



A phylogenetic analysis of the Arecoid Line of palms based on plastid DNA sequence data

William J. Hahn*

Center for Environmental Research and Conservation, Columbia University, New York, NY 10027, USA

Received 17 July 2001; received in revised form 30 November 2001

Abstract

A phylogenetic analysis of the Arecoid Line (*sensu* Moore) of palms was conducted using 7 kb of coding and noncoding plastid DNA sequence data. Recovered maximum-parsimony and maximum-likelihood phylogenies support monophyly for the Arecoid Line relative to the rest of the family but paraphyly for subfamily Arecoideae and polyphyly for subfamily Ceroxyloideae (*sensu* Dransfield and Uhl). Tribes Cocoeae, Geonomeae, Hyophorbeae, and Iriarteae and subfamily Phytelephantoideae were identified as monophyletic as were subfamily Phytelephantoideae + *Ravenia* (tribe Ceroxyleae of Ceroxyloideae), *Podococcus* (tribe Podococceae of Arecoideae) + *Pseudophoenix* (tribe Cyclospatheae of Ceroxyloideae), *Reinhardtia* (tribe Malortieinae) + tribe Cocoeae (both of Arecoideae), and a clade containing all IndoPacific pseudomonomerous genera of tribe Areceae (Arecoideae). A few taxa show spurious resolution with noncoding plastid DNA data but noncoding data are generally congruent with protein-coding data. Biogeographic interpretation suggests a Gondwanan origin for the Arecoid Line with several lineages found on more than one fragment of the former supercontinent and primary diversification in these groups possibly due to continental breakup vicariance. Three groups involving *Cocos*, *Orania*, and the IndoPacific clade demonstrate independent dispersals into the IndoPacific region from a Gondwanan origin. © 2002 Elsevier Science (USA). All rights reserved.

1. Introduction

An extensive array of molecular phylogenetic studies of the palms (Asmussen et al., 2000; Asmussen and Chase, 2001; Baker et al., 1999; Lewis and Doyle, 2001; Hahn, 2002; Uhl et al., 1995) has identified four major groups corresponding to: (1) Calamoideae, (2) Nypoideae, (3) Coryphoideae + Caryoteae, and (4) Arecoideae (minus Caryoteae) + Ceroxyloideae + Phytelephantoideae (all *sensu* Dransfield and Uhl, 1998). Calamoideae are monophyletic in all studies of palms and the monotypic Nypoideae is usually resolved as branching near the base of the family (e.g., Asmussen and Chase, 2001; Hahn, 2002). A general grouping of tribe Caryoteae (subfamily Arecoideae *sensu* Dransfield and Uhl, 1998) and Coryphoideae appears in all molecular studies but a monophyletic grouping of all sampled genera of Coryphoideae + Caryoteae has been evidenced only in the study of Hahn (2002). The fourth group, identified as

the informal “Arecoid Line” by Moore (1973), is the largest of the four groups with approximately 60% of the genera in the family. Support for monophyly of the Arecoid Line is not strong and internal relationships are not completely resolved in any published phylogeny.

A range of molecular data types has been employed in palm family-wide phylogenetic studies including plastid DNA restriction fragment length polymorphisms (RFLPs) (Uhl et al., 1995; Wilson et al., 1990), plastid DNA coding and noncoding sequence data (Asmussen et al., 2000; Asmussen and Chase, 2001; Baker et al., 1999; Hahn, 2002), and nuclear DNA sequence data (Lewis and Doyle, 2001; Hahn, 2002). When well supported, phylogenetic results among these studies are generally congruent and, as more data are added, the degree of phylogenetic resolution and support has continued to increase. This pattern suggests that additional character sampling will ultimately result in a fully resolved phylogeny for the family, particularly for the Arecoid Line taxa where relationships are among the least understood of all higher taxa of palms (Hahn, 1999).

* Fax: +212-854-8188.

E-mail address: wh55@columbia.edu (W.J. Hahn).

Moore (1973) delimited the Arecoideae Line as one of his five principal groups in the family and recognized three main internal groups. Dransfield and Uhl (1986, 1998) and Uhl and Dransfield (1987, 1999) formally described the three internal groups as subfamilies Arecoideae (including the 3 genera of subfamily Caryotoideae *sensu* Moore, 1967), Ceroxyloideae, and Phytelephantoideae (Table 1). In their most recent comprehensive taxonomic summary, Dransfield and Uhl (1998) circumscribed Phytelephantoideae to include 3 genera, Ceroxyloideae with 10 genera assigned to three tribes, and Arecoideae (including Caryoteae) with 113 genera in six tribes and 22 subtribes.

The Phytelephantoideae are very distinctive and defined by several morphological synapomorphies (Barfod et al., 1999). All molecular phylogenetic analyses (e.g., Asmussen et al., 2000; Asmussen and Chase, 2001;

Hahn, 2002; Uhl et al., 1995) support monophyly of the group and indicate an affinity with tribe Ceroxyleae of subfamily Ceroxyloideae although concerns over branch lengths have questioned this latter result (Barfod et al., 1999).

Ceroxyloideae *sensu* Dransfield and Uhl (1998) were considered polyphyletic by these authors as all molecular and morphological phylogenetic studies resolve the subfamily as interdigitating with subfamilies Arecoideae and Phytelephantoideae (e.g., Uhl et al., 1995). The specific relationships of the three Ceroxyloid tribes (Ceroxyleae, Cyclospatheae, and Hyophorbeae) to other palms are uncertain. Ceroxyleae (with four genera) usually group with Phytelephantoideae (e.g., Asmussen and Chase, 2001; Hahn, 2002) but Cyclospatheae (with the single genus *Pseudophoenix*) and Hyophorbeae (the five genera of Chamaedoreoid palms *sensu* Moore, 1973)

Table 1
Overview of the Palmae *sensu* Moore (1973) and Dransfield and Uhl (1986, 1998), Uhl and Dransfield (1986, 1999)

Moore	Dransfield and Uhl
Coryphoid Line	Coryphoideae
Lepidocaryoid Line	Calamoideae
Nypoid Line	Nypoideae
Caryotoid Line	Caryoteae (Arecoideae)
Arecoideae Line	Arecoideae, Ceroxyloideae, and Phytelephantoideae
Pseudophoenicoid palms	Cyclospatheae (Ceroxyloideae)
Ceroxyloid palms	Ceroxyleae (Ceroxyloideae)
Chamaedoreoid palms	Hyophorbeae (Ceroxyloideae)
Iriartoid palms	Iriarteae (Arecoideae)
Podococoid palms	Podococceae (Arecoideae)
Arecoideae palms	Areceae pro parte (Arecoideae)
Triovulate genera	
<i>Manicaria</i> alliance	Manicariinae
<i>Leopoldinia</i> alliance	Leopoldiinae
<i>Reinhardtia</i> alliance	Malortieinae
<i>Orania</i> alliance	Oraniinae
Uniovulate genera	
<i>Euterpe</i> alliance	Euterpeinae
<i>Roystonea</i> alliance	Roystoneinae
<i>Dypsis</i> alliance	Dypsidinae
n/a	Lemurophoenicinae
<i>Masoala</i> and <i>Marojejya</i> alliances	Masoalinae
<i>Beccariophoenix</i> alliance	Cocoeae—Beccariophoenicinae
<i>Archontophoenix</i> alliance	Archontophoenicinae
<i>Cyrtostachys</i> alliance	Cyrtostachydinae
<i>Linospadix</i> alliance	Linospadicinae
<i>Ptychosperma</i> alliance	Ptychospermatinae
<i>Areca</i> alliance	Arecinae
<i>Clinostigma</i> alliance	Iguanurinae
<i>Oncosperma</i> alliance	Oncospermatinae
<i>Sclerosperma</i> alliance	Sclerospermatinae
Cocosoid palms	Cocoeae, pro parte (Arecoideae)
<i>Cocos</i> alliance	Butiinae and Attaleinae
<i>Elaeis</i> alliance	Elaeidinae
<i>Bactris</i> alliance	Bactridinae
Geonomoid palms	Geonomeae (Arecoideae)
<i>Pholidostachys</i> alliance	
<i>Geonoma</i> alliance	
Phytelephantoid palms	Phytelephantoideae

Note. The genus *Lemurophoenix* (Lemurophoenicinae) was unknown to Moore.

are variable in position (e.g., Asmussen and Chase, 2001 vs Lewis and Doyle, 2001).

Subfamily Arecoideae, with over half of the genera in the family, were divided by Moore (1973) and by Dransfield and Uhl (1998) into a series of triovulate and uniovulate (or pseudomonomerous) groups that the latter authors formally recognized as tribes and subtribes (Table 1). Among the most notable differences between the taxonomies of Moore (1973) and Dransfield and Uhl (1998) is the treatment of the Caryotoid palms. Moore (1960) accepted this group as a subfamily more closely allied with his Coryphoid Line (=subfamily Coryphoideae of Dransfield and Uhl, 1998) while Dransfield and Uhl (1998) included the group in their subfamily Arecoideae as tribe Caryoteae.

Morphological phylogenies of the Arecoid Line (Uhl et al., 1995) are poorly resolved and identify only a few patterns of relationship. Studies of plastid DNA data (Asmussen et al., 2000; Asmussen and Chase, 2001; Hahn, 2002; Uhl et al., 1995) only moderately support the Arecoid Line as monophyletic but subfamilies Arcoideae and Ceroxyloideae are always resolved as paraphyletic or polyphyletic. Within the Arecoid Line monophyly is supported for many of the recognized tribes including Cocoeae, Geonomeae, Hyophorbeae, and Iriarteae and some patterns of relationship among major lineages are resolved in these studies.

Arecoid Line relationships estimated from nuclear sequence data (e.g., Lewis and Doyle, 2001; Hahn, 2002) differ from plastid DNA trees in several respects but the differences are not well supported in the nuclear DNA trees. Phylogenetic analysis of malate synthase exon data (Lewis and Doyle, 2001) identified many of the same major lineages as the plastid data but relationships among these lineages were only partly congruent with the plastid DNA results. In a comparison of malate synthase and plastid DNA *rps16* and *trnL-trnF* sequences of 13 palm genera, Lewis and Doyle (2001) found different but poorly supported positions for Calamoideae and Caryoteae. Analyses of malate synthase data for a larger taxon sampling resulted in a loss of basal resolution for the family, leaving relationships among most of the Arecoid Line ambiguous.

Phylogenetic analysis of 65 genera of palms using nuclear 18S nrDNA data (Hahn, 2002) resolved trees that were mostly congruent with the plastid DNA trees in that study, differing mainly in the position of Iriarteae. These relationships, however, were not well supported by 18S data alone.

Although the placement of the Caryotoid palms is one of the most important differences between the taxonomies of Moore (1973) and those of Dransfield and Uhl (1998), all molecular phylogenetic studies have placed the group either among members of Coryphoideae (e.g., Asmussen et al., 2000; Asmussen and Chase, 2001; Hahn, 2002; Uhl et al., 1995) or as equivocal in

position (Lewis and Doyle, 2001). Given this evidence, the Caryotoid palms clearly do not form part of subfamily Arecoideae (*sensu* Dransfield and Uhl, 1998) and are herein taken as an outgroup to the Arecoid Line.

Molecular phylogenetic studies of Arecoid Line relationships generally concur with the taxonomy of Moore (1973) but the results are often poorly supported and several limitations suggest the need for further study. First, monophyly of the Arecoid palms has been supported with plastid DNA sequence data but support is not always strong. Second, monophyly of many of the major lineages within the Arecoid Line has not been firmly established and relationships among the major Arecoid lineages are poorly understood. Further character sampling is needed to resolve these issues and work toward a more completely resolved phylogeny for the Arecoid palms. In this study, a phylogenetic analysis of 7 kb of coding and noncoding plastid DNA sequences was conducted to help resolve relationships among the tribal and subtribal taxa that comprise the Arecoid Line (*sensu* Moore, 1973).

2. Methods and materials

2.1. Plant materials and DNA extraction

Representatives of 51 genera of the Arecoid Line (*sensu* Moore, 1973) were included in this study as well as two representatives of Calamoideae, six of Coryphoideae + Caryoteae, and the single species of Nypoideae as outgroups (Table 1). Outgroup choice was guided by the results of previous molecular phylogenetic studies (Asmussen and Chase, 2001; Hahn, 2002).

Difficulties in alignment for noncoding plastid DNA have created problems in rooting family-wide studies of the palms (Asmussen et al., 2000; Asmussen and Chase, 2001; Baker et al., 1999). For the current study, representatives of the Calamoideae, Nypoideae, Coryphoideae, and Caryoteae were chosen as functional outgroups to test monophyly of the Arecoid Line and to explore internal phylogenetic relationships. Although controversy exists over the basal branching order of palms (e.g., Asmussen and Chase, 2001; Lewis and Doyle, 2001; Hahn, 2002), none of the alternate hypotheses of basal branching elements include the Arecoid palms, thereby justifying use of the three other major groups of palms as outgroups to the Arecoid Line.

Ingroup taxon sampling included representatives of all three subfamilies and all tribes assignable to Moore's (1973) Arecoid Line. Within subfamily Arecoideae, 19 of the 22 subtribes delimited by Dransfield and Uhl (1998) were sampled. When possible, more than one species of each higher taxon was included to test monophyly of that taxon and to control for spurious

resolution caused by solitary long-branch taxa. Material of the West African subtribe Sclerospermatinae and the Malagasy subtribes Lemurophoenicinae and Masoalinae was not available for this study.

Leaf tissue was collected from wild or cultivated material as summarized in Table 2. Methods for collection and storage are as found in Sytsma et al. (1993). Nucleic acids were extracted using a modified CTAB method (Hahn and Sytsma, 2000; Taylor and Powell, 1982) and the same extraction was used to generate all gene sequences analyzed.

2.2. PCR amplification and DNA sequencing

The plastid genes *atpB*, *rbcL*, and *ndhF* were sequenced according to the protocols of Hoot et al. (1995), Olmstead et al. (1992), and Olmstead and Reeves (1995), respectively. The plastid intergenic spacers for *trnD-trnT* (Demesure et al., 1995) and *trnQ-rps16* (W.J. Hahn and K.J. Sytsma, unpublished) were amplified and sequenced using the primer sequences listed in Table 3. The intergenic spacer between *trnQ* (UUG) and *rps16* is composed entirely of noncoding DNA while the span between *trnD* (GUC) and *trnT* (GGU) includes two additional transfer RNA genes: *trnY* (GUA) and *trnE* (UUC).

PCR amplifications were conducted using the Fail-Safe premix (Epicentre, Madison, WI) and *Taq* Poll (Promega, Madison, WI). Sequencing was performed on an Applied Biosystems 377 automated sequencer using BigDye terminator cycle sequencing chemistry (PE Biosystems, Foster City, CA) following the manufacturer's protocols.

2.3. Sequence alignment and gap handling

Contigs of individual sequence fragments were assembled using Sequencher ver.3.1 (GeneCodes, Ann Arbor, MI) and compared against a consensus template to facilitate contig assembly and editing. Multiple sequence alignments were constructed using various gap:substitution weights with ClustalW (Thompson et al., 1994). Sequences of the plastid intergenic spacers were heavily populated with indels and repetitive sequences requiring special attention to alignments. Alignment-ambiguous regions of noncoding DNA were manually adjusted and several regions were deleted following suggestions outlined by Kelchner (2000) and Asmussen and Chase (2001).

All simple sequence (microsatellite) repeat regions based on one to six nucleotide repeat units were excluded from the analyses. This type of sequence variation is probably due to slipped-strand mispairing (Levison and Gutman, 1987), a process that may lead to homoplastic indel events and difficulties in homology assessment. Additionally, the number of repeat units is

known to vary among individuals within some species of palms (G. Second, W.J. Hahn, and J.-C. Pintaud, in prep.).

Indels involving more complex sequence structure were handled in two ways (Simmons and Ochoterena, 2000): (1) with all nonsimple sequence gap sites included and gap character states coded as missing data, and (2) the same but with all potentially parsimony-informative gaps coded as additional binary characters appended to the data matrix. A copy of the data matrix is posted at <http://cerc.columbia.edu/hahn/datasets>.

2.4. Phylogenetic analysis

Phylogenetic relationships were estimated using the maximum-parsimony (MP) and maximum-likelihood (ML) criteria as implemented in PAUP* ver. 4.0b8 (Swofford, 2001). Although many authors have argued that the linked nature of plastid genes is *a priori* evidence of data combinability among different plastid DNA data types (e.g., RFLPs vs sequences; coding vs noncoding sequences), several instances of phylogenetic incongruence between coding and noncoding data have been noted. For example, while no formal incongruence test was used by Asmussen and Chase (2001), inspection of their coding and noncoding plastid DNA consensus trees shows differences in the position of *Nypa* and monophyly of the Arecoideae Line, suggesting the need for additional investigation. In the current study, potential incongruence between coding and noncoding sequence data was examined using the Partition Homogeneity test (PH) (=ILD test; Farris et al., 1994) implemented in PAUP*. Certain limitations of this test are noted (e.g., Yoder et al., 2001) but the test nonetheless provides a reasonable indication of difference in phylogenetic signal between two different data types (in this case, coding vs noncoding plastid DNA sequence data).

Parsimony analyses were conducted on the partitioned coding and noncoding data and on these data combined with simple sequence repeat regions removed and remaining gaps scored as missing data. Analysis was also conducted on the combined data set with additional binary characters as described above. All analyses used a heuristic search strategy with tree bisection-reconnection (TBR), saving all shortest trees at each step (MULPARS), and branch swapping on all trees saved (STEEPEST descent). One hundred random order of taxon entry replicates were conducted to search for islands of most parsimonious trees. Relative support for each clade was assessed by bootstrap analysis (Felsenstein, 1985; Sanderson, 1995) with 1000 resamplings of the data and TBR branch swapping. Bootstrap analyses for the partitioned data sets limited each replicate to a maximum of 200 trees.

Maximum-likelihood analysis was conducted using a subset of the MP trees for model parameter estimation

Table 2
List of specimens used in this study. Taxonomy from Dransfield and Uhl (1998)

	Species	Voucher	GenBank accession numbers				
			<i>atpB</i>	<i>rbcL</i>	<i>ndhF</i>	<i>trnQ-rsp16</i>	<i>trnD-trnT</i>
Calamoideae							
Calameae	<i>Calamus caesius</i> Blume	FTG 64 129	AF233081	AY044619	AY044523	AY04572	AY044474
Lepidocarpaceae	<i>Mauritia flexuosa</i> L. f.	FTG 88 576	AY012416	AY012473	AY044524	AY044573	AY044475
Nypoideae	<i>Nypa fruticans</i> Wurmbr.	FTG s.n.	AY012414	AY012471	AY044525	A044574	AY044476
Coryphoideae							
Corypheae	<i>Thrinax radiata</i> Lodd. ex Schult. & Schult. f.	H7340	AY012402	AY012459	AY044526	AY044575	AY044477
	<i>Livistona speciosa</i> Kurz	H5918	AY012406	AY012463	AY044528	AY044577	AY044479
	<i>Washingtonia filifera</i> (Linden) H. Wend.	H6941	AY012408	AY012465	AY044527	AY044576	AY044478
Phoeniceae	<i>Phoenix dactylifera</i> L.	H6899	AY012411	AY012468	AY044529	AY044578	AY044480
Borasseae	<i>Borassus flabellifer</i> L.	FTG 74 201	AY012412	AY012469	AY044530	AY044579	AY044481
Ceroxyloideae							
Ceroxyleae	<i>Ravenea hildebrandtii</i> C.D. Bouché	FTG 71A	AY012418	AY012475	AY044544	AY044591	AY044495
Cyclospatheae	<i>Pseudophoenix vinifera</i> (Mart.) Becc.	H7732 MONT	AY012417	AY02474	AY044543	AY044590	AY044494
Hyophorbeae	<i>Chamaedorea seifrizii</i> Burret	Hahn s.n.	AF233083	AF206748	AY044540	n.a.	AY044491
	<i>Hyophorbe lagenocaulis</i> (L.H. Bailey) H.E. Moore	FTG 80 509	AY012419	AY012476	AY044542	AY044589	AY044493
	<i>Wendlandiella polyclada</i> Burret	FTG 77 248	AY012420	AY012477	AY044541	n.a.	AY044492
Arecoideae							
Caryoteae	<i>Caryota mitis</i> Lour.	H6627	AF233082	AY044620	AY044531	AY044580	AY044482
Iriarteae	<i>Iriartea deltoidea</i> Ruiz & Pav.	H6340	AF233084	AF233088	AY044545	AY044592	AY044496
	<i>Wettinia hirsuta</i> Burret	FTG 86 409	AY012424	AY012481	AY044546	AY044593	AY044497
Podococceae	<i>Podococcus barteri</i> G. Mann & J. Wendl.	FTG 88 480	AF233086	AF233090	AY004450	AY044597	AY044501
Areceae							
Oraniinae	<i>Orania trispatha</i> (J. Dransf. & N. Uhl) Beentje & J. Dransf.	FTG 92 313	AY012425	AY012482	AY044549	AY044596	AY044500
Manicariinae	<i>Manicaria saccifera</i> J. Gaertn.	H7641	AY012426	AY012483	AY044548	AY044595	AY044499
Leopoldiniinae	<i>Leopoldinia pulchra</i> Mart.	H7642	AY012427	AY012484	AY044547	AY044594	AY044498
Malortieinae	<i>Reinhardtia simplex</i> (Wendl.) Drude ex Dammer	H7811	AY012428	AY012485	AY044551	AY044598	AY044502
Dypsidinae	<i>Dypsis lastelliana</i> (Baill.) Beentje & J. Dransf.	FTG 88 171A	AY012429	AY012486	AY044534	AY044583	AY044485
Euterpeinae	<i>Prestoea acuminata</i> (Willd.) H.E. Moore	H7650 MONT	AY012430	AY012487	AY044553	AY044600	AY044504
	<i>Oenocarpus bataua</i> Mart.	FTG 88 134	AY044461	AY044624	AY044552	AY044599	AY044503
Roystoneinae	<i>Roystonea regia</i> (Kunth) O.F. Cook	FTG 92 386	AY012431	AY012488	AY044554	AY044601	AY044505
Archontophoenicinae	<i>Archontophoenix alexandrae</i> F.v. Mueller	FTG 83 119	AF449170	AF449156	AF449142	AF449145	AF449159
	<i>Chambeyronia macrocarpa</i> (Brongn.) Vieill. ex Becc.	FTG 77 147B	AY012432	AY012489	AY044536	AY044585	AY044487
Cyrstostachyidinae	<i>Cyrstostachys renda</i> Blume	FTG 87 688	AY012434	AY012491	AF453470	AF449149	AF449163
Linospadicinae	<i>Howea belmoreana</i> (C. Moore & F.v. Mueller) Becc.	FTG 73 337	AY012435	AY012492	AF453472	AF449151	AF449165
	<i>Linospadix longicruris</i> (Becc.) Burret	FTG 81 613	AF449172	AF449158	AF449144	AF449153	AF449167
Ptychospermatinae	<i>Drymophloeus beguinii</i> (Burret) H.E. Moore	FTG 91 31A	AY012437	AY012494	AY044537	AY044586	AY044488
	<i>Ptychosperma burretiana</i> Essig	FTG 81 592	AY012438	AY012495	AF453475	AF449155	AF449169
Arecinae	<i>Areca vestiaria</i> Giseke	FTG 73 451	AY012440	AY012497	AY044535	AY044584	AY044486
	<i>Gronophyllum pinangoides</i> (Becc.) Essig & B.E. Young	FTG 81 618	AY012441	AY012498	AF453471	AF449150	AF449164
	<i>Hydriastele wendlandiana</i> (F.v. Mueller) H.A. Wend. & Drude	FTG 57 793	AY012447	AY012504	AF453473	AF449152	AF449166
Iguanurinae	<i>Bentinckia nicobarica</i> (Kurz) Becc.	FTG 88 1007	AY012442	AY012499	AF453468	AF449146	AF449160
	<i>Burretioekentia hapala</i> H.E. Moore	FTG 89 689	AY012443	AY012500	AF453469	AF449147	AF449161
	<i>Clinostigma savoryanum</i> (Rehd. & Wilson) H.E. Moore & Fosberg	FTG 90 391	AF449171	AF449157	AF449145	AF449148	AF449162

Table 2 (continued)

	Species	Voucher	GenBank accession numbers				
			<i>atpB</i>	<i>rbcL</i>	<i>ndhF</i>	<i>trnQ-rsp16</i>	<i>trnD-trnT</i>
Oncospermatinae	<i>Oncosperma tigillarum</i> (Jack) Ridley	FTG 152	AY012448	AY012505	AF453474	AF449154	AF449155
Cocoeae							
Beccariophoenicinae	<i>Beccariophoenix madagascariensis</i> Jum. & H. Perrier	FTG 91 308	AY012449	AY012506	AY044563	AY044610	AY044514
Butiinae	<i>Allagoptera arenaria</i> (Gomes) Kuntze	FTG 71 442	AY044468	AY044631	AY044564	AY044611	AY044515
	<i>Butia eriospatha</i> (Mart.) Becc.	FTG 92 271	AY044469	AY044632	AY044565	AY044612	AY044526
	<i>Cocos nucifera</i> L.	FTG 80 798	AY012450	AY012507	AY044566	AY044613	AY044517
	<i>Lytocaryum weddelianum</i> (H. Wend.) Tol.	FTG 88 556	AY044470	AY044633	AY044567	AY044614	AY044518
	<i>Syagrus glaucescens</i> Glaz. ex Becc.	FTG 93 109	AY044471	AY044634	AY04458	AY044615	AY044519
	<i>Voanaiola gerardii</i> J. Dransf.	FTG 90 80	AY044472	AY044635	AY044569	AY044616	AY044520
Attaleinae	<i>Orbignya barbosiana</i> Burret	FTG 75 644	AY012451	AY012508	AY044570	AY044617	AY044521
	<i>Scheelea butyracea</i> (Mutis ex Mart.) Karst. ex Wendl.	FTG 93 105	AY044473	AY044636	AY044571	AY044618	AY044522
Elaeidinae	<i>Barcella odora</i> (Trail) Drude	Pintaud 15.1	AY044467	AY044630	AY044561	AY044608	AY044512
	<i>Elaeis oleifera</i> (Kunth) Cortés	FTG 87 117	AY012452	AY012509	AY044562	AY044609	AY044513
Bactridinae	<i>Acrocomia aculeata</i> Mart.	Noblick 5019	AY044462	AY044625	AY044555	AY044602	AY044506
		MONT					
	<i>Aiphanes aculeata</i> Willd.	FTG 90 136	AY044463	AY044626	AY044556	AY044603	AY044507
	<i>Astrocaryum alatum</i> Loomis	FTG 78 424	AY012453	AY012510	AY044557	AY044604	AY044508
	<i>Bactris humilis</i> (Wall.) Burret	FTG 94 1212	AY044464	AY044627	AY044558	AY044605	AY044509
	<i>Desmoncus orthacanthos</i> Mart.	H7900	AY044465	AY044628	AY044559	AY044606	AY044510
	<i>Gastrococcus crisper</i> (Kunth) H.E. Moore	Perry s.n.	AY044466	AY044629	AY044560	AY044607	AY044511
		MONT					
Genomeae	<i>Calyptronoma occidentalis</i> (Sw.) H.E. Moore	FTG 75 159	AY044459	AY044622	AY044538	AY044587	AY044489
	<i>Geonoma oxycarpa</i> Mart.	FTG 86 408	AY044460	AY044623	AY044539	AY044688	AY044490
Phytelephantoideae	<i>Phytelephas aequatorialis</i> Spruce	H6247	AY012455	AY012512	AY044533	AY044582	AY044484
	<i>Aphandra natalia</i> (Balslev & Henderson) Barfod	Pintaud 20	AY044458	AY044619	AY044532	AY044581	AY044483

Note. Vouchers (H, Hahn) deposited at the University Wisconsin-Madison unless otherwise indicated (MONT, Montgomery Botanical Center). Cultivated plants from Fairchild Tropical Garden annotated with FTG and accession number.

Table 3
Plastid spacer primers used in this study

<i>trnD–trnT</i>	
<i>trnD</i> (GUC)	ACC AAT TGA ACT ACA ATC CC
D + 420	AAC CAG CGT AGA CAT ATC GT
D + 420R	ACG ATA TGT CTA CGC TGG TT
TR + 480	CGG TTA ATG GGG ACG GAC TG
TR + 480R	CAG TCC GTC CCC ATT AAC CG
<i>trnT</i> (GGU)	CTA CCA CTG AGT TAA AAG GG
<i>trnQ–rps16</i>	
<i>trnQ</i> (UUG)	TCG GAG GTT CGA ATC CTT CCG TCC CAG A
PQ + 300	TAC TTA TGG AAA CTT TAT GTC
P16 + 300	GTT CGG GAC GAT CAA TCA TGA
P16 + 400	GAT TTG CAA TCC CGG AAC AA
<i>rps16</i> 5'	CAA GTC CGA CGT TGC TTT CTA CCA CAT CGT TT

and as starting points for more extensive ML tree searches. A full general time reversible model (GTR; Lanave et al., 1984; Yang, 1994; Yang et al., 1994) was used and among-site rate variability was modeled assuming a fixed proportion of invariable sites (PInvar estimated) with gamma-distributed rate variation (discrete approximation of the shape parameter with eight rate categories) across sites free to vary ($I + \Gamma$; Gu et al., 1995). Previous studies of plastid DNA sequence data from palms (Hahn, 2002) have identified GTR + $I + \Gamma$ as the simplest acceptable model to describe patterns of variation (Goldman, 1993; Whelan and Goldman, 1999).

All ML model parameter estimates were calculated by averaging across the combined data set without the added binary gap characters. Although there are potential problems with averaging parameter estimates across mixed data sets (Sullivan, 1996; Yang, 1996), mixed likelihood models are currently not possible with the software used in this analysis. All optimal MP trees were used as starting points for subsequent ML tree searches in an iterative procedure (Fрати et al., 1997; Mallatt and Sullivan, 1998; Sullivan et al., 1997; Swofford et al., 1996). Parameter estimates were averaged across a sample of the MP starting trees for each ML search. Heuristic searches were conducted with TBR branch swapping to completion. All model parameter values were reestimated on the new tree and a new

search was initiated with the new parameter values held fixed. This procedure was repeated until no new topologies were recovered. Branch support was estimated by ML bootstrap using the FASTSTEP procedure and 100 replicates with parameter estimates from the ML tree were held fixed for each replicate.

The ML trees were compared with each other and with the various MP trees to determine whether significant differences existed. Significance ($P < 0.05$) was tested using the Shimodaira–Hasegawa (SH) test as implemented in PAUP* (Shimodaira and Hasegawa, 1999; Swofford, 2001). All model parameters were reestimated for each tree. The RELL method of Kishino et al. (1990) was used for bootstrap replications (Goldman et al., 2000; Shimodaira and Hasegawa, 1999).

3. Results

3.1. Sequence data and patterns of molecular evolution

A summary of sequence characteristics for each of the five loci is presented in Table 4. Sequence data for the *trnQ–rps16* spacers were not recovered for the closely related *Chamaedorea* and *Wendlandiella*. All attempts to amplify the region were unsuccessful, suggesting that modifications of the *trnQ–rps16* region have occurred in these closely related taxa. Further study is needed to

Table 4
Characteristics of the five DNA sequences used in this study

Locus	Aligned length	Number of gap sites	Number of variable sites; 1st, 2nd, and 3rd positions for protein-coding sequences	Parsimony-informative sites	Range of P -values for non-repeat sites	Mean GC content: all sites; all variable sites; 1st, 2nd, and 3rd positions for protein-coding sequences
<i>atpB</i>	1515	0	110 (7.3%); 17, 21, 72	33 (2.1%)	0.0–1.0%	43; 51; 64, 34, 54
<i>ndhF</i>	2095	6	381 (18.2%); 112, 93, 176	189 (9.0%)	0.2–4.0%	32; 47; 54, 53, 39
<i>rbcL</i>	1428	0	177 (12.4%); 38, 34, 105	55 (3.9%)	0.2–2.6%	44; 44; 65, 31, 41
<i>trnQ–rps16</i>	941	415	182 (19.3%)	51 (5.4%)	0.2–6.1%	35; 52
<i>trnD–trnT</i>	1150	753	303 (26.3%)	79 (6.9%)	0.0–6.3%	30; 52

Note. Values based on 60-taxon dataset (*Aphandra*, *Chamaedorea*, and *Wendlandiella* removed from *trnQ–rps16* calculations).

determine the nature and extent of these changes. The genus *Aphandra* possessed a stretch of highly divergent DNA in the *trnQ-rps16* spacer and was removed from sequence data calculations for that locus.

The three protein-coding loci were mostly invariant in length with one insertion of six nucleotides detected in the *ndhF* sequence of *Phytelephas* and only minor variation in the position of the first stop codon in some species for *atpB*. In contrast, considerable length variation was noted for both of the plastid DNA intergenic spacers as reported for other plastid DNA intergenic spacers and introns for palms (Asmussen et al., 2000; Baker et al., 1999). Some of the length variation was due to simple sequence repeats and two alignment-ambiguous regions were removed from each of the two non-coding sequences as noted above. Numerous additional gaps involving more complex sequence remained and 10 parsimony-informative indels were scored as binary characters for the *trnD-trnT* spacer and 11 for *trnQ-rps16*. Several indels restricted to a single taxon were noted but were not included as independent characters.

Base composition for variable sites was relatively uniform among taxa and across loci (GC = 44–52%). When constant sites were considered also, GC content was as low as 30% for the *trnD-trnT* spacer and 32% for *ndhF* (Table 4). The noncoding data had the highest percentage of variable sites (24.2% vs 13.3% for coding data) but the difference in parsimony-informative sites was less severe (6.5% for noncoding vs 5.5% for coding sequence data). Removal of all gap sites had only a small effect on the percentage of variable and parsimony-informative noncoding characters.

3.2. Phylogenetic results

Maximum-parsimony tree searches of the noncoding sequence data with simple sequence repeat sites removed and gaps scored as missing data were stopped at 25,000 shortest trees of length (L) = 629, a consistency index (CI) = 0.83, and a retention index (RI) = 0.74 (Table 5). A strict consensus of these trees (Fig. 1A) demonstrates some structure among the main groups of Arecoids but with comparatively little bootstrap support (BS). Several sets of relationships contradict current taxonomic

systems, particularly the positions of *Butia*, *Calyptronoma*, *Pseudophoenix*, and *Wettinia*.

Parsimony analysis of the coding sequence data (*atpB*, *ndhF*, and *rbcL*) recovered 13,019 trees (L = 1150, CI = 0.64, RI = 0.62). A strict consensus of these (Fig. 1B) identified several recognized tribes and subtribes as monophyletic but relationships among the major lineages of the Arecoid Line are poorly resolved and weakly supported. There were comparatively few relationships that contradicted current taxonomies.

Partition homogeneity tests identified incongruence between coding and noncoding data ($P = 0.01$), suggesting that different phylogenetic signal is present in each of these subsets of the data. Incongruence was not detected ($P = 0.42$) when *Butia*, *Calyptronoma*, *Pseudophoenix*, and *Wettinia* were removed from the PH tests. Inspection of the noncoding data for these taxa revealed several unique indels and some stretches of highly divergent sequence but the spurious relationships appeared to be due to homoplastic point mutations.

Parsimony analysis of the combined data with simple sequence repeat regions removed and gaps scored as missing data recovered 2016 shortest trees (L = 1705 steps, CI = 0.69, RI = 0.62). A strict consensus of these trees (Fig. 1C) showed considerable resolution with most basal branching patterns showing moderate to strong levels of bootstrap support (>60% BS). Several differences were noted between the combined data trees and the noncoding data trees (which were 30 parsimony steps longer) but the principal difference between the shortest combined data and the coding data MP trees (which were 2 parsimony steps longer) was in the position of *Acrocomia* + *Gastrococos*.

Maximum-parsimony analysis of the combined sequence data with added binary indel characters recovered 480 shortest trees (L = 7017, CI = 0.69, RI = 0.63) that constitute a subset of the trees found in the combined data analyses without added binary characters. Considerable structure is seen in the strict consensus (Fig. 2), including a monophyletic Arecoid Line (68% BS), a paraphyletic Arecoideae, a polyphyletic Ceroxyloideae, indications of relationship among many of the major Arecoid Line lineages, monophyly of most tribal and subtribal groups, and a grouping of the Caryotoid palms with *Borassus* of subfamily Coryphoideae.

Table 5
Character statistics on the maximum-parsimony trees recovered in this study

Data set	No. characters	No. variable characters	No. informative characters	No. trees	Length	CI	RI
Sequence data, all gap sites removed	5970	912 (15.3%)	343 (5.7%)	2940	1502	0.67	0.62
Sequence data, repeat DNA removed	6996	1142 (16.3%)	405 (5.8%)	2016	1705	0.69	0.62
Coding sequence only	5038	668 (13.3%)	277 (5.5%)	13,019	1150	0.64	0.62
Noncoding sequence only	1958	474 (24.2%)	128 (6.5%)	>25,000	629	0.83	0.74
Sequence data + binary gap characters	7017	1163 (16.6%)	426 (6.1%)	480	1848	0.69	0.62

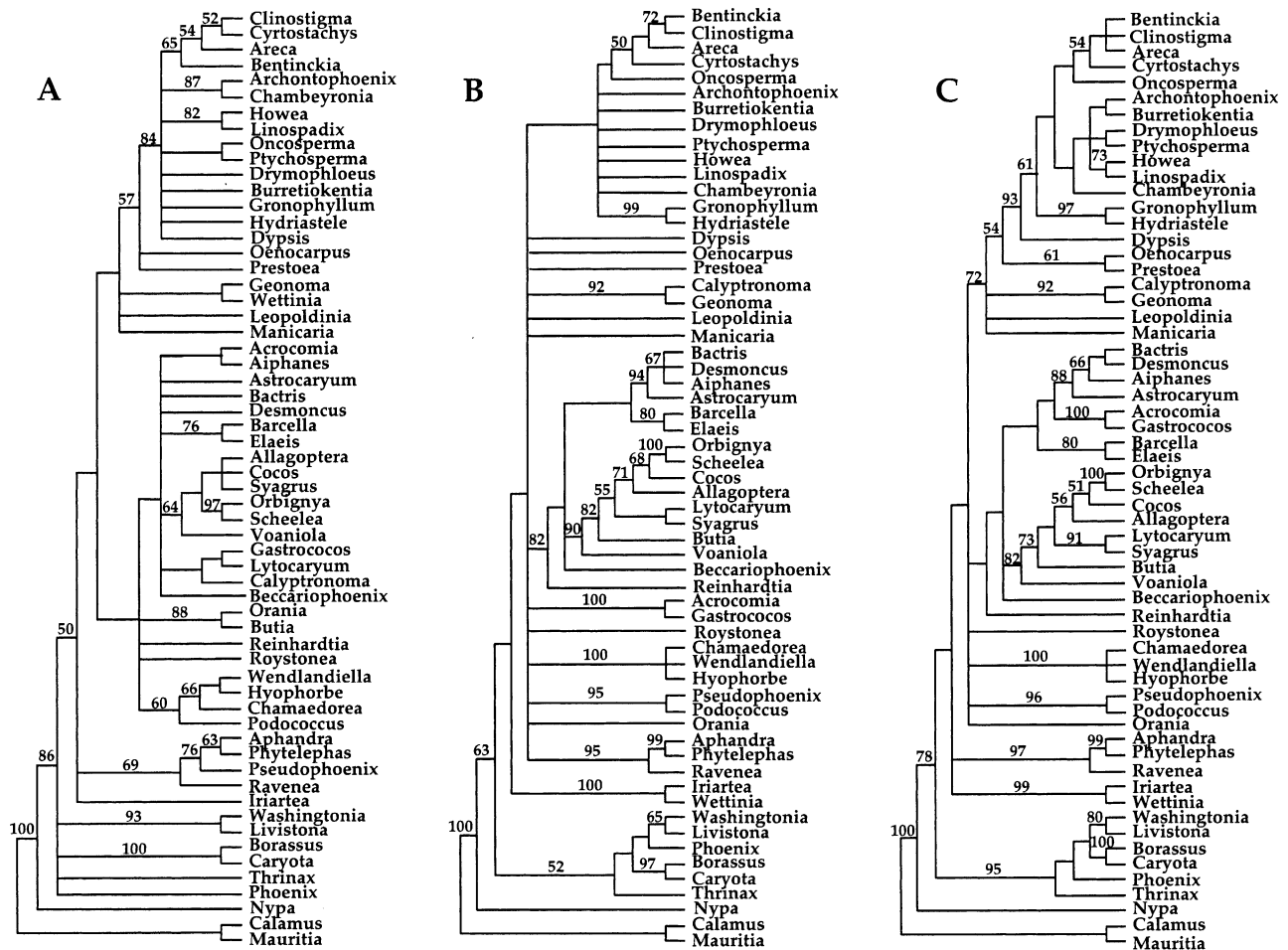


Fig. 1. Strict consensus trees of maximum-parsimony analysis on sequence data with simple sequence data removed. (A) Consensus of 25,000 MP trees from analysis of 1958 bp of noncoding data ($L = 629$, $CI = 0.83$, $RI = 0.74$). (B) Consensus of 12,019 MP trees from analysis of 5038 bp of coding data ($L = 1150$, $CI = 0.64$, $RI = 0.62$). (C) Consensus of 2016 MP trees from analysis of 6996 bp of combined coding and noncoding data ($L = 1705$, $CI = 0.69$, $RI = 0.62$). Bootstrap support values $> 50\%$ are indicated above branches.

Relationships among the major Arecoideae lineages are resolved but not well supported. Iriarteae are sister to the remainder of the Arecoideae with *Phytelephantoidae* + *Ravenea* as the next branching element (65% BS). Arecoideae relationships that are well supported include monophyly of tribes *Hyophorbeae* (100% BS), *Iriarteae* (100% BS), and *Geonomeae* (97% BS) and subfamily *Phytelephantoidae* (97% BS). Additional relationships that are supported include *Phytelephantoidae* with *Ravenea* (tribe *Ceroxyleae* of *Ceroxyloideae*; 92% BS), *Podococcus* (tribe *Podococceae* of *Arecoideae*) with *Pseudophoenix* (tribe *Cyclospatheae* of *Ceroxyloideae*; 92% BS), and *Reinhardtia* (tribe *Malortieinae*) with tribe *Cocoeae* (both of *Arecoideae*; 80% BS).

With the exception of *Roystonea*, all pseudomonomerous *Areceae* (Table 1) form a clade with the trivulate *Leopoldinia* and *Manicaria* plus tribe *Geonomeae* (77% BS). Within this clade, a well-supported (93% BS) grouping of the IndoPacific pseudomonomerous *Arecoideae* taxa (*Archontophoenicinae*, *Areceae*,

Cyrtostachyinae, *Dypsidinae*, *Iguanurinae*, *Linospadicinae*, *Oncospermatinae*, and *Ptychospermatinae*) is also recovered. Of these subtribes that were sampled with more than one taxon, *Ptychospermatinae* and *Linospadicinae* were resolved as monophyletic (55 and 84% BS, respectively). Subtribes *Archontophoenicinae*, *Areceae*, and *Iguanurinae* were polyphyletic but with mixed support for the recovered resolution.

Cocoeae were resolved as monophyletic (62% BS). Subtribe *Elaeidinae* was monophyletic (97% BS) but support for monophyly of *Bactridinae* was only 56%. These two subtribes were resolved as sister taxa but without much support (52% BS). Subtribe *Beccario-phoenicinae* is sister to the other nonspiny *Cocoeae* (61% BS) and subtribes *Attaleinae* and *Butiinae* form a well-supported clade (87% BS). However, subtribe *Butiinae* is paraphyletic with *Attaleinae* embedded within.

Maximum-likelihood analysis of the complete DNA data set recovered a topology that was one parsimony step longer than the complete data MP trees (Fig. 3)

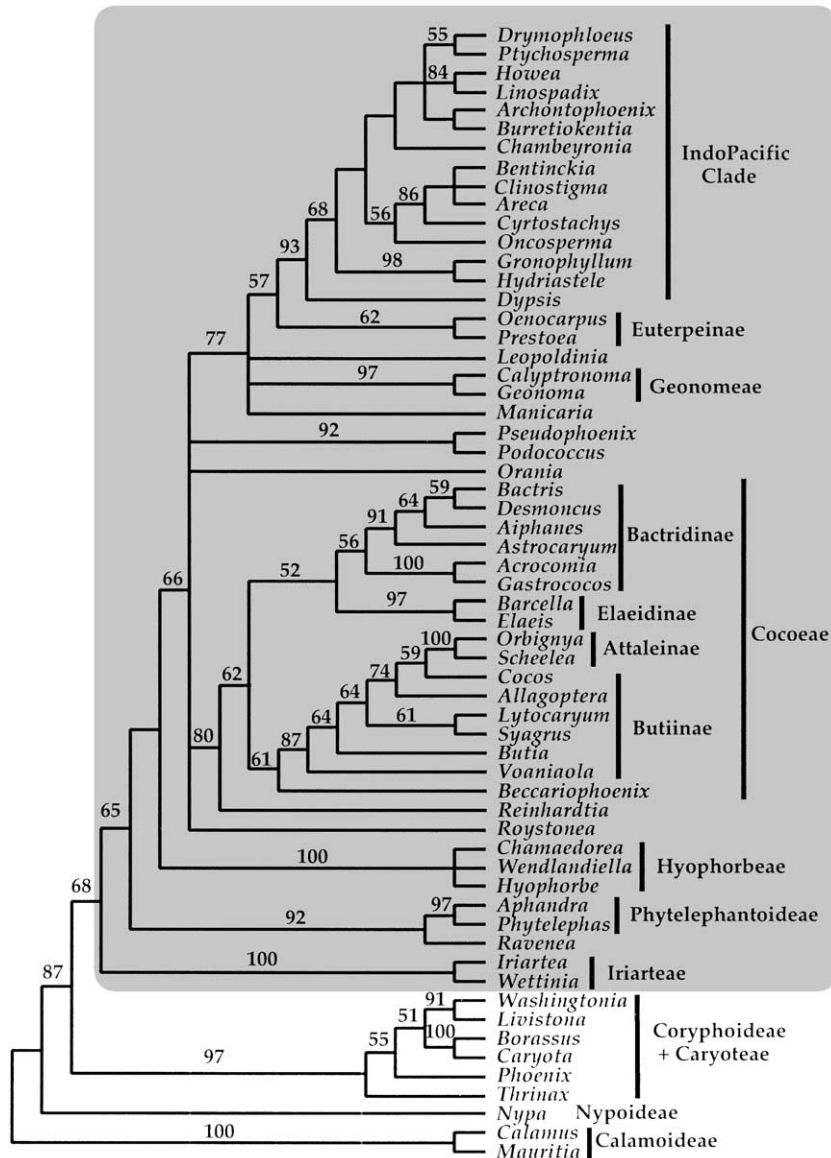


Fig. 2. Strict consensus of 480 maximum-parsimony trees from the combined coding and noncoding cpDNA data set with binary indel characters included. The Arecoideae Line is in the shaded area. Length = 1848; CI = 0.69; RI = 0.62. Bootstrap support values > 50% are indicated above branches.

and, except for identifying a monophyletic Archontophoenicinae, was identical to one of the combined data MP trees. Measures of support generally reflect those observed in the MP tree and the ML tree was not significantly different from any of the complete DNA data MP trees as determined by the SH test.

Shimodaira–Hasegawa tests on each data set and each MP tree set (coding, noncoding, and combined sequence data) indicated that the noncoding and coding data were different in phylogenetic signal. The noncoding data MP trees were rejected by both the coding data and the combined sequence data while the coding and combined data trees were rejected by the noncoding data (all $P \ll 0.01$). In contrast, the coding and combined sequence data MP trees were not rejected by the

reciprocal data set. Given the larger number of variable and informative characters in the coding data, the combined data trees were probably more heavily influenced by coding data than by noncoding data but, unlike the results in the PH tests, removal of *Butia*, *Calyptronoma*, *Pseudophoenix*, and *Wettinia* did not change the SH test results.

Estimates of ML model parameters demonstrate several forms of biased evolution that differed between coding and noncoding DNA (Table 6). In particular, differences are seen between coding and noncoding DNA in their substitution dynamics. For example, a stronger transition:transversion bias is seen in coding DNA (3.36:1) relative to noncoding DNA (1.77:1). At a more detailed level, AC changes were more frequent in

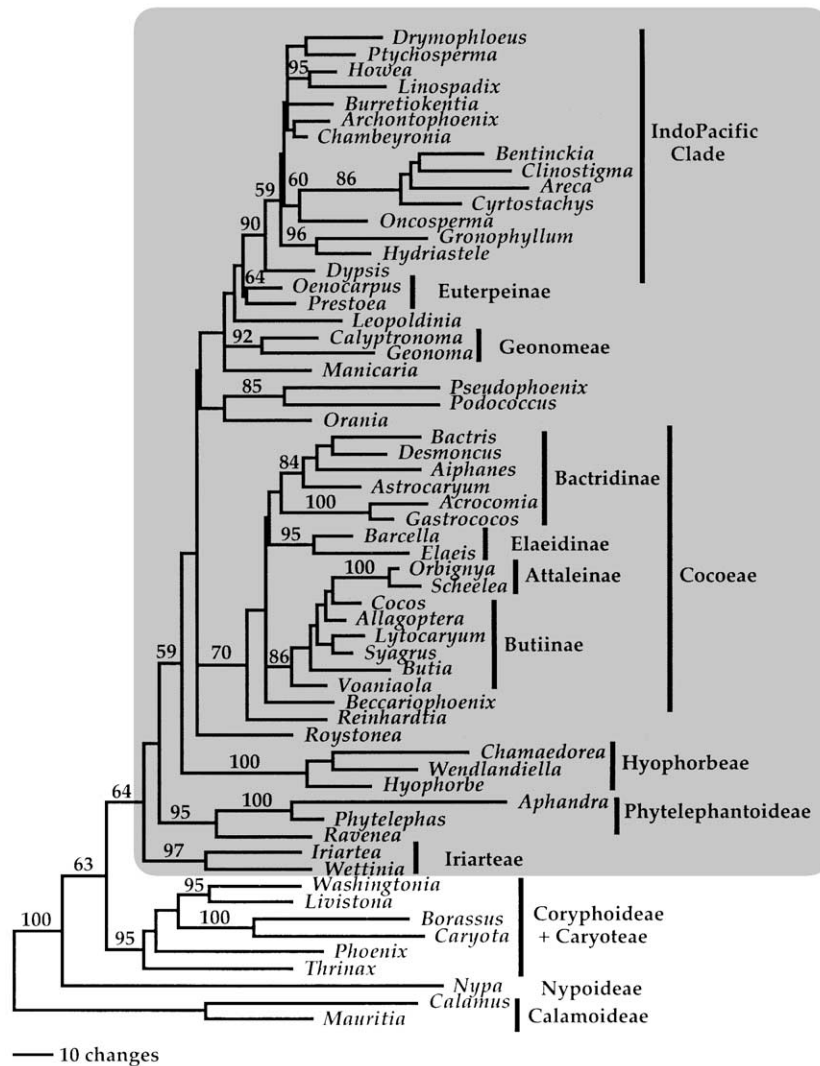


Fig. 3. Maximum-likelihood tree from combined coding and noncoding cpDNA data set. The Arecoideae Line is in the shaded area. ML bootstrap support values $> 50\%$ are indicated above branches. $-\log$ likelihood = 22280.09. Substitution model parameters: AC = 2.001; AG = 3.205; AT = 0.620; CG = 1.015; CT = 2.943; GT = 1; PInvar = 0.512; shape parameter alpha = 0.623.

Table 6
Likelihood scores and model parameter estimates for coding and noncoding data for the respective ML tree

Data set	$-\log$ likelihood ML tree	R-matrix (dAC, dAG, dAT, dCG, cCT, dGT)	Percentage invariable sites	Alpha shape parameter
Combined sequence data	22280.09	2.001, 3.205, 0.620, 1.015, 2.943, 1	0.512	0.623
Coding DNA	14949.43	2.657, 4.622, 0.584, 0.982, 4.142, 1	0.645	0.557
Noncoding DNA	6833.03	1.206, 1.634, 0.474, 1.052, 1.672, 1	0.201	1.088

coding regions than in noncoding regions (2.66 vs 1.21) and a paucity of AT changes is observed in both coding and noncoding DNA (each ca. 50% of the total transversion average for the respective data set). The percentage of invariant sites was higher in coding regions (65% vs 20% for coding) and the coding data showed greater among-site rate variation (shape parameter = 0.56 for coding data vs 1.09 for noncoding).

4. Discussion

The primary results of this study include evidence of monophyly for the Arecoideae of Moore (1973), paraphyly of Arecoideae, polyphyly of Ceroxyloideae (*sensu* Dransfield and Uhl, 1998), and improved phylogenetic resolution for the major groups of Arecoideae palms. Additionally, although several patterns of biased

molecular evolution were noted, the results of this study support the use of noncoding plastid DNA for palm phylogeny reconstruction. Several of these results have been observed in other studies of palm relationships but this study demonstrates generally higher levels of resolution and support.

4.1. Molecular evolution

Both nuclear protein-coding (Lewis and Doyle, 2001; Morton et al., 1997) and nuclear ribosomal (Baker et al., 2000; Hahn, 2002) data have been used in palm phylogeny reconstruction, but most molecular phylogenetic studies of palms have relied on plastid DNA information (e.g., Asmussen et al., 2000; Asmussen and Chase, 2001; Baker et al., 1999; Hahn, 2002; Uhl et al., 1995; Wilson et al., 1990). Noncoding plastid DNA sequence data have been used by several authors (e.g., Asmussen et al., 2000; Baker et al., 1999) and a comparison of coding and noncoding plastid DNA in palms was conducted by Asmussen and Chase (2001).

In their study, Asmussen and Chase (2001) noted that the coding sequence *rbcL* and the noncoding introns and spacers for *rps16* and *trnL-trnF* demonstrated roughly the same number of variable sites per total number of sites examined (6.58, 10.83, and 7.38%, respectively), but that the noncoding sequences demonstrated lower homoplasy (RI of 0.69, 0.89, and 0.89, respectively) and higher branch support (14, 23, and 19 clades >50% BS, respectively). These authors concluded that the noncoding regions, specifically *rps16* introns, were more “useful” than *rbcL* although they cautioned that the exclusion of alignment-ambiguous sites in *rps16* can be at least partly responsible for this result.

In the current study, comparisons of coding and noncoding DNA echo many of the observations made by Asmussen and Chase (2001) and other workers who have examined plastid DNA evolution in other taxa (e.g., Kelchner and Clark, 1997, for grasses). The coding sequences for *atpB*, *ndhF*, and *rbcL* (5038 bp in total) demonstrated roughly the same percentage (5.5%) of parsimony-informative sites per bases sampled as the 1958 bp of noncoding, nonrepeat DNA sampled (6.5%) but the RI for noncoding data was higher than that for coding data (0.74 and 0.62, respectively).

The current study sampled more than twice the number of characters (6996 vs 3186) and more than twice as many parsimony-informative characters (272 vs 139) than Asmussen and Chase (2001) if only Arecoide Line taxa are considered. Of the trees recovered, the coding data trees are more similar to the combined data trees than are the noncoding data trees, probably because of the greater number of parsimony-informative coding data characters than noncoding (277 vs 128). The apparently anomalous positions of four taxa in the noncoding trees (*Butia*, *Calyptrotrichia*, *Pseudophoenix*,

and *Wettinia*) are at least partly responsible for these differences.

Additional factors that might be involved in the differences in resolution include patterns of molecular evolution in the plastid DNA spacer sequences such as indels, inversions, and inverted repeats. The frequent occurrence of indel events in noncoding plastid DNA sequences is well known (e.g., Golenberg et al., 1993; Kelchner, 2000) and the sequences in the current study show variation within the range described by these authors. The distributions of indel types also follow general patterns with a mix of different simple repeat sequences (1- to 6-bp repeat units), indels associated with inversions and other rearrangements, and larger indels involving more complex sequence structure. Despite these potential problems, most of the indels found in the current study are confined to clades identified with base substitution data alone such as *Caryota* + *Borassus*, *Acrocomia* + *Gastrococcus*, Bactridinae + Elaeidinae, Elaeidinae, Euterpeinae, Geonomeae, Iriarteae, Calamoidae, Nypoideae, and Phytelphantoideae.

Patterns of base substitution are also different between coding and noncoding DNA in the current study. Because of the lack of codon-based constraints, sequences of noncoding plastid DNA are known to show less transition:transversion bias than coding plastid DNA (Morton, 1995; Morton et al., 1997). This pattern is maintained in the present analysis as transitions are 3.36 times more frequent than transversions for coding DNA but only 1.77 times more frequent for noncoding DNA (Table 6). Elevated AC transversions in coding DNA are most probably due to three-fold degenerate sites for isoleucine—the only instance of a silent transversion in the genetic code. The low levels of AT changes found in coding and noncoding DNA are seen in most studies of plastid DNA and are probably associated with the general AT bias of the plastid genome.

Although the patterns of nucleotide substitution for the coding DNA show many features common to plastid genes in general, some patterns of variation for *ndhF* are highly unusual (Table 4). For example, while *atpB* and *rbcL* show a typical pattern of 3rd position variation ca. three to four times greater than 1st and 2nd position variation, *ndhF* has 3rd position variation less than twice that of 1st and 2nd position variation. Furthermore, most of these substitutions are found on the more variable 3' half of *ndhF*. This pattern has been noted in other studies of *ndhF* (e.g., Kim and Jansen, 1995).

4.2. Phylogenetic resolution

Two of the chief goals of this study were to test monophyly of Moore's (1973) Arecoide Line and to resolve broad internal Arecoide Line relationships. Evidence of monophyly for the Arecoide Line was uncovered and additional internal resolution was

achieved. Many of the internal relationships have been identified in previous plastid DNA studies but some differences are noted.

A grouping of Phytelphantoideae + Ceroyxyleae + Cyclospatheae was resolved by Asmussen and Chase (2001) as the first branching element of the Arecoid Line (67% BS) with Iriarteae branching next (<50% BS). In the current study, Iriarteae are resolved as the first branching element (65% BS; Fig. 2) followed by Phytelphantoideae + Ceroyxyleae (<50% BS; Fig. 2) with the single genus of Cyclospatheae (*Pseudophoenix*) placed as sister to *Podococcus* (92% BS; Fig. 2). Differences in taxon sampling and character sample size might explain some of the differences between the results of Asmussen and Chase (2001) and those of the current study. The former study sampled four Phytelphantoideae, three Ceroyxyleae, and two Cyclospatheae versus only two, one, and one species, respectively, in the current analyses. Nonetheless, the same two species of Iriarteae were sampled and the bootstrap support for the alternate resolutions is weak. Long-branch attraction (Felsenstein, 1978; Hendy and Penny, 1989) is also a possibility in this group of taxa as discussed for these and other palms (Barfod et al., 1999; Hahn, 2002).

Resolved relationships for the remainder of the Arecoid Line include monophyly of tribes Cocoeae, Geonomeae, and Hyophorbeae and a strongly supported clade consisting of all IndoPacific pseudomononamous genera. A well-supported sister group relationship between *Reinhardtia* (tribe Malortieinae) and Cocoeae is also identified.

Geonomeae have been the focus of study by Asmussen (1999) and subsequent analyses of the tribe have indicated monophyly with moderate support (e.g., 76% BS in Asmussen and Chase, 2001; 81% BS in Hahn, 2002). In the current study, strong bootstrap support for monophyly is seen (97% BS; Fig. 2) and several synapomorphic indels that further emphasize monophyly of Geonomeae are identified. Asmussen and Chase (2001) recovered a weakly supported (<50% BS) sister group relationship between Geonomeae and Euterpeinae but the current study places this latter group as sister to the IndoPacific clade, albeit with weak support (57% BS; Fig. 2). The current study agrees with Asmussen and Chase (2001) in grouping *Leopoldinia* and *Manicaria* with Geonomeae, Euterpeinae, and the IndoPacific clade but with greater support (77% BS vs <50% BS).

Hyophorbeae are a distinctive group of Ceroyxylid palms characterized by long branches in all molecular analyses to date. The current study repeats this pattern and identifies several indels restricted to the tribe. Along with other basal-branching Arecoid Line elements, the position of Hyophorbeae among other Arecoids is not entirely clear, although a relationship with *Roystonea* has been recovered in previous studies (Asmussen and Chase, 2001; Hahn, 2002). This relationship is not re-

covered in the current study but both taxa show long branches and the possibility of long-branch attraction must be considered as with other Ceroyxloideae and Phytelphantoideae as mentioned above.

The grouping of all IndoPacific pseudomononamous Arecoids has been recovered in most other molecular phylogenies of palms (Asmussen et al., 2000; Asmussen and Chase, 2001; Hahn, 2002) and poses several interesting evolutionary and biogeographic questions that are discussed below. Relationships within the IndoPacific clade are less well resolved and monophyly for some of the larger subtribes (e.g., Archontophoenicinae, Arecinae, and Iguanurinae) is not supported. Additional character and taxon sampling is needed to fully resolve relationships within this complex group.

Cocoeae (*sensu* Dransfield and Uhl, 1998) are one of the more distinctive lineages of Arecoid palms and are characterized by the absence of a crown shaft, the usually woody peduncular bract that exceeds the prophyll, staminodes connate in a short ring, a gynoeceum with three or more locules and ovules, and an endocarp with three or more pores. Of these characters, the endocarp pores are considered the best evidence of monophyly for the group. The tribe was recovered as monophyletic in the current study with only weak support (62% BS; Fig. 2). A similar result was obtained in previous molecular phylogenetic studies (Asmussen and Chase, 2001; Hahn, 2002) but the large number of morphological characters that identify the group would argue for recognition of monophyly for the Cocoeae. A novel result of this study is identification of *Reinhardtia* as strongly supported sister group to the Cocoeae (80% BS; Fig. 2).

As seen in previous studies (Asmussen and Chase, 2001; Hahn, 2002), the five recognized subtribes of Cocoeae were resolved in the current study as two major clades. The first, a grouping of Bactridinae + Elaeidinae, is not well supported in the molecular analyses (52% BS; Fig. 2) but is strongly corroborated by several morphological characters including fibrous peduncular bracts, pistillate flowers sunken in pits, the pistillate sepals connate, endocarp pores at or above the equator, and lateral ovules (Uhl and Dransfield, 1987). Subtribe Elaeidinae consists of two genera, *Barcella* and *Elaeis*, and is always recovered as monophyletic (97% BS; Fig. 2).

Subtribe Bactridinae is not always recovered as monophyletic in molecular analyses but is defined by the presence of spines on at least some parts of the plant body and endocarps with deeply impressed pores at or above the middle and usually covered with adherent fibers. Relationships among the six genera of Bactridinae have been inferred based on the presence/absence of staminate floral pits on the rachillae and the nature of perianth fusion (Hahn, 1991). Moore (1967) considered *Acrocomia* the least specialized, *Aiphanes* and *Gastrococos* (with partly fused perianth parts) intermediate,

and *Astrocaryum*, *Bactris*, and *Desmoncus* the most derived with connate pistillate petals. Uhl and Dransfield (1987) identified *Acrocomia* and *Gastrococos* as the least specialized, *Aiphanes* and *Bactris* intermediate, and *Desmoncus* and *Astrocaryum* as the most highly derived. These interpretations have been challenged, however, as some of the character states involved are inconsistent within some of the genera (e.g., *Aiphanes*; Borchsenius and Bernal, 1996).

In the current study, *Acrocomia* + *Gastrococos* form a clade (100% BS; Fig. 2) sister to the remainder of the subtribe (52% BS; Fig. 2). Several indels support these two genera as a clade and distinct from the other genera of Bactridinae. *Astrocaryum* is the next-branching element (91% BS) followed by *Aiphanes* (64% BS) then a grouping of *Bactris* + *Desmoncus* (59% BS; all Fig. 2).

The second major clade of Cocoeae recovered includes subtribes Attaleinae, Beccariophoenicinae, and Butiinae and is diagnosed by endocarp pores below the equator, the lack of spines, and pistillate flowers borne superficially. This group is moderately well supported by in the MP trees (61% BS; Fig. 2) but *Beccariophoenix* is unresolved in the ML tree. Attaleinae + Butiinae are strongly supported as monophyletic (87% BS; Fig. 2) but Butiinae is paraphyletic with Attaleinae embedded within. Relationships among members of Butiinae are not strongly supported (59–74% BS in the MP trees; <50% in the ML trees). Additional character and taxon sampling is needed to further resolve relationships within this clade.

4.3. Biogeography

An analysis of the biogeographic patterns present among Arecoideae taxa suggests that the group is Gondwanan in origin with several independent dispersals into the IndoPacific region (Asia and the South Pacific plus Indian Ocean islands including Madagascar). Several of the first-branching lineages, including Iriarteae (Neotropics), Hyophorbeae (Neotropics and Mascarenes), Phytelephantoideae (Neotropics), and Ceroyxyleae (Neotropics, Madagascar, Juan Fernandez Islands, and Australia), are distributed across one or more fragments of the former supercontinent. Additionally, some of the remaining taxa such as Cocoeae are present on one or more fragments, indicating that primary diversification in these groups might have coincided with continental breakup. If diversification of these lineages was coupled with Gondwanan breakup, phylogenetic branching patterns should reflect the timing and sequence of the tectonic events.

Of the remaining diversity, only three clades have representatives outside of Gondwanaland and each of these has at least one member in Madagascar. Subtribe Oraniinae, with the single genus *Orania*, is found in Madagascar, Southeast Asia, and New Guinea. Only a

single species (the Malagasy *Orania trispatha*) was included in the present analysis but if a Gondwanan origin and dispersal to the IndoPacific region is correct, this species should be sister to the remainder of the genus.

Tribe Cocoeae is distributed in Madagascar (*Beccariophoenix* and *Voaniaola*), Africa (one species of *Eleais* and *Jubaeopsis*), the IndoPacific (*Cocos*), and the Neotropics (the remaining 15 genera). The recovered phylogeny for Cocoeae shows a close match to proposed patterns of continental breakup with Malagasy elements (*Beccariophoenix* and *Voaniaola*) branching first followed by diversification of African and Neotropical taxa. The African *Jubaeopsis* was not sampled in this study but would presumably branch from the remaining Butiinae just after *Voaniaola* if the Gondwanan breakup biogeographic model were correct. The lone non-Gondwanan element *Cocos* (the coconut) is a well-known long-distance dispersal specialist and, given the phylogenetic resolution uncovered in this study, is most readily explained by westward dispersal from South America rather than from Madagascar as postulated for *Orania*. Fossil endocarps ascribed to subtribe Attaleinae have been found in the South Pacific (Dransfield et al., 1984) and suggest a complicated story of repeated dispersals of Butiinae and/or Attaleinae into the region followed by extinction.

The 67 genera of palms ascribed to subtribes Archontophoenicinae, Arecinae, Cyrtostachydinae, Dyspodiinae, Iguanurinae, Lemurophoenicinae, Linospadicinae, Masoalinae, Oncospermatinae, and Ptychospermatinae are completely restricted to the IndoPacific region and in the current and most previous analyses (e.g., Asmussen et al., 2000; Asmussen and Chase, 2001; Hahn, 2002) are resolved as a single, relatively derived clade with moderate to strong bootstrap support for monophyly. Previous studies (e.g., Asmussen and Chase, 2001; Hahn, 2002) have resolved this clade with most of the Malagasy taxa (all but *Masoala*) sister to the remainder, suggesting that dispersal was from Madagascar into Asia and the south Pacific. Despite the high levels of morphological diversity among the genera in this clade, molecular diversity is very low (J.-C. Pintaud and W. Hahn, in prep.). If a molecular clock is in operation, morphological diversification must have been comparatively recent and very rapid in this clade, emphasizing the relative plasticity of palm morphology at several scales (e.g., Tomlinson, 1990). Much additional study is required to fully understand the patterns of molecular and morphological diversification in palms.

Acknowledgments

I thank the Fairchild Tropical Garden and the Montgomery Botanical Center for access to collections. Bill Baker and two anonymous reviewers offered many

comments that improved the manuscript. Conny Asmussen, Brian Morton, Larry Noblick, Jean-Christophe Pintaud, and Gerard Second provided many helpful comments on palm taxonomy and molecular evolution. This work was supported by NSF Grant NEB 9303266.

References

- Asmussen, C.B., 1999. Toward a chloroplast DNA phylogeny of the Tribe Geonomeae. *Mem. N.Y. Bot. Gard.* 83, 121–130.
- Asmussen, C.B., Chase, M.W., 2001. Coding and noncoding plastid DNA in palm systematics. *Am. J. Bot.* 88, 1103–1117.
- Asmussen, C.B., Baker, W.J., Dransfield, J., 2000. Phylogeny of the palm family (Arecaceae) based on *rps16* intron and *trnL-trnF* plastid DNA sequences. In: Wilson, K.L., Morrison, D.A. (Eds.), *Monocots: Systematics and Evolution*. CSIRO, Melbourne, Australia, pp. 525–535.
- Baker, W.J., Asmussen, C.B., Barrow, S.C., Dransfield, J., Hedderson, T.A., 1999. A phylogenetic study of the palm family (Palmae) based on DNA sequences from the *trnL-trnF* region. *Pl. Syst. Evol.* 219, 111–126.
- Baker, W.J., Hedderson, T.A., Dransfield, J., 2000. Molecular phylogenetics of subfamily Calamoideae (Palmae) based on nrDNA ITS and cpDNA *rps16* intron sequence data. *Mol. Phylogenet. Evol.* 14, 195–217.
- Barfod, A.S., Ervik, F., Bernal, R., 1999. Recent evidence on the evolution of Phytelphantooid palms (Palmae). *Mem. N.Y. Bot. Gard.* 83, 265–277.
- Borchsenius, F., Bernal, R., 1996. Aiphanes (Palmae). *Fl. Neotrop. Monogr.* 70, 1–95.
- 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.
- Dransfield, J., Flenley, J.R., King, S.M., Harkness, D.D., Rapu, S., 1984. A recently extinct palm from Easter Island. *Nature* 312, 750–752.
- Dransfield, J., Uhl, N.W., 1986. An outline of a classification of palms. *Principes* 30, 3–11.
- Dransfield, J., Uhl, N.W., 1998. Palmae. In: Kubitzki, K. (Ed.), *Families and Genera of Vascular Plants: vol. IV. Flowering Plants. Monocotyledons*. Springer, Berlin, pp. 306–389.
- Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1994. Testing significance of incongruence. *Cladistics* 10, 315–319.
- Felsenstein, J., 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* 27, 401–410.
- Felsenstein, J., 1985. Confidence limits on phylogeny: an approach using the bootstrap. *Evolution* 39, 783–791.
- Frati, F., Simon, C., Sullivan, J., Swofford, D.L., 1997. Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. *J. Mol. Evol.* 44, 145–158.
- Goldman, N.J., 1993. Statistical tests of models of DNA substitution. *J. Mol. Evol.* 36, 182–198.
- Golenberg, E.M., Clegg, M.T., Durbin, M.L., Doebley, J., Ma, D.P., 1993. Evolution of a noncoding region of the chloroplast genome. *Mol. Phylogenet. Evol.* 5, 460–466.
- Goldman, N.J., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* 49, 652–670.
- Gu, X., Fu, Y.-X., Li, W.-H., 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol. Biol. Evol.* 12, 546–557.
- Hahn, W.J., 1991. Notes on the genus *Acanthococos* Barbosa Rodrigues, (Palmae). *Principes* 35, 167–171.
- Hahn, W.J., 1999. Molecular systematic studies of the Palmae. *Mem. N.Y. Bot. Gard.* 83, 47–60.
- Hahn, W.J., Sytsma, K.J., 2000. Molecular systematics and biogeography of the Southeast Asian genus *Caryota* (Palmae). *Syst. Bot.* 24, 558–580.
- Hahn, W.J., 2002. A molecular phylogenetic study of the Palmae (Arecaceae) based on *atpβ*, *rbcL*, and 18S nrDNA sequences. *Syst. Biol.* 51, 92–112.
- Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* 38, 297–309.
- 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.
- Kelchner, S.A., 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Ann. Missouri Bot. Gard.* 87, 482–498.
- Kelchner, S.A., Clark, L.G., 1997. Molecular evolution and phylogenetic utility of the chloroplast. *Mol. Phylogenet. Evol.* 8, 385–397.
- Kim, K.-J., Jansen, R.K., 1995. *ndhF* sequence evolution and the major clades in the sunflower family. *Proc. Natl. Acad. Sci. USA* 92, 10379–10383.
- Kishino, H., Miyata, T., Hasegawa, M., 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.* 30, 151–160.
- Lanave, C., Preparata, G., Saccone, C., Serio, G., 1984. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* 20, 86–93.
- Levison, G., Gutman, G.A., 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4, 203–221.
- Lewis, C., Doyle, J.J., 2001. Phylogenetic utility of the nuclear gene Malate Synthase in the Palm family (Arecaceae). *Mol. Phylogenet. Evol.* 19, 409–420.
- Mallatt, J., Sullivan, J., 1998. 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. *Mol. Biol. Evol.* 15, 1706–1718.
- Moore Jr., H.E., 1960. A new subfamily of palms—the Caryotoideae. *Principes* 4, 102–117.
- Moore Jr., H.E., 1967. The genus *Gastrococos* (Palmae—Cocoideae). *Principes* 11, 114–121.
- Moore Jr., H.E., 1973. The major groups of palms and their distribution. *Gentes Herb.* 11, 27–141.
- Morton, B.R., 1995. Neighboring base composition and transversion/transition bias in a comparison of rice and maize chloroplast non-coding regions. *Proc. Natl. Acad. Sci. USA* 92, 9717–9721.
- Morton, B.R., Oberholzer, V.M., Clegg, M.T., 1997. The influence of specific neighboring bases on substitution bias in noncoding regions of the plant chloroplast genome. *J. Mol. Evol.* 45, 227–231.
- Olmstead, R.G., Reeves, P.A., 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbcL* and *ndhF* sequences. *Ann. Missouri Bot. Gard.* 82, 176–193.
- 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.
- Sanderson, M.J., 1995. Objections to bootstrapping phylogenies: a critique. *Syst. Biol.* 44, 299–320.
- Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49, 369–381.
- Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114–1116.
- Sullivan, J., 1996. Combining data with different distributions of among-site rate variation. *Syst. Biol.* 45, 375–380.
- Sullivan, J., Markert, J.A., Kirkpatrick, C.W., 1997. Phylogeography and molecular systematics of the *Peromyscus aztecus* species group (Rodentia: Muridae) inferred using parsimony and likelihood. *Syst. Biol.* 46, 426–440.

- Swofford, D., 2001. PAUP*. Phylogenetic analysis using parsimony (* and other methods), version 4b8. Sinauer, Sunderland, MA.
- Swofford, D., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), *Molecular Systematics*, second ed. Sinauer, Sunderland, MA, pp. 407–514.
- Sytsma, K.J., Givnish, T.J., Smith, J.F., Hahn, W.J., 1993. Obtaining and storing land plant samples for macromolecular comparisons. *Methods Enzymol.* 224, 23–37.
- Taylor, B., Powell, A., 1982. Isolation of plant DNA and RNA. *Focus* 4, 4–6.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Tomlinson, P.B., 1990. *The Structural Biology of Palms*. Clarendon, Oxford.
- Uhl, N.W., Dransfield, J., 1987. *Genera Palmarum: A Classification of Palms Based on the Work of Harold E. Moore, Jr.* Alien Press, Lawrence, KS.
- Uhl, N.W., Dransfield, J., 1999. *Genera Palmarum after ten years*. *Mem. N.Y. Bot. Gard.* 83, 245–253.
- Uhl, N.W., Dransfield, J., Davis, J.I., Luckow, M.A., Hansen, K.S., Doyle, J.J., 1995. Phylogenetic relationships among palms: cladistic analyses of morphological and chloroplast DNA restriction site variation. In: Rudall, P., Cribb, P.J., Cutler, D.F., Humphries, C.J. (Eds.), *Monocotyledons: Systematics and Evolution*. Royal Botanic Gardens, Kew, England, pp. 623–661.
- Whelan, S., Goldman, N., 1999. Distributions of statistics used for the comparison of models of sequence evolution in phylogenetics. *Mol. Biol. Evol.* 16, 1292–1299.
- Wilson, M.A., Gaut, B., Clegg, M.T., 1990. Chloroplast DNA evolves slowly in the palm family (Arecaceae). *Mol. Biol. Evol.* 7, 303–314.
- Yang, Z., 1994. Estimating the pattern of nucleotide substitution. *J. Mol. Evol.* 39, 105–111.
- Yang, Z., 1996. Maximum-likelihood models for combined analyses of multiple sequence data. *J. Mol. Evol.* 42, 587–596.
- Yang, Z., Goldman, N., Friday, A., 1994. Comparison of models for nucleotide substitution used in maximum likelihood phylogenetic estimation. *Mol. Biol. Evol.* 11, 316–324.
- Yoder, A.D., Irwin, J.A., Payseur, B.A., 2001. Failure of the ILD to determine data combinability for Slow Loris phylogeny. *Syst. Biol.* 50, 408–424.