nucleotides of individual sequences in regions that could be unambiguously aligned for the remaining sequences were scored as uncertain ("?").

Reading frames of the four protein-coding genes, and boundaries between tRNA genes and their adjacent spacers or introns, were determined using MacClade ver. 4.03 (Maddison and Maddison, 2001), with reference to annotated sequences in GenBank. Six singleton deletions were observed in *ndhF* sequences generated for this study at four different positions (*Combretocarpus* at position 7; *Schizopepon* at position 30; *Begonia herbacea*, *B. oxyloba*, and *Hillebrandia* at position 187; *Brexia* at position 483). The singleton deletion at position 187 was observed in all three species of the Begoniaceae sampled. Among the 213 amino acid positions in the region of *ndhF* that was sequenced, there were 13 unique amino acids (relative to the 28 sequences lacking singleton deletions) for *Combretocarpus*, five unique amino acids for *Schizopepon*, and 18 unique amino acids for *Brexia* (the only member of the Celastrales sampled as an outgroup). A singleton deletion and a two-nucleotide insertion were observed in *matK* sequences generated for this study (*Schizopepon* at position 142; *Octomeles* at positions 38–39). Of the 410 amino acid positions in the region of *matK* that was sequenced, there was only a single unique amino acid in the *Octomeles* sequence, and three unique amino acids in the *Schizopepon* sequence. Note that the inferred *Schizopepon matK* deletion is

adjacent to 11 thymines and may be a sequencing artifact. The base calls for nucleotides adjacent to these eight indels were confirmed in the original electropherograms.

#### 2.4. Phylogenetic analysis

Gap characters, whose inclusion often affects the inferred tree topology and increases branch-support values (Simmons et al., 2001), were scored using modified complex indel coding (Müller, 2005; Simmons and Ochoterena, 2000). This modification only applies when asymmetrical step matrices would have been coded by the original complex indel coding. None of the step matrices violated the three-point-condition or triangle inequality (Farris, 1981). Parsimony-informative gap characters were scored from unambiguously aligned regions. Individual sequences for which the gap was considered ambiguously aligned relative to gaps in other sequences were coded as uncertain for the gap character in question. However, ambiguously aligned gaps that moved together in each of the alternative equally optimal alignments were included in the analysis, following Davis et al. (1998). A total of 110 gap characters were included (18S rDNA, 3; 26S rDNA, 2; *atpB*, 0; *matK*, 4; *nadl*, 40; *ndhF*, 5; *rbcL*, 0; *rpl20-rps12*, 19; *trnL-F* region, 37). Data matrices have been deposited in TreeBASE (<http://www.treebase.org/treebase/>) with accession numbers S1392 and M2494–M2504.

As a means of data exploration, several potential process partitions (Bull et al., 1993) were analyzed separately, although actual delimitation of process partitions is often arbitrary (Siddall, 1997). (1) Each of the nine loci was analyzed separately (using nucleotide and gap characters). (2) The *trnL* intron and *trnL-F* spacer, including their respective gap characters, were analyzed separate from one another and from the tRNA genes. (3) The four protein-coding plastid loci (*atpB*, *matK*, *ndhF*, and *rbcL*) were analyzed using amino acid characters (together with their gap characters, when applicable). (4) Putatively coalescent genes (nuclear rDNA, mitochondrial *nadl*, plastid genes; Doyle, 1992) were analyzed in groups and their trees were compared to check for well supported, contradictory signal that might have been caused by lineage sorting, potential introgression of the plastid genome or nrDNA (Doyle, 1992; Wendel et al., 1995), unrecognized paralogy problems with nrDNA caused by incomplete concerted evolution (Álvarez and Wendel, 2003; Bailey et al., 2003), or horizontal gene transfer (of *nadl*; Bergthorsson et al., 2003; Won and Renner, 2003). The three putatively coalescent groups were: *nadl*, from the (generally) non-recombining (between genomes) mitochondrial genome; *atpB*, *matK*, *ndhF*, *rbcL*, the *rps20-rps12* spacer, and the *trnL-F* region, from the (generally) non-recombining plastid genome; and the linked nuclear genes, 18S and 26S nrDNA.

To compare the phylogenetic signal in nucleotide, amino acid, and gap characters, separate analyses were conducted on the nucleotide characters (both, from all loci or only the

protein-coding genes), the amino acid characters (from all protein-coding genes), and the gap characters (from all loci). Note that the amino acid-based analyses of *matK* and *ndhF* are problematic due to the inclusion of potential pseudogenes caused by frameshift mutations (see above). Separate analyses were also conducted on third codon positions versus first plus second codon positions to compare their phylogenetic signals. Finally, separate analyses were conducted with all nucleotide and gap characters from the plastid protein-coding genes and from the plastid introns and spacers. A simultaneous ('total evidence') analysis (Kluge, 1989; Nixon and Carpenter, 1996) included nucleotide and gap characters from all loci. In a second simultaneous analysis, individual sequences that showed patterns consistent with long-branch attraction were removed (see below), following Lecointre and Deleporte (2005). The simultaneous-analysis matrix included 12% missing/inapplicable characters.

Equally weighted parsimony tree searches were conducted for each data matrix using 1000 tree-bisection–reconnection (TBR) searches in PAUP\* ver. 4.0b10 (Swofford, 2001) with a maximum of 1000 trees held per TBR search. Parsimony jackknife analyses (Farris et al., 1996) were conducted using PAUP\* with the removal probability set to approximately  $e^{-1}$  (36.7879%), and "jac" resampling emulated. One thousand replicates were performed with 10 TBR searches per replicate and a maximum of 100 trees held per TBR search.

Maximum likelihood (Felsenstein, 1973) analyses of nucleotide characters from each of the nine loci were used to test for long-branch attraction (Felsenstein, 1978), although this test is not infallible (Gaut and Lewis, 1995; Sanderson and Kim, 2000; Siddall, 1998). One hundred jackknife replicates were performed with one TBR search per replicate and a maximum of 100 trees held per TBR search. Well-supported ( $\geq 70\%$  jackknife support) clades that conflicted with one another in the parsimony and likelihood jackknife trees were tested for long-branch attraction in the parsimony analyses by alternatively removing the terminals in question. If the terminal(s) remaining in the parsimony analysis moved to a different part of the tree (with  $\geq 70\%$  jackknife support) when the potentially attracting terminal(s) was/were removed, the result was consistent with the explanation of long-branch attraction (Siddall and Whiting, 1999).

Modeltest ver. 3.6 (Posada and Crandall, 1998) was used to select the best-fit likelihood model for maximum likelihood analyses. The Akaike Information Criterion (Akaike, 1974) served to select among models instead of the hierarchical likelihood ratio test, following Pol (2004) and Posada and Buckley (2004). The models selected were GTR + I + G (18S rDNA, 26S rDNA, *atpB*, *ndhF*, *rbcL*, and the simultaneous analysis of all nucleotide characters), TVM + I + G (*trnL-F* region), TVM + G (*matK*, *rpl20-rps12*), or TVMef + I + G (both *nadl* analyses; see below). The selected model and parameter estimates were then used for tree searches from the respective data partitions.

## 2.5. Parsimony reconstruction of sexual systems

Character evolution was studied with MacClade, which optimizes character changes on a tree based on the principle of parsimony. The sexual systems of the species included in the analysis were categorized as follows: (1) flowers bisexual, (2) flowers unisexual, plants monoecious, (3) flowers unisexual, populations dioecious, (4) some flowers bisexual, some unisexual, populations andromonoecious, (5) some flowers bisexual, some unisexual, populations androdioecious, and (6) ambiguous. Information on sexual states was taken from taxonomic treatments, augmented by personal observations for the Cucurbitaceae. Taxa were coded as follows (starting with the outgroups): *Larrea*, bisexual; *Glycine*, bisexual; *Oxalis*, bisexual; *Brexia*, bisexual; *Prunus*, *Ziziphus*, bisexual; *Fagus*, *Juglans*, monoecious; *Anisophyllea corneri*, *A. fallax*, flowers unisexual, plants monoecious; bisexual flowers occur occasionally, but have fewer ovules or anthers than normal flowers (Tobe and Raven, 1988); *Combretocarpus rotundatus*, flowers bisexual (Tobe and Raven, 1988); *Corynocarpus laevigatus*, flowers bisexual (Matthews and Endress, 2004); *Coriaria ruscifolia*, *C. sarmentosa*, bisexual (Thompson and Gornall, 1995), *C. myrtifolia*, andromonoecious (Thompson and Gornall, 1995), *C. nepalensis*, andromonoecious, occasionally monoecious (Thompson and Gornall, 1995); *Octomeles sumatrana*, *Tetrameles nudiflora*, dioecious (Matthews and Endress, 2004 and personal observation, SSR); *Datisca cannabina*, dioecious, *D. glomerata*, androdioecious (Wolf et al., 2001); *Begonia herbacea*, *B. oxyloba*, monoecious; *Hillebrandia sandwicensis*, monoecious (Matthews and Endress, 2004). Sexual systems of the included Cucurbitaceae are: *Coccinia sessilifolia*, dioecious; *Dendrosicyos socotranus*, monoecious; *Ecballium elaterium* has a dioecious and a monoecious subspecies that hybridize easily and for which it has been shown that sexual system is determined by a single gene (Costich and Galan, 1988); we sequenced the monoecious subspecies, but coded *Ecballium* as ambiguous for monoecy/dioecy; *Gurania tubulosa*, probably monoecious and sex-changing (Condon and Gilbert, 1988; this species has not been studied, but close relatives all monoecious and sex-changing), coded as monoecious; *Gynostemma pentaphyllum*, dioecious; *Lagenaria brevifolia*, monoecious; *Marah macrocarpus*, monoecious; *Nealsomitra sarcophylla*, dioecious (Matthews and Endress, 2004); *Seyrigia humbertii*, dioecious; *Schizopepon bryoniifolius*, androdioecious (Fukuhara and Akimoto, 1999); *Xerosicyos danguyi*, dioecious.

## 3. Results and discussion

### 3.1. Process partitions

The jackknife trees with parsimony jackknife values above branches, and maximum likelihood jackknife values below branches for each of the nine loci are presented in Figs. S1–S9 as supplemental data at: <http://www.biology.colostate.edu/Research/>. Fig. 1 shows the simultaneous-

analysis jackknife tree. The most parsimonious, parsimony jackknife, and likelihood jackknife trees for all analyses are available as supplemental data. Data-matrix and tree statistics for all analyses are presented in Table 2. Congruence of data matrices with the primary simultaneous analysis of all characters was assessed by mapping each matrix's parsimony-informative characters onto the six equally parsimonious trees found in the simultaneous analyses (Table 2). The amino acid-based *atpB* and *rbcL* matrices were outliers in requiring a minimum of 13 and 15% additional steps on the simultaneous-analysis trees. Among the nucleotide-based analyses of individual loci, 18S rDNA showed the least congruence (6% increase) with the topology from the simultaneous analysis, and *rbcL* the most (0.3% increase). Of the three putatively coalescent genes, the plastid genes showed the most congruence (0.02% increase), whereas nrDNA and *nadl* both required a 3% increase in steps.

Two contradicting clades in the coalescent-gene jackknife trees (available as supplemental data), involving members of different families, received  $\geq 70\%$  jackknife support. First, *Glycine* (Fabales) was resolved as the sister group to *Oxalis* (Oxalidales) and *Brexia* (Celastrales) as sister to *Larrea* (Zygophyllales) on the nrDNA jackknife tree. Both placements contradict the plastid and *nadl* gene trees as well as previous studies (Hilu et al., 2003; Savolainen et al., 2000a,b; Soltis et al., 2000; Zhang and Simmons, 2006). Although these placements may result from unrecognized paralogs that have not been homogenized by concerted evolution among the nrDNA loci, we think it more likely results from long-branch attraction between *Glycine* and *Oxalis*. When *Glycine* was removed from the nrDNA matrix, *Oxalis* and *Brexia* were resolved as sister groups with 65% jackknife support (available as supplemental data), corresponding with the other coalescent genes and other studies.

Second, *Corynocarpus* was resolved as the sister group of Anisophylleaceae by *nadl* with 78% jackknife support (Fig. S9), whereas nrDNA and plastid genes resolved it as the sister group of Coriariaceae as found in previous studies (Hilu et al., 2003; Savolainen et al., 2000a; Soltis et al., 2000; Wagstaff and Dawson, 2000). Although this could represent horizontal gene transfer in the mitochondrial genome (Berghthorsson et al., 2003; Won and Renner, 2003), we believe that it more likely represents long-branch attraction between *Corynocarpus* and Anisophylleaceae. To test this hypothesis, Anisophylleaceae were removed from the *nadl* matrix and the parsimony analyses were repeated. *Corynocarpus* was then placed as the sister group of Coriariaceae with 91% jackknife support (available as supplemental data), in agreement with the other coalescent genes and previous studies.

### 3.2. Nucleotide-frequency heterogeneity

None of the nine matrices for the separate loci exhibited significant heterogeneity for the parsimony-informative nucleotide characters among different terminals, as measured by the  $\chi^2$  test implemented in PAUP\*, which,



Table 2  
Data-matrix and tree statistics for each of the analyses

Matrix	# characters	# PI characters	MPT length	# MPTs	# jackknife clades	Average jackknife support (%)	CI	RI	Increase in steps
18S nrDNA	1,789	135	524	555	18/18	80/81	.46	.60	22/6%
26S nrDNA	966	133	567	35	17/15	80/80	.41	.65	20/4%
<i>atpB</i> DNA	1,473	238	907	42	22/24	92/85	.50	.63	7/1%
<i>atpB</i> amino acid	490	40	179	2599	12	72	.65	.71	16/13%
<i>matK</i> DNA	1,234	425	1673	1	28/27	87/89	.52	.67	10/1%
<i>matK</i> amino acid	414	211	985	226	24	89	.62	.71	12/1%
<i>nadI</i> DNA	2,117	253	758	15	25/25	84/84	.65	.79	16/3%
<i>ndhF</i> DNA	645	220	951	23	24/24	86/86	.51	.64	18/2%
<i>ndhF</i> amino acid	218	91	498	352	21	79	.66	.65	13/3%
<i>rbcL</i> DNA	1,421	217	891	4	20/23	88/85	.44	.63	2/0.3%
<i>rbcL</i> amino acid	473	43	212	982	10	72	.42	.62	24/15%
<i>rpl20-rps12</i>	906	270	1015	8	28/25	86/86	.58	.67	8/1%
<i>trnL-F</i> region	1,220	305	936	198	26/26	84/85	.65	.80	11/1%
<i>trnL</i> intron	597	142	549	3919	21	85	.59	.71	10/2%
<i>trnL-F</i> spacer ingroup	520	153	337	18	18	90	.76	.89	8/3%
nrDNA genes	2,755	268	1107	7	22	81	.42	.62	27/3%
DNA exons	4,773	1100	4458	1	29	90	.49	.64	2/0.06%
Plastid non-coding	1,989	531	1833	2	31	88	.60	.72	8/1%
Plastid genes	6,899	1675	6432	1	31	92	.52	.70	1/0.02%
First and second positions only	3176	456	1841	4	25	92	.52	.65	8/1%
Third positions only	1,588	635	2576	6	27	91	.48	.64	7/0.3%
Gap characters only	110	110	264	370	22	84	.74	.85	7/3%
DNA characters only	11,661	2085	8066	4	29/29	95/93	.51	.66	1/0.02%
AA characters only	1,586	376	1899	1	24	91	.59	.66	14/1%
Simultaneous	11,771	2196	8348	6	29	94	.52	.67	N/A
Simultaneous—LBA	11,771	2110	7759	6	29	95	.54	.69	N/A

The number of characters for each data matrix is reported after exclusion of alignment-ambiguous regions, if any. “PI,” parsimony-informative. “MPT,” most parsimonious tree(s). “CI,” ensemble consistency index (Kluge and Farris, 1969) on the most parsimonious tree(s) for the parsimony-informative characters. “RI,” ensemble retention index (Farris, 1989). “Increase in steps” was determined by mapping the PI characters from the respective data matrix onto the most parsimonious trees from the simultaneous analysis; the lowest increases in steps were reported. The number of, and average support for, jackknife clades for the nucleotide-based analyses of individual genes (and all nucleotide characters together) are reported for the parsimony analyses, followed by the same information for the likelihood analyses.

however, ignores phylogenetic correlations. By far, the most heterogeneous gene was 26S rDNA ( $\chi^2 = 0.13$ ; 131 parsimony-informative nucleotide characters) with 9–22% adenine, 23–40% cytosine, 21–37% guanine, and 15–34% thymine among the terminals. The matrix for the simultaneous analysis of 2085 parsimony-informative nucleotide characters was significantly heterogeneous ( $\chi^2 = 0.02$ ) with 21–26% adenine, 24–29% cytosine, 20–26% guanine, and 24–31% thymine among the terminals. No obvious nucleotide-frequency outliers were found for either the 26S rDNA or simultaneous-analysis matrices.

### 3.3. Nucleotide vs. gap characters

In the jackknife tree constructed using the 110 parsimony-informative gap characters (available as supplemental data), 22 clades were resolved with an average of 84% jackknife support (Table 2). Seventeen (77%) of these clades were identical to those resolved using the 2085 parsimony-informative nucleotide characters (available as supplemental data) and five (23%) were contradictory. The five contradictory clades received lower average jackknife support (64%) than did the 17 matching clades (89%).

Of these five contradictory clades, four ([*Begoniaceae* + *Tetramelaceae*], [*Dendrosicyos* + *Gurania*], [*Dend-*

*rosicyos* + *Gurania* + *Lagenaria* + *Seyrigia*], [*Coriaria myrtifolia* + *C. nepalensis* + *C. ruscifolia*]) were resolved in the simultaneous analysis of all nucleotide and gap characters as they were in the nucleotide-characters-only analysis. In contrast, one clade was due only to gap characters (Fagales + Cucurbitales; 73% jackknife support) that contradicted the nucleotide-characters-only resolution (Rosales + Cucurbitales; 55% jackknife support) but not the simultaneous analysis (Fig. 1). Therefore, although there were only 5% as many parsimony-informative gap characters as there were nucleotide characters, they were not “swamped” by the larger data matrix. The Fagales + Cucurbitales clade revealed by the gap characters is consistent with independent evidence (Chase et al., 1993; Setoguchi et al., 1999; Schwarzbach and Ricklefs, 2000; Soltis et al., 2000, 2003; Swensen et al., 1994), in contrast to the nucleotide-characters-only resolution. Given the general taxonomic congruence and relatively weak branch support for the five contradictory clades, we consider the nucleotide and gap characters to have congruent phylogenetic signal.

### 3.4. Nucleotide vs. amino acid characters

Of the 67 total clades resolved across the four gene trees obtained from amino acid characters (available as supple-

mental data), seven (10%; *atpB*, 2; *matK*, 1; *ndhF*, 4; *rbcL*, 0) contradicted clades resolved in the corresponding trees constructed using nucleotide characters. As with the conflicts between nucleotide and gap characters, the seven contradictory clades received lower average jackknife support (69%) than the 60 matching or congruent clades (81%).

As in Simmons et al. (2002), the amino acid-based *atpB* sequences yielded some unusual clades. *Begonia herbacea* appeared as the sister group of *Datisca glomerata* with 64% jackknife support, whereas *Begonia oxyloba* was placed in a clade with *Gynostemma* and *Neosalsmitra*, both of which conflicts with the delimitation of Begoniaceae and well-supported clades in the nucleotide-based *atpB* gene tree (Fig. S3). Furthermore, *Brexia* was resolved with *Larrea* in a well-supported (78% jackknife) clade by the amino acid characters, but was placed in a basal polytomy by the nucleotide characters (Fig. S3). No cases of convergent amino acids (Simmons, 2000) or artifacts caused by composite characters (Simmons and Freudenstein, 2002) were identified upon inspection of the amino acid and nucleotide data matrices, suggesting that these were simply cases in which silent substitutions were more phylogenetically informative than replacement substitutions (Simmons et al., 2002). The other five contradicted clades were less striking, with independent evidence being equivocal in discriminating between the nucleotide- and amino acid-based contradictory resolution (*matK*), or questionable contradictory resolution in both the nucleotide- and amino acid-based gene trees (*ndhF*).

### 3.5. Potential long-branch attraction

Comparison of the nucleotide-based gene trees from the nine loci revealed no instances of well-supported ( $\geq 70\%$  jackknife support) conflict between trees or between parsimony and likelihood analyses. As in the parsimony *nadI* analysis, *Corynocarpus* was resolved as the sister group of Anisophylleaceae in the likelihood jackknife tree (74% jackknife support; Fig. S9). As with parsimony, when Anisophylleaceae were removed from the *nadI* matrix and the likelihood analysis was repeated, *Corynocarpus* was resolved as the sister of Coriariaceae (90% jackknife support; available as supplemental data), in agreement with the other coalescent genes and previous studies. Therefore, this appears to be a case in which parsimony and likelihood were both affected by long-branch attraction, reinforcing the finding of Gaut and Lewis (1995) and Chang (1996) that likelihood analyses can suffer from long-branch effects when the model is mis-specified, which is probably often the case.

In the *ndhF* parsimony tree, Datisceae and Tetramelaceae formed a clade with Coriariaceae and Corynocarpaceae (62% jackknife support; Fig. S5), while in the likelihood tree Coriariaceae and Corynocarpaceae were sister to Begoniaceae (but only with 64% jackknife support). Both of these resolutions contrast with most other gene trees and the simultaneous analysis. To test whether the parsimony reso-

lution was caused by long-branch attraction between Datisceae and Tetramelaceae and Coriariaceae + Corynocarpaceae, the parsimony analysis was repeated after excluding Coriariaceae + Corynocarpaceae. Datisceae and Tetramelaceae were still resolved as a clade separate from Begoniaceae (with 65% jackknife support; tree available as supplemental data). This result provides no evidence for long-branch attraction between Datisceae and Tetramelaceae, and Coriariaceae and Corynocarpaceae.

In comparing the likelihood-based tree inferred from all nucleotide characters with the parsimony-based tree inferred from the simultaneous analysis of all nucleotide and gap characters, there was relatively well-supported incongruence among some outgroups. In the parsimony tree, the Rosales were sister of Cucurbitales + Fagales with 87% jackknife support and *Glycine* was sister to *Larrea* with 95% jackknife support. In the likelihood tree, however, the Rosales were sister to *Larrea* with 61% jackknife support, while *Glycine* was sister to *Brexia* + *Oxalis* with 69% jackknife support. Unfortunately, other studies do not yet conclusively support either the parsimony or likelihood resolutions for these taxa. For example, the likelihood resolution of *Glycine* is consistent with Savolainen et al.'s (2000b; <50% bootstrap) results, but not with the trees presented by Soltis et al. (2000; 68% jackknife support) and Savolainen et al. (2000a; <50% bootstrap).

### 3.6. Simultaneous analysis

Following the analysis of the combined nucleotide and gap characters, we performed a second combined analysis in which sequences suspected of causing long-branch attraction were recoded as uncertain. To minimize data loss (i.e., the number of additional cells in a matrix scored as uncertain), in each case, the lineage with the fewest terminals was chosen for re-coding, for example, the *Corynocarpus nadI* sequence rather than all three Anisophylleaceae sequences. The following four sets of sequences (both nucleotide and gap characters, when applicable) were re-scored as uncertain for blocks of characters: *Corynocarpus* for *nadI*, *Glycine* for both nrDNA genes (see above), *Glycine* for *ndhF*, and *Brexia* and *Oxalis* for third codon positions from all four protein-coding genes. *Glycine* was resolved as more closely related to *Oxalis* than *Oxalis* was to *Brexia* in both the parsimony- and likelihood-based *ndhF* gene trees (78 and 76% jackknife support, respectively). These relative positions of *Brexia*, *Glycine*, and *Oxalis* in the *ndhF* gene tree relative to their expected positions based on independent evidence (Hilu et al., 2003; Savolainen et al., 2000a,b; Soltis et al., 2000, 2003; Zhang and Simmons, 2006) did not allow using Siddall and Whiting's (1999) test of long-branch attraction by alternate exclusion of the putatively attracting lineages. The long branches leading to *Glycine* and *Oxalis* in the parsimony trees, however, are consistent with long-branch attraction between *Glycine* and *Oxalis*.

In the jackknife tree inferred from third codon positions of *atpB*, *matK*, *ndhF*, and *rbcL* (available as supplemental

data), the clade of *Brexia* + *Oxalis* was resolved as the sister group of Cucurbitales with 68% support. Because of the position of *Brexia* and *Oxalis* relative to the other outgroup terminals in this tree it was again not possible to apply Siddall and Whiting's (1999) test of long-branch attraction. Other studies with denser sampling (Hilu et al., 2003; Savolainen et al., 2000a,b; Soltis et al., 2000, 2003; Zhang and Simmons, 2006), however, suggest that the clustering of *Brexia*, *Oxalis*, and Anisophylleaceae (plus the other Cucurbitales) in the six most parsimonious trees may result from long-branch attraction. In an attempt to reduce the amount of excluded data, jackknife trees were constructed from third codon positions for each of the four protein-coding genes independently of one another to identify which gene(s) was (were) the source of this postulated long-branch attraction. However, none of the jackknife trees resolved *Brexia* + *Oxalis* as the sister clade of Cucurbitales (the outgroup orders were generally resolved in a polytomy; trees available as supplemental data).

The jackknife tree from the second combined analysis (with the recoded sequences, above) was identical in topology obtained from the first combined analysis. Therefore, our conclusions are not dependent upon the re-coding approach to overcoming suspected long-branch attraction. Although topologies were identical, the second combined analysis provided substantially higher jackknife support values for several clades (Fig. 1).

### 3.7. Monophyly, affinities, and synapomorphies of the Cucurbitales

Our findings strongly support the monophyly of Cucurbitales (100% jackknife support in all simultaneous analyses and  $\geq 66\%$  jackknife support in all individual gene trees except those from 18S and 26S rDNA). This contrasts with earlier morphological (Cronquist, 1981, 1988; Dahlgren, 1983, 1988; Engler, 1896; Hutchinson, 1973; Melchior, 1964; Müller and Pax, 1894; Takhtajan, 1969, 1980, 1997; Thorne, 2000; Tobe and Raven, 1987) and anatomical studies (Matthews et al., 2001; Schönberger et al., 2001), which have (strongly) disagreed about the circumscription of the order. However, our results do not conclusively resolve the placement of Cucurbitales relative to Fabales, Fagales, and Rosales. The moderately supported sister group relationship between Cucurbitales and Fagales, with both together sister to Rosales, is consistent with most earlier studies (e.g., Chase et al., 1993; Setoguchi et al., 1999; Schwarzbach and Ricklefs, 2000; Soltis et al., 2000, 2003; Swensen et al., 1994, with one to four of the same genes sampled here), but contrasts with studies that have placed Zygophyllales as sister to Cucurbitales (Savolainen et al., 2000b) or Cucurbitales as sister to Fabales + Fagales + Rosales + Zygophyllales (Hilu et al., 2003), albeit in both cases with weak support. Our second combined analysis, in which sequences suspected of causing long-branch attraction were re-coded, provided higher jackknife support (96%) for the sister relationship between

Rosales and Cucurbitales + Fagales (Fig. 1). In combination with the findings of Soltis et al. (2000), which supported the monophyly of Cucurbitales + Fabales + Fagales + Rosales (68% jackknife support), our results are suggestive of the following relationships among these four orders: (Fabales (Rosales (Cucurbitales, Fagales))).

Considering morphological evidence, it is noteworthy that the first-branching family in the Rosales, Rosaceae (Savolainen et al., 2000a), often has numerous stamens and perfect flowers (Cronquist, 1981; Hutchinson, 1973; Takhtajan, 1997), while the first-branching families in Cucurbitales and Fagales (i.e., Anisophylleaceae and Nothofagaceae (Li et al., 2002)) usually have few (4–15) stamens and unisexual flowers (Cronquist, 1981; Hutchinson, 1973; Matthews and Endress, 2004; Takhtajan, 1997). Parsimony reconstruction (Fig. 1) suggests that unisexual flowers evolved independently within Fagales and Cucurbitales, and this agrees with paleobotanical and morphological data suggesting that early Fagales may have had bisexual flowers (Friis, 1983; Manos et al., 2001; Sims et al., 1998). If the relationships among Cucurbitales, Fagales, and Rosales suggested above are correct, synapomorphies of Cucurbitales and Fagales may include the tendency for flowers to evolve from a bisexual to a unisexual state. These ordinal relationships would also imply that the large, colorful, and insect-pollinated flowers of Cucurbitales and Rosales are plesiomorphic in the (Fabales (Rosales (Cucurbitales, Fagales))) lineage and that the numerous stamens of most Rosaceae are derived (since Fabales ancestrally have stamens). Large cotyledons may be another synapomorphy of Fagales and Cucurbitales (see Table 3 and discussion of this character in the next section). A root symbiosis with nitrogen-fixing *Frankia* actinomycetes is known from families in all four orders (Soltis et al., 1995; Swensen, 1996; Swensen and Mullin, 1997).

Anatomical features that may be synapomorphic for Cucurbitales are absence of mucilage cavities or cells, absence of fasciculate or stellate hairs, presence of libriform fibers, presence of storied rays in the wood, a slightly oblique vessel end wall angle (Nandi et al., 1998), banded wood parenchyma in some members (Baas et al., 2000), alternate circular to polygonal pits on lateral walls of vessels, non-bordered or minimally bordered perforation plates, and wide multiseriate rays (Carlquist and Miller, 2001). However, knowledge of the distribution of these characters is such that it is still unclear whether they really are synapomorphic for the order. An anatomical study of the flowers of representatives of all seven families (Matthews and Endress, 2004) found that Begoniaceae, Datisceae, Tetramelaceae, and Cucurbitaceae share (in various combinations) basifixed and extrorse or latrorse anthers, trimerous gynoecea, bifurcate free carpel parts, an extended roof over the ovary formed by the ventral parts of the carpels, and parietal placentae. Important trends identified by Matthews and Endress include pointed petals (if present), a 2-cell-layered inner integument, which is delayed in development, and lacking or scant tanniferous tissues in the flowers.

Table 3  
Taxonomically important character states in the Cucurbitales sensu APG (1998, 2003)

Character	Anisophylleaceae	Begoniaceae	Coriariaceae	Corynocarpaceae	Cucurbitaceae	Datisceaeae	Tetramelaceae
Life form	Trees or shrubs	Herbs	Shrubs, rarely trees	Trees or shrubs	Climbers, herbs, rarely shrubs	Herbs	Trees
Leaf margin	Entire	Entire	Entire	Entire	Entire or toothed	Toothed	Entire
Stipules	Present	Present	Present	Present	Absent	Absent	Absent
Corolla/calyx	Distinct or no corolla	Distinct or indistinct	Distinct	Distinct	Sympetalous	Distinct or indistinct	Distinct or no corolla
Ovary	Inferior	Inferior	Superior	Superior	Inferior	Inferior	Inferior
Disc nectary	Present	Absent	Absent	Absent	Absent	Absent	Absent
Placentation	Parietal	Parietal	Apical	Apical	Parietal	Parietal	Parietal
Stigma	Short	Elongated	Short	Short	Elongated	Elongated	Elongated
Inflorescence	Paniculate, racemose or spicate	Cymose	Raceme	Paniculate	Axillary	Fasciculate	Spicate
Fruit	Drupe or samara	Capsule	Achene	Drupe	Baccate (pepo)	Capsule	Capsule
# seeds per fruit	One	Many	One	One	Many	Many	Many
Cotyledon	Mostly small	Moderate	Large	Very large	Large	Moderate	Moderate
Sexual system	Andromonoecious, monoecious	Monoecious, dioecious	Bisexual, monoecious, andromonoecious	Bisexual	Dioecious, monoecious, androdioecious	Androdioecious, dioecious	Dioecious

Information mainly from Hutchinson (1973), Cronquist (1981, 1988), Takhtajan (1997), and Matthews and Endress (2004), except for sexual systems for which sources are listed in Section 2.

A synapomorphy for the order could not be identified (cf. our Table 3). A trend that might be added to the ones listed above is asymmetrical leaf bases.

### 3.8. Familial relationships within Cucurbitales

Although the current circumscription of the Cucurbitales (APG, 1998, 2003) is based only on molecular evidence (Setoguchi et al., 1999; Schwarzbach and Ricklefs, 2000), the core of the Cucurbitales—Begoniaceae, Cucurbitaceae, Datisceaeae, and Tetramelaceae—was established based on morphology (Cronquist, 1981, 1988; Dahlgren, 1983, 1988; Hutchinson, 1973; Melchior, 1964; Takhtajan, 1969). A clade comprising these four families received 77, 88, and 78% jackknife support in our analyses of the combined data (Fig. 1). A morphological synapomorphy for the four families is a fruit with numerous seeds. Inferior ovaries and parietal placentation are plesiomorphic (below). Our data placed the Cucurbitaceae as sister to Begoniaceae, Datisceaeae, and Tetramelaceae (Fig. 1). This contrasts with Carlquist and Miller's (2001) wood-anatomical study, which suggested that Cucurbitaceae are close to Coriariaceae and Corynocarpaceae with which they share axial parenchyma that is banded, vasicentric scanty, and ray adjacent.

The close affinity among Begoniaceae, Datisceaeae, and Tetramelaceae found here agrees with flower morphological studies (Matthews and Endress, 2004), earlier classifications that united Begoniaceae and Datisceaeae (including Tetramelaceae) as a suborder (of Violales; Melchior, 1964) or order (Takhtajan, 1969), and molecular studies based on different gene regions than used here (Swensen et al., 1998: ITS and 18S rDNA; Clement et al., 2004: ITS, 18S, and *rbcL*). The three families share scanty vasicentric parenchyma while lacking banded axial

parenchyma (Carlquist and Miller, 2001). Morphological synapomorphies for Begoniaceae, Datisceaeae, and Tetramelaceae are capsular fruits and moderate-sized cotyledons, while Cucurbitaceae have pepos and large-sized cotyledons (Cronquist, 1981; Table 3). Large cotyledons may be plesiomorphic in the Cucurbitales, since all members of the Fagales, as the likely sister group of the Cucurbitales, all Coriariaceae and Corynocarpaceae, and some Anisophylleaceae have large-sized cotyledons (Cronquist, 1981; Takhtajan, 1997). The precise relationship among Begoniaceae, Datisceaeae, and Tetramelaceae remains unclear, although a sister group relationship of Tetramelaceae to Datisceaeae and Begoniaceae (Fig. 1) agrees with the morphological resemblance of Datisceaeae and Tetramelaceae (they share exstipulate leaves and small and yellow-green flowers, but differ in growth form, leaf margins, and inflorescences; Table 3).

Anisophylleaceae are clearly the sister group to all other Cucurbitales. Morphological synapomorphies for the Cucurbitales except the Anisophylleaceae include filaments shorter than anthers in bud, basifixed anthers, and lack of disc nectaries (Stevens, 2001). The next-branching clade, Corynocarpaceae + Coriariaceae, is also strongly supported (100% jackknife support in the three combined analyses; Fig. 1), and the exceptional placement of Corynocarpaceae as sister to Anisophylleaceae in the *nad1* tree (Fig. S9) is best explained as a long-branch-attraction artifact (see above). Previous molecular and anatomical studies also found a *Corynocarpus* + Coriariaceae clade (Carlquist and Miller, 2001; Wagstaff and Dawson, 2000). The synapomorphies for the Coriariaceae + Corynocarpaceae include superior ovary and apical placentation. Members of these two families occur sympatrically in the Pacific islands (Wagstaff and Dawson, 2000; Yokoyama et al., 2000).

### 3.9. Intra-familial relationships in Cucurbitales

All families of Cucurbitales are strongly supported as monophyletic (Corynocarpaceae are monogeneric, and only one species was included here), and studies of these families that include denser taxon sampling are not in conflict with the relationships found here. Specifically, *Combretocarpus* is sister to all other Anisophylleaceae (Zhang, Simmons, and Renner, unpublished data); Eurasian species of *Coriaria* are sister to the remaining Coriariaceae (Yokoyama et al., 2000); *Hillebrandia* is sister to all other Begoniaceae (Forrest et al., 2005; Plana et al., 2004); and subfamily Nhandiroboideae (represented by *Gynostemma*, *Neoalsomitra*, and *Xerosicyos*) is sister to subfamily Cucurbitaceae (Kocyan, Zhang, Schaefer, and Renner, unpublished data).

### 3.10. Character evolution in Cucurbitales

Based on the phylogeny presented here (Fig. 1), and accepting the Fagales as the sister group of the Cucurbitales, character-state transformations can be inferred as follows.

#### 3.10.1. Stipules

Stipules characterize Anisophylleaceae, Begoniaceae, Coriariaceae, and Corynocarpaceae, and are lacking in Cucurbitaceae, Datisceae, and Tetramelaceae (Cronquist, 1981; Matthews and Endress, 2004; Table 3). Stipules are also present in six of the eight families of Fagales (sensu APG, 1998, 2003; Fagales lacking stipules are the Casuarinaceae and Myricaceae; Cronquist, 1981; Takhtajan, 1997), and, importantly, stipules are present in the first-branching families Nothofagaceae and Fagaceae (Li et al., 2002; Manos and Steele, 1997). Parsimony reconstruction of the stipule character on our tree suggests that the stipulate leaves in the Cucurbitales represent the ancestral state, with exstipulate leaves being derived. Begoniaceae appear to have regained stipulate leaves from an exstipulate ancestor.

#### 3.10.2. Ovary

Ovary position is inferior except in Coriariaceae and Corynocarpaceae (Table 3). *Hillebrandia* (Begoniaceae) was formerly thought to have semi-inferior ovaries, but Matthews and Endress (2004) found it to have inferior ovaries. Among Fagales, Betulaceae have a nude ovary, Casuarinaceae and Rhoipteleaceae have a superior ovary, and Juglandaceae, Tiodendraceae, and the earliest-diverging lineages, Nothofagaceae and Fagaceae, have an inferior ovary (Cronquist, 1981; Hutchinson, 1973; Takhtajan, 1997). Parsimony reconstruction on the phylogeny (Fig. 1) suggests that the inferior ovary is plesiomorphic in the Cucurbitales and that superior ovaries evolved from inferior ones along the stem lineage of Coriariaceae and Corynocarpaceae.

#### 3.10.3. Placentation

There are two types of placentation in extant taxa of the Cucurbitales. Parietal placentation characterizes most families except Coriariaceae and Corynocarpaceae, which have apical placentation (Cronquist, 1981; Hutchinson, 1973; Matthews and Endress, 2004). Axile and basal placentations are dominant in Fagales, but the first-branching families Nothofagaceae and Fagaceae have axile placentation (Cronquist, 1981). Tracing of this character on the phylogeny (Fig. 1) suggests that parietal placentation is ancestral and that apical placentation evolved once in the ancestor of Coriariaceae and Corynocarpaceae (also within the occasional Cucurbitaceae). Notably, the characters apical placentation and a superior ovary appear to be coupled.

#### 3.10.4. Fruit type

The Anisophylleaceae have drupes or samaras, the Coriariaceae achenes, the Corynocarpaceae drupes, the Cucurbitaceae pepos, and the Begoniaceae, Datisceae, and Tetramelaceae capsules (Hutchinson, 1973; Cronquist, 1981; Table 3). With respect to seed number per fruit, there are only two kinds of fruits, with the Anisophylleaceae, Coriariaceae, and Corynocarpaceae having one seed per fruit, and Begoniaceae, Cucurbitaceae, Datisceae, and Tetramelaceae having (very) numerous seeds per fruit. Since the Fagales have one-seeded fruits (Cronquist, 1981; Hutchinson, 1973; Takhtajan, 1997), fruits with one seed are inferred as representing the ancestral state in the Cucurbitales; fruits with numerous seeds represent the derived state.

#### 3.10.5. Sexual systems

Sexual systems in the Cucurbitales are exceptionally labile (e.g., Condon and Gilbert, 1988; Costich and Galan, 1988; Fukuhara and Akimoto, 1999; Roy and Saran, 1990; Thompson and Gornall, 1995; Tobe and Raven, 1988; Wolf et al., 2001), and parsimony reconstruction may underestimate the true number of switches between monoecy, andromonoecy, dioecy, and androdioecy in this order. In addition, sexual systems are still poorly known, especially in the large family Cucurbitaceae (800 species), where possible sex change during ontogeny (documented in at least one genus, *Gurania*) complicates the situation further (Condon and Gilbert, 1988). Population-level variation between monoecy and dioecy, apparently depending on habitat, is well documented in *Ecballium* and *Bryonia*, and crossing studies have shown that dioecy is controlled by a single locus (Correns, 1903; Galán, 1946). Tracing of sexual systems on the phylogeny found for the order (Fig. 1) suggests that bisexual flowers are ancestral in the Cucurbitales, but that dioecy had already evolved prior to the divergence of Begoniaceae, Datisceae, Tetramelaceae, and Cucurbitaceae. It was later lost within Begoniaceae and Cucurbitaceae. Both instances of androdioecy (*Datisca glomerata* and *Schizopepon bryoniifolius*) likely evolved from dioecious ancestors, although our taxon sampling does not show this for *Schizopepon* because we included only the androdioecous species of this genus.

*Schizopepon* has eight species, of which one is androdioecious, four or five are dioecious, and one or two monoecious (Lu, 1985). This frequency distribution makes it likely that *S. bryoniifolius* may have dioecious closest relatives. The closest relatives of *Schizopepon* are several dioecious genera occurring in the Himalayas and China (Kocyan, Zhang, Schaefer, and Renner, unpublished data), all of them dioecious. If androdioecy, which is only known from a handful of species of flowering plants (Pannell, 2002), indeed evolved from dioecy twice within the Cucurbitales, this constitutes a remarkable corroboration of recent hypotheses, which suggest that androdioecy evolves from dioecy, usually when opportunities for cross-fertilization among unisexual plants are scarce.

More detailed work on sexual system evolution within the Cucurbitaceae is ongoing, but based on the present results (Fig. 1) it seems safe to conclude that lineages in this family over the course of their history evolved from dioecy to monoecy and back, mirroring processes at the population level (see above) and cautioning against simple scenarios of one-way paths to dioecy.

## Acknowledgments

We thank K. Pickett for helpful discussions; D. Ama, M. Chase, D.-F. Chen, R. Chung, N. Fujii, T. Fukuhara, G. Gerlach, L. G. Saw, S. Lee, H. Li, E. Mirmanto, M. Olson, V. Plana, H. Setoguchi, S. Swensen, H. Takahashi, W. Takeuchi, S. Wagstaff, and curators of KUN, MO, and NY for plant material; D. Nicolson, W. Greuter, and J. Reveal for advice on the gender of *Corynocarpus*, *Dendrosicyos*, and *Marah*, M. Islam and P. Stevens for suggestions, associate editor Vicent Savolainen and two anonymous reviewers for comments, and the University of Missouri—St. Louis for a research award to S.S.R.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.ympev.2005.10.002](https://doi.org/10.1016/j.ympev.2005.10.002).

## References

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans. Auto. Control* 19, 716–723.
- Akimoto, J., Fukuhara, T., Kikuzawa, K., 1999. Sex ratio and genetic variation in a functionally androdioecious species, *Schizopepon bryoniifolius* (Cucurbitaceae). *Am. J. Bot.* 86, 880–886.
- Albert, V.A., Williams, S.F., Chase, M.W., 1992. Carnivorous plants, phylogeny and structural evolution. *Science* 257, 1491–1494.
- Álvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29, 417–434.
- Anderson, J.A.R., Muller, J., 1975. Palynological study of a Holocene peat and a Miocene coal deposit from NW Borneo. *Rev. Paleobot. Palynol.* 19, 291–317.
- Angiosperm Phylogeny Group (APG), 1998. An ordinal classification for the families of flowering plants. *Ann. Missouri Bot. Gard.* 85, 531–553.
- Angiosperm Phylogeny Group (APG), 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* 141, 399–436.
- Baas, P., Wheeler, E., Chase, M.W., 2000. Dicotyledonous wood anatomy and the APG system of angiosperm classification. *Bot. J. Linn. Soc.* 134, 3–17.
- Bailey, C.D., Carr, T.G., Harris, S.A., Hughes, C.E., 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Mol. Phylogenet. Evol.* 29, 435–455.
- Barkman, T.J., Chenery, G., McNeal, J.R., Lyons-Weiler, J., Elisens, W.J., Moore, G., Wolfe, A.D., dePamphilis, C.W., 2000. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proc. Natl. Acad. Sci. USA* 97, 13166–13171.
- Bergthorsson, U., Adams, K.L., Thomason, B., Palmer, J.D., 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424, 197–201.
- Bremer, K., Bremer, B., Thulin, M., 1997. *Introduction to Phylogeny and Systematics of Flowering Plants*, third ed. Department of Systematic Botany, Uppsala University, Uppsala.
- Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L., Waddell, P.J., 1993. Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* 42, 384–397.
- Campbell, J.D., 2002. Angiosperm fruit and leaf fossils from Miocene silcrete, Landslip Hill, northern Southland, New Zealand. *J. R. Soc. New Zeal.* 32, 149–154.
- Carlquist, S., Miller, R.B., 2001. Wood anatomy of Corynocarpaceae is consistent with its cucurbitalean placement. *Syst. Bot.* 26, 54–65.
- Chandler, M.E.J., 1964. *The Lower Tertiary Floras of Southern England IV: A Summary and Survey of Findings in the Light of Recent Botanical Observations*. British Museum, London.
- Chang, J.T., 1996. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Math. Biosci.* 134, 189–215.
- Chase, M.W., Soltis, D.E., Olmstead, R.G., Morgan, D., Les, D.H., Mishler, B.D., Duvall, M.R., Price, R.A., Hills, H.G., Qiu, Y.-L., Kron, K.A., Rettig, J.H., Conti, E., Palmer, J.D., Manhart, J.R., Sytma, K.J., Michaels, H.J., Kress, W.J., Karol, K.G., Clark, W.D., Hedrén, M., Gaut, B.S., Jansen, R.J., Kim, K.-J., Wimpee, C.F., Smith, J.F., Furnier, G.R., Strauss, S.H., Xiang, Q.-Y., Plunkett, G.M., Soltis, P.S., Swensen, S.M., Williams, S.E., Gadek, P.A., Quinn, C.J., Eguiarte, L.E., Golenberg, E., Learn Jr., G.H., Graham, S.W., Barrett, S.C.H., Dayanandan, S., Albert, V.A., 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Gard.* 80, 528–580.
- Clement, W.L., Tebbitt, M.C., Forrest, L.L., Blair, J.E., Brouillet, L., Eriksson, T., Swensen, S.M., 2004. Phylogenetic position and biogeography of *Hillebrandia sandwicensis* (Begoniaceae): a rare Hawaiian relict. *Am. J. Bot.* 91, 905–917.
- Collinson, M.E., 1986. The Felpham flora—a preliminary report. *Tertiary Res.* 8, 29–32.
- Collinson, M.E., Boulter, M.C., Holmes, P.R., 1993. Magnoliophyta (Angiospermae). In: Benton, M.J. (Ed.), *The Fossil Record 2*. Chapman and Hall, London, pp. 809–841, 864.
- Condon, M.A., Gilbert, L.E., 1988. Sex expression of *Gurania* and *Psiguria* (Cucurbitaceae): Neotropical vines that change sex. *Am. J. Bot.* 75, 875–884.
- Correns, C., 1903. Weitere Beiträge zur Kenntnis der dominierenden Merkmale und der Mosaikbildung der Bastarde. *Ber. Deut. Bot. Ges.* 21, 195–201.
- Costich, D.E., Galan, F., 1988. The ecology of the monoecious and dioecious subspecies of *Ecballium elaterium* (L.) A. Rich. (Cucurbitaceae) I. Geographic distribution and its relationship to climatic conditions in Spain. *Lagascalia* 15 (Suppl.), 697–710.
- Cronquist, A., 1981. *An Integrated System of Classification of the Flowering Plants*. Columbia University Press, New York.
- Cronquist, A., 1988. *The Evolution and Classification of Flowering Plants*, second ed. The New York Botanical Garden, New York.

- Dahlgren, R., 1983. General aspects of angiosperm evolution and macrosystematics. *Nord. J. Bot.* 3, 119–149.
- Dahlgren, R.M.T., 1988. Rhizophoraceae and Anisophylleaceae: summary statement, relationships. *Ann. Missouri Bot. Gard.* 75, 1259–1277.
- Davis, J.I., Simmons, M.P., Stevenson, D.W., Wendel, J.F., 1998. Data decisiveness, data quality, and incongruence in phylogenetic analysis: an example from the monocotyledons using mitochondrial *atpA* sequences. *Syst. Biol.* 47, 282–310.
- Demesure, B., Sodzi, N., Petit, R.J., 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol. Ecol.* 4, 129–131.
- Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Doyle, J.J., 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst. Bot.* 17, 144–163.
- Engler, A., 1896. Coriariaceae. In: Engler, A., Prantl, K. (Eds.), *Die natürlichen Pflanzenfamilien nebst ihren Gattungen und wichtigeren Arten insbesondere den Nutzpflanzen*, vol. 3 (5). W. Engelmann, Leipzig, pp. 8–9.
- Farris, J.S., 1981. Distance data in phylogenetic analysis. In: Funk, V.A., Brooks, D.R. (Eds.), *Advances in Cladistics: Proceedings of the First Meeting of the Willi Hennig Society*. New York Botanical Garden, Bronx, New York, pp. 3–23.
- Farris, J.S., 1989. The retention index and the rescaled consistency index. *Cladistics* 5, 417–419.
- Farris, J.S., Albert, V.A., Källersjö, M., Lipscomb, D., Kluge, A.G., 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12, 99–124.
- Fay, M.F., Swensen, S.M., Chase, M.W., 1997. Taxonomic affinities of *Medusagyne oppositifolia* (Medusagynaceae). *Kew Bull.* 52, 111–120.
- Felsenstein, J., 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst. Zool.* 22, 240–249.
- Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27, 401–410.
- Forrest, L.L., Hughes, M., Hollingsworth, P.M., 2005. A phylogeny of *Begonia* using nuclear ribosomal sequence data and morphological characters. *Syst. Bot.* 30, 671–682.
- Friis, E.M., 1983. Upper Cretaceous (Senonian) floral structures of juglandalean affinity containing *Normapolles* pollen. *Rev. Palaeobot. Palynol.* 39, 161–188.
- Fritsch, P., Rieseberg, L.H., 1992. High outcrossing rates maintain male and hermaphrodite individuals in populations of the flowering plant *Datisca glomerata*. *Nature* 359, 633–636.
- Fukuhara, T., Akimoto, J., 1999. Flora morphology and vasculature of *Schizopepon bryoniaefolius* (Cucurbitaceae). *Acta Phytotax. Geobot.* 50, 59–73.
- Galán, F., 1946. Sur la génétique de la monoécie et la dioécie zygotique chez *Ecballium elaterium* Rich. C.R. Séan. Acad. Sci. Paris 222, 1130–1131.
- Gaut, B.S., Lewis, P.O., 1995. Success of maximum likelihood phylogeny inference in the four-taxon case. *Mol. Biol. Evol.* 12, 152–162.
- Graybeal, A., 1994. Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. *Syst. Biol.* 43, 174–193.
- Gregor, H.-J., 1980. Seeds of the genus *Coriaria* Linne (Coriariaceae) in the European Neogene. *Tertiary Res.* 3, 61–69.
- Hamilton, M.B., 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Mol. Ecol.* 8, 521–523.
- Hillis, D.M., 1987. Molecular versus morphological approaches. *Annu. Rev. Ecol. Syst.* 18, 23–42.
- Hilu, K.W., Borsch, T., Müller, K., Soltis, D.E., Soltis, P.S., Savolainen, V., Chase, M.W., Powell, M.P., Alice, L.A., Evans, R., Sauquet, H., Neinhuis, C., Slotta, T.A.B., Rohwer, J.G., Campbell, C.S., Chatrou, L.W., 2003. Angiosperm phylogeny based on *matK* sequence information. *Am. J. Bot.* 90, 1758–1776.
- Hoot, S.B., Culham, A., Crane, P.R., 1995. The utility of *atpB* gene sequences in resolving phylogenetic relationships: comparison with *rbcL* and 18S ribosomal DNA sequences in the Lardizabalaceae. *Ann. Missouri Bot. Gard.* 82, 194–207.
- Hutchinson, J., 1973. *The Families of Flowering Plants Arranged According to a New System Based on Their Probable Phylogeny*. Clarendon Press, Oxford.
- Jeffrey, C., 2005. A new system of Cucurbitaceae. *Bot. Zhurn.* 90, 332–335.
- Johnson, L.A., Soltis, D.E., 1995. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *matK* sequences. *Ann. Missouri Bot. Gard.* 82, 149.
- Kato, T., Kaneko, T., Sato, S., Nakamura, Y., Tabata, S., 2000. Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. *DNA Res.* 7 (6), 323–330.
- Kluge, A.G., 1989. A concern for evidence and a phylogenetic hypothesis for relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38, 7–25.
- Kluge, A.G., Farris, J.S., 1969. Quantitative phyletics and the evolution of Anurans. *Syst. Zool.* 18, 1–32.
- Kocyan, A., Zhang, L.-B., Renner, S.S., 2004. Towards a densely sampled phylogeny of Cucurbitaceae worldwide. Botany 2004 abstract at <http://www.botanyconference.org/engine/search/index.php?func=detail&aid=326>.
- Kumar, J.K., Sinha, A.K., 2004. Resurgence of natural colourants: a holistic view. *Nat. Prod. Res.* 18, 59–84.
- Kuzoff, R.K., Sweere, J.A., Soltis, D.E., Soltis, P.S., Zimmer, E.A., 1998. The phylogenetic potential of entire 26S rDNA sequences in plants. *Mol. Biol. Evol.* 15, 151–163.
- Lakhanpal, R.N., Verma, J.K., 1965 [1966]. Fossil wood of *Tetrameles* from the Deccan Intertrappean beds of Mohgaonkalan, Madhya Pradesh. *Paleobotanist* 14, 209–213, +2 plates.
- Lakhanpal, R.N., 1970. Tertiary floras of India and their bearing on the historical geology of the region. *Taxon* 19, 675–694.
- Lecointre, G., Deleporte, P., 2005. Total evidence requires exclusion of phylogenetically misleading data. *Zool. Scr.* 34, 101–117.
- Li, R.-Q., Chen, Z.-D., Hong, Y.-P., Lu, A.-M., 2002. Phylogenetic relationships of the “higher” hamamelids based on chloroplast *trnL-F* sequences. *Acta Bot. Sin.* 44, 1462–1468.
- Liston, A.L., Rieseberg, L.H., Hanson, M.A., 1992. Geographic partitioning of chloroplast DNA variation in the genus *Datisca* (Datisceae). *Plant Syst. Evol.* 181, 121–132.
- Lledó, M.D., Crespo, M.B., Cameron, K.M., Fay, M.F., Chase, M.W., 1998. Systematics of Plumbaginaceae based upon cladistic analysis of *rbcL* sequence data. *Syst. Bot.* 23, 21–29.
- Lu, A.-M., 1985. Studies on the genus *Schizopepon* Max. (Cucurbitaceae). *Acta Phytotax. Sin.* 23, 106–120.
- Maddison, D.R., Maddison, W.P., 2001. *MacClade: analysis of phylogeny and character evolution version 4.03*. Sinauer, Sunderland, Massachusetts.
- Manos, P.S., Steele, K.P., 1997. Phylogenetic analysis of “higher” Hamamelididae based on plastid sequence data. *Am. J. Bot.* 84, 1407–1419.
- Manos, P.S., Zhou, Z.-K., Cannon, C.H., 2001. Systematics of Fagaceae: phylogenetic tests of reproductive trait evolution. *Int. J. Plant Sci.* 162, 1361–1379.
- Matthews, M.L., Endress, P.K., Schönenberger, J., Friis, E.M., 2001. A comparison of floral structures of Anisophylleaceae and Cunoniaceae and the problem of their systematic position. *Ann. Bot.* 88, 439–455.
- Matthews, M.L., Endress, P.K., 2004. Comparative floral structure and systematics in Cucurbitales (Corynocarpaceae, Coriariaceae, Tetramelaceae, Datisceae, Begoniaceae, Cucurbitaceae, Anisophylleaceae). *Bot. J. Linn. Soc.* 145, 129–185.
- Melchior, H., 1964. Cucurbitales. In: Melchior, H. (Ed.) *unter Mitarbeit von Buchheim, G., Eckardt, T., Hamann, U., Potzta, E., Scholz, H., Schultze-Motel, W., Schulze-Menz, G.K., Wagenitz, G.A.* Engler’s Syllabus der Pflanzenfamilien, mit besonderer Berücksichtigung der Nutzpflanzen nebst einer Übersicht über die Florenreiche und Florengebiete der Erde, ed. 12. Gebrüder Borntraeger, Berlin-Nikolassee, pp. 341–345.

- Mirza, M.S., Hahn, D., Dobritsa, S.V., Akkermans, A.D., 1994. Phylogenetic studies on uncultured *Frankia* populations in nodules of *Datisca cannabina*. *Can. J. Microbiol.* 40, 313–318.
- Morley, R.J. 1977. Palynology of tertiary and quaternary sediments in Southeast Asia. *Proc. Indonesian Petroleum Assoc., 6th Ann. Conv., 255–276.*
- Müller, E.G.O., Pax, F., 1894. Cucurbitaceae. In: Engler, A., Prantl, K. (Eds.), *Die natürlichen Pflanzenfamilien nebst ihren Gattungen und wichtigeren Arten insbesondere den Nutzpflanzen*, vol. IV, 5 (34). W. Engelmann, Leipzig, pp. 1–39.
- Müller, K., 2005. Incorporating information from length-mutational events into phylogenetic analysis. *Mol. Phylogenet. Evol.*
- Nandi, O.I., Chase, M.W., Endress, P.K., 1998. A combined cladistic analysis of angiosperms using rbcL and non-molecular data sets. *Ann. Missouri Bot. Gard.* 85, 137–212.
- Nickrent, D.L., Starr, E.M., 1994. High rates of nucleotide substitution in nuclear small-subunit (18S) rDNA from holoparasitic flowering plants. *J. Mol. Evol.* 39, 62–70.
- Nixon, K.C., Carpenter, J.M., 1996. On simultaneous analysis. *Cladistics* 12, 221–242.
- Olmstead, R.G., Sweere, J.A., 1994. Combining data in phylogenetic systematics—an empirical approach using three molecular data sets in the Solanaceae. *Syst. Biol.* 43, 467–481.
- Olmstead, R.G., Michaels, H.J., Scott, K.M., Palmer, J.D., 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of rbcL. *Ann. Missouri Bot. Gard.* 79, 249–265.
- Ooi, K.A., Endo, Y., Yokoyama, J., Murakami, N., 1995. Useful primer designs to amplify DNA fragments of the plastid gene *matK* from angiosperm plants. *J. Jap. Bot.* 70, 328.
- Ozaki, K., 1991. Late Miocene and Pliocene floras in Central Honshu, Japan. *Bull. Kanagawa Pref. Mus., Nat. Sci., special issue*, 244 pp.
- Pannell, J., 2002. The evolution and maintenance of androdioecy. *Ann. Rev. Ecol. Syst.* 33, 397–425.
- Pennington, R.T., 1996. Molecular and morphological data provide phylogenetic resolution at different hierarchical levels in *Andira*. *Syst. Biol.* 45, 496–515.
- Plana, V., Gascoigne, A., Forrest, L.L., Harris, D., Pennington, R.T., 2004. Pleistocene and pre-Pleistocene *Begonia* speciation in Africa. *Mol. Phylogenet. Evol.* 31, 449–461.
- Pol, D., 2004. Empirical problems of the hierarchical likelihood ratio test for model selection. *Syst. Biol.* 53, 949–962.
- Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* 53, 793–808.
- Raven, P.H., Axelrod, D.I., 1974. Angiosperm biogeography and past continental movements. *Ann. Missouri Bot. Gard.* 61, 539–637.
- Renner, S.S., Ricklefs, R.E., 1995. Dioecy and its correlates in the flowering plants. *Am. J. Bot.* 82, 596–606.
- Reveal, J.L., 1997. Reveal system of classification. Norton-Brown Herbarium, University of Maryland. At <<http://www.inform.umd.edu/PBIO/pb250/revel.html>>.
- Rieseberg, L.H., Hanson, M.A., Philbrick, C.T., 1992. Androdioecy is derived from dioecy in Datisceae: evidence from restriction site mapping of PCR-amplified chloroplast DNA fragments. *Syst. Bot.* 17, 324–336.
- Roy, R.P., Saran, S., 1990. Sex expression in the Cucurbitaceae. In: Bates, D.M., Robinson, R.W., Jeffrey, C. (Eds.), *Biology and Utilization of the Cucurbitaceae*. Comstock Publ. Associates, Cornell Univ. Press, Ithaca, New York, pp. 251–268.
- Sanderson, M.J., Kim, J., 2000. Parametric phylogenetics? *Syst. Biol.* 49, 817–829.
- Sanjur, O.I., Piperno, D.R., Andres, T.C., Wessel-Beaver, L., 2002. Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: implications for crop plant evolution and areas of origin. *Proc. Natl. Acad. Sci. USA* 99, 535–540.
- de Saporta, G., 1865. Etudes sur la végétation du Sud-Est de la France à l'époque tertiaire. *Ann. Sci. Nat. Bot. Ser* 5 (4), 5–264, 13 plates.
- Savolainen, V., Chase, M.W., Hoot, S.B., Morton, C.M., Soltis, D.E., Bayer, C., Fay, M.F., De Bruijn, A., Sullivan, S., Qiu, Y.-L., 2000a. Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences. *Syst. Biol.* 49, 306–362.
- Savolainen, V., Fay, M.F., Albach, D.C., Backlund, A., van der Bank, M., Cameron, K.M., Johnson, S.A., Lledó, M.D., Pinaud, J.-C., Powell, M., Sheenan, M.C., Soltis, P.S., Soltis, D.E., Weston, P., Whitten, W.M., Wurdack, K.J., Chase, M.W., 2000b. Phylogeny of the eudicots: a nearly complete familial analysis based on rbcL gene sequences. *New Bull.* 55, 257–309.
- Schönenberger, J., Friis, E.M., Matthews, M.L., Endress, P.K., 2001. Cunoniaceae in the Cretaceous of Europe: evidence from fossil flowers. *Ann. Bot.* 88, 423–437.
- Schwarzbach, A.E., Ricklefs, R.E., 2000. Systematic affinities of Rhizophoraceae and Anisophylleaceae, and intergeneric relationships within Rhizophoraceae, based on chloroplast DNA, nuclear ribosomal DNA, and morphology. *Am. J. Bot.* 87, 547–564.
- Setoguchi, H., Kosuge, K., Tobe, H., 1999. Molecular phylogeny of Rhizophoraceae based on rbcL gene sequences. *J. Plant Res.* 112, 443–455.
- Siddall, M.E., 1997. Prior agreement: arbitration or arbitrary? *Syst. Biol.* 46, 765–769.
- Siddall, M.E., 1998. Success of parsimony in the four-taxon case: long-branch repulsion by likelihood in the Farris Zone. *Cladistics* 14, 209–220.
- Siddall, M.E., Whiting, M.F., 1999. Long-branch abstractions. *Cladistics* 15, 9–24.
- Sims, H.J., Herendeen, P.S., Crane, P.R., 1998. New genus of fossil Fagaceae from the Santonian (Late Cretaceous) of Central Georgia, USA. *Int. J. Plant Sci.* 159, 391–404.
- Simmons, M.P., 2004. Independence of alignment and tree search. *Mol. Phylogenet. Evol.* 31, 874–879.
- Simmons, M.P., 2000. A fundamental problem with amino-acid-sequence characters for phylogenetic analyses. *Cladistics* 16, 274–282.
- Simmons, M.P., Freudenstein, J.V., 2002. Artifacts of coding amino acids and other composite characters for phylogenetic analysis. *Cladistics* 18, 354–365.
- Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49, 369–381.
- Simmons, M.P., Ochoterena, H., Freudenstein, J.V., 2002. Conflict between amino acid and nucleotide characters. *Cladistics* 18, 200–206.
- Simmons, M.P., Ochoterena, H., Carr, T.G., 2001. Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. *Syst. Biol.* 50, 454–462.
- Soltis, D.E., Soltis, P.S., Morgan, D.R., Swensen, S.M., Mullin, B.C., Dowd, J.M., Martin, P.G., 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. Sci. USA* 92, 2647–2651.
- Soltis, D.E., Sinters, A.E., Zanis, M.J., Kim, S., Thompson, J.D., Soltis, P.S., Ronse de Craene, L.P., Endress, P.K., Farris, J.S., 2003. Gunnerales are sister to other core eudicots: implications for the evolution of pentamery. *Am. J. Bot.* 90, 461–470.
- Soltis, D.E., Soltis, P.S., Chase, M.W., Mort, M.E., Albach, D.C., Zanis, M., Savolainen, V., Hahn, W.H., Hoot, S.B., Fay, M.F., Axtell, M., Swensen, S.M., Nixon, K.C., Farris, J.S., 2000. Angiosperm phylogeny inferred from a combined data set of 18S rDNA, rbcL, and *atpB* sequences. *Bot. J. Linn. Soc.* 133, 381–461.
- Steele, K.P., Vilgalys, R., 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. *Syst. Bot.* 19, 126–142.
- Stevens, P.F., 2001 (onwards). Angiosperm Phylogeny Website. Version 5, May 2005 (and more or less continuously updated since). <<http://www.mobot.org/MOBOT/research/APweb/>>.

- Swensen, S.M., 1996. The evolution of actinorhizal symbioses: evidence for multiple origins of the symbiotic association. *Am. J. Bot.* 83, 1503–1512.
- Swensen, S.M., Mullin, B.C., 1997. Phylogenetic relationships among actinorhizal plants: the impact of molecular systematics and implications for the evolution of actinorhizal symbioses. *Physiol. Plant* 99, 565–573.
- Swensen, S.M., Mullin, B.C., Chase, M.W., 1994. Phylogenetic affinities of Datisceae based on an analysis of nucleotide sequences from the plastid *rbcL* gene. *Syst. Bot.* 19, 157–168.
- Swensen, S.M., Luthi, J.N., Rieseberg, L.H., 1998. Datisceae revisited: monophyly and the sequence of breeding system evolution. *Syst. Bot.* 23, 157–169.
- Swofford, D.L., 2001. PAUP\*: phylogenetic analysis using parsimony (\* and other methods). Sinauer, Sunderland, Massachusetts.
- Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* 17, 1105–1109.
- Takhtajan, A.L., 1969. Flowering Plants: Origin and Dispersal (Translated by C. Jeffrey). Smithsonian Institution Press, Washington, DC.
- Takhtajan, A.L., 1997. Diversity and Classification of Flowering Plants. Columbia University Press, New York.
- Takhtajan, A.L., 1980. Outline of the classification of flowering plants (Magnoliophyta). *Bot. Rev.* 46, 225–359.
- Thin, N.N., 1997. The vegetation of Cuc Phuong National Park, Vietnam. *Sida* 17, 719–759.
- Thompson, P.N., Gornall, R.J., 1995. Breeding systems in *Coriaria* (Coriariaceae). *Bot. J. Linn. Soc.* 117, 293–304.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL\_X windows interface, flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* 25, 4876–4882.
- Thorne, R.F., 2000 [2001]. The classification and geography of flowering plants: Dicotyledons of the class Angiospermae (subclasses Magnoliidae, Ranunculidae, Caryophyllidae, Dilleniidae, Rosidae, Asteridae, and Lamiidae). *Bot. Rev.* 66, 441–647.
- Tobe, H., Raven, P.H., 1987. Systematic embryology of the Anisophylleaceae. *Ann. Missouri Bot. Gard.* 74, 1–26.
- Tobe, H., Raven, P.H., 1988. Floral morphology and evolution of Anisophylleaceae. *Bot. J. Linn. Soc.* 98, 1–25.
- Wagstaff, S.J., Dawson, M.I., 2000. Classification, origin, and patterns of diversification of *Corynocarpus* (Corynocarpaceae) inferred from DNA sequences. *Syst. Bot.* 25, 134–149.
- Wendel, J.F., Schnabel, A., Seelanan, T., 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc. Natl. Acad. Sci. USA* 92, 280–284.
- Wolf, D.E., Satkoski, J.A., White, K., Rieseberg, L.H., 2001. Sex determination in the androdioecious plant, *Datisca glomerata*, and its dioecious sister species, *D. cannabina*. *Genetics* 159, 1243–1257.
- Won, H., Renner, S.S., 2003. Horizontal gene transfer from flowering plants to *Gnetum*. *Proc. Natl. Acad. Sci. USA* 100, 10824–10829.
- Yokoyama, J., Suzuki, M., Iwatsuki, K., Hasebe, M., 2000. Molecular phylogeny of *Coriaria*, with special emphasis on the disjunct distribution. *Mol. Phylogenet. Evol.* 14, 11–19.
- Zhang, L.-B., Renner, S.S., 2003. The deepest splits in Chloranthaceae as resolved by chloroplast sequences. *Int. J. Plant Sci.* 164 (5 Suppl.), S383–S392.
- Zhang, L.-B., Simmons, M.P., 2006. Phylogeny and delimitation of the Celastrales inferred from nuclear and plastid genes. *Syst. Bot.*, 31.
- Zurawski, G., Clegg, M.T., 1987. Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure–function and phylogenetic studies. *Annu. Rev. Plant Physiol.* 38, 391–418.