SYSTEMATICS OF THE PALM SUBFAMILY ARECOIDEAE (ARECACEAE) BASED ON CHLOROPLAST AND NUCLEAR SEQUENCE DATA

by

JASON R. COMER

(Under the Direction of Wendy B. Zomlefer and James H. Lebeens-Mack)

ABSTRACT

The palms, Arecaceae/Palmae (183 genera; ca. 2,600 species), are distributed throughout the tropics and subtropics. The family comprises key species in tropical ecosystems and some (e.g. coconut, date palm, oil palm) have significant economic importance. As currently circumscribed, Arecaceae comprise five subfamilies. The largest subfamily, Arecoideae, has 14 tribes, including coconut and oil palm. The tribal relationships of Arecoideae are not well understood. Slow rates of molecular evolution have made phylogenetic inferences difficult, prompting researchers to propose the use of larger data sets. For this study, a phylogenomics approach was used to generate the largest sequence data set to date for Arecoideae with representatives of all 14 tribes. Plastid sequence data was generated using whole genome shotgun sequencing and two targeted sequencing approaches, long range PCR and hybrid gene capture. Both long range PCR and hybrid gene capture were successful in enriching for the plastid genome and provided similar sequencing coverage. Hybrid gene capture was also used to enrich for 176 nuclear genes. One hundred and fourteen plastid and 168 nuclear genes were used for phylogenetic analyses. The resulting phylogenies were largely congruent with each

other and with previous studies. Tribes Chamaedoreeae and Iriarteeae represented the earliest diverging lineages within subfamily Arecoideae. The POS clade (tribes Podococceae, Oranieae, and Sclerospermeae) was recovered as sister to the RRC clade (tribes Roystoneeae, Reinhardtieae, and Cocoseae) and the core arecoids clade (tribes Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae). Within the core arecoids, Areceae and Euterpeae (AE clade) were consistently and strongly supported as monophyletic. Analyses inferred two radiation events in the evolutionary history of Arecoideae. The first occurred after subfamilies Arecoideae and Ceroxyloideae diverged 86 to 80 million years ago. The second radiation event was within the core arecoids during the early Eocene (56 to 50 million years ago). Ancestral area analyses supported North America as the range of origin for Arecoideae with subsequent dispersals into South America, Africa, and the Indopacific. The current distribution of tribe Oranieae may be explained by the “out of India” hypothesis, i.e., early ancestors of the tribe rafted from Africa/Madagascar to the Indopacific via India.

INDEX WORDS: ancestral biogeography, Arecaceae, Arecoideae, phylogenomics, systematics

# SYSTEMATICS OF THE PALM SUBFAMILY ARECOIDEAE (ARECACEAE) BASED ON CHLOROPLAST AND NUCLEAR SEQUENCE DATA

by

JASON R. COMER

BS, College of the Ozarks, 2007

MS, Missouri State University, 2009

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2015

# SYSTEMATICS OF THE PALM SUBFAMILY ARECOIDEAE (ARECACEAE) BASED ON CHLOROPLAST AND NUCLEAR SEQUENCE DATA

by

JASON R. COMER

| | |
|---|---|
| Major Professors: | Wendy B. Zomlefer<br>James H. Lebeens-Mack |
| Committee: | Shu-Mei Chang<br>Jerrold I. Davis<br>David E. Giannasi |

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2015

# DEDICATION

This work is dedicated to my family, whose understanding and support has provided the foundation for completing my dissertation.

# ACKNOWLEDGEMENTS

I first would like to thank my co-advisers Wendy B. Zomlefer and James H. Leeben-Mack. Without their guidance and support this work would not have been completed. I also thank my advisory committee Shu-Mei Chang, Jarrold I. Davis, David Giannasi, and Dorset Trapnell. I must also thank collaborators Craig F. Barrett, Karolina Heyduk, and Dennis Wm. Stevenson for providing critiques and technical support. William Baker provided invaluable insight during the initial stages of my research and for being a reviewer for the second chapter. The Leebens-Mack lab and University of Georgia Herbarium staff, past and present, provided feedback and ideas for a number of aspects with my dissertation. I would also like to thank Larry Noblick and Patrick Griffith (Montgomery Botanical Center) and Brett Jestrow (Fairchild Tropical Botanic Garden) for assistance collecting many of the palms used in this study. Anders Lindstrom and Kampon Tansacha were gracious enough to host and allow me to sample the palm collection at Nong Nooch Tropical Botanical Garden (Pattaya, Thailand). Several species of *Reinhardtia* were kindly provided by Jeff Marcus (Floribunda Palms and Exotics), and Thomas Couvreur supplied tissue for *Podococcus* and *Sclerosperma*. Funding was provided by the National Science Foundation (DEB-083009; J. H. Leebens-Mack, PI and W. B. Zomlefer, co-PI). Additional travel funds were provided to the first author by the Department of Plant Biology, University of Georgia (Palfrey Grant for Graduate Student Research).

TABLE OF CONTENTS

Page

# CHAPTER 1

# INTRODUCTION AND LITERATURE REVIEW

## Family Arecaceae

Arecaceae (Palmae) comprise 183 genera and ca. 2,600 species mainly distributed in the tropics and subtropics (Dransfield et al., 2008; Baker et al., 2009; Trias-Blasi et al., 2015). Palms are "woody" perennials (from primary growth, sclerenchyma), with habits that include creeping, climbing, or erect and solitary to colonial stems that are usually unbranched (Moore, 1973; Dransfield et al,, 2008). Palm inflorescences are usually solitary in the leaf axils and subtended by one to many bracts. Flowers are usually imperfect, typically sessile, and solitary or in various specialized arrangements such as triads (central pistillate and two lateral staminate flowers) and acervuli (two ranked lines of flowers) (Moore, 1973; Dransfield, 2008). The fruit type is a berry or a drupe with typically one seed but with as many as 10 (Moore, 1973).

The most recognizable synapomorphies for the family are the perennial habit with "wood" from primary growth; the plicate (folded) leaves in bud; and the inflorescence always subtended by a bract called a prophyll (Uhl et al., 1995; Baker et al., 2009). Other synapomorphies include uniovulate carpels (single ovule per carpel), usually indehiscent baccate fruit, and stegmata (cells that contain silica bodies) next to vascular and non-vascular fibers (Uhl et al., 1995; Dransfield et al., 2008; Baker et al., 2009).

**Challenges of palm systematics**

Palms are well known for their slow rates of molecular evolution. Wilson et al. (1990) found a five- to thirteen-fold decrease in substitution rates in palms relative to annual species using restriction site (RFLP) and *rbcL* sequence data, with the average of 0.009 substitutions per base for palms. Compared to the grasses, for example, the substitution rate for the nuclear *Adh* gene was 2.5-fold slower in the palms (Gaut et al., 1996). Clegg et al. (1994) found palms with the lowest substitution rates among the Orchidales, Liliales, and Bromeliales. This led authors (Uhl et al., 1995; Baker et al., 1999; Asmussen et al., 2000; Asmussen et al., 2006; Loo et al., 2006; Norup et al., 2006; Dransfield et al., 2008; Baker et al., 2009) to suggest that many markers were needed to identify enough informative characters for analyses. Whole chloroplast genome sequencing generates large amounts of sequence data and has been shown to have utility in resolving more difficult taxa (Moore et al., 2007; Guisinger et al., 2010).

**Systematics of subfamily Arecoideae**

Moore (1973) established a framework for the subfamilial taxa of the Arecaceae with a suite of morphological characters. He did not assign formal ranks to these groupings or to his five proposed evolutionary lines (Table 1.1). The largest group, the Arecoid line, includes the Arecoid, Chamaedoreoid, Ceroxyloid, Cocosoid, Geonomoid, Iriarteoid, Phytelephantoid, Podococcoid, and Pseudophoenicoid groups.

Moore (1973) defined the Arecoid line as having reduplicately paripinnate and usually deciduous leaves (by abscission zones), inflorescences with a

prophyll plus one to several smaller bracts (bracteoles); and branches of the inflorescence subtended by small bracts with generally imperfect flowers. The four main flower arrangements of the Arecoid line are solitary flowers, flowers in adnate cincinni (a cluster of flowers with one flower arising from a bracteole of the previous flower; Baker et al. 2009), carpellate flowers in heads and staminate flowers in spikes or heads, and flowers in triads (two staminate to one carpellate, structurally a cincinnus; Baker et al. 2009).

Uhl and Dransfield (1987) used morphological characters to split the Arecoid line into three subfamilies: Arecoideae, Ceroxyloideae, and Phytelephantoideae. The Arecoideae comprised six tribes (Areceae, Caryoteae, Cocoeae, Geonomeae, Iriarteeae, and Podococceae) characterized by flowers in triads or clusters derived from triads. This circumscription of the Arecoideae removed the Pseudophoenicoid, Ceroxyloid, Chamaedoreoid and Phytelephantoid lines from Moore's (1973) Arecoid line but includes the Caryotoid group.

Uhl et al. (1995) used chloroplast restriction fragment length polymorphisms (RFLP) and morphological data to examine representatives from all palm tribes (67 taxa), including 10 from the subfamily Arecoideae. The Arecoideae plus the tribe Hyophorbeae (subfamily Ceroxyloideae) formed one clade in the strict consensus tree of the restriction site data and in strict consensus tree of the combined morphological and restriction site data. Tribes Caryoteae and Iriarteeae were a sister group to the rest of subfamily Arecoideae, and tribe Areceae was polyphyletic. *Orania* (Areceae) had a closer affinity to Podococceae than to the other representatives of Areceae.

Several subsequent studies utilized two to five molecular markers, including chloroplast (e.g., *rbcL*, *rps16*, *trnL-trnF*) and low copy nuclear genes (e.g., *18S*, *PKR*, *RPB2*). Several relationships were congruent: (1) the tribe Caryoteae did not group with the rest of the subfamily Arecoideae; (2) a *Podococcus*/*Orania* or a *Podococcus*/*Orania*/*Sclerosperma* clade was strongly supported; and (3) a Indo-Pacific clade within tribe Areceae was well supported (Baker et al., 1999; Asmussen et al., 2000; Asmussen and Chase, 2001; Lewis and Doyle, 2002; Hahn, 2002a, 2002b).

A new classification of the family by Dransfield et al. (2005) was the foundation for the revision of *Genera Palmarum* (Dransfield et al., 2008). This volume was based on Asmussen et al. (2006) who used all previously published molecular data and added the *matK* gene data for 178 species. Subfamily Arecoideae was circumscribed with following tribes (Fig. 1.1): Areceae, Chamaedoreeae (formerly Hyophorbeae, Ceroxyloideae), Cocoseae, Euterpeae, Geonomateae, Iriarteeae, Leopoldinieae, Manicarieae, Oranieae, Pelagodoxeae, Podococceae, Reinhardtieae, Roystoneae, and Sclerospermeae. Caryoteae was removed from subfamily Arecoideae and placed in subfamily Coryphoideae. There were also several unplaced genera within the Arecoideae (e.g. *Dictyosperma*, *Heterospathe*, *Hydriastele*). Ceroxyloideae were sister to Arecoideae (bs 85) and the Arecoideae were supported as monophyletic (bs 70). Tribal relationships were not well resolved.

A study focused on tribe Areceae using nuclear DNA (Norup et al., 2006) found strong support for the monophyly of the Indo-Pacific tribe, Areceae (bs

100). The bootstrap support for Arecoideae, however, was less than 50, and there are many unresolved nodes. Loo et al. (2006) focused on subtribe Arecinae within the Areceae using nuclear DNA and found strong support for a monophyletic Arecoideae in one data set (*PRK* gene, bs 100, posterior probability 0.95). The combined data set (*PKR* and *RPB2*) strongly supported the Areceae Indo-Pacific clade (bs 100, posterior probability 1).

A comprehensive genus-level analysis (including all 192 genera except *Tahina*) by Baker et al. (2009) incorporated all published molecular data and a new morphological data set to construct trees based on super matrix and super tree approaches (Fig. 1.2). The Arecoideae were supported as monophyletic (bs 93) with Iriarteeae as sister to the rest of the Arecoideae. An *Orania/Podococcus/Sclerosperma* (POS) clade had strong support (bs 98) and was sister to a core Arecoid group including tribes Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae. Relationships within this core group were not well supported except for tribe Areceae (bs 84) but the nodes within the tribe are not well resolved.

Baker et al. (2011) used two nuclear makers (*PRK* and *RPB2*) to evaluate relationships within subfamily Arecoideae. In the combined analyses, the subfamily and tribes were resolved as monophyletic with strong. One exception was tribe Reinhardtieae that was embedded within Cocoseae. Three major clades (POS clade, the RRC (Roystoneeae/Reinhardtieae/Cocoseae) clade, and the core arecoids clade were well supported.

**Purpose of study**

For this study a phylogenomics approach was used to resolve relationships within the palm subfamily Arecoideae and to explore the historical biogeographical implications of the phylogenies. Next-generation sequencing was used to generate sequence data for chloroplast and nuclear markers. Two targeted sequencing approaches, long range PCR (designed for this study) and hybrid gene capture used to generate chloroplast sequence data, were compared, and hybrid gene capture was used to sequence a set of 176 nuclear genes.

**Literature cited**

ASMUSSEN, C. B., W. J. BAKER, AND J. DRANSFIELD. 2000. Phylogeny of the palm family (Arecaceae) based on *rps16* intron and *trnL-trnF* plastid DNA sequences. *In* K. L. Wilson AND D. A. Morrison [eds.], Monocots: Systematics and Evolution, 525–537. CSIRO Publishing, Collingwood VIC, Australia.

ASMUSSEN, C. B., AND M. W. CHASE. 2001. Coding and noncoding plastid DNA in palm systematics. *American Journal of Botany* 88: 1103–1117.

ASMUSSEN, C. B., J. DRANSFIELD, V. DEICKMANN, A. S. BARFOD, J.-C. PINTAUD, AND W. J. BAKER. 2006. A new subfamily classification of the palm family (Arecaceae): Evidence from plastid DNA phylogeny. *Botanical Journal of the Linnean Society* 151: 15–38.

BAKER, W. J., C. B. ASMUSSEN, S. C. BARROW, J. DRANSFIELD, AND T. A. HEDDERSON. 1999. A phylogenetic study of the palm family (Palmae)

based on chloroplast DNA sequences from the *trnL-trnF* region. *Plant Systematics and Evolution* 219: 111–126.

BAKER, W. J., M. V. NORUP, J. J. CLARKSON, T. L. P. COUVREUR, J. L. DOWE, C. E. LEWIS, J.-C. PINTAUD, et al. 2011. Phylogenetic relationships among arecoid palms (Arecaceae: Arecoideae). *Annals of Botany* 108: 1417–1432.

BAKER, W. J., V. SAVOLAINEN, C. B. ASMUSSEN-LANGE, M. W. CHASE, J. DRANSFIELD, F. FOREST, M. M. HARLEY, et al. 2009. Complete generic-level phylogenetic analyses of palms (Arecaceae) with comparisons of supertree and supermatrix approaches. *Systematic Biology* 58: 240–256.

CLEGG, M. T., B. S. GAUT, G. H. LEARN, AND B. R. MORTON. 1994. Rates and patterns of chloroplast DNA evolution *Proceedings of the National Academy of Sciences, USA* 91: 6795–6801.

DRANSFIELD, J., N. W. UHL, C. B. ASMUSSEN, W. J. BAKER, M. M. HARLEY, AND C. E. LEWIS. 2005. A new phylogenetic classification of the palm family, Arecaceae. *Kew Bulletin* 60: 559–569.

DRANSFIELD, J., N. W. UHL, C. B. ASMUSSEN, W. J. BAKER, M. M. HARLEY AND C. LEWIS 2008. Genera Palmarum. The evolution and classification of palms. Royal Botanical Gardens, Kew, Richmond, Surrey, England.

GAUT, B. S., B. R. MORTON, B. C. MCCAIG, AND M. T. CLEGG. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proceedings of the National Academy of Sciences* 93: 10274–10279.

GUISINGER, M., T. CHUMLEY, J. KUEHL, J. BOORE, AND R. JANSEN. 2010. Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. *Journal of Molecular Evolution* 70: 149–166.

HAHN, W. J. 2002a. A molecular phylogenetic study of the Palmae (Arecaceae) based on *atpB*, *rbcL*, and *18S* nrDNA sequences. *Systematic Biology* 51: 92–112.

_____. 2002b. A phylogenetic analysis of the Arecoid Line of palms based on plastid DNA sequence data. *Molecular Phylogenetics and Evolution* 23: 189–204.

HEYDUK, K., D. W. TRAPNELL, C. F. BARRETT, AND J. LEEBENS-MACK. 2015. Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biological Journal of the Linnean Society* 115: n/a–n/a. (doi: 10.1111/bij.12551)

LEWIS, C. E., AND J. J. DOYLE. 2002. A phylogenetic analysis of tribe Areceae (Arecaceae) using two low-copy nuclear genes. *Plant Systematics and Evolution* 236: 1–17.

LOO, A. H. B., J. DRANSFIELD, M. W. CHASE, AND W. J. BAKER. 2006. Low-copy nuclear DNA, phylogeny and the evolution of dichogamy in the betel nut palms and their relatives (Arecinae; Arecaceae). *Molecular Phylogenetics and Evolution* 39: 598–618.

MOORE, H. E. 1973. The major groups of palms and their distribution. *Gentes Herbarum* 11: 27–141.

MOORE, M. J., C. D. BELL, P. S. SOLTIS, AND D. E. SOLTIS. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences* 104: 19363–19368.

NORUP, M. V., J. DRANSFIELD, M. W. CHASE, A. S. BARFOD, E. S. FERNANDO, AND W. J. BAKER. 2006. Homoplasious character combinations and generic delimitation: a case study from the Indo-Pacific arecoid palms (Arecaceae: Areceae). *American Journal of Botany* 93: 1065–1080.

TRIAS-BLASI, A., W. J. BAKER, A. L. HAIGH, D. A. SIMPSON, O. WEBER, AND P. WILKIN. 2015. A genus-level phylogenetic linear sequence of monocots. *Taxon* 64: 552–581. (doi:10.12705/643.9)

UHL, N. W., AND J DRANSFIELD 1987. Genera palmarum: a classification of palms based on the work of Harold E.Moore, Jr. Allen Press, Lawrence, Kansas, USA.

UHL, N. W., J. DRANSFIELD, J. I. DAVIS, M. A. LUCKOW, K. S. HANSEN, AND J. J. DOYLE. 1995. Phylogenetic relationships among palms: cladistic analyses of morphological and chloroplast DNA restriction site variation. *In* P. J. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries [ed.], Monocotyledons: Systematics and Evolution, vol. 2, 623–662. Whitstable Litho Printers Ltd., Kent, England.

WILSON, M. A., B. GAUT, AND M. T. CLEGG. 1990. Chloroplast DNA evolves slowly in the palm family (Arecaceae). *Molecular Biology and Evolution* 7: 303–314

**Table 1.1.** The informal groupings of the Arecaceae, and the five proposed evolutionary lines, the distribution of genera, and number of species of palms. Modified version from Moore (1973). Highlighted sub-groupings are included in the Arecoideae.

| Major Evolutionary Line | Sub-groupings | Moore's designator | Western Hemisphere | | | | American Total | | Africa Arabia Europe | | Madagascar Mascarenes Seychelles | | Eastern Tropics | | Grand Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | North America | | South America | | | | | | | | | | | |
| | | | Genera | Spp. | Genera | Spp. | Genera | Spp. | Genera | Spp. | Genera | Spp. | Genera | Spp. | Genera | Spp. |
| Coryphiod | | | | | | | | | | | | | | | | |
| | Coryphoid | I | 13 | 91 | 7 | 15 | 16 | 105 | 3 | 3 | | | 14 | 214 | 32 | 322 |
| | Phoenicoid | II | | | | | | | 1 | 5 | 1 | 1 | 1 | 12 | 1 | 17 |
| | Borassoid | III | | | | | | | 3 | 42 | 5 | 8 | 3 | 6 | 6 | 56 |
| Lepidocaroid | | IV | 1 | 1 | 3 | 31 | 3 | 31 | 5 | 60 | 1 | 1 | 16 | 573 | 22 | 664 |
| Nypoid | | V | | | | | | | | | | | 1 | 1 | 1 | 1 |
| Caryotoid | | VI | | | | | | | | | | | 3 | 35 | 3 | 35 |
| Arecoid | | | | | | | | | | | | | | | | |
| | Pseudophoenicoid | VII | 1 | 4 | | | 1 | 4 | | | | | | | 1 | 4 |
| | Ceroxyloid | VIII | | | 2 | 18 | 2 | 18 | | | 2 | 12 | | | 4 | 30 |
| | Chamaedoreoid | IX | 4 | 103 | 3 | 39 | 5 | 141 | | | 1 | 5 | | | 6 | 146 |
| | Iriarteoid | X | 3 | 3 | 8 | 50 | 8 | 52 | | | | | | | 8 | 52 |
| | Podococcoid | XI | | | | | | | 1 | 2 | | | | | 1 | 2 |
| | Arecoid | XII | 8 | 24 | 9 | 91 | 10 | 109 | 1 | 3 | 19 | 105 | 58 | 543 | 88 | 760 |
| | Cocosoid | XIII | 11 | 78 | 24 | 507 | 26 | 580 | 2 | 2 | | | 1 | 1 | 28 | 583 |
| | Geonomoid | XIV | 6 | 29 | 4 | 77 | 6 | 92 | | | | | | | 6 | 92 |
| | Phytelephantoid | XV | 1 | 6 | 4 | 9 | 4 | 15 | | | | | | | 4 | 15 |
| Totals | | | 48 | 339 | 64 | 837 | 81 | 1147 | 16 | 117 | 29 | 132 | 97 | 1385 | 212 | 2779 |

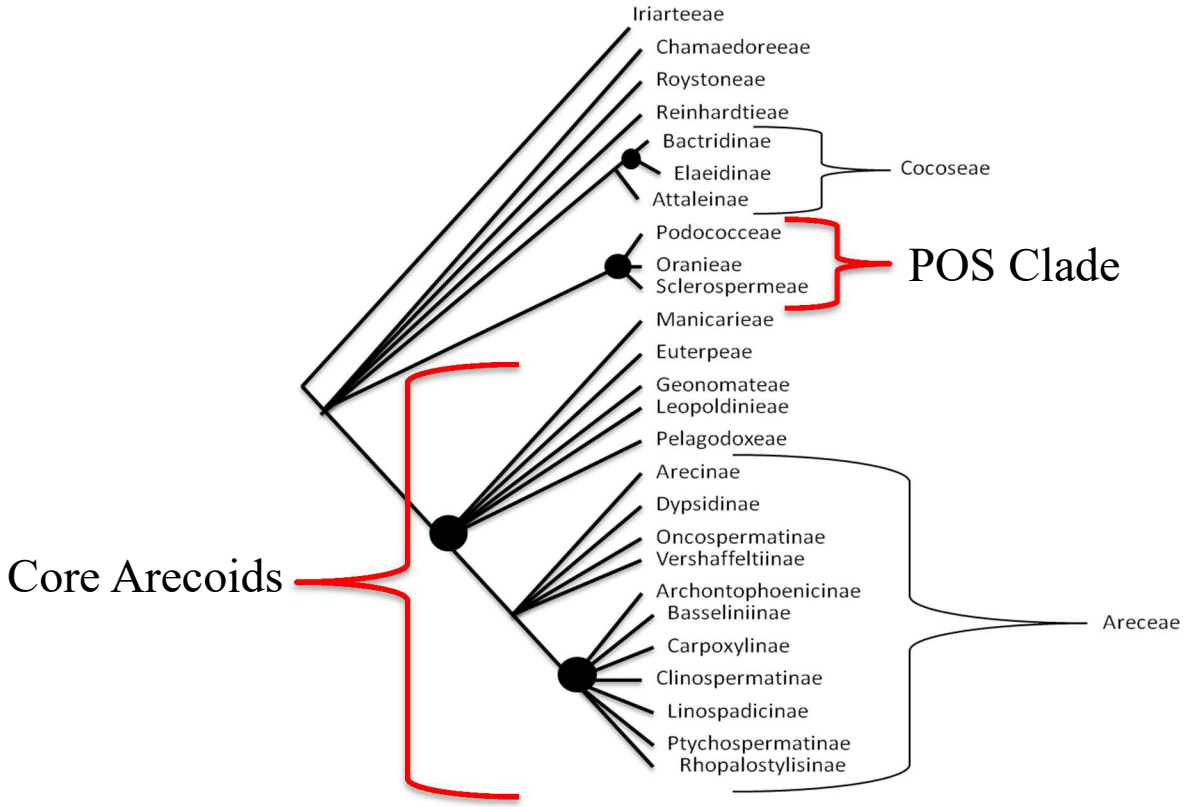



**Figure 1.1.** The phylogeny of the palm subfamily Arecoideae (Dransfield, 2008) shows the tribes, subtribes, and a few major groupings. The circled nodes are well supported clades. From Dransfield et al. (2008).

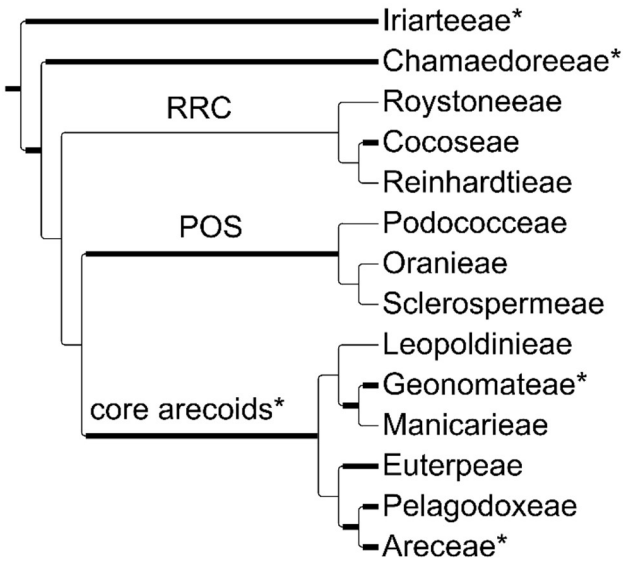



**Figure 1.2.** Tribal phylogeny of subfamily Arecoideae modified from the most congruent supertree (Fig. 3 in Baker et al. [2009]) and the summary tree ("Supertree," Fig. 5 in Baker et al. [2011]). All branches were supported by at least one input tree. Bold lines = branches supported by five or more input trees; * = clades supported by 10 or more input trees.

# CHAPTER 2

# RESOLVING RELATIONSHIPS WITHIN SUBFAMILY ARECOIDEAE (ARECACEAE) USING PLASTID SEQUENCES DERIVED FROM NEXT-GENERATION SEQUENCING[1]

[1] Comer JR, Zomlefer WB, Barrett CF, Davis JI, Stevenson DW, Heyduk K, Leebens-Mack J. 2015. *Am. J. Bot*. **102**, 888 – 899. (doi:10.3732/ajb.1500057) Reprinted here with permission of the publisher

**Abstract**

- *Premise of the study*: Several studies have incorporated molecular and morphological data to study the phylogeny of the palms (Arecaceae) but some relationships within the family remain ambiguous—particularly those within Arecoideae, the most diverse subfamily including coconut and oil palm. Here two next-generation targeted plastid enrichment methods were compared and used to elucidate Arecoideae phylogeny.
- *Methods*: Next-generation sequencing techniques were used to generate a plastid genome data set. Long range PCR and hybrid gene capture were used to enrich for chloroplast targets. Ten taxa were enriched using both methods for comparison. Chloroplast sequence data were generated for 31 representatives of the 14 Arecoideae tribes and five outgroup taxa. The phylogeny was reconstructed using maximum likelihood, maximum parsimony, and Bayesian analyses.
- *Key results*: Long range PCR and hybrid gene capture were both successful in enriching the plastid genome and provided similar sequencing coverage. Subfamily Arecoideae was resolved as monophyletic with tribe Chamaedoreeae as the earliest diverging lineage, implying that the development of flowers in triads defines a synapomorphy for the Arecoideae clade excluding Chamaedoreeae. Three major clades within this group were recovered: Roystoneeae/Reinhardtieae/Cocoseae (RRC), Areceae/Euterpeae/Geonomateae/Leopoldinieae/Manicarieae/Pelagodoxeae (core arecoids), and Podococceae/Oranieae/Sclerospermeae (POS). An Areceae + Euterpeae clade was resolved within the core arecoids. The POS clade was sister to a RRC + core arecoids clade, implying a shared ancestral area in South America for these three clades.

- *Conclusions*: The plastome phylogeny recovered here provides robust resolution of previously ambiguous studies and new insights into palm evolution.

**Introduction**

Arecaceae (Palmae; 183−188 genera, ca. 2600 species) are mainly distributed in the tropics and subtropics (Dransfield et al., 2008; Baker et al., 2009; Palmweb, 2015; Trias-Blasi et al., 2015). The most recognizable synapomorphies for the family are the perennial habit with "wood" derived from primary growth; plicate (folded) leaves in bud; and inflorescences subtended by the prophyll (Uhl et al., 1995; Baker et al., 2009). Other synapomorphies include uniovulate carpels, usually indehiscent baccate fruit, and stegmata (cells that contain silica bodies) adjacent to vascular and non-vascular fibers (Uhl et al., 1995; Dransfield et al., 2008; Baker et al., 2009). The Arecoideae, the largest palm subfamily (107 genera, ca. 1300 species; Table 2.1), are characterized by reduplicately pinnate leaves and flowers arranged as triads, acervuli, or their derivatives (Dransfield et al., 2008).

Moore's (1973) revision of Arecaceae established a framework for subsequent subfamilial classifications using a suite of morphological characters, but he did not assign formal ranks (Table 2.1). The largest grouping, the "Arecoid line," included the Arecoid, Chamaedoreoid, Ceroxyloid, Cocosoid, Geonomoid, Iriarteoid, Phytelephantoid, Podococcoid, and Pseudophoenicoid groups. Dransfield and Uhl (1986; Uhl and Dransfield, 1987) split the Arecoid line into three formal subfamilies: Arecoideae, Ceroxyloideae, and Phytelephantoideae (Table 2.1). The Arecoideae comprised six tribes (Areceae, Caryoteae, Cocoeae, Geonomeae, Iriarteeae, and Podococceae) characterized by the flowers in triads or clusters derived from triads. This circumscription included

Moore's (1973) Caryotoid line but removed the Pseudophoenicoid, Ceroxyloid, Chamaedoreoid and Phytelephantoid groups from his Arecoid line.

The first molecular study of the palms (Uhl et al., 1995) used chloroplast restriction fragment length polymorphisms (RFLP) and morphological data to examine representatives (67 taxa) from all tribes [*sensu* Dransfield and Uhl (1986)], including 10 from subfamily Arecoideae. Several subsequent studies (Baker et al., 1999; Asmussen et al., 2000; Asmussen and Chase, 2001; Hahn, 2002a, 2002b; Lewis and Doyle, 2002) utilized two to five molecular markers, including chloroplast regions (e.g. *rbcL*, *rps16* intron) and nuclear genes (e.g. 18SrDNA, *PRK*). Several relationships were congruent among these studies: (1) the tribe Caryoteae did not group with the rest of subfamily Arecoideae; (2) the *Podococcus*/*Orania* or the *Podococcus*/*Orania*/*Sclerosperma* clade had strong support; and (3) an Indo-Pacific clade within tribe Areceae was well supported.

A new classification of the Arecaceae (see Table 2.1) by Dransfield et al. (2005) was the foundation for the revision of *Genera Palmarum* (Dransfield et al., 2008). This classification was supported by the phylogeny of Asmussen et al. (2006) based on a comprehensive analysis of *matK* sequences and all previously published molecular data for 178 species. The Arecoideae were circumscribed to include the following tribes (see Table 2.2 and Fig. 2.1): Areceae, Chamaedoreeae (formerly Hyophorbeae, Ceroxyloideae), Cocoseae, Euterpeae, Geonomateae, Iriarteeae, Leopoldinieae, Manicarieae, Oranieae, Pelagodoxeae, Podococceae, Reinhardtieae, Roystoneeae, and Sclerospermeae. Caryoteae was removed from the subfamily Arecoideae and placed in subfamily Coryphoideae.

A comprehensive generic level analysis by Baker et al. (2009) incorporated all published molecular data and a new morphological data set to construct trees based on supermatrix and supertree approaches (Fig. 2.2). The Arecoideae were supported as monophyletic with tribe Iriarteeae placed sister to the rest of the subfamily. An *Orania/Podococcus/Sclerosperma* (POS) clade was strongly supported (bootstrap values [bsv] 98) and placed sister to the core arecoid group that included tribes Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae. Relationships within this core group were not well supported except for tribe Areceae (84 bsv), and considerable ambiguity remained for the relationships within tribe Areceae. These supertree analyses served as the basis for several biogeographical studies focusing on the ancestral areas and diversification in palms (Couvreur et al., 2011; Baker and Couvreur, 2013a, b).

Baker et al. (2011) used the *PRK* and *RPB2* nuclear genes/spacers to study relationships within the Arecoideae. In the combined analyses, Arecoideae was resolved as monophyletic with strong support. All tribes represented by multiple taxa were resolved as monophyletic with strong support, except for Reinhardtieae embedded within Cocoseae. The POS clade, the Roystoneeae/Reinhardtieae/Cocoseae clade (RRC), and the core arecoids were all well supported.

Palms are well known for their slow rates of chloroplast evolution. Using restriction site (RFLP) and *rbcL* sequence data, Wilson et al. (1990) found a five- to thirteen-fold decrease in substitution rates in 22 representatives of all five palm subfamilies relative to annual species of Asteraceae, Brassicaceae, Gentianaceae, Onagraceae, and Poaceae. The average for palms was 0.009 substitutions per base, with

estimated substitution rates of $1.3 \times 10^{-10}$ substitutions/site/year between *Calamus* (Calamoideae) compared to all the other palms, and $5.2 \times 10^{-11}$ for *Ceroxylon* (Ceroxyloideae). The estimates were calculated using a minimum divergence time of 60 mya based on fossil data (Daghlian, 1981; Muller, 1981). Clegg et al. (1994) found that palms had the lowest substitution rates among the Bromeliales, Liliales and Orchidales. This led subsequent authors (Uhl et al., 1995; Baker et al., 1999; Asmussen et al., 2000, 2006; Loo et al., 2006; Norup et al., 2006; Dransfield et al., 2008; Baker et al., 2009) to suggest that many markers would be needed to provide enough informative characters for plastome-based analyses. A large amount of data generated from next-generation (next-gen) sequencing for a large number of taxa may resolve these polytomies (see Figure 2.1; Jansen et al., 2007; Shendure and Ji, 2008; Givnish et al., 2010; Metzker, 2010; Steele et al., 2012).

In this study the objectives were (1) to compare the utility of two targeted DNA enrichment methods (long range PCR and hybrid gene capture) for whole plastid genome sequencing or assembly, (2) to resolve the deep relationships within the subfamily Arecoideae, particularly among the three major clades (core arecoids, POS clade, and RRC clade), and (3) to use the resulting phylogeny to estimate the evolution of floral arrangements and to infer ancestral areas.

**Materials and methods**

***Taxon sampling*—**Thirty-six taxa were included in the analyses: 31 from the Arecoideae and five from other subfamilies (Appendix 2.1). Sampling included at least one species of all tribes within the Arecoideae (Table 2.2). The sequences for 29 taxa were newly generated for this study (Appendix 2.1). Plastid sequences for *Chamaedorea*

*seifrizii* (Chamaedoreeae), *Elaeis oleifera* (Cocoseae), *Bactris major*, and *Dictyosperma album* were obtained from previous studies (Jansen et al. 2007, Givnish et al. 2010, Heyduk et al., 2015; Appendix 2.1). As also documented in Appendix 2.1, plastome sequences were obtained from previous studies for outgroup taxa *Calamus caryotoides* (Calamoideae, Barrett et al., 2013), *Bismarckia nobilis* and *Phoenix dactylifera* (Coryphoideae, Yang et al., 2010; Barrett et al., 2013), *Pseudophoenix vinifera* and *Ravenea hildebrandtii* (Ceroxyloideae, Barrett et al., 2013).

***Extraction to assembly*—**Three methods were used to extract DNA. Plastid isolation using a sucrose gradient (Jansen et al., 2005) was used initially. Total genomic DNA was also extracted using a modified CTAB method (Doyle and Doyle, 1987), and for problematic taxa (extraction or amplification), the Qiagen DNeasy Plant Kit (Valencia, California, USA) was used with Blattner and Kadereit's (1999) modifications. Plastid isolation was not as reliable as direct sequencing (direct shotgun sequencing or targeted sequencing) from total genomic DNA. Seven species (see Appendix 2.1) were sequenced using the Roche 454 sequencing platform, and all other species were sequenced on the Illumina platform.

*Long range PCR*—A long range PCR (LPCR) protocol was developed to enrich for the chloroplast genome for samples with low concentration of total genomic DNA (13 species, Appendix 2.1). Primers appropriate for LPCR (see Table 2.3) were designed using Primer 3 version 0.4.0 (Koressaar and Remm, 2007; Untergasser et al., 2012); the published *Phoenix dactylifera* (Yang et al., 2010) plastome and plastomes assembled from 454 sequencing were aligned and used as references. Each primer was at least 25 bp long with a Tm greater than 60ºC. Primers were designed within genes and with a

minimum of 20 base pair (bp) overlap between primer pairs. A total of 12 primer pairs (Integrated DNA Technologies, Coralville, Iowa, USA) were used to amplify the entire plastome with each primer pair amplifying 10–20 kilobases (kb). LPCR amplification utilized New England BioLabs LongAmp$^{R}$ Taq PCR Kit (Ipswich, Massachusetts, USA) at one-quarter reactions (12.5 µL final volume). Thermocycler protocols were optimized according to the manufacturer's recommendations. PCR products were cleaned using a 96 well plate with 2 µL 125mM EDTA, 2 µL 3M sodium acetate, and 50 µL 100% ethanol added to each sample. After incubating for 15 min at room temperature, the plate was centrifuged for 30 min at 3000 × *g*. The plate was then turned over and spun for one min at 200 × *g*. Seventy µL of 70% ethanol were added to each well and the plate was then centrifuged at 1700 ×*g* for 15 min. The plate was inverted and centrifuged again at 200 × *g* to dry the pellets. Ten µL of TE was added to each well to resuspend the samples. Concentrations were estimated by nanodrop and then normalized (estimated nanodrop concentration/estimated amplicon size). Amplicon size was estimated for each primer pair (Table 2.3) using the *Phoenix dactylifera* plastome as the reference. All 12 regions were pooled (in equal concentrations) for each taxon. Pooled samples were sheared to 400 bp for Illumina library preparation. Libraries were prepared using the University of Georgia Genomics Facility's (http://dna.uga.edu) modified version of Fisher et al.'s (2011) protocol. Each taxon received a unique barcode to allow pooled samples to be sequenced on the Illumina MiSeq platform (http://www.illumina.com) as 150 bp paired end reads. For one sequencing run, LPCR and gene capture samples were pooled to a final concentration of 10 nM. One-third was from LPCR samples, and two-thirds, from gene capture samples, for a total of 20 taxa.

*Gene capture*—For nine taxa, DNA quality was insufficient for LPCR, and an RNA baits set designed by Heyduk et al. (2015) was used to enrich for the chloroplast genome. Total genomic DNA was sheared to 400 or 600 bp, and Illumina libraries were prepared (see *Long range PCR* above). The RNA baits were designed from 101 732 bp of the *Sabal domingensis* Becc. plastome (Heyduk et al., 2015). The entire sequence was sent to MYcroarray (www.mycroarray.com) for custom oligonucleotide design. Complementary RNA baits were 120 bp long and overlapped by 60 bp against the targeted region. A set of nuclear baits was also included in the hybridization reaction, and the final baits concentration ratio was 1: 100 (plastid: nuclear; Heyduk et al., 2015). Three to five libraries were pooled per hybridization reaction that was carried out according to the MYbaits (Ann Arbor, Michigan, USA) protocol. Pooled samples were sequenced on Illumina MiSeq with 150 bp or 250 bp paired end reads. Ten taxa were enriched using both LPCR and gene capture methods.

*Assembly and annotation*— Reads were first quality trimmed on the 3' end to remove base-pairs with Phred scores (a measure of base call quality) less than 20, following Heyduk et al. (2015). Reads were removed if they were less than 40 bp or if more than 20% of the bases had a Phred score lower than 20. Remaining reads were assembled with both de novo and reference-based assemblers. Velvet version 1.2.03 (Zerbino and Birney, 2008) or EDENA version 3 (Hernandez et al., 2008) were used for de novo assemblies, and YASRA version 2.3 (Ratan, 2009) or AMOScmp-shortReads version 3.1.0 (Pop et al., 2004) were used for reference based assemblies. The plastome of *Phoenix dactylifera* (Yang et al., 2010) served as the reference. Sequencher v5.1

(http://www.genecodes.com) was used to merge these assemblies and for manual editing where merged contigs were in disagreement (e.g. differing bases, insertion/deletions).

The sequence files were uploaded to DOGMA (Dual Organellar Genome Annotator) web server (http://dogma.ccbb.utexas.edu) (Wyman et al., 2004) for gene annotation. Start and stop codons were manually selected within DOGMA. Sequencing coverage and other descriptive statistics were obtained from the YASRA outputs and from Bowtie 2 version 2.2.3 (Langmead and Salzberg, 2012) and BEDTools version 2.21.1 (Quinlan and Hall, 2010). The minimum alignment score function in Bowtie 2 was changed for a more conservative estimate of coverage (setting: score-min L,-0.3,-0.3). Eight taxa enriched by LPCR were also enriched using the RNA baits set as part of another study (Appendix 2.1; Comer et al., in prep.). These taxa, combined with two Arecoideae representatives (*Bactris major* and *Dictyosperma album*) from Heyduk et al. (2015; Appendix 2.1), allowed comparisons between target enrichment methods. Average coverage for the large single copy region (LSC, targeted by baits) and the small single copy region (SSC, not targeted by baits) were calculated for the 10 taxa that were enriched for the plastome using both methods. Paired t-tests were used to determine significant differences in average coverage.

***Phylogenetic analyses*—**Chloroplast genes were aligned in MUSCLE version 3.7 (Edgar, 2004). Mean entropy (Shannon's entropy) was used to assess the variability of each alignment. Alignments with a high entropy value, relative to other alignments, were visually inspected, and poorly aligned genes were excluded (data not shown; Shenkin et al., 1991; Capriotti et al., 2004; Ahola et al., 2006). One hundred and fifteen genes were aligned, and *ycf1* was excluded due to poor alignment. Two data sets (114 genes and 85

genes; Appendix 2.2) were assembled by concatenating aligned genes; the 85 gene set was restricted to those targeted by the RNA baits. Both data sets were uploaded to the CIPRES Science Gateway version 3.3 for analysis (Miller et al., 2010).

Subfamily Calamoideae (*Calamus*) was used as the outgroup based on Dransfield et al. (2008) and Baker et al. (2009). PAUPRat (Nixon, 1999; Sikes and Lewis, 2001) implementing PAUP* version 4.0b10 (Swofford, 2002) was used for the maximum parsimony analyses with the following options selected in the CIPRES portal: seed value randomly generated, 500 replicates with 20% of the informative characters perturbed, uniform weight, increase set to auto, tree bisection-reconnection (branch swapping algorithm), and no rearrangement limit, time limit or reconnection limit specified (sets value to infinity). Two additional runs were conducted similar to the preceding except with 25% of the informative characters perturbed. Bootstrapping analyses for both data sets utilized Phylip version 3.69 (Felsenstein, 1989, 2009) to perform 1 000 blocked bootstrap replicates (block size 597 [85 genes] and 640 [114 genes] bases) of both data sets implementing Seqboot, followed by the parsimony search (Dnapars) with the following options: more thorough search, five trees saved, and the input order jumbled twice each search. The most parsimonious trees and the bootstrap replicates were summarized in a majority rule consensus tree for both data sets. Maximum likelihood analyses were implemented in RAxML version 8.1.11 (Stamatakis, 2006, 2014) with the GTRGAMMA substitution model. The "–f a" option was implemented to conduct a rapid bootstrap analysis (1 000 replicates) and to search for the best scoring tree using the rapid hill-climbing tree search algorithm (Stamatakis et al., 2007). The settings for MrBayes version 3.2.3 (Huelsenbeck and Ronquist, 2001; Ronquist et al., 2012) were:

number of runs, two; number of chains, four; number of substitution types, six; among site rate variation, gamma; number of generations, 50 000 000; sampling frequency, 1000; minimum partition frequency, 0.10; burn-in, 0.20; stoprule, yes; and stopval, 0.01 average standard deviation of split frequencies (Huelsenbeck and Ronquist, 2001; Ronquist et al., 2012). The Bayesian analyses ran for 19 840 000 (114 genes) and 6 230 000 (85 genes) generations before reaching the convergence diagnostic stop value (0.01), and then 20% was discarded as burn-in. For maximum likelihood and Bayesian analyses, data were partitioned by each gene.

***Ancestral area reconstruction*—**To explore the implications of the chloroplast phylogeny on the inferred ancestral distributions of Arecoideae, Lagrange version 20130526 (Ree et al., 2005; Ree and Smith, 2008) was used following Couvreur et al. (2011) and Baker and Couvreur (2013a, b). Geographic distributions were divided into seven areas (Fig. 2.4) based on Couvreur et al. (2011) and Baker and Couvreur (2013a). Taxa were coded based on current general geographic ranges (Dransfield et al., 2008). Geographic assignment of outgroup taxa was based on their inferred ancestral areas (Baker and Couvreur, 2013a), and a few ingroup taxon distributions (e.g., the disjunct *Elaeis guineensis* [Cocoseae]) were modified to simplify estimations (see Lagrange $M_1$ input file, deposited in the Dryad data repository [DOI: http://dx.doi.org/10.5061/dryad.4tn05]). The 85 gene ML tree served as the input tree with the root age set at 100 my (Baker and Couvreur, 2013a). Two models of dispersal were implemented: equal dispersal between all areas ($M_0$) and dispersal probabilities restricted based on geographic constraints at five geological time frames ($M_1$), as described in Baker and Couvreur (2013a).

**Results**

The results of the analyses are summarized in Tables 2.4 and 2.5 and Figures 2.3 and 2.4. Data sets (Appendix 2.2; Dryad data repository, DOI: to be deposited upon manuscript acceptance) of 114 genes (protein, rRNA, and tRNA) and 85 genes (protein and tRNA) comprised 72 957 and 57 312 characters, respectively. The 114 gene matrix had ca. 8% missing data and gaps, and the 85 gene set had 5%. Assemblies derived from long range PCR had an average of 59 contigs per taxon with an average maximum length of 13 119 base pairs and an average coverage of 423× (Table 2.4). Assemblies derived from gene capture averaged 43 contigs per taxon with an average maximum length of 15 991 base pairs and average coverage of 153×. The baits were designed for the LSC region, and the targeted region's average coverage was significantly higher than the non-target region (SSC; Table 2.5; $P < 0.05$). There was no significant difference in average coverage between LPCR and gene capture for the LSC (Table 2.5; $P > 0.05$). The coverage of the SSC regions was significantly greater for the LPCR samples than gene capture samples that were not enriched for the SSC (Table 2.5; $P < 0.05$).

Figure 2.3 shows the ML tree of the 114 gene matrix with support values for all three analyses. Analyses of the 114 gene data set were mainly congruent with those of the 85 gene set (Fig. 5). The ML trees had identical topologies for the tribal relationships, and the three major clades (POS, RRC, and the core arecoids) were well supported (bsv > 79, 114 genes; bsv > 80, 85 genes). Tribe Chamaedoreeae was the earliest diverging lineage within the Arecoideae, followed by Iriarteeae. The POS clade was resolved as sister to a RRC + core arecoids clade. Roystoneae was sister to a Reinhardtieae + Cocoseae clade. Within the core arecoids, Pelagodoxeae +

Leopoldinieae (bsv 78) was sister to a [(Manicarieae + Geonomateae) + (Areceae + Euterpeae)] clade. Tribes Areceae and Euterpeae formed a strongly supported clade. However, the relationships among Areceae + Euterpeae, Leopoldinieae, Geonomateae, Manicarieae, and Pelagodoxeae were not well supported (bsv < 75).

The Bayesian analyses ran for 19 840 000 (114 genes) and 6 230 000 (85 genes) generations before reaching convergence. Topologies and support values were generally congruent with the best maximum likelihood tree. The POS clade was strongly supported (posterior probability [pp] 1.0 both analyses) with Oranieae as sister to a Podococceae + Sclerospermeae (pp 1, 114 genes; pp 0.91, 85 genes). While the monophyly of the RRC clade and tribe Cocoseae were strongly supported in both analyses (pp 1.0), the other tribal relationships within the RRC were weakly supported (pp < 0.90) in the 85 gene set. *Reinhardtia gracilis* (Reinhardtieae) was recovered within *Reinhardtia* (pp 1.0) with the 114 gene set but was the basal branch of the RRC clade with the 85 gene set. The Areceae + Euterpeae clade was also strongly supported (pp 1.0 for both data sets).

For the maximum parsimony analyses, 1 138 (114 genes) and 830 (85 genes) characters were informative, with 1 376 and 1 501 most parsimonious trees recovered, respectively. All subfamilies and most tribal relationships were recovered in all of the most parsimonious trees, but bootstrap support values were very weak (≤ 50; Fig. 2.3). Several elements have been shown to negatively affect bootstrap support, such as data sets with relatively few informative characters compared to the full data matrices and relative to the number of constant characters, and other factors related to the phylogenetic reconstruction program, such as too few parsimony informative characters and/or equal substitution rates across sites (Stewart, 1993; Soltis and Soltis, 2003).

Ancestral areas were mapped onto the summary tree, (Fig. 2.4). Inferred ancestral areas were similar for both dispersal models, but the $M_1$ model (lnL = −102.9), used for ancestral area inference, was a better fit than the $M_0$ model (lnL = −109.8). The raw output for the $M_1$ model has been deposited to Dryad data repository (DOI: to be deposited upon manuscript acceptance). The ancestral areas with relative probabilities greater than 10% of the sum of likelihoods are included in Fig. 2.4. Ancestral areas inferred within two log-likelihood values are provided in the Lagrange output.

**Discussion**

***Methodology comparison*—**Both long range PCR and gene capture were effective methods of targeted enrichment for next generation sequencing with average coverage generally greater than 100× (Table 2.4). Gene capture significantly enriched for the targeted plastid region (LSC) but even without targeted enrichment reads, mapping to the SSC was observed for these samples (Table 2.5). Both methods performed equally well in enrichment of the LSC, with no significant difference (Table 2.5) in average coverage. Coverage was very high for both methods, indicating that at least five times as many libraries could have been pooled for sequencing, and coverage for most species would still be greater than 50×. The variability of coverage among taxa (Table 2.5) is likely due to unequal pooling: LPCR amplicons were pooled for library construction and sequencing, and genomic libraries were pooled for hybrid enrichment and sequencing. Overrepresentation of a region (amplicon pooling) or taxon (library pooling) would result in higher coverage for that sample. As detailed below, DNA quality and quantity are primary considerations for choosing between the two methods, as well as overall time and expense.

*Long range PCR*—Long range PCR primer design requires determination of conserved regions flanking the target and an estimate of amplicon size. Selecting appropriate LPCR reagents (or kits) depends on target size since the maximum amplicon length varies between vendors. The LPCR kit (100 50 μL reactions, $110.00) and 12 primer pairs used here (see Methods for vendor information) cost about $350.00 (primers ranged from 25–30 bp and synthesis cost $ 0.35 per base). Reaction volumes were reduced one-fourth, resulting in 400 reactions per LPCR kit and about 1 800 reactions per primer (about 900 μL at 10 μM, 0.5 μL per reaction). This was sufficient to amplify each chloroplast region and produce enough PCR product for Illumina library construction. Five LPCR kits and one set of the 12 primer pairs would be adequate for at least 100 taxa (one reaction per region, 12 regions per taxon) for PCR optimization.

Long range PCR requires high quality genomic DNA (for this study: intact DNA > 10 kb), with the goal of amplifying relatively large amplicons (≥ 10 kb). However, this method requires relatively small amounts of DNA: in this study, less than 50 ng was needed to amplify all 12 chloroplast regions and provide more than the minimum of 1μg template DNA for the Illumina library construction. Total LPCR thermocycling time, which can exceed 10 hours, was curtailed by including reactions with primers of similar $T_m$s and amplicon size. Verifying (gel electrophoresis) and cleaning (standard ethanol precipitation method) PCR products required about five additional hours.

*Gene capture*—For gene capture, full reference sequences are needed to design tiled baits, and a reference genome is required for intron size estimation when transcriptome data are used for bait design. Initial costs were much higher for the gene capture MYbaits kits. A twelve-reaction kit with a maximum of 20 000 bait sequences

cost $2 400 (not including reagents for post capture). The MYbaits protocol was scaled back one-twentieth from the 2 Mb target size because our target was about 100 kb (see Heyduk et al., 2015). With this scaling and pooling five libraries per hybridization reaction, over 1 000 genomic libraries can be enriched with the chloroplast baits.

Gene capture enrichment requires 1 µg of genomic DNA for library construction. However, gene capture is less sensitive to degraded DNA than LPCR. For example, baits have been used for large scale sequencing of degraded mammoth DNA (Enk, 2014), indicating gene capture as a promising method for amplifying DNA from herbarium material. In this study, one DNA sample (*Podococcus barteri*) was very degraded (visual estimate by gel electrophoresis) but had good coverage (about 90×). The largest fragments of the this sample were ca. 1 kb, with the highest density of fragments ca. 0.5 kb, which is similar to fragment size ranges from DNA extracted from other arecoid herbarium specimens (J. Comer, personal observation). Gene capture thermocycler duration (including PCR after target recovery) was about 40 hours, with most of this time for bait/library hybridization (36 hours). Approximately four additional hours were required to recover and clean targets following hybridization. The rate of evaluating library enrichment (gel electrophoresis and qPCR) was ca. three hours per ninety-six-well plate.

***Phylogeny*—**The approach taken here (whole plastid genome sequencing and selected taxon sampling) is complementary to the denser taxon sampling of previous studies (Baker et al., 2009, 2011). While some plastid markers (e.g., *rbcL*) were the same as those used previously, the data used here were generated independently, and each terminal taxon was represented by a single individual. Relationships recovered here (Fig.

2.3) were largely congruent with previous studies with most differences located at deeper nodes. Three major clades were strongly supported by most of the analyses: (1) POS (Podococceae, Oranieae, Sclerospermeae, (2) RRC (Roystoneae, Reinhardtieae, Cocoseae), and (3) core arecoids (Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae).

*Tribe Chamaedoreeae*—Earlier studies recovered tribe Iriarteeae or Iriarteeae + Chamaedoreeae as the earliest diverging lineage (Asmussen and Chase, 2001; Hahn, 2002b; Asmussen et al., 2006; Baker et al., 2009, 2011), and others suggested alternative placements for tribe Iriarteeae (Lewis and Doyle, 2002; Loo et al., 2006). Here Chamaedoreeae was recovered as the earliest diverging lineage, with Iriarteeae as sister to the rest of the arecoids (Fig. 2.3). Tribe Chamaedoreeae shares several morphological features with subfamily Ceroxyloideae (sister to Arecoideae; see Moore [1973] and Uhl and Dransfield [1987]), and the tribe was previously placed within subfamily Ceroxyloideae (as Hyophorbeae sensu Dransfield and Uhl, 1986), with Arecoideae characterized by flowers in triads or triad derivatives (Dransfield et al., 2005; Asmussen et al., 2006). The flowers of subfamily Ceroxyloideae are predominantly solitary. Tribe Chamaedoreeae, however, has flowers arranged in acervuli or acervulus derivatives (Dransfield et al., 2008). Therefore, the position of Chamaedoreeae as sister to all other arecoids suggests that the triad is a synapomorphy for the first major node of Arecoideae, rather than for the subfamily as a whole.

*POS clade*—The resolution of the relationship between this clade and other tribes has varied (Baker et al., 2009, 2011). With maximum likelihood and Bayesian analyses, the POS clade here was recovered as sister to an RRC + core arecoids clade with

moderate to strong support. Within the POS clade, Oranieae was recovered as sister to a Podococceae + Sclerospermeae clade with varying degrees of support (114 genes: ML bsv 79 and Bayesian pp 1; 85 genes: ML bsv 63 and Bayesian pp 0.91). While some analyses (Baker et al., 2009, 2011) have shown Oranieae and Sclerospermeae as strongly supported sister tribes, Lewis and Doyle (2002) recovered a weakly supported Podococceae + Sclerospermeae clade with a two nuclear gene data set (bsv 69). This clade had some of the shortest internal branches in the subfamily (Fig. 2.3B), and the relatively recent diversification of the clade (about 43 mya; Couvreur et al., 2011; Baker and Couvreur, 2013a) may explain lower support values.

The current geographic distribution of members of the POS clade is disjunct: Podococceae and Sclerospermeae are restricted to the equatorial rainforests of Africa, and Oranieae occurs predominantly in the Malesian region with a few species in Madagascar (Dransfield et al., 2008). Baker and Couvreur's (2013a) analyses suggested that this clade diverged from the core arecoids in Eurasia and then expanded into Africa and the Indo-Pacific. In this study, the placement of POS relative to the RRC + core arecoids clade suggests an alternative hypothesis (Fig. 2.4). The POS clade here was inferred to have dispersed from South America into Africa prior to the diversification of the tribes, with tribe Oranieae later spreading into the Indo-Pacific region, potentially through India (see Morley [2003] and Dransfield et al. [2008]).

*RRC clade*—The RRC clade has been recovered in two previous studies using nuclear genes *PRK* and *RPB2* (Baker et al., 2009, 2011), and other analyses have supported alternative topologies (Asmussen and Chase, 2001; Hahn, 2002a, 2002b; Lewis and Doyle, 2002; Loo et al., 2006). Here the RRC clade was well supported with

Roystoneae as sister to Cocoseae + Reinhardtieae, except in the 85 gene Bayesian analysis where *Reinhardtia* was not monophyletic due to the position of *R. gracilis*. (Assembling sequence data for this taxon was problematic due to low coverage [ca. 9×; 24% gaps and missing data; see Table 2.4].) Both Cocoseae and Reinhardtieae were supported as monophyletic. As with previous studies (Asmussen and Chase, 2001; Baker et al., 2009), subtribe Attaleinae (Cocoseae) was sister to a Bactridinae + Elaeidinae clade.

Reinhardtieae—Reinhardtieae comprises six species of *Reinhardtia* (Henderson, 2002; Dransfield et al., 2008). The species vary considerably in morphology: from *R. paiewonskiana* (tall solitary stems; leaves with many divisions) to *R. koschnyana* (short clustered stems; simple leaves), with the other species forming a morphological grade between these two extremes (Moore, 1957; Henderson, 2002). *Reinhardtia* is monophyletic based on morphological data (Henderson, 2002), and two species (*R. gracilis* and *R. simplex*) were supported as monophyletic based on phylogenetic analysis of two nuclear genes (Baker et al., 2011). This study included four *Reinhardtia* species, and the genus was recovered as monophyletic with two clades: *R. paiewonskiana* + *R. latisecta* and *R. gracilis* + *R. simplex* (Fig. 2.3). These clades correspond to Moore's (1957) subgenera *Reinhardtia* (*R. paiewonskiana* and *R. latisecta*) and *Malortiea* (*R. gracilis* and *R. simplex*) that were based on morphology.

*Core arecoids*—As discussed in the introduction, the core arecoids have been recovered in several studies with varying degrees of support. Here, this group was recovered in most analyses with strong support (Fig. 2.3), in addition to a sister relationship between Areceae and Euterpeae (see below). While topologies between the

likelihood analyses were congruent between data sets, support was generally weak (85 genes bsv 56–60; Fig. 2.4; 114 genes, bsv 50–78; Fig. 2.5). As with the POS clade, the core arecoids had short internal branches (Fig. 2.3B) and relatively recent diversification (Couvreur et al., 2011; Baker and Couvreur, 2013a), which may contribute to the difficulties in resolving the tribal relationships.

Previous analyses (Baker and Couvreur, 2013a) inferred that the core arecoids diverged from the POS clade in Eurasia, with Euterpeae, Geonomateae, Leopoldinieae, and Manicarieae expanding into South America, and Areceae and Pelagodoxeae dispersing into the Indo-Pacific. The ancestral area reconstruction analysis based on our phylogeny (Fig. 2.4) suggested South America as the most likely ancestral area for the core arecoids, with subsequent dispersals into North America (Euterpeae, Geonomateae, and Manicarieae), the Pacific (Pelagodoxeae), and Eurasia with later expansion into the Indo-Pacific (Areceae).

Tribes Areceae and Euterpeae—A clade comprised of Areceae + Euterpeae, recovered in all analyses (Fig. 2.3), was the only clade within the core arecoids well supported in this study. This clade (or this clade + Pelagodoxeae) has been recovered in several studies (Hahn, 2002b; Baker et al., 2011). Morphologically, Areceae and Euterpeae are very similar, sharing an infra- and interfoliar inflorescence, a pseudomonomerous gynoecium (Areceae type: conspicuous sterile ovaries), and fruit with a smooth epicarp (Hahn, 2002b; Dransfield et al., 2008; Baker et al., 2011). However, these characters are not restricted to Areceae and Euterpeae. For example, the Areceae type of pseudomonomery occurs in Pelagodoxeae, as well as some taxa outside

the core arecoids (Roystoneae and Sclerospermeae; Stauffer et al., 2004; Dransfield et al., 2008).

In this study both tribes were represented by multiple taxa and were monophyletic in all analyses. In Euterpeae, *Prestoea* was sister to a monophyletic *Oenocarpus* as in previous studies (Hahn, 2002b; Baker et al., 2009). While Areceae was recovered as monophyletic here, relationships within this tribe were not resolved—most likely due to limited taxon sampling. Areceae is the largest tribe in Arecaceae (Table 2.3 and Appendix 2.1), and ca. 20% of the genera were sampled here (> 2% of the species).

***Conclusions***—The difficulties in recovering well supported resolution of relationships within the core arecoids may have been due to insufficient phylogenetic signal in a limited number of chloroplast genes with low substitution rates (Fig. 2.3; Wilson et al., 1990; Clegg et al., 1994). For this study, long range PCR and gene capture were successful next-generation sequencing tools for generating a large data set of plastid genes for subfamily Arecoideae. Tribal relationships were largely congruent with previous studies, and three major clades (POS, RRC, and core arecoids) were recovered with high support in the maximum likelihood and Bayesian analyses. In light of the short internodes estimated for portions of the trees (Fig. 2.3), caution should be taken in equating the inferred plastome history with the species phylogeny. Incomplete lineage sorting between speciation events may result in species tree/gene tree discordance (e.g., Maddison, 1997), and the chloroplast genome represents a single non-recombining locus. Future work (J. Comer, in prep.) will test the plastid-based phylogenetic inference described here through coalescence-based analysis of numerous nuclear genes.

**Acknowledgements**

We thank Larry Noblick and Patrick Griffith (Montgomery Botanical Center) and Brett Jestrow (Fairchild Tropical Botanic Garden) for assistance collecting many of the palms used in this study. We are also grateful to Anders Lindstrom and Kampon Tansacha (Nong Nooch Tropical Botanical Garden) for hosting the first author and allowing him to sample the palm collection. Several species of *Reinhardtia* were kindly provided by Jeff Marcus (Floribunda Palms and Exotics), and Thomas Couvreur supplied tissue for *Podococcus* and *Sclerosperma*. William Baker and an anonymous reviewer gave constructive criticisms of the manuscript. Funding was provided by the National Science Foundation (DEB-083009; J. H. Leebens-Mack, PI and W. B. Zomlefer, co-PI). Additional travel funds were provided to the first author by the Department of Plant Biology, University of Georgia (Palfrey Grant for Graduate Student Research).

**Literature cited**

AHOLA, V., T. AITTOKALLIO, M. VIHINEN, AND E. UUSIPAIKKA. 2006. A statistical score for assessing the quality of multiple sequence alignments. *BMC Bioinformatics* 7: 484.

ASMUSSEN, C. B., W. J. BAKER, AND J. DRANSFIELD. 2000. Phylogeny of the palm family (Arecaceae) based on *rps16* intron and *trnL-trnF* plastid DNA sequences. *In* K. L. Wilson and D. A. Morrison [eds.], Monocots: Systematics and evolution, 525–537. CSIRO Publishing, Collingwood, Victoria, Australia.

ASMUSSEN, C. B., AND M. W. CHASE. 2001. Coding and noncoding plastid DNA in palm systematics. *American Journal of Botany* 88: 1103–1117.

ASMUSSEN, C. B., J. DRANSFIELD, V. DEICKMANN, A. S. BARFOD, J. -C. PINTAUD, AND W. J. BAKER. 2006. A new subfamily classification of the palm family (Arecaceae): Evidence from plastid DNA phylogeny. *Botanical Journal of the Linnean Society* 151: 15–38.

BAKER, W. J., C. B. ASMUSSEN, S. C. BARROW, J. DRANSFIELD, AND T. A. HEDDERSON. 1999. A phylogenetic study of the palm family (Palmae) based on chloroplast DNA sequences from the *trnL-trnF* region. *Plant Systematics and Evolution* 219: 111–126.

BAKER, W. J., AND T. L. P. COUVREUR. 2013a. Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical biogeography. *Journal of Biogeography* 40: 274–285.

BAKER, W. J., AND T. L. P. COUVREUR. 2013b. Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. II. Diversification history and origin of regional assemblages. *Journal of Biogeography* 40: 286–298.

BAKER, W. J., M. V. NORUP, J. J. CLARKSON, T. L. P. COUVREUR, J. L. DOWE, C. E. LEWIS, J. -C. PINTAUD, et al. 2011. Phylogenetic relationships among arecoid palms (Arecaceae: Arecoideae). *Annals of Botany* 108: 1417–1432.

BAKER, W. J., V. SAVOLAINEN, C. B. ASMUSSEN-LANGE, M. W. CHASE, J. DRANSFIELD, F. FOREST, M. M. HARLEY, et al. 2009. Complete generic-level phylogenetic analyses of palms (Arecaceae) with comparisons of supertree and supermatrix approaches. *Systematic Biology* 58: 240–256.

BARRETT, C. F., J. I. DAVIS, J. LEEBENS-MACK, J. G. CONRAN, AND D. W. STEVENSON. 2013. Plastid genomes and deep relationships among the commelinid monocot angiosperms. *Cladistics* 29: 65–87.

BLATTNER, F., AND J. KADEREIT. 1999. Morphological evolution and ecological diversification of the forest-dwelling poppies (Papaveraceae: Chelidonioideae) as deduced from a molecular phylogeny of the ITS region. *Plant Systematics and Evolution* 219: 181–197.

CAPRIOTTI, E., P. FARISELLI, I. ROSSI, AND R. CASADIO. 2004. A Shannon entropy-based filter detects high- quality profile–profile alignments in searches for remote homologues. *Proteins: Structure, Function, and Bioinformatics* 54: 351–360.

CLEGG, M. T., B. S. GAUT, G. H. LEARN, AND B. R. MORTON. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences, USA* 91: 6795–6801.

COUVREUR, T. L. P., F. FOREST, AND W. J. BAKER. 2011. Origin and global diversification patterns of tropical rain forests: Inferences from a complete genus-level phylogeny of palms. *BMC Biology* 9: 44.

DAGHLIAN, C. 1981. A review of the fossil record of monocotyledons. *The Botanical Review* 47: 517–555.

DOYLE, J. J., AND J. L. DOYLE 1987. Genomic plant DNA preparation from fresh tissue—CTAB method. *Phytochemical Bulletin* 19: 11–15.

DRANSFIELD, J., AND N. W. UHL. 1986. An outline of a classification of palms. *Principes* 30: 3–11.

DRANSFIELD, J., N. W. UHL, C. B. ASMUSSEN, W. J. BAKER, M. M. HARLEY, AND C. E. LEWIS. 2005. A new phylogenetic classification of the palm family, Arecaceae. *Kew Bulletin* 60: 559–569.

DRANSFIELD, J., N. W. UHL, C. B. ASMUSSEN, W. J. BAKER, M. M. HARLEY, AND C. LEWIS. 2008. Genera palmarum. The evolution and classification of palms. Royal Botanical Gardens, Kew, Richmond, Surrey, England.

EDGAR, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

ENK, J. M. 2014. Mammoth phylogeography south of the ice: Large-scale sequencing of degraded DNA from temperate deposits. Ph.D. dissertation, McMaster University, Hamilton, Ontario, Canada.

ENK, J. M., A. M. DEVAULT, M. KUCH, Y. E. MURGHA, J. -M. ROUILLARD, AND H. N. POINAR. 2014. Ancient whole genome enrichment using baits built from modern DNA. *Molecular Biology and Evolution* 31: 1292–1294.

FELSENSTEIN, J. 1989. PHYLIP—Phylogeny inference package (version 3.2). *Cladistics* 5: 164–166.

FELSENSTEIN, J. 2009. PHYLIP (phylogeny inference package) version 3.7a. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, Washington, U.S.A.

FISHER, S., A. BARRY, J. ABREU, B. MINIE, J. NOLAN, T. DELOREY, G. YOUNG, et al. 2011. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biology* 12: R1.

GIVNISH, T. J., M. AMES, J. R. MCNEAL, M. R. MCKAIN, P. R. STEELE, C. W. DEPAMPHILIS, S. W. GRAHAM, et al. 2010. Assembling the tree of the monocotyledons: Plastome sequence phylogeny and evolution of Poales. *Annals of the Missouri Botanical Garden* 97: 584–616.

HAHN, W. J. 2002a. A molecular phylogenetic study of the Palmae (Arecaceae) based on *atpB*, *rbcL*, and *18S* nrDNA sequences. *Systematic Biology* 51: 92–112.

HAHN, W. J. 2002b. A phylogenetic analysis of the arecoid line of palms based on plastid DNA sequence data. *Molecular Phylogenetics and Evolution* 23: 189–204.

HENDERSON, A. J. 2002. Phenetic and phylogenetic analysis of *Reinhardtia* (Palmae). *American Journal of Botany* 89: 1491–1502.

HERNANDEZ, D., P. FRANCOIS, L. FARINELLI, M. OSTERAS, AND J. SCHRENZEL. 2008. De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Research* 18: 802–809.

HEYDUK, K., D. W. TRAPNELL, C. F. BARRETT, AND J. LEEBENS-MACK. 2015. Phylogenomic analyses of *Sabal* (Arecaceae) species relationships using targeted sequence capture. Biological Journal of the Linnean Society 115: in press.

HUELSENBECK, J. P., AND F. RONQUIST. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

JANSEN, R. K., Z. CAI, L. A. RAUBESON, H. DANIELL, C. W. DEPAMPHILIS, J. LEEBENS-MACK, K. F. MÜLLER, et al. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences, USA* 104: 19369–19374.

JANSEN, R. K., L. A. RAUBESON, J. L. BOORE, C. W. DEPAMPHILIS, T. W. CHUMLEY, R. C. HABERLE, S. K. WYMAN, et al. 2005. Methods for obtaining and analyzing whole chloroplast genome sequences. *In* E. A. Zimmer and E. H. Roalson [eds.], Methods in enzymology, vol. 395, 348–384. Academic Press, Waltham, Massachusetts, USA.

KORESSAAR, T., AND M. REMM. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23: 1289–1291.

LANGMEAD, B., and S. L. SALZBERG. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359.

LEWIS, C. E., AND J. J. DOYLE. 2002. A phylogenetic analysis of tribe Areceae (Arecaceae) using two low-copy nuclear genes. *Plant Systematics and Evolution* 236: 1–17.

LOO, A. H. B., J. DRANSFIELD, M. W. CHASE, AND W. J. BAKER. 2006. Low-copy nuclear DNA, phylogeny and the evolution of dichogamy in the betel nut palms and their relatives (Arecinae; Arecaceae). *Molecular Phylogenetics and Evolution* 39: 598–618.

MADDISON, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.

METZKER, M. L. 2010. Sequencing technologies—the next generation. *Nature Reviews Genetics* 11: 31–46.

MILLER, M. A., W. PFEIFFER, AND T. SCHWARTZ. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE) in New Orleans, Louisiana, USA, 2010, 1–8. Also at website http://www.phylo.org/sub_sections/portal/cite.php.

MOORE, H. 1957. *Reinhardtia*. *Gentes Herbarum* 8: 541–576.

MOORE, H. 1973. The major groups of palms and their distribution. *Gentes Herbarum* 11: 27–141.

MORLEY, R. J. 2003. Interplate dispersal paths for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics* 6: 5–20.

MULLER, J. 1981. Fossil pollen records of extant angiosperms. *The Botanical Review* 47: 1–142.

NIXON, K. C. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* 15: 407–414.

NORUP, M. V., J. DRANSFIELD, M. W. CHASE, A. S. BARFOD, E. S. FERNANDO, AND W. J. BAKER. 2006. Homoplasious character combinations and generic delimitation: A case study from the Indo-Pacific arecoid palms (Arecaceae: Areceae). *American Journal of Botany* 93: 1065–1080.

PALMWEB. 2015. Palmweb: Palms of the world online. Website http://www.palmweb.org/ [accessed 31 January 2015].

POP, M., A. PHILLIPPY, A. L. DELCHER, AND S. L. SALZBERG. 2004. Comparative genome assembly. *Briefings in Bioinformatics* 5: 237–248.

QUINLAN, A. R., AND I. M. HALL. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842.

RATAN, A. 2009. Assembly algorithms for next-generation sequence data. Ph.D. dissertation, Pennsylvania State University, State College, Pennsylvania, USA.

REE, R. H., B. R. MOORE, C. O. WEBB, AND M. J. DONOGHUE. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 59: 2299–2311.

REE, R. H., AND S. A. SMITH. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4–14.

RONQUIST, F., M. TESLENKO, P. VAN DER MARK, D. L. AYRES, A. DARLING, S. HOHNA, B. LARGET, et al. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542.

SHENDURE, J., AND H. JI. 2008. Next-generation DNA sequencing. *Nature Biotechnology* 26: 1135–1145.

SHENKIN, P. S., B. ERMAN, AND L. D. MASTRANDREA. 1991. Information-theoretical entropy as a measure of sequence variability. *Proteins: Structure, Function, and Bioinformatics* 11: 297–313.

SIKES, D., AND P. O. LEWIS. 2001. PAUPRat: PAUP* implementation of the parsimony ratchet. Distributed by the authors. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut , U.S.A.

SOLTIS, P. S., AND D. E. SOLTIS. 2003. Applying the bootstrap in phylogeny reconstruction. *Statitstical Science* 18: 256–267.

STAMATAKIS, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.

STAMATAKIS, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

STAMATAKIS, A., F. BLAGOJEVIC, D. S. NIKOLOPOULOS, AND C. D. ANTONOPOULOS. 2007. Exploring new search algorithms and hardware for phylogenetics: RAxML meets the IBM cell. *The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology* 48: 271–286.

STAUFFER, F. W., W. J. BAKER, J. DRANSFIELD, AND P. K. ENDRESS. 2004. Comparative floral structure and systematics of *Pelagodoxa* and *Sommieria* (Arecaceae). *Botanical Journal of the Linnean Society* 146: 27–39.

STEELE, P. R., K. L. HERTWECK, D. MAYFIELD, M. R. MCKAIN, J. LEEBENS-MACK, and J. C. PIRES. 2012. Quality and quantity of data recovered from massively parallel sequencing: Examples in Asparagales and Poaceae. *American Journal of Botany* 99: 330–348.

STEWART, C. -B. 1993. The powers and pitfalls of parsimony. *Nature* 361: 603–607.

SWOFFORD, D. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, Massachusetts, USA.

TRIAS-BLASI, A., W. J. BAKER, A. L. HAIGH, D. A. SIMPSON, O. WEBER, AND P. WILKIN. 2015. A genus-level phylogenetic linear sequence of monocots. *Taxon* 64: in press.

UHL, N. W., AND J. DRANSFIELD 1987. Genera palmarum: A classification of palms based on the work of Harold E.Moore, Jr. Allen Press, Lawrence, Kansas, USA.

UHL, N. W., J. DRANSFIELD, J. I. DAVIS, M. A. LUCKOW, K. S. HANSEN, AND J. J. DOYLE. 1995. Phylogenetic relationships among palms: Cladistic analyses of

morphological and chloroplast DNA restriction site variation. *In* P. J. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries [eds.], Monocotyledons: Systematics and evolution, vol. 2, 623–662. Whitstable Litho Printers, Kent, England.

UNTERGASSER, A., I. CUTCUTACHE, T. KORESSAAR, J. YE, B. C. FAIRCLOTH, M. REMM, AND S. G. ROZEN. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research* 40: e115.

WILSON, M. A., B. GAUT, AND M. T. CLEGG. 1990. Chloroplast DNA evolves slowly in the palm family (Arecaceae). *Molecular Biology and Evolution* 7: 303–314.

WYMAN, S. K., R. K. JANSEN, AND J. L. BOORE. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.

YANG, M., X. ZHANG, G. LIU, Y. YIN, K. CHEN, Q. YUN, D. ZHAO, et al. 2010. The complete chloroplast genome sequence of date palm *Phoenix dactylifera* L. *PLoS ONE* 5: e12762.

ZERBINO, D. R., AND E. BIRNEY. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18: 821–829.

**Table 2.1.** Current subfamilial circumscriptions of Arecaceae (Dransfield et al., 2005, 2008).

| Subfamily | Number of genera | Number of species |
|---|---|---|
| Calamoideae | 21 | 600 |
| Nypoideae | 1 | 1 |
| Coryphoideae | 46 | 450 |
| Ceroxyloideae | 8 | 40 |
| Arecoideae | 107 | 1 300 |

*Notes:* For a comprehensive summary of the history of subfamilial classification, see Table 9.1 in Dransfield et al. (2008).

**Table 2.2.** Arecoideae tribes with number of genera and species (Dransfield et al., 2005, 2008) and species sampled for this study (see Appendix 2.1 for voucher information).

| Tribe | Genera (species) | Sampled species |
|---|---|---|
| Areceae | 59 (630) | *Areca vestiaria*; *Burretiokentia grandiflora*; *Dictyosperma album*; *Drymophloeus litigiosus*; *Dypsis decaryi*; *Heterospathe cagayanensis*; *Hydriastele microspadix*; *Kentiopsis piersoniorum*; *Satakentia liukiuensis*; *Veitchia spiralis* |
| Chamaedoreeae | 5 (120) | *Chamaedorea seifrizii* |
| Cocoseae | 18 (360) | *Attalea speciosa*; *Bactris major*; *Beccariophoenix madagascariensis*; *Elaeis oleifera* |
| Euterpeae | 5 (30) | *Oenocarpus bataua*; *O. minor*; *Prestoea acuminata* var. *montana* |
| Geonomateae | 6 (80) | *Geonoma undata* subsp. *dussiana* |
| Iriarteeae | 5 (30) | *Iriartea deltoidea* |
| Leopoldinieae | 1 (3) | *Leopoldinia pulchra* |
| Manicarieae | 1 (1) | *Manicaria saccifera* |
| Oranieae | 1 (25) | *Orania palindan* |
| Pelagodoxeae | 2 (2) | *Pelagodoxa henryana* |
| Podococceae | 1 (2) | *Podococcus barteri* |

| | | |
|---|---|---|
| Reinhardtieae | 1 (6) | *Reinhardtia gracilis*; *R. latisecta*; *R. paiewonskiana*; *R. simplex* |
| Roystoneeae | 1 (10) | *Roystonea regia* |
| Sclerospermeae | 1 (3) | *Sclerosperma profizianum* |

**Table 2.3.** Long range PCR forward (F) and reverse (R) primer sequences and approximate amplicon size (kb) for each primer pair.

| Name | Sequence (5' to 3') | Size (kb) |
|---|---|---|
| psbA F | CGATTGATGATATCAGCCCAAGTGT | |
| attpH R | CCAAGCTGTAGAAGGTATTGCGAGA | 14 |
| atpH F | GCCACGACCAGTCCATAAATTGTTA | |
| rpoC1 R | AGGGCTTGACGGAAGAATTTCATAA | 10 |
| rpoC1 F | CATATTTCGTCGACCAATCCTTCCT | |
| IhbA R | TGACCAACCATCAGAAGAAGCAAAT | 12 |
| IhbA F | CCGTTGTATTTGCTTCTTCTGATGG | |
| ndhK R | TCCCAATTGTTGGTTCAGTTTATGC | 15 |
| ndhK F | ACTGTGCCGGCTGTTAAAATTAGGT | |
| petA R | GAAGTGAACGTCTTTCCTCGTAGCA | 13 |
| petA F | TAGTGAAATCGCCTTTCCCATTCTT | |
| petB R | CAATTTGGTCCCGAGGTAAGGAATA | 14 |
| petB F | GTTTGAGGAACGTCTCGAGATTCAG | |
| rpl23 R | TTCGGTTATTGGGGAACAATCAATA | 10 |
| rpl23 F | ACACCAAAGAAGAGTTCGACCCAAT | |
| rps7 R | ACGTCGAGGTACTGCAGAAGAAAAA | 12 |
| rps7 F | AATTGGATCGGATTTTGCAGTTTTT | |
| ndhF R | TCCATAATAATGGGGTCAGCTCCTT | 15 |
| ndhF F | GCCAACTCCATTTGTAATTCCATCA | |
| rps15 R | TCAAGTATTAAGTTTCACCAGTAAGATACG | 12 |

| | | |
|---|---|---|
| rps15 F | CTTTTGTGCAATTCCAAATGTGAAG | |
| rrn16 R | AACAACAACTGGAAACGGTTGCTAA | 15 |
| rrn16 F | TTCCAGTACGGCTACCTTGTTACGA | |
| psbA R | CGTCCTTGGATTGCTGTTGCATATT | 20 |

*Note:* Overlapping primer pairs are listed as starting in the psbA gene, around the chloroplast, and ending back at psbA.

**Table 2.4.** Comparison of averages for the Illumina platform (gene capture, long range PCR) and 454 platform (shotgun sequencing) summary statistics from YASRA assemblies, using *Phoenix dactylifera* as the reference.

| Method | Number of contigs | Total length of assembled contigs | Maximum individual contig length | N50 | N90 | Total number of mapped reads | Average coverage [range] |
|---|---|---|---|---|---|---|---|
| Gene capture | 43 | 92516 | 15991 | 7823 | 3523 | 89574 | 153 [57–346] |
| LPCR | 59 | 95962 | 13119 | 5911 | 1200 | 280953 | 423 [9–587] |
| 454 | 58 | 155497 | 18755 | 7346 | 2848 | 80092 | 49 [19–170] |

**Table 2.5.** Comparisons of average coverages of the large single copy region (LSC, targeted by baits) and the small single copy region (SSC, not targeted by baits) for taxa enriched by long range PCR (LPCR) and gene capture.

| | LSC | | SSC | | Gene capture enrichment | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Taxa | Gene capture | LPCR | Gene capture | LPCR | LSC/SSC | Comparison | Degrees of freedom | *t* | *P* |
| *Attalea* | 240 | 542 | 1 | 2626 | 225.03 | Target | 9 | 0.26 | 0.80 |
| *Burretiokentia* | 52 | 811 | 0.5 | 6 | 110.62 | Non-target | 9 | 6.62 | 0.001** |
| *Geonoma* | 313 | 307 | 1 | 2976 | 233.84 | Gene capture | 9 | 3.01 | 0.01* |
| *Leopoldinia* | 754 | 1204 | 4 | 1815 | 209.79 | | | | |
| *Manicaria* | 876 | 299 | 5 | 3345 | 174.52 | | | | |
| *Orania* | 2309 | 907 | 13 | 2014 | 175.25 | | | | |
| *Pelagodoxa* | 713 | 2046 | 6 | 2620 | 126.25 | | | | |
| *Roystonea* | 1870 | 916 | 11 | 1009 | 166.31 | | | | |
| *Bactris* | 284 | 701 | 46 | 2497 | 6.17 | | | | |
| *Dictyosperma** | 48 | 400 | 19 | 1825 | 2.53 | | | | |

*Notes: Sabal domingensis* was used as the reference, and reads were mapped using Bowtie 2. Statistical tests are shown on the

right. Paired two-tailed t-tests were used to determine significant departures of average coverage, between methods for each region, and between regions for the gene capture method. For *Dictyosperma* (*) a tenfold lower concentration of plastid baits was used (Heyduk et al., 2015). Comparison: t-test comparing average coverages of gene capture and LPCR methods for the LSC (target), SSC (non-target), and between regions for gene capture; ** denotes significant differences in coverage, $P < 0.05$.

**Appendix 2.1.** Taxa included in this study including voucher information, GenBank accession number, and enrichment/sequencing method.

---

Subfamily (tribe); ***Species***; *Voucher specimen* (herbarium); GenBank accessions; Enrichment/sequencing method.

Arecoideae (Areceae); ***Areca vestiaria*** Giseke*; Zomlefer 2310* (FTG, NY); KP221698; 454. ***Burretiokentia grandiflora*** Pintaud & Hodel*; Comer 297* (BKF); KP221702; LPCR/GC. ***Dictyosperma album*** (Bory) H.L. Wendl. & Drude ex Scheff.*; Noblick 5069* (FTG); KP221703; LPCR/GC*. ***Drymophloeus litigiosus*** (Becc.) H. E. Moore*; Comer 299* (BKF); KP221704; LPCR. ***Dypsis decaryi*** (Jum.) Beentje & J. Dransf.*; Noblick 5056* (FTG); KP221705; 454. ***Heterospathe cagayanensis*** Becc.*; Kyburz s.n. [31 May 1995]* (FTG); KP221707; 454. ***Hydriastele microspadix*** (Warb. ex K. Schum. & Lauterb.) Burret*; Noblick 5667* (FTG); KP221708; 454. ***Kentiopsis piersoniorum*** Pintaud & Hodel*; Comer 274* (GA); KP221710; GC. ***Satakentia liukiuensis*** (Hatus.) H. E. Moore*; Comer 275* (GA); KP221695; LPCR. ***Veitchia spiralis*** H. Wendl.*; Zona 724* (FTG); KP221697; 454.

Arecoideae (Chamaedoreeae); ***Chamaedorea seifrizii*** Burret*; Zomlefer 2358* (FTG, GA, NY; Givnish et al., 2010); Givnish et al. (2010); 454.

Arecoideae (Cocoseae); ***Attalea speciosa*** Mart. ex Spreng.*; Noblick 4950* (FTG); KP221699; LPCR/GC. ***Bactris major*** Jacq.*; Noblick 5467* (FTG); KP221700; LPCR/GC*. ***Beccariophoenix madagascariensis*** Jum. & H.

Perrier*; Jestrow 2014-FTG-022* (FTG); KP221701; 454. ***Elaeis oleifera*** (Kunth) Cortés*;* Jansen et al., 2007; EU016883-EU016962; 454.

Arecoideae (Euterpeae); ***Oenocarpus bataua*** Mart.*; Comer 294* (BKF); KP221713; GC. ***O. minor*** Mart.*; Comer 300* (BKF); KP221714; GC. ***Prestoea acuminata*** (Willd.) H.E. Moore var. *montana* (Graham) A. J. Hend. & Galeano; *Comer 317* (GA); KP221689; GC.

Arecoideae (Geonomateae); ***Geonoma undata*** Klotzsch subsp. *dussiana* (Becc.) A. J. Hend.*; Roncal 025* (FTG); KP221706; LPCR/GC.

Arecoideae (Iriarteeae); ***Iriartea deltoidea*** Ruiz & Pav.*; Stevenson s.n. [July 2009]* (GA); KP221709; 454.

Arecoideae (Leopoldinieae); ***Leopoldinia pulchra*** Mart.*; Comer 325* (GA); KP221711; LPCR/GC.

Arecoideae (Manicarieae); ***Manicaria saccifera*** Gaertn.*; Noblick 5482* (FTG); KP221712; LPCR/GC.

Arecoideae (Oranieae); ***Orania palindan*** (Blanco) Merr.*; Horn 4981*(FTG); KP221686; LPCR/GC.

Arecoideae (Pelagodoxeae); ***Pelagodoxa henryana*** Becc.*; Comer 276* (GA); KP221687; LPCR/GC.

Arecoideae (Podococceae); ***Podococcus barteri*** Mann & H. Wendl.*; Sunderland 1803* (K); KP221688; GC

Arecoideae (Reinhardtieae); ***Reinhardtia gracilis*** (H. Wendl.) Drude ex Dammer*; Comer 295* (BKF); KP221690; LPCR. ***R. latisecta*** (H. Wendl.) Burret*; Comer 232* (GA); KP221691; GC. ***R. paiewonskiana*** Read, Zanoni & M.

Mejía*; Comer 324* (GA); KP221693; GC. ***R. simplex*** (H. Wendl.) Drude ex Dammer*; Comer 320* (GA); KP221694; GC.

Arecoideae (Roystoneeae); ***Roystonea regia*** (Kunth) O. F. Cook*; Noblick 5248* (GA); KP221692; LPCR/GC.

Arecoideae (Sclerospermeae); ***Sclerosperma profizianum*** Valk. & Sunderl.*; Stauffer & Ovattara 5-010* (G); KP221696; GC.

Calamoideae (Calameae); ***Calamus caryotoides*** A. Cunn. ex Mart.*; Perry s.n. [14 July 1997]* (FTG; Barrett et al., 2013); NC_020365; 454.

Coryphoideae (Borasseae); ***Bismarckia nobilis*** Hildebrandt & H. Wendl.*; Noblick 5054* (FTG; Barrett et al., 2013); NC_020366; 454.

Coryphoideae (Phoeniceae); ***Phoenix dactylifera*** L.*;* Yang et al., 2010; GU811709; 454.

Ceroxyloideae (Cyclospatheae); ***Pseudophoenix vinifera*** (Mart.) Becc.*; Zomlefer 2355* (FTG; Barrett et al., 2013); NC_020364; 454.

Ceroxyloideae (Ceroxyleae); ***Ravenea hildebrandtii*** C. D. Bouché.*; Zomlefer 2357* (FTG; Givnish et al., 2010); Givnish et al., 2010; 454.

---

*Notes:* For data generated from other studies, the voucher location (herbarium) includes the publication citation. Both long range PCR (LPCR) and gene capture (GC) used the Illumina sequencing platform, and genome shotgun sequencing utilized the 454 sequencing platform. **Bactris* and *Dictyosperma* gene capture data from Heyduk et al. (2015).

**Appendix 2.2.** A list of the 114 chloroplast genes analyzed for this study. **Boldface font** = 85 gene data set.

---

***accD, atpA, atpB, atpE, atpF, atpH, atpI,*** *ccsA,* ***cemA, clpP, infA, lhbA, matK,*** *ndhA,* ***ndhB, ndhC,*** *ndhD, ndhE, ndhF, ndhG, ndhH, ndhI,* ***ndhJ, ndhK, petA, petB, petD, petG, petL, petN, psaA, psaB,*** *psaC,* ***psaI, psaJ, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, rbcL, rpl14, rpl16, rpl2, rpl20, rpl22, rpl23,*** *rpl32,* ***rpl33, rpl36, rpoA, rpoB, rpoC1, rpoC2, rps11, rps12, rps14,*** *rps15,* ***rps16, rps18, rps19, rps2, rps3, rps4, rps7, rps8,*** *rrn16, rrn23, rrn4.5, rrn5, trnA-UGC,* ***trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU,*** *trnG-GCC,* ***trnG-UCC, trnH-GUG,*** *trnI-CAU, trnI-GAU,* ***trnK-UUU,*** *trnL-CAA, trnL-UAA, trnL-UAG,* ***trnM-CAU,*** *trnN-GUU,* ***trnP-UGG, trnQ-UUG,*** *trnR-ACG, trnR-UCU,* ***trnS-GCU,*** *trnS-GGA,* ***trnS-UGA, trnT-GGU, trnT-UGU,*** *trnV-GAC, trnV-UAC,* ***trnW-CCA, trnY-GUA, ycf15, ycf2, ycf3, ycf4,*** *ycf68*

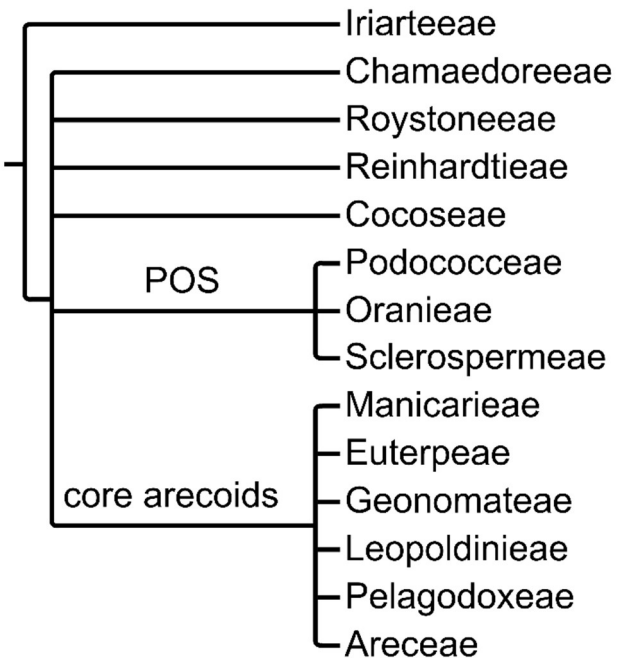



**Figure 2.1.** The phylogeny of the palm subfamily Arecoideae showing tribal relationships according to Dransfield et al. (2008).

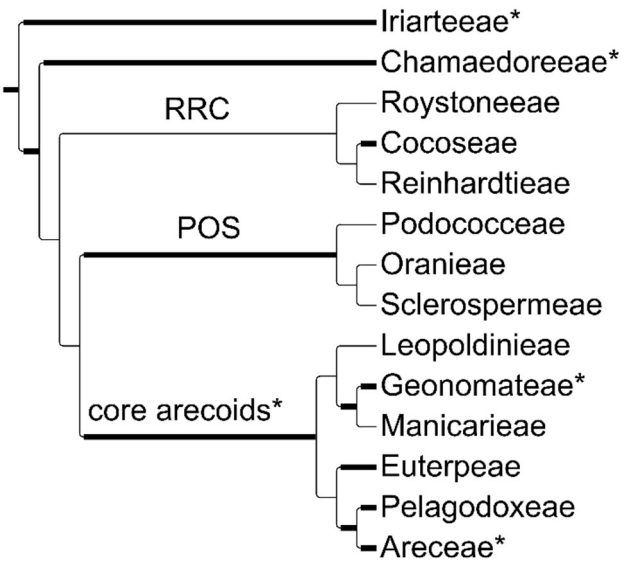
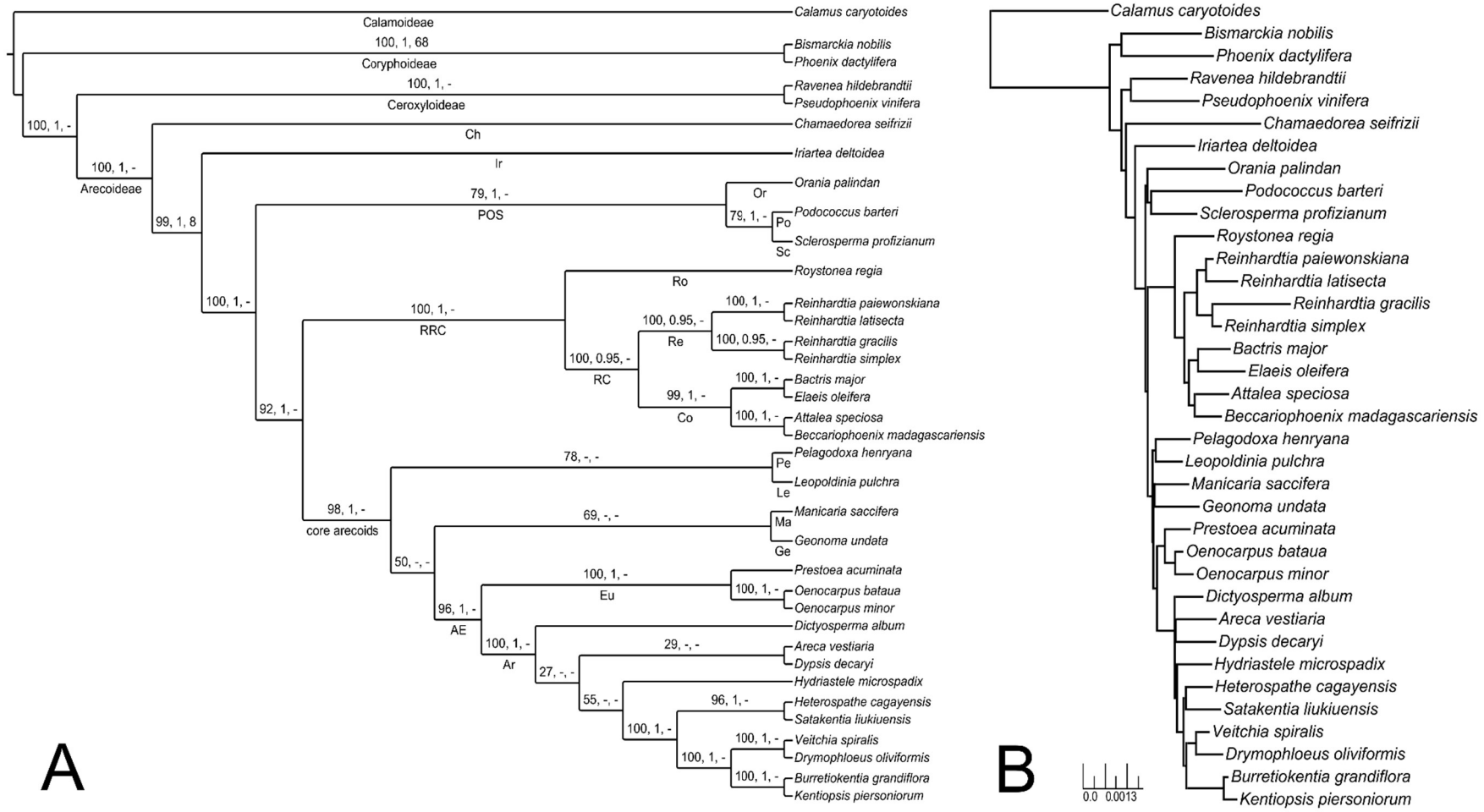



**Figure 2.2.** Tribal phylogeny of subfamily Arecoideae modified from the most congruent supertree (Fig. 3 in Baker et al. [2009]) and the summary tree ("Supertree," Fig. 5 in Baker et al. [2011]). All branches were supported by at least one input tree. Bold lines =

branches supported by five or more input trees; * = clades supported by 10 or more input trees.

**Figure 2.3. (Next page)** (A) Maximum likelihood best tree from the 114 chloroplast gene data set. Numbers above branches = branch support from maximum likelihood, Bayesian, and maximum parsimony analyses; – = clades with bsv ≤ 50 or not supported in the respective analysis. Labels below branches = subfamily, tribe, or major clade. Tribes: Ar = Areceae, Ch = Chamaedoreeae, Co = Cocoseae, Eu = Euterpeae, Ge = Geonomateae, Ir = Iriarteeae, Le = Leopoldinieae, Ma = Manicarieae, Or = Oranieae, Pe = Pelagodoxeae, Po – Roystoneeae, Po = Podococceae, Re = Reinhardtieae, Sc = Sclerospermeae. Major clades: AE (Areceae, Euterpeae); POS (Podococceae, Oranieae, Sclerospermeae); RC (Reinhardtieae, Cocoseae); RCC (Roystoneeae, Reinhardtieae, Cocoseae); core arecoids (Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, Pelagodoxeae). (B) Phylogram of A, showing the slow substitution rates of the chloroplast genome within the palms. The core Arecoids and the POS clade have some of the shortest branch lengths.


Calamus caryotoides
Calamoideae
100, 1, 68
Bismarckia nobilis
Coryphoideae
Phoenix dactylifera
100, 1, -
Ravenea hildebrandtii
Ceroxyloideae
Pseudophoenix vinifera
100, 1, -
Chamaedorea seifrizii
Ch
Iriartea deltoidea
Ir
100, 1, -
Arecoideae
Orania palindan
Or
79, 1, -
POS
99, 1, 8
79, 1, -
Podococcus barteri
Po
Sclerosperma profizianum
Sc
Roystonea regia
Ro
100, 1, -
Reinhardtia paiewonskiana
Reinhardtia latisecta
100, 1, -
RRC
100, 0.95, -
Re
100, 0.95, -
Reinhardtia gracilis
Reinhardtia simplex
100, 0.95, -
RC
100, 1, -
Bactris major
Elaeis oleifera
99, 1, -
Co
92, 1, -
100, 1, -
Attalea speciosa
Beccariophoenix madagascariensis
78, -, -
Pelagodoxa henryana
Pe
Leopoldinia pulchra
Le
98, 1, -
core arecoids
Manicaria saccifera
69, -, -
Ma
Geonoma undata
Ge
50, -, -
Prestoea acuminata
100, 1, -
Eu
100, 1, -
Oenocarpus bataua
Oenocarpus minor
96, 1, -
AE
Dictyosperma album
100, 1, -
Ar
29, -, -
Areca vestiaria
Dypsis decaryi
27, -, -
Hydriastele microspadix
55, -, -
96, 1, -
Heterospathe cagayensis
Satakentia liukiuensis
100, 1, -
100, 1, -
Veitchia spiralis
Drymophloeus oliviformis
100, 1, -
100, 1, -
Burretiokentia grandiflora
Kentiopsis piersoniorum
A
B
0.0 0.0013

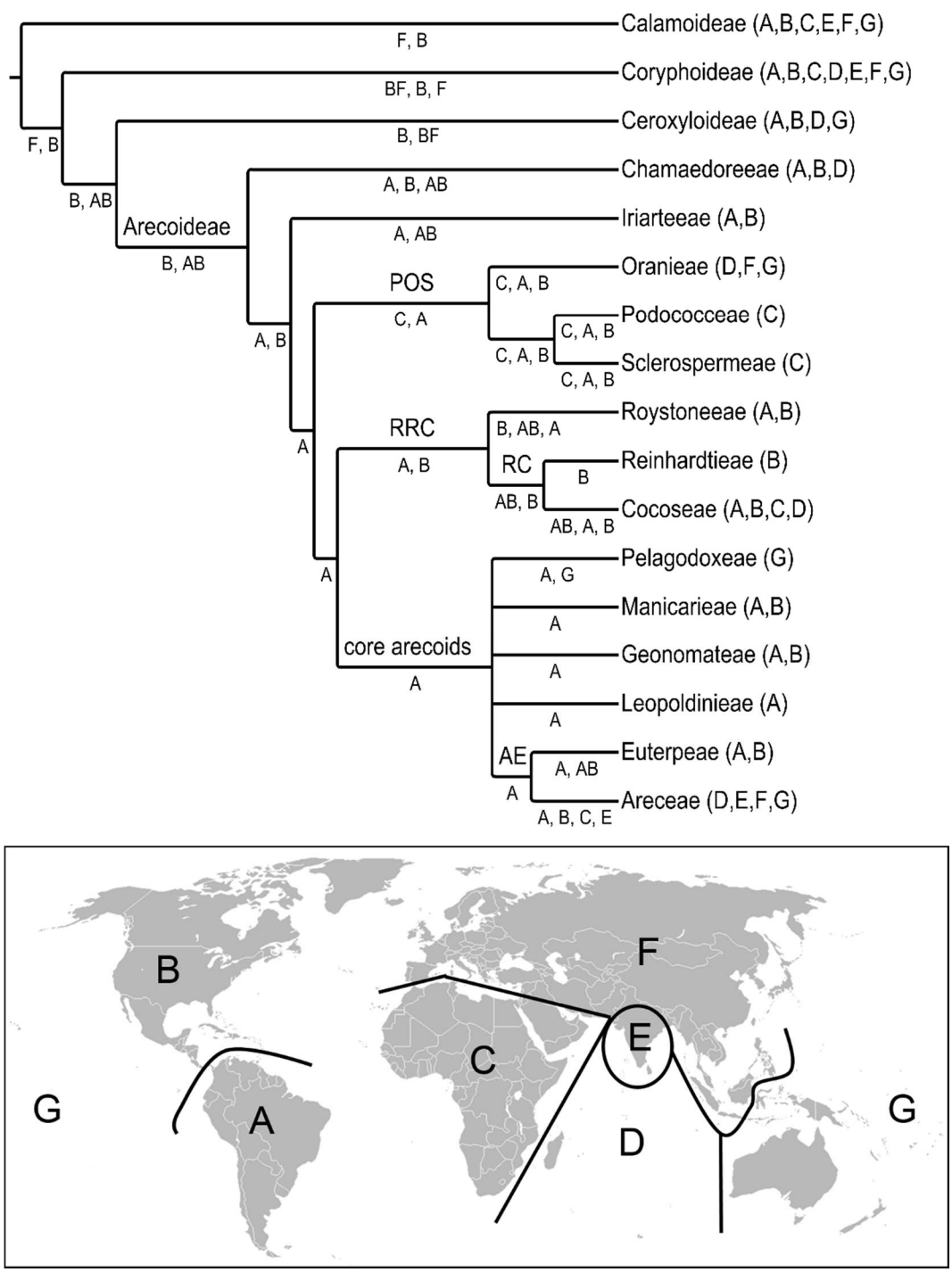



**Figure 2.4.** Summary tree of the tribal relationships in subfamily Arecoideae from all analyses of both data sets (85 and 114 chloroplast genes), with inferred ancestral geographic distributions (below branches; relative probabilities > 10%) and current geographic range following tribal name. Labels above branches = subfamily Arecoideae

and the major clades: AE (Areceae, Euterpeae), POS (Podococceae, Oranieae, Sclerospermeae), RC (Reinhardtieae, Cocoseae), RCC (Roystoneeae, Reinhardtieae, Cocoseae), and core arecoids (Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, Pelagodoxeae). Geographic areas shown in the map inset (Couvreur et al., 2011; Baker and Couvreur, 2013a): A = South America; B = North America, Central America, and the Caribbean; C = Africa and Arabia; D = Indian Ocean Islands and Madagascar; E = India and Sri Lanka; F = Eurasia to Wallace's line; G = Australia and Pacific east of Wallace's line.

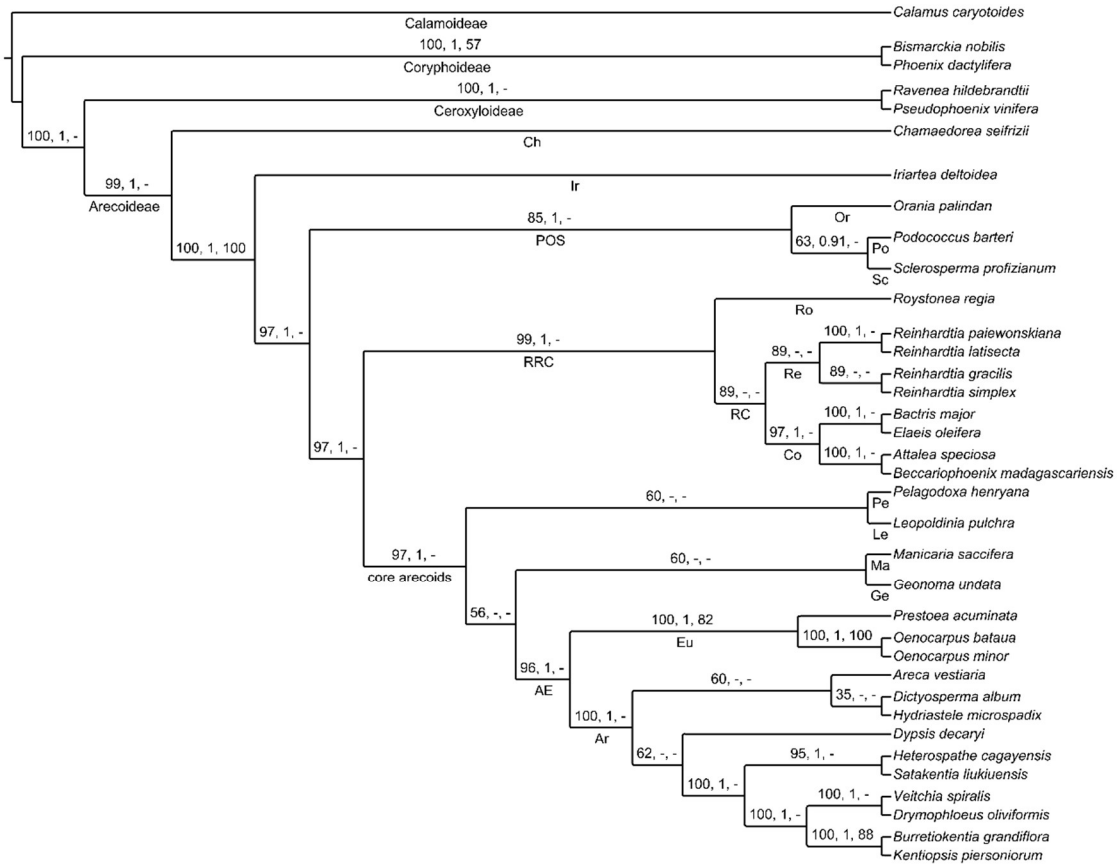



**Figure 2.5.** Maximum likelihood best tree from the 85 chloroplast gene data set. Numbers above branches = branch support from maximum likelihood, Bayesian, and maximum parsimony analyses; – = clades with bsv ≤ 50 or not supported in the respective analysis. Tribes Chamaedoreeae [Ch], Iriarteeae [Ir], and the major clades: POS (Podococceae [Po], Oranieae [Or], Sclerospermeae [Sc]), RCC (Roystoneeae [Ro], Reinhardtieae [Re], Cocoseae [Co]), and core arecoids (Areceae [Ar], Euterpeae [Eu], Geonomateae [Ge], Leopoldinieae [Le], Manicarieae [Ma], Pelagodoxeae [Pe]). AE = Areceae + Euterpeae clade; RC = Reinhardtieae + Cocoseae; labels below branches = subfamily, tribe, or major clade.

# CHAPTER 3

# NUCLEAR PHYLOGENOMICS OF THE PALM SUBFAMILY ARECOIDEAE (ARECACEAE)[2]

[2] Comer JR, Zomlefer WB, Barrett CF, Stevenson DW, Heyduk K, Leebens-Mack J. Submitted to *Molecular Phylogenetics and Evolution* 9/16/2015

**Abstract**

Palms (Arecaceae) include economically important species such as coconut, date palm, and oil palm. Resolution of the palm phylogeny has been problematic due to rapid diversification and slow rates of molecular evolution. The focus of this study is on relationships of the 14 tribes of subfamily Arecoideae and their inferred ancestral areas. A targeted sequencing approach was used to generate a data set of 168 single/low copy nuclear genes for 34 species representing the Arecoideae tribes and the other palm subfamilies. Species trees from the concatenated and coalescent based analyses recovered largely congruent topologies. Three major tribal clades were recovered: the POS clade (Podococceae, Oranieae, Sclerospermeae), the RRC clade (Roystoneeae, Reinhardtieae, Cocoseae), and the core arecoid clade (Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, Pelagodoxeae). Leopoldinieae was sister to the rest of the core arecoids (Geonomateae, Manicarieae + Pelagodoxeae, and Areceae + Euterpeae). The nuclear phylogeny supported a North American origin for subfamily Arecoideae, with most tribal progenitors diversifying within the Americas. The POS clade may have dispersed from the Americas into Africa, with tribe Oranieae subsequently spreading into the Indo-Pacific. Two independent dispersals into the Indo-Pacific were inferred for two tribes within the core arecoids (tribes Areceae and Pelagodoxeae).

**Introduction**

Arecaceae (palm family, 183 genera/ca. 2,600 spp.) occur throughout the tropical to subtropical regions of the world (Baker et al., 2009; Dransfield et al., 2008; Trias-Blasi et al., 2015) and comprise one of the most morphologically diverse angiosperm families (Dransfield et al., 2008). The family is readily identified by the ‘woody’ type growth

(from primary growth and not from a vascular cambium), plicate leaves, and prominent first bract (prophyll) of the inflorescence (Baker et al., 2009; Dransfield et al., 2008; Moore, 1973). The diversity of palms is reflected in their range of habit and habitats, from those with crowns in the canopy to the rattan vines twining throughout forest layers (Dransfield et al., 2008; Moore, 1973). Palms are often key components of pantropical forests, especially in the neotropics (Gentry, 1988; Peters et al., 2004). The fruits of many species are a staple for frugivorous animals and a food source for indigenous peoples. Coconut (*Cocos nucifera* L.), date palm (*Phoenix dactylifera* L.), and oil palm (*Elaeis guineensis* Jacq.) have significant economic importance at the global scale (Balick, 1988; Dransfield et al., 2008; Fadini et al., 2009).

***Arecaceae systematics*—**Studies on the Arecaceae have been summarized in two editions of *Genera Palmarum* (Dransfield et al., 2008; Uhl and Dransfield, 1987), which also provide a formal taxonomic classification system. The first edition (Uhl and Dransfield, 1987) was based largely on the work of H. E. Moore (1973), who devised an informal unranked classification for the palms. More recent classifications have incorporated a variety of molecular marker data, such as random fragment length polymorphisms (RFLP) and data sets with few (one to five markers) nuclear and/or plastid loci (Asmussen and Chase, 2001; Asmussen et al., 2006; Baker et al., 2009, 2011; Loo et al., 2006; Meerow et al., 2009; Uhl et al., 1995). The family is now divided into five subfamilies (see Table 3.1) (Dransfield et al., 2008). The largest generic level study by Baker et al. (2009) included molecular data (six nuclear and nine plastid markers) and a morphological data set (105 characters) for nearly all palm genera. The results of their supertree and super matrix analyses supported many relationships recovered in previous

studies (often with stronger support), provided greater resolution throughout the phylogeny, and identified several problematic clades.

Resolving the phylogeny of palms with molecular data has been challenging due to the slow rates of molecular evolution for both plastid and nuclear markers (Clegg et al., 1994; Gaut et al., 1992, 1996; Wilson et al., 1990; Uhl et al., 1995). When compared to the grasses, the palm substitution rates were about 5-fold and 2.5-fold lower for the plastid (*rbcL*) and nuclear (*Adh*) genes, respectively (Gaut et al., 1992, 1996). This has led authors (Asmussen and Chase, 2001; Asmussen et al., 2006; Baker et al., 2009, 2011) to emphasize the need for larger molecular character data sets. Studies employing high-throughput next-generation sequencing (NGS) approaches have shown that large data sets can improve phylogenetic resolution (Barrett et al., in press; Givnish et al., 2010; Jansen et al., 2007; Metzker, 2010; Steele et al., 2012). Targeted sequencing approaches have been developed to capitalize on NGS methods to sequence large numbers of low copy nuclear genes and small genomes (e.g. plastids and mitochondria) (Comer et al., 2015; Heduk et al., 2015; Lemmon and Lemmon, 2013; Lemmon et al., 2012).

***Subfamily Arecoideae*—**The largest palm subfamily Arecoideae (109 genera/1300 spp.) is distinguished by reduplicately pinnate leaves and unisexual flowers arranged in triads (clusters of one female and two male flowers) with the exception of tribe Chamaedoreeae that has an acervulus (a line of flowers generally comprising several male and a single female flower) or a solitary flower (Dransfield et al., 2008; Moore, 1973; Trias-Blasi et al., 2015; Uhl and Dransfield, 1987). The subfamily encompasses much of the diversity of palms, from diminutive understory species to those reaching the canopy (Dransfield et al., 2008; Moore, 1973; Peters et al., 2004). Arecoideae is

currently divided into 14 tribes (Table 3.1; Dransfield et al., 2008). Previous studies (Asmussen and Chase, 2001; Asmussen et al., 2006; Baker et al., 2009, 2011; Comer et al., 2015; Dransfield et al., 2005, 2008; Loo et al., 2006,) supported the monophyly of the subfamily and identified three major clades: (1) the POS clade (Podococceae, Oranieae, and Sclerospermeae); (2) the RRC clade (Roystoneae, Reinhardtieae, and Cocoseae); and (3) the core arecoid clade (Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae). These studies differed in the resolution and support of relationships within and among these three clades and placement of tribes Chamaedoreeae and Iriarteeae.

Comer et al. (2015) assessed relationships within subfamily Arecoideae with a data set of 114 plastid genes (36 taxa). The results were largely congruent with previous studies, with the following exceptions (Figs. 3.1a, b): (1) tribe Chamaedoreeae was placed as the earliest diverging lineage within the subfamily, (2) the POS clade was recovered as sister to an RRC + core arecoid clade, and (3) an Areceae + Euterpeae clade was well supported.

***Historical biogeography*—**The historical biogeography of palms has been of particular interest due to their primary restriction to equatorial regions (Baker and Couvreur, 2013a,b; Corner, 1966; Couvreur et al., 2011; Dransfield et al., 2008; Gentry, 1988; Moore, 1973). Moreover, palms have been used as a model system to investigate the origin and evolution of tropical rainforests (Baker and Couvreur, 2013a, 2013b; Couvreur and Baker, 2013; Couvreur et al., 2011). Recent studies on the biogeography of palms (Baker and Couvreur, 2013a,b; Couvreur et al., 2011) have been based on the supertree phylogeny from Baker et al. (2009) that infers a Laurasian origin for Arecaceae

about 100 million years ago (Ma) (mid-Cretaceous), with subfamily Arecoideae diverging from subfamily Ceroxyloideae in North America in the Late Cretaceous (ca. 78 Ma). Arecoideae subsequently dispersed into South America where crown node divergence occurred about 74 Ma. While much of the diversification occurred within the Americas, the POS and core arecoid clades were inferred to have diverged in Eurasia with Podococceae and Sclerospermeae spreading into Africa, Oranieae into the Indo-Pacific, and four tribes of the core arecoids (Euterpeae, Leopoldinieae, Geonomateae, and Manicarieae) dispersing back into the Americas. The ancestral area reconstruction based on the plastid phylogeny of Comer et al. (2015) supported similar hypotheses for Chamaedoreeae, Iriarteeae, and the RRC clade. However, the most recent common ancestral area for both the core arecoids and POS clades was inferred as South America, with the POS clade later dispersing into Africa and the Indo-Pacific and some of the core arecoids spreading into North America and the Indo-Pacific.

***Study objectives*—**Our goals were to: (1) produce a robust multi-locus phylogeny of subfamily Arecoideae using more than 100 nuclear genes to test the current phylogenetic hypotheses of Comer et al. (2015) and Baker et al. (2009), and (2) use this phylogeny to reassess the historical biogeographical hypotheses proposed for the Arecoideae by Comer et al. (2015) and Baker and Couvreur (2013a,b).

**Methods**

***Taxon sampling*—**Thirty-four species were sampled, representing the five palm subfamilies and the 14 tribes of subfamily Arecoideae (see Appendix 3.1). *Kingia australis* R. Br. (Dasypogonaceae) was used as the outgroup based on the findings of Givnish et al. (2010), Barrett et al. (2013, 2014), and others.

***Hybrid gene capture*—**Total genomic DNA was extracted using a modified CTAB method (Doyle and Doyle, 1987) and for some problematic taxa Qiagen's DNeasy Plant Kit (Valencia, California, USA) was used with the modifications of Blattner and Kedereit (1999). The total genomic DNA was sheared with a Covaris sonicator (Woburn, MA, USA) to 400 base pairs (bp) or 600 bp for MiSeq 150 or 250, respectively, and then used for Illumina library construction with the protocol of the University of Georgia Genomics Facility (http://dna.uga.edu/services/illumina-sequencing/sample-preparation) modified from Fisher et al. (2011; see also Heyduk et al. [2015] and Comer et al. [2015]).

The resulting genomic libraries were enriched for putatively single copy nuclear genes through hybridization to RNA baits (MYcoarray, Ann Arbor, Michigan, USA) for target exons following protocols described by Heyduk et al. (2015) and Comer et al. (2015). Three to five genomic libraries were pooled (equal concentrations) per hybridization reaction, and the plastid and nuclear RNA baits sets designed by Heyduk et al. (2015) were included at a 1:100 ratio (plastid to nuclear). Following enrichment verification by quantitative PCR, hybridization reactions were pooled for paired-end sequencing with 150 or 250 bp reads on the Illumina MiSeq platform (Comer et al., 2015).

***Assembly*—**Sequence reads were demultiplexed and quality trimmed from the 3′ ends to remove bases with quality scores (Phred scores) less than 20. Following trimming, reads were removed if the length was less than 40 bp, a read had a Phred score less than 20, or if more than 20% of the bases had a Phred score less than 20 (Comer et al., 2015; Heyduk et al., 2015). The *de novo* assembler Trinity v. 2.06 (Grabherr et al.,

2011) was used to assemble the cleaned reads, and CAP3 v. 102011 (Huang and Madan, 1999) was used to collapse assembled contigs with 95% or greater identity. Assembled contigs with segments matching the target exons were identified using BLAST (Basic Local Alignment Search Tool; Expect value $1\times10^{-20}$; Altschul et al., 1990). Following Heyduk et al. (2015), duplicate contigs (two contigs with best hits to the same exon) were removed to reduce the potential for paralogy (see Fig. 3.2b: note genes for which all exons had multiple copies). While some assemblies spanned the intervening intergenic region, exons from the same gene but separated by large unsequenced introns, were concatenated into super scaffolds.

Assemblies for each locus were aligned using PRANK v. 100802 (Löytynoja and Goldman, 2005). Gblocks v. 0.91b (Castresana, 2000) was used to filter poorly aligned and non-conserved regions (-b3=50; minimum length of non-conserved positions) as well as short regions after cleaning (-b4=100; minimum block length after cleaning, 100 bp). Genes were excluded if a significant amount of data was missing (less than 12 taxa with at least 50% of the sequence present) or if the aligned gene exhibited an average pair-wise genetic distance of more than 0.15 (possible paralogy or assembly artifact). Assembly statistics were generated using Bowtie 2 v. 2.2.3 (Langmead and Salzberg, 2012) and BEDtools v. 2.21.2 (Quinlan and Hall, 2010) following Comer et al. (2015) and Heyduk et al. (2015). Scripts used for this study's assembly pipeline can be found at: https://github.com/kheyduk/reads2trees.

***Phylogenetic reconstruction*—**Phylogenetic analyses were performed using supermatrix and coalescence-based species tree estimation approaches. For the supermatrix analyses, aligned genes were concatenated into a single supermatrix

alignment. RAxML v. 8.1.11 (Stamatakis, 2006, 2014) was used to estimate the maximum likelihood tree (ML) with the GTRGAMMA substitution model. Additionally, the concatenated data were partitioned by gene to allow individual parameter estimation. The '-f a' option was used to conduct 500 bootstrap replicates and implement the rapid hill-climbing tree search algorithm (Stamatakis et al., 2007) to find the best scoring ML tree for the concatenated matrix.

The concatenated data were uploaded to the CIPRES Science Gateway v. 3.3 (Miller et al., 2010), and MrBayes v. 3.2.3 (Huelsenbeck and Ronquist, 2001; Ronquist et al., 2012) was used for Bayesian analysis. Preliminary runs indicated that convergence was reached (average standard deviation of split frequencies < 0.01) within 2,000,000 generations. The following settings were used: number of runs, two; number of chains, four; number of substitution types, six; among site rate variation, gamma; number of generations, 2,000,000; sampling frequency, 40; minimum partition frequency, 0.10; burn-in, 0.20; and stoprule, no.

A maximum parsimony analysis of the concatenated data was carried out with TNT v. 1.1 (Tree Analysis Using New Technology, Willi Hennig Society edition; Goloboff [1999] and Goloboff et al. [2008]). The TNT "one-shot" analysis script (consecutively ran random addition sequences, TBR, sectorial searches, and tree fusing each iteration for 20 iterations) was modified to include 100 random addition replications and 1000 standard bootstrap replicates.

We used ASTRAL, a coalescent based species tree estimation method utilizing unrooted gene trees, to estimate the species tree (Mirarab et al., 2014). Individual gene trees and bootstrap replicates were estimated with RAxML (GTRGAMMA, '-f a', and

500 bootstrap replicates) and used as input for ASTRAL v. 4.7.8 (Mirarab et al., 2014). The heuristic version of ASTRAL implements a multi-locus bootstrapping analysis for both the ML best scoring gene trees and the ML bootstrap replicates.

Incongruence between gene trees and the species tree was assessed following Heyduk et al. (2015) (script available at: https://github.com/kheyduk/reads2trees). For each node of the species tree, the number of gene trees supporting or conflicting with the clade were grouped according to bootstrap support (BS), BS ≥ 75, 75 < BS ≥ 50, 50 < BS > 20, and BS ≤ 20.

**Results**

Results are summarized in Figures 3.1–3.6 and Table 3.2. Raw reads have been submitted to the Sequence Read Archive (SRA Study accession: SRP061467). The concatenated genes nexus file with appended gene positions and additional materials (e.g. gene trees) have been deposited to the Dryad Data Repository [DOI: to be deposited upon acceptance of manuscript].

Average coverage for exons was 25.08× (range: 3.07–63.7×), and for introns, was 33.71× (2.42–241.43×) (see Table 3.2). The total length for assembled genes was 165,444 (range: 28,201–307,152 bp). Comparison of exon baits and captured exon sequences showed variable average pairwise genetic distances across genes and taxa (Fig. 3.2a). After filtering, 168 genes were included for the phylogenetic analyses (Fig. 3.2). Figures 3.3–3.6 provide the number of gene trees supporting and conflicting (by BS ≥ 75) with the inferred species tree clade.

***ASTRAL analyses*—**The ASTRAL species trees from the ML bootstrap replicates (BR) analysis (Fig. 3.3) had a normalized quartet score of 73.65% (percent of quartets

induced from the gene trees matching the species tree) (Mirarab et al., 2014). Subfamily Calamoideae was sister to the rest of the palms. A Nypoideae + Coryphoideae clade (BS 96) was well supported (BS 85) as sister to an Arecoideae + Ceroxyloideae clade (BS 48). Subfamily Arecoideae was recovered as monophyletic (BS 57). The POS clade (Podococceae, Oranieae, and Sclerospermeae), the RRC clade (Roystoneae, Reinhardtieae, and Cocoseae), and the core arecoid clade (Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae) were all strongly supported as monophyletic (BS 100). The POS clade was recovered as sister to a RRC + core arecoid clade (BS 51). Within the POS clade, Podococceae was sister to a strongly supported Oranieae + Sclerospermeae clade (BS 100). Roystoneeae was sister to Reinhardtieae + Cocoseae (BS 100), and both Cocoseae and Reinhardtieae were monophyletic (BS 100). Leopoldinieae was sister to the rest of the core arecoids. A clade comprising Geonomateae, Manicarieae, and Pelagodoxeae was well supported as sister to an Areceae + Euterpeae clade (BS 99). Geonomateae was weakly supported (BS 62) as sister to a Manicarieae + Pelagodoxeae clade (BS 73), and the Areceae + Euterpeae clade was strongly supported (BS 100).

The best tree (BGT) recovered in the ASTRAL analysis of the ML gene trees differed from the BR consensus tree in resolution of the placements for the Chamaedoreeae and Iriarteeae, and the RRC lineages, although the differences were poorly supported (BS < 50). Otherwise the BGT analysis was largely congruent with the species tree recovered by the BR analysis (supplemental material Fig. 3.5) and had a comparable normalized quartet score, 73.56%.

***Supermatrix analyses*—**The topologies for concatenated maximum likelihood (CML) and Bayesian (BI) analyses were congruent (Fig. 3.4, CML species tree). Most nodes in the ML species tree were supported by BS ≥ 90 and all nodes in the Bayesian tree had posterior probabilities (PP) ≥ 0.95. The results of the concatenated analyses differed from the BR species tree in the placement of subfamily Nypoideae as sister to a Coryphoideae, Ceroxyloideae, and Arecoideae clade (BS 100, PP 1.0). Ceroxyloideae and Arecoideae were strongly supported as sister subfamilies (BS 100, PP 1.0). Arecoideae was strongly supported as monophyletic (BS 94, PP 1.0), with tribe Iriarteeae as the basal lineage and Chamaedoreeae as sister to the rest of the subfamily (BS 92, PP 1.0). The POS clade was supported as sister to a RRC + core arecoid clade (BS ≥ 95, PP 1.0). Relationships within the POS and the RRC clades were identical to those recovered by ASTRAL analyses, with strong support (BS ≥ 98). Within the core arecoids, two sister clades were supported: the Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae clade (BS 73, PP 1.0), and the Areceae + Euterpeae clade (BS 100, PP 1.0).

The maximum parsimony analysis (MP) of the concatenated data recovered one most parsimonious tree that was consistent topologically with those from the preceding analyses (see Fig. 3.3 and supplemental material Fig. 3.6). The Iriarteeae + Chamaedoreeae clade and relationships between subfamilies Arecoideae, Ceroxyloideae, and Coryphoideae received little support (BS 35 and BS < 40, respectively). Most nodes within the Arecoideae were well supported (BS ≥ 80).

## Discussion

***Phylogenomic implications*—**Two different approaches for estimating the species tree, concatenation and multi-species coalescent model (MSC; Kingman, 1982), were employed in this study. Concatenation combines multiple genes into a single 'supergene' with increased phylogenetic signal over individual genes to estimate an often well-supported species tree (Gadagkar et al., 2005; Rokas and Carroll, 2005). Approaches based on coalescence theory have been used to model incomplete lineage sorting (ILS) and estimate species relationships in the face of ILS (Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007; Maddison, 1997; Mirarab et al., 2014). The advantages and disadvantages of both approaches have been well documented. Simulation studies have shown that concatenation can provide moderate to strong support for incorrect topology under high ILS (Kubatko and Degnan, 2007; Mirarab et al., 2014). Coalescent based approaches may be sensitive to poorly estimated gene trees and missing data (Bayzid and Warnow, 2013; DeGiorgio and Degnan, 2010; Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007; Mirarab et al., 2014; Rosenberg, 2013; Rosenberg and Nordborg, 2002; Springer and Gatesy, 2014).

Congruence between concatenation (CML/BI) and coalescent-based (BR/BGT) species trees (Fig. 3.3, Fig. 3.4 and supplemental material Fig. 3.5) with our palm data set indicates that both approaches can provide robust species trees despite low phylogenetic signal and missing data (including gaps) for some genes and apparent ILS. Topological disagreement between species trees highlight conditions producing weak phylogenetic reconstruction. For example, tribe Leopoldinieae was recovered as the sister lineage to the remaining core arecoid lineages in the ASTRAL species tree (BS 100, Fig. 3.3) but

the tribe was embedded within a Geonomateae, Manicarieae, and Pelagodoxeae clade (BS 73, Fig. 3.4) in the supermatrix tree. One gene tree supported the placement of Leopoldinieae (BS ≥ 75) within the Geonomateae, Manicarieae, and Pelagodoxeae clade, whereas seven gene trees support the placement of Leopoldinieae as sister to the remaining arecoid lineages (BS ≥ 75). This may be due to artifacts of high ILS leading the supermatrix analysis to converge on the inconsistent topology. Inconsistencies and weak support at deeper nodes, e.g. placement of tribes Chamaedoreeae and Iriarteeae, are possibly the result of a combination of factors that are not well modelled with current approaches (Bayzid and Warnow, 2013; Mirarab et al., 2014; Springer and Gatesy, 2014).

The nuclear species trees (Figs. 3.3–3.6) were more congruent with the plastid phylogeny of Comer et al. (2015), particularly at deeper nodes (Fig. 3.3), than with the supertree of Baker et al. (2009). Tribal relationships within the core arecoid clade were largely unresolved in the plastid phylogeny, except for the Areceae + Euterpeae clade. The nuclear phylogeny here provided better support for tribal relationships within the core arecoids. Leopoldinieae was well supported as the earliest diverging lineage in this clade. A clade of Geonomateae, Manicarieae + Pelagodoxeae was supported as sister to the Areceae + Euterpeae clade. A moderately supported clade, comprised of tribes Manicarieae and Pelagodoxeae, was recovered in all analyses. Although morphologically and biogeographically distinct, tribes Manicarieae and Pelagodoxeae are both characterized by fruit with a corky-warted epicarp (a synapomorphy for the clade) not found in other core arecoids (Dransfield et al., 2008; Stauffer et al., 2004).

***Gene tree discordance and summary tree synthesis*—**The origin of the Arecoideae is marked by rapid diversification following divergence from the subfamily Ceroxyloideae lineage ca. 78 Mya (Baker and Couvreur, 2013a; Couvreur and Baker, 2011). The Iriarteeae and Chamaedoreeae lineages diverged from the remaining Arecoideae in rapid succession with only 5 My separating these early events and similar rapid radiations within the core arecoids (Baker and Couvreur, 2013a; Couvreur et al., 2011). The multi-species coalescent model predicts that such speciation events would result in gene tree discordance (DeGiorgio and Degnan, 2010; Degnan and Rosenberg, 2006; Maddison, 1997; Rosenberg and Nordborg, 2002); this is shown by the resolution of the relationships between Chamaedoreeae and Irarteeae and within the core arecoids in the species tree (Fig. 3.3). The rapid diversification of these lineages can be interpreted as hard or near-hard polytomies (Humphries and Winker, 2010; Kodandaramaiah et al., 2010; Lewis et al., 2005; Maddison, 1989). The possibility of a true radiation resulting in a hard polytomy (Fig. 3.1c) cannot be rejected due to the weak support for the resolution of the branching events for the earliest Arecoideae and within the core arecoids in the bifurcating ASTRAL BR tree (Fig. 3.3).

Additional taxon and gene sampling may help resolve this ambiguity (Maddison and Knowles, 2006; Stanley et al., 2011). Targeting genes relating to morphological synapomorphies may also be useful; for example, the history of genes influencing floral organization may provide insight into the importance of these morphological differences in the divergence of the Ceroxyloideae (solitary flowers), Chamaedoreeae (distinct floral clusters [acervuli]), and the remaining Arecoideae (floral clusters in triads; Dransfield et al., 2008; Moore, 1973; Uhl and Dransfield, 1987). Triads and acervuli were both

derived from a cincinnus, a condensed helicoid cyme (Castaño et al., 2014; Ortega-Chávez and Stauffer, 2011; Uhl and Moore, 1978). Tracing the evolutionary history of regulatory genes for inflorescence development may also help resolve relationships involving the earliest branches of the Arecoideae, although these genes may have a history discordant with the species tree, reflecting the plurality of gene trees (e.g. Fig. 3.3).

***Ancestral biogeography*—**The inferred species trees (Fig. 3.1c and Fig. 3.3) support a shared North American origin for subfamilies Arecoideae and Ceroxyloideae, with most of the tribal diversification occurring within the Americas as suggested by Baker and Couvreur (2013a) and Comer et al. (2015). As outlined in the Introduction (section 1.3), these studies provided alternative hypotheses for the shared ancestral area of the core arecoid and the POS clades. Congruence between the nuclear species trees (Fig. 3.1c) and the plastid phylogeny of Comer et al. (2015; Fig. 3.1b) supports the Americas as the most recent shared ancestral area of the core arecoid and POS clades. This also implies the Americas as the ancestral area of the core arecoid clade, with independent dispersal events into the Indo-Pacific for tribes Areceae and Pelagodoxeae. While independent dispersal for these tribes was inferred with the plastid phylogeny, relationships within the core arecoid clade were largely unresolved, limiting those inferences. The placement of Pelagodoxeae as sister to Manicarieae in the nuclear phylogeny provides additional support for the independent dispersal of Areceae and Pelagodoxeae into the Indo-Pacific.

The nuclear species phylogeny supports the hypothesis of Comer et al. (2015), with the diversification of the POS clade taking place as the earliest ancestors dispersed

from the Americas into Africa. Tribes Oranieae and Sclerospermeae diverged within Africa, and subsequently Oranieae dispersed into the Indo-Pacific, with a few extant species restricted to Madagascar and the remaining species distributed in Malesia (Dransfield et al., 2008). This inferred dispersal pattern for Oranieae may be explained by the 'out of India hypothesis', which proposes that India served as a 'raft' to Asia, after it split from Africa and Madagascar during the Maastrichtian (Late Cretaceous, 72−66 Ma; Conti et al., 2002; Morley, 2003). Fossilized fruit, attributed to *Cocos* (Cocoseae), from central India and dated between 65−61 Ma (Srivastava and Srivastava, 2014), provide additional support for this hypothesis. Based on the placement of the POS clade in the nuclear and plastid (Comer et al., 2015) phylogenies, the stem lineage of the POS clade diverged before 65 Ma, and the climate of India was then also habitable to palms (see also Morley [2003]).

**Conclusions**

Resolving the phylogeny of the palms is challenging due to their slow rates of molecular evolution, coupled with recent and rapid diversification that can result in low phylogenetic signal and incomplete lineage sorting (Asmussen and Chase, 2001; Baker and Couvreur, 2013a, b; Baker et al., 2009; Clegg et al., 1994; Couvreur et al., 2011; Gaut et al., 1992, 1996; Wilson et al., 1990). This phylogenomic study incorporated the largest number of nuclear loci generated to date for the Arecaceae and also confirmed the utility across the palm family of the nuclear baits probes designed by Heyduk et al. (2015). The nuclear phylogeny presented here, combined with the work of Heyduk et al. (2015) and Comer et al. (2015) incorporating plastid data, provide a framework for future phylogenomic studies of the palm family.

**Acknowledgments**

The authors thank Larry Noblick and Patrick Griffith (Montgomery Botanical Center) and Brett Jestrow (Fairchild Tropical Botanic Garden) for assistance collecting many of the palms used in this study. They are also grateful to Anders Lindstrom and Kampon Tansacha (Nong Nooch Tropical Botanical Garden) for hosting J.R.C. and allowing him to sample the palm collection. Several species of *Reinhardtia* were kindly provided by Jeff Marcus (Floribunda Palms and Exotics), and Thomas Couvreur supplied tissue for *Podococcus* and *Sclerosperma*. Jerrold I. Davis provided constructive criticism of the manuscript.

**References**

Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, Al-Azwani EK, Chaluvadi S, Pontaroli AC, DeBarry J, et al. 2011 De novo genome sequencing and comparative genomics of date palm (Phoenix dactylifera). *Nat. Biotech.* **29**, 521-527. (doi:http://www.nature.com/nbt/journal/v29/n6/abs/nbt.1860.html#supplementary-information)

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403 – 410.

Asmussen CB, Chase MW. 2001 Coding and noncoding plastid DNA in palm systematics. *Am. J. Bot.* **88**, 1103 – 1117.

Asmussen CB, Dransfield J, Deickmann V, Barfod AS, Pintaud J-C, Baker WJ. 2006 A new subfamily classification of the palm family (Arecaceae): evidence from

plastid DNA phylogeny. *Bot. J. Linn. Soc.* **151**, 15 – 38. (doi:10.1111/j.1095-8339.2006.00521.x)

Baker WJ, Couvreur TLP. 2013a Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical biogeography. *J. Biogeogr.* **40**, 274 – 285. (doi:10.1111/j.1365-2699.2012.02795.x)

Baker WJ, Couvreur TLP. 2013b Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. II. Diversification history and origin of regional assemblages. *J. Biogeogr.* **40**, 286 – 298. (doi:10.1111/j.1365-2699.2012.02794.x)

Baker WJ, Norup MV, Clarkson JJ, Couvreur TLP, Dowe JL, Lewis CE, Pintaud J-C, Savolainen V, Wilmot T, Chase MW. 2011 Phylogenetic relationships among arecoid palms (Arecaceae: Arecoideae). *Ann. Bot.-London* **108**, 1417 – 1432. (doi:10.1093/aob/mcr020)

Baker WJ, Savolainen V, Asmussen-Lange CB, Chase MW, Dransfield J, Forest F, Harley MM, Uhl NW, Wilkinson M. 2009 Complete generic-level phylogenetic analyses of palms (Arecaceae) with comparisons of supertree and supermatrix approaches. *Syst. Biol.* **58**, 240 – 256. (doi:10.1093/sysbio/syp021)

Balick MJ. 1988 *Jessenia* and *Oenocarpus* : neotropical oil palms worthy of domestication. *FAO Plant P.* **88**, 191 p.

Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR, Perez LG, *et al*. In press 2015 Plastid genomes reveal deep phylogenetic support and extensive rate variation among palms and other commelinid monocots. *New Phytol*. (doi: 10.1111/nph.13617)

Barrett CF, Davis JI, Leebens-Mack J, Conran JG, Stevenson DW. 2013 Plastid genomes and deep relationships among the commelinid monocot angiosperms. *Cladistics* **29**, 65 – 87. (doi:10.1111/j.1096-0031.2012.00418.x)

Barrett CF, Specht CD, Leebens-Mack J, Stevenson DW, Zomlefer WB, Davis JI. 2014 Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales)? *Ann. Bot.-London* **113**, 119–133. (doi:10.1093/aob/mct264)

Bayzid MS, Warnow T. 2013 Naive binning improves phylogenomic analyses. *Bioinformatics* **29**, 2277 – 2284.

Blattner F, Kadereit J. 1999 Morphological evolution and ecological diversification of the forest-dwelling poppies (Papaveraceae: Chelidonioideae) as deduced from a molecular phylogeny of the ITS region. *Plant Syst. Evol.* **219**, 181 – 197. (doi:10.1007/BF00985578)

Bourgis F, Kilaru A, Cao X, Ngando-Ebongue G-F, Drira N, Ohlrogge JB, Arondel V. 2011 Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *P. Natl. Acad. Sci. USA* **108**, 12527-12532.

Castaño F, Stauffer F, Marquinez X, Crèvecoeur M, Collin M, Pintaud J-C, Tregear J. 2014 Floral structure and development in the monoecious palm *Gaussia attenuata* (Arecaceae; Arecoideae). *Ann. Bot.-London* **114**, 1483 – 1495.

Castresana J. 2000 Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540 – 552.

Clegg MT, Gaut BS, Learn Jr BH, Morton. BR 1994 Rates and patterns of chloroplast DNA evolution. *P. Natl. Acad. Sci. USA* **91**, 6795 – 6801.

Comer JR, Zomlefer WB, Barrett CF, Davis JI, Stevenson DW, Heyduk K, Leebens-Mack J. 2015 Resolving relationships within the palm subfamily Arecoideae (Arecaceae) using next-gen derived plastid sequences. *Am. J. Bot*. **102**, 888 – 899. (doi:10.3732/ajb.1500057)

Conti E, Eriksson T, Schönenberger J, Sytsma KJ, Baum DA. 2002 Early Tertiary out-of-India dispersal of Crypteroniaceae: evidence from phylogeny and molecular dating. *Evolution* **56**, 1931 – 1942. (doi:10.1111/j.0014-3820.2002.tb00119.x)

Corner EJH. 1966 *The natural history of palms.* Berkley California, USA: University of California Press, 393 p.

Couvreur TLP, Baker WJ. 2013 Tropical rain forest evolution: palms as a model group. *BMC Biol.* **11**, 48.

Couvreur TLP, Forest F, Baker WJ. 2011 Origin and global diversification patterns of tropical rain forests: inferences from a complete genus-level phylogeny of palms. *BMC Biol.* **9**, 44.

DeGiorgio M, Degnan JH. 2010 Fast and consistent estimation of species trees using supermatrix rooted triples. *Mol. Biol. Evol.* **27**, 552 – 569.

Degnan JH, Rosenberg NA. 2006 Discordance of species trees with their most likely gene trees. *PLoS Genet.* **2**, e68.

Doyle JJ, Doyle JL. 1987 Genomic plant DNA preparation from fresh tissue–CTAB method. *Phytochem. Bull.* **19**, 11 – 15.

Dransfield J, Uhl NW, Asmussen CB, Baker WJ, Harley MM, Lewis CE. 2005 A new phylogenetic classification of the palm family, Arecaceae. *Kew Bull.* **60**, 559 – 569.

Dransfield J, Uhl NW, Asmussen CB, Baker WJ, Harley MM, Lewis CE. 2008 *Genera palmarum. The evolution and classification of palms*. Richmond, Surrey, England: Royal Botanical Gardens, Kew; 744 p.

Fadini RF, Fleury M, Donatti CI, Galetti M. 2009 Effects of frugivore impoverishment and seed predators on the recruitment of a keystone palm. *Acta Oecol.* **35**, 188 – 196. (doi:http://dx.doi.org/10.1016/j.actao.2008.10.001)

Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey T, Young G, Fennell T, Allen A, Ambrogio L, *et al*. 2011 A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* **12**, R1.

Gadagkar SR, Rosenberg MS, Kumar S. 2005 Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *J. Exp. Zool. Part B* **304B**, 64 – 74. (doi:10.1002/jez.b.21026)

Gaut BS, Morton BR, McCaig BC, Clegg MT. 1996 Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. *P. Natl. Acad. Sci. USA* **93**, 10274 – 10279.

Gaut BS, Muse SV, Clark WD, Clegg MT. 1992 Relative rates of nucleotide substitution at the rbcL locus of monocotyledonous plants. *J. Mol. Evol.* **35**, 292 – 303.

Gentry AH. 1988 Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Ann. Mo. Bot. Gard.* **75**, 1 – 34. (doi:10.2307/2399464)

Givnish TJ, Ames M, McNeal JR, McKain MR, Steele PR, dePamphilis CW, Graham SW, Pires JC, Stevenson DW, Zomlefer WB, *et al*. 2010 Assembling the tree of the monocotyledons: plastome sequence phylogeny and evolution of Poales 1. *Ann. Mo. Bot. Gard.* **97**, 584 – 616. (doi:10.3417/2010023)

Goloboff PA. 1999 Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics* **15**, 415 – 428.

Goloboff PA, Farris JS, Nixon KC. 2008 TNT, a free program for phylogenetic analysis. *Cladistics* **24**, 774 – 786.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q. 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. biotechnol.* **29**, 644 – 652.

Heyduk K, Trapnell DW, Barrett CF, Leebens-Mack J. 2015 Phylogenomic analyses of *Sabal* (Arecaceae) species relationships using targeted sequence capture. *Bot. J. Linn. Soc*. n/a-n/a. (doi: 10.1111/bij.12551)

Huang X, Madan A. 1999 CAP3: a DNA sequence assembly program. *Genome Res.* **9**, 868 – 877.

Huelsenbeck JP, Ronquist F. 2001 MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754 – 755. (doi:10.1093/bioinformatics/17.8.754)

Humphries EM, Winker K. 2010 Working through polytomies: Auklets revisited. *Mol. Phylogenet. Evol.* **54**, 88 – 96. (doi:http://dx.doi.org/10.1016/j.ympev.2009.07.023)

Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, *et al*. 2007 Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *P. Natl. Acad. Sci. USA* **104**, 19369 – 19374. (doi:10.1073/pnas.0709121104)

Johnson MT, Carpenter EJ, Tian Z, Bruskiewich R, Burris JN, Carrigan CT, Chase MW, Clarke ND, Covshoff S, Edger PP. 2012 Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS ONE* **7**, e50226. (doi:10.1371/journal.pone.0050226)

Kingman JFC. 1982 The coalescent. *Stoch. Proc Appl.* **13**, 235 – 248.

Kodandaramaiah U, Peña C, Braby MF, Grund R, Müller CJ, Nylin S, Wahlberg N. 2010 Phylogenetics of Coenonymphina (Nymphalidae: Satyrinae) and the problem of rooting rapid radiations. *Mol. Phylogenet. Evol.* **54**, 386 – 394. (doi:http://dx.doi.org/10.1016/j.ympev.2009.08.012)

Kubatko LS, Degnan JH. 2007 Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* **56**, 17 – 24.

Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357 – 359. (doi:10.1038/nmeth.1923)

Lemmon AR, Emme SA, Lemmon EM. 2012 Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst. Biol.* **61**, 727 – 744. (doi:10.1093/sysbio/sys049)

Lemmon EM, Lemmon AR. 2013 High-throughput genomic data in systematics and phylogenetics. *Annu. Rev. Ecol. Evol. S.* **44**, 99 – 121. (doi:10.1146/annurev-ecolsys-110512-135822)

Lewis PO, Holder MT, Holsinger KE. 2005 Polytomies and Bayesian phylogenetic inference. *Syst. Biol.* **54**, 241 – 253. (doi:10.1080/10635150590924208)

Loo AHB, Dransfield J, Chase MW, Baker WJ. 2006 Low-copy nuclear DNA, phylogeny and the evolution of dichogamy in the betel nut palms and their relatives (Arecinae; Arecaceae). *Mol. Phylogenet. Evol.* **39**, 598 – 618. (doi:10.1016/j.ympev.2005.12.006)

Löytynoja A, Goldman N. 2005 An algorithm for progressive multiple alignment of sequences with insertions. *P. Natl. Acad. Sci. USA* **102**, 10557 – 10562. (doi:10.1073/pnas.0409137102)

Maddison WP. 1989 Reconstructing character evolution on polytomous cladograms. *Cladistics* **5**, 365 – 377.

Maddison WP. 1997 Gene trees in species trees. *Syst. Biol.* **46**, 523 – 536.

Maddison WP, Knowles LL. 2006 Inferring phylogeny despite incomplete lineage sorting. *Syst. Biol.* **55**, 21 – 30.

Meerow AW, Noblick L, Borrone JW, Couvreur TLP, Mauro-Herrera M, Hahn WJ, Kuhn DN, Nakamura K, Oleas NH, Schnell RJ. 2009 Phylogenetic analysis of

seven *WRKY* genes across the palm subtribe Attaleinae (Arecaceae) identifies *Syagrus* as sister group of the coconut. *PLoS ONE* **4**, e7353.

Metzker ML. 2010 Sequencing technologies—the next generation. *Nat. Rev. Genet.* **11**, 31 – 46.

Miller MA, Pfeiffer W, Schwartz T. 2010 Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Gateway Computing Environments Workshop (GCE), 2010,* pp. 1 – 8.

Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014 ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, i541 – i548. (doi:10.1093/bioinformatics/btu462)

Moore HE. 1973 The major groups of palms and their distribution. *Gentes Herb.* **11**, 27 – 141.

Morley RJ. 2003 Interplate dispersal paths for megathermal angiosperms. *Perspect. Plant Ecol.* **6**, 5 – 20. (doi:http://dx.doi.org/10.1078/1433-8319-00039)

Ortega-Chávez N, Stauffer FW. 2011 Ontogeny and structure of the acervulate partial inflorescence in *Hyophorbe lagenicaulis* (Arecaceae; Arecoideae). *Ann. Bot.-London*, mcr149. (doi:10.1093/aob/mcr149)

Peters HA, Pauw A, Silman MR, Terborgh JW. 2004 Falling palm fronds structure Amazonian rainforest sapling communities. *Proc. R. Soc. B* **271**, S367 – S369.

Quinlan AR, Hall IM. 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841 – 842. (doi:10.1093/bioinformatics/btq033)

Rokas A, Carroll SB. 2005 More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Mol. Biol. Evol.* **22**, 1337 – 1344.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012 MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539 – 542. (doi:10.1093/sysbio/sys029)

Rosenberg NA. 2013 Discordance of species trees with their most likely gene trees: a unifying principle. *Molecular Biol. Evol.* **30**, 2709 – 2713. (doi:10.1093/molbev/mst160)

Rosenberg NA, Nordborg M. 2002 Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat Rev Genet* **3**, 380 – 390.

Singh R, Ong-Abdullah M, Low E-TL, Manaf MAA, Rosli R, Nookiah R, Ooi LC-L, Ooi SE, Chan K-L, Halim MA. 2013 Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* **500**, 335-339.

Springer MS, Gatesy J. 2014 Land plant origins and coalescence confusion. *Trends Plant Sci.* **19**, 267 – 269.

Srivastava R, Srivastava G. 2014 Fossil fruit of *Cocos* L. (Arecaceae) from Maastrichtian-Danian sediments of central India and its phytogeographical significance. *Acta Palaeobot.* **54**, 67 – 75. (doi:10.2478/acpa-2014-0003)

Stamatakis A. 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688 – 2690. (doi:10.1093/bioinformatics/btl446)

Stamatakis A. 2014 RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313. (doi:10.1093/bioinformatics/btu033)

Stamatakis A, Blagojevic F, Nikolopoulos DS, Antonopoulos CD. 2007 Exploring new search algorithms and hardware for phylogenetics: RAxML meets the IBM cell. *J VLSI Sign Process Syst Sign Im* **48**, 271 – 286. (doi:10.1007/s11265-007-0067-4)

Stanley EL, Bauer AM, Jackman TR, Branch WR, Mouton PLFN. 2011 Between a rock and a hard polytomy: rapid radiation in the rupicolous girdled lizards (Squamata: Cordylidae). *Mol. Phylogenet. Evol.* **58**, 53 – 70. (doi:http://dx.doi.org/10.1016/j.ympev.2010.08.024)

Stauffer FW, Baker WJ, Dransfield J, Endress PK. 2004 Comparative floral structure and systematics of Pelagodoxa and Sommieria (Arecaceae). *Bot. J. Linn. Soc*. **146**, 27 – 39. (doi:10.1111/j.1095-8339.2004.00307.x)

Steele PR, Hertweck KL, Mayfield D, McKain MR, Leebens-Mack J, Pires JC. 2012 Quality and quantity of data recovered from massively parallel sequencing: examples in Asparagales and Poaceae. *Am. J. Bot.* **99**, 330 – 348. (doi:10.3732/ajb.1100491)

Trias-Blasi A, Baker WJ, Haigh AL, Simpson DA, Weber O, Wilkin P. 2015 A genus-level phylogenetic linear sequence of monocots. *Taxon* **64**, 552 – 581. (doi: 10.12705/643.9)

Uhl NW, J Dransfield. 1987 *Genera palmarum: a classification of palms based on the work of Harold E. Moore, Jr*. Lawrence, Kansas, USA: Allen Press; 610 p.

Uhl NW, Dransfield J, Davis JI, Luckow MA, Hansen KS, Doyle JJ. 1995 Phylogenetic relationships among palms: cladistic analyses of morphological and chloroplast DNA restriction site variation. In *Monocotyledons: systematics and evolution* (ed. PJ Rudall, PJ Cribb, DF Cutler, CJ Humphries), pp. 623 – 662. Kent, England: Whitstable Litho Printers Ltd.

Uhl NW, Moore Jr HE. 1978 The structure of the acervulus, the flower cluster of Chamaedoreoid palms. *Am. J. Bot.* **65**, 197 – 204. (doi:10.2307/2442453)

Wilson MA, Gaut B, Clegg MT. 1990 Chloroplast DNA evolves slowly in the palm family (Arecaceae). *Mol. Biol. Evo.* **7**, 303 – 314.

**Appendix 3.1.** Taxa included in this study including voucher information. For data generated from previous studies, the voucher location (herbarium) includes the publication citation.

Subfamily (tribe); ***species***; *voucher specimen* (herbarium).

Arecoideae (Areceae); ***Burretiokentia grandiflora*** Pintaud & Hodel*; Comer 297* (BKF). ***Dictyosperma album*** (Bory) H. L. Wendl. & Drude ex Scheff.*; Noblick 5069* (FTG; Heyduk et al. [2015]). ***Dypsis decaryi*** (Jum.) Beentje & J. Dransf.*; Noblick 5056* (FTG). ***Heterospathe cagayanensis*** Becc.*; Kyburz s.n. [31 May 1995]* (FTG). ***Hydriastele microspadix*** (Warb. ex K. Schum. & Lauterb.) Burret*; Noblick 5667* (FTG). ***Kentiopsis piersoniorum*** Pintaud & Hodel*; Comer 274* (GA). ***Veitchia spiralis*** H. Wendl.*; Zona 724* (FTG).

Arecoideae (Chamaedoreeae); ***Chamaedorea seifrizii*** Burret*; Zomlefer 2358* (FTG, GA, NY; Givnish et al. [2010]).

Arecoideae (Cocoseae); ***Attalea speciosa*** Mart. ex Spreng.*; Noblick 4950* (FTG). ***Bactris major*** Jacq.*; Noblick 5467* (FTG; Heyduk et al. [2015]). ***Elaeis guineensis*** Jacq.*;* (Bourgis et al. [2011]). ***E. oleifera*** (Kunth) Cortés*;* (Singh et al. [2013]).

Arecoideae (Euterpeae); ***Oenocarpus bataua*** Mart.*; Comer 294* (BKF). ***O. minor*** Mart.*; Comer 300* (BKF). ***Prestoea acuminata*** (Willd.) H. E. Moore var. *montana* (Graham) A. J. Hend. & Galeano; *Comer 317* (GA).

Arecoideae (Geonomateae); ***Geonoma undata*** Klotzsch subsp. *dussiana* (Becc.) A. J. Hend.*; Roncal 025* (FTG).

Arecoideae (Iriarteeae); ***Iriartea deltoidea*** Ruiz & Pav.*; Stevenson s.n. [July 2009]* (GA).

Arecoideae (Leopoldinieae); ***Leopoldinia pulchra*** Mart.*; Comer 325* (GA).

Arecoideae (Manicarieae); ***Manicaria saccifera*** Gaertn.*; Noblick 5482* (FTG).

Arecoideae (Oranieae); ***Orania palindan*** (Blanco) Merr.*; Horn 4981* (FTG).

Arecoideae (Pelagodoxeae); ***Pelagodoxa henryana*** Becc.*; Comer 276* (GA).

Arecoideae (Podococceae); ***Podococcus barteri*** Mann & H. Wendl.*; Sunderland 1803* (K).

Arecoideae (Reinhardtieae); ***Reinhardtia gracilis*** (H. Wendl.) Drude ex Dammer*; Comer 295* (BKF). ***R. latisecta*** (H. Wendl.) Burret*; Comer 323* (GA). ***R. paiewonskiana*** Read, Zanoni & M. Mejía*; Comer 324* (GA). ***R. simplex*** (H. Wendl.) Drude ex Dammer*; Comer 320* (GA).

Arecoideae (Roystoneeae); ***Roystonea regia*** (Kunth) O. F. Cook*; Noblick 5248* (GA).

Arecoideae (Sclerospermeae); ***Sclerosperma profizianum*** Valk. & Sunderl.*; Stauffer & Ouattara 5-010* (G).

Calamoideae (Calameae); ***Calamus caryotoides*** A. Cunn. ex Mart.*; Perry s.n. [14 July 1997]* (FTG).

Coryphoideae (Phoeniceae); ***Phoenix dactylifera*** L.*;* (Al-Dous et al. [2011]).

Coryphoideae (Sabaleae); ***Sabal domingensis*** Becc.; *Jestrow 2012-207* (FTG; Heyduk et al. [2015]).

Coryphoideae (Trachycarpeae); ***Serenoa repens*** (W. Bartram) Small; *Soltis & Miles 2935* (FLAS; Johnson et al. [2012]).

Ceroxyloideae (Cyclospatheae); ***Pseudophoenix vinifera*** (Mart.) Becc.*; Zomlefer 2355* (FTG).

Nypoideae; ***Nypa fruticans*** Wurmb; *Chase 34461* (K; Johnson et al. [2012]).

Dasypogonaceae; ***Kingia australis*** R.Br.; *Thiele 3703* (PERTH; Givnish et al. [2010]).

**Table 3.1.** Current subfamilial circumscription of the Arecaceae and tribes of subfamily Arecoideae (Dransfield et al., 2008, Dransfield et al., 2009 and Trias-Blasi, 2015) and the species sampled for this study (see Appendix 3.1 for voucher information).

| Subfamily | Tribe | Genera (Species) | Species Sampled |
|---|---|---|---|
| Calamoideae | | 18 (600) | *Calamus caryotoides* |
| Ceroxyloideae | | 8 (40) | *Pseudophoenix vinifera* |
| | | 47 (450) | *Phoenix dactylifera; Sabal* |
| Coryphoideae | | | *domingensis; Serenoa repens* |
| Nypoideae | | 1 (1) | *Nypa fruticans* |
| Arecoideae | | 109 (1300) | |
| | Areceae | 61 (630) | *Burretiokentia grandiflora*; *Dictyosperma album*; *Dypsis decaryi; Heterospathe cagayanensis; Hydriastele microspadix; Kentiopsis piersoniorum; Veitchia spiralis* |
| | Chamaedoreeae | 5 (120) | *Chamaedorea seifrizii* |
| | Cocoseae | 18 (360) | *Attalea speciosa*; *Bactris major*; *Elaeis guineensis*; *Elaeis oleifera* |

| | | |
|---|---|---|
| Euterpeae | 5 (30) | *Oenocarpus bataua*; *O. minor*; *Prestoea acuminata* var. *montana* |
| Geonomateae | 6 (80) | *Geonoma undata* subsp. *Dussiana* |
| Iriarteeae | 5 (30) | *Iriartea deltoidea* |
| Leopoldinieae | 1 (3) | *Leopoldinia pulchra* |
| Manicarieae | 1 (1) | *Manicaria saccifera* |
| Oranieae | 1 (25) | *Orania palindan* |
| Pelagodoxeae | 2 (2) | *Pelagodoxa henryana* |
| Podococceae | 1 (2) | *Podococcus barteri* |
| Reinhardtieae | 1 (6) | *Reinhardtia gracilis*; *R. latisecta*; *R. paiewonskiana*; *R. simplex* |
| Roystoneeae | 1 (10) | *Roystonea regia* |
| Sclerospermeae | 1 (3) | *Sclerosperma profizianum* |

**Table 3.2.** Summary statistics for nuclear genes generated by this study, after removing non-target reads and duplicate exons. The baits set total target length was ca. 200kb for 176 genes (837 exons). E = number exons assembled, EC = average coverage of exons, EL = total assembled length for exons, G = number of genes assembled, IC = average coverage of introns, IL = total length of assembled introns, TL = total assembled length of exons and introns in bp.

| Taxon | TL | EL | IL | E | G | EC | IC |
|---|---|---|---|---|---|---|---|
| *Attalea speciose* | 188126 | 101231 | 86895 | 504 | 167 | 17.79 | 32.76 |
| *Burretiokentia grandiflora* | 51366 | 35488 | 15878 | 261 | 137 | 3.07 | 2.44 |
| *Calamus caryotoides* | 71069 | 39646 | 31423 | 268 | 139 | 3.83 | 2.42 |
| *Dypsis decaryi* | 113890 | 64216 | 49674 | 346 | 145 | 11.54 | 27.76 |
| *Geonoma undata* | 207351 | 97993 | 109358 | 471 | 160 | 14.75 | 5.94 |
| *Heterospathe cagayanensis* | 162701 | 93521 | 69180 | 485 | 166 | 9.97 | 5.80 |
| *Hydriastele microspadix* | 152542 | 88732 | 63810 | 484 | 161 | 26.57 | 14.73 |
| *Iriartea deltoidea* | 100714 | 62106 | 38608 | 437 | 160 | 18.92 | 9.08 |
| *Kentiopsis piersoniorum* | 157805 | 93787 | 64018 | 498 | 163 | 34.74 | 110.49 |
| *Leopoldinia pulchra* | 74168 | 40833 | 33335 | 265 | 135 | 19.74 | 21.86 |
| *Manicaria saccifera* | 226605 | 102990 | 123615 | 442 | 158 | 44.57 | 14.05 |
| *Oenocarpus bataua* | 194533 | 101816 | 92717 | 480 | 163 | 63.70 | 21.49 |
| *Oenocarpus minor* | 165913 | 95452 | 70461 | 485 | 159 | 42.92 | 142.98 |
| *Orania palindan* | 214807 | 98045 | 116762 | 467 | 167 | 26.92 | 8.67 |
| *Pelagodoxa henryana* | 170366 | 94001 | 76365 | 481 | 168 | 20.20 | 8.33 |
| *Podococcus barteri* | 87748 | 56472 | 31276 | 449 | 162 | 30.84 | 241.43 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Prestoea acuminate* | 169125 | 93518 | 75607 | 469 | 163 | 26.60 | 12.57 |
| *Pseudophoenix vinifera* | 169331 | 91539 | 77792 | 457 | 166 | 9.71 | 17.04 |
| *Reinhardtia gracilis* | 28201 | 18376 | 9825 | 200 | 120 | 3.32 | 9.72 |
| *Reinhardtia latisecta* | 272145 | 127591 | 144554 | 499 | 165 | 46.82 | 12.91 |
| *Reinhardtia paiewonskiana* | 265446 | 128701 | 136745 | 497 | 167 | 39.14 | 74.09 |
| *Reinhardtia simplex* | 301752 | 127914 | 173838 | 470 | 161 | 34.82 | 10.18 |
| *Roystonea regia* | 259472 | 120962 | 138510 | 484 | 164 | 34.38 | 11.93 |
| *Sclerosperma profizianum* | 175139 | 97134 | 78005 | 489 | 167 | 23.41 | 11.33 |
| *Veitchia spiralis* | 155775 | 94717 | 61058 | 495 | 163 | 18.71 | 12.68 |
| Average overall | 165444 | 86671 | 78772 | 435 | 158 | 25.08 | 33.71 |

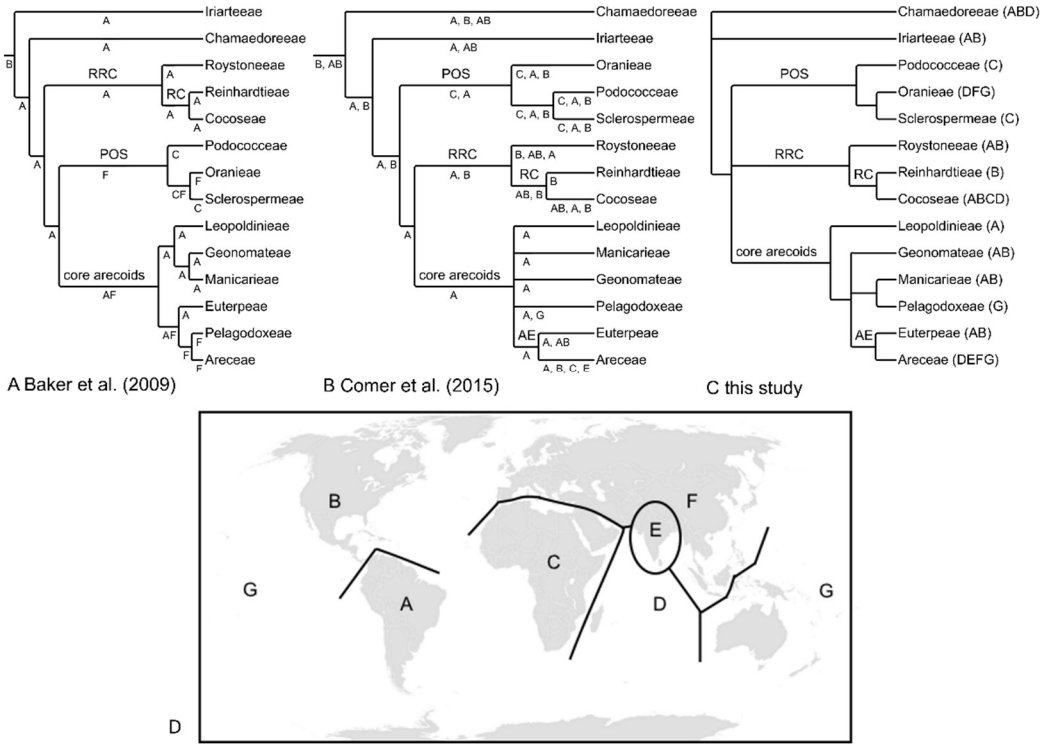



**Figure 3.1.** Tribal phylogenies of subfamily Arecoideae and geographic areas: A) most congruent supertree with all branches supported by one or more input trees (modified from Fig. 3 in Baker et al. [2009]) with inferred ancestral areas below branches (from Fig. 1 in Baker and Couvreur [2013a]), B) chloroplast summary tree with branches supported by BS ≥ 79 and PP ≥ 0.95 and inferred ancestral areas below branches (modified from Fig. 4 in Comer et al. [2015]), C) nuclear summary tree (this study) with branches supported by BS ≥ 70 in the bootstrap replicate ASTRAL analysis with current geographic ranges (see D) following tribal names, D) Map of geographic areas (Baker and Couvreur, 2013a and Couvreur et al., 2011): A = South America; B = North America, Central America, and the Caribbean; C = Africa and Arabia; D = Indian Ocean Islands and Madagascar; E = India and Sri Lanka; F = Eurasia to Wallace's line; G = Australia and Pacific east of Wallace's line. Labels above branches are major clades: AE (Areceae + Euterpeae); core arecoids (Areceae, Euterpeae, Geonomateae, Leopoldinieae,

Manicarieae, and Pelagodoxeae); POS (Podococceae, Oranieae, and Sclerospermeae); RC (Reinhardtieae + Cocoseae); and RRC (Roystoneeae, Reinhardtieae, and Cocoseae).

**Figure 3.2. (Next page)** Graphic comparisons of assembled genes. Panel A shows comparison of pairwise genetic distance from target taxon to baits reference sequence. Distances were averaged for tribes represented by multiple taxa. Black dots = non-arecoid palms; Arecoideae: yellow dots = core arecoids (Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, Pelagodoxeae); blue dots = RRC (Roystoneeae, Reinhardtieae, Cocoseae); red dots = POS (Podococceae, Oranieae, Sclerospermeae); green dots = tribe Iriarteeae. Panel B shows genes that were captured (black), not captured (white), or were in multiple copies (red). * = Genes excluded from further analyses.

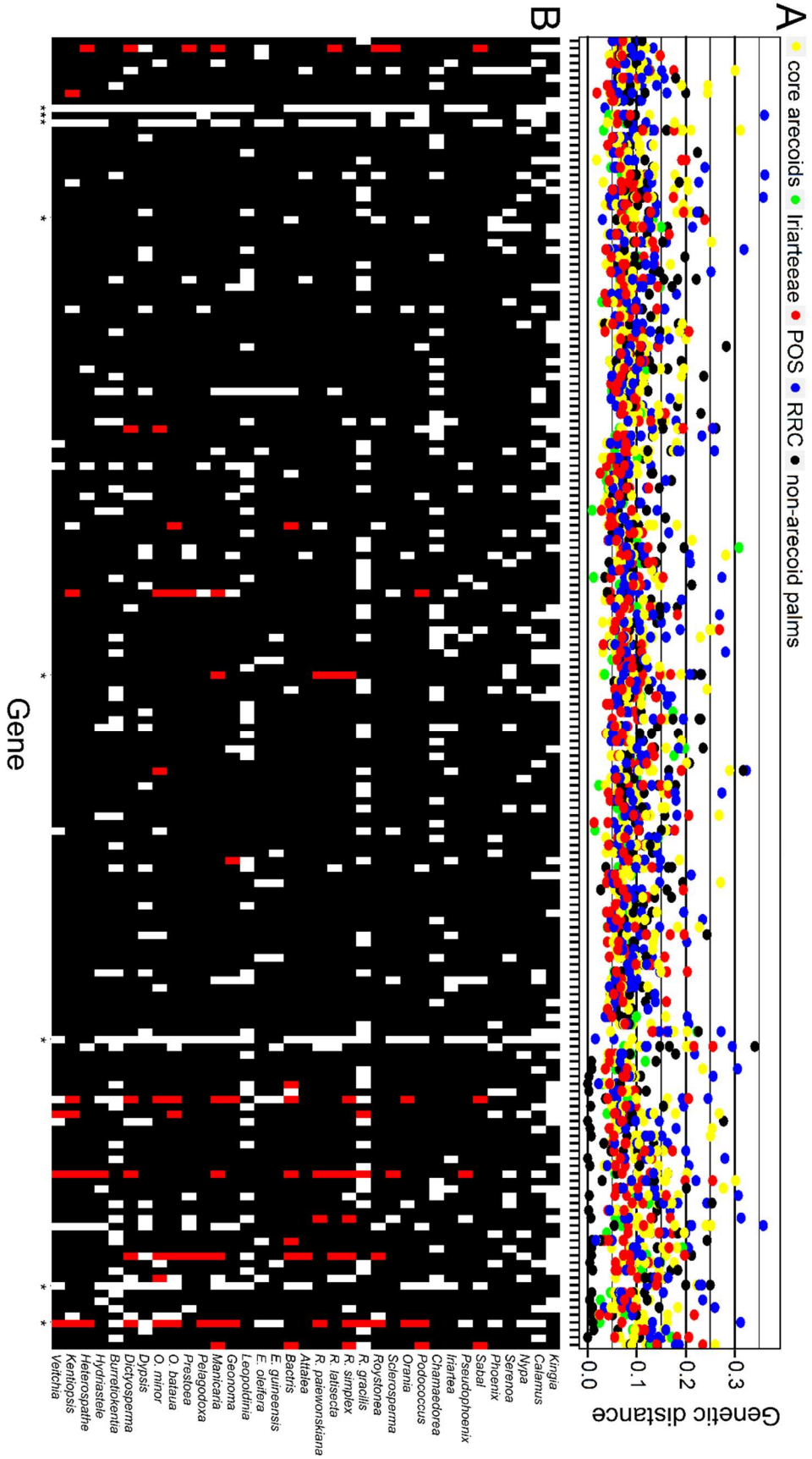
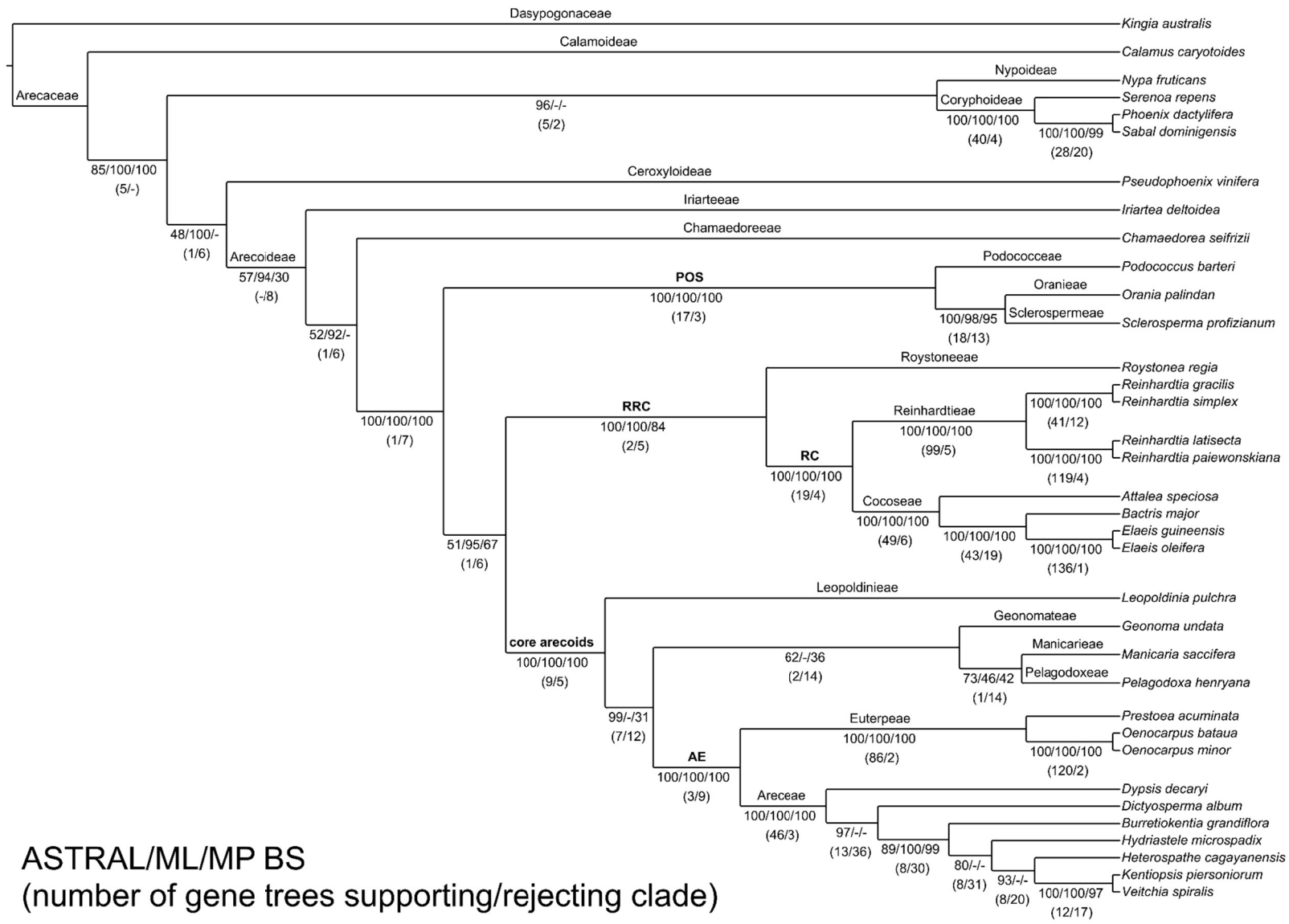

A
core arecoids
Iriarteeae
POS
RRC
non-arecoid palms
Genetic distance
0.0
0.1
0.2
0.3
B
Gene
Kingia
Calamus
Nypa
Serenoa
Phoenix
Sabal
Pseudophoenix
Iriartea
Chamaedorea
Podococcus
Orania
Sclerosperma
Roystonea
R. gracilis
R. simplex
R. latisecta
R. palewonskiana
Attalea
Bactris
E. guineensis
E. oleifera
Leopoldinia
Geonoma
Manicaria
Pelagodoxa
Prestoea
O. bataua
O. minor
Dypsis
Dictyosperma
Burretiokentia
Hydriastele
Heterospathe
Kentiopsis
Veitchia

**Figure 3.3. (Next page)** Bootstrap consensus tree recovered in the ASTRAL analysis of 168 nuclear genes. Labels above branches = family, subfamily, tribe, and major clade (boldface). Numbers below branches = bootstrap support from the ASTRAL/concatenated ML/concatenated MP analyses. The number of gene trees that support the ASTRAL clade (monophyletic) or reject the clade (polyphyletic) with a bootstrap value ≥ 75 are in parentheses. A dash (–) indicates a clade not recovered in the ML/MP analyses or no genes trees with a bootstrap value of ≥ 75. Major clades: AE (Areceae + Euterpeae); core arecoids (Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae); POS (Podococceae, Oranieae, and Sclerospermeae); RC (Reinhardtieae + Cocoseae); and RRC (Roystoneeae, Reinhardtieae, and Cocoseae).




ASTRAL/ML/MP BS
(number of gene trees supporting/rejecting clade)

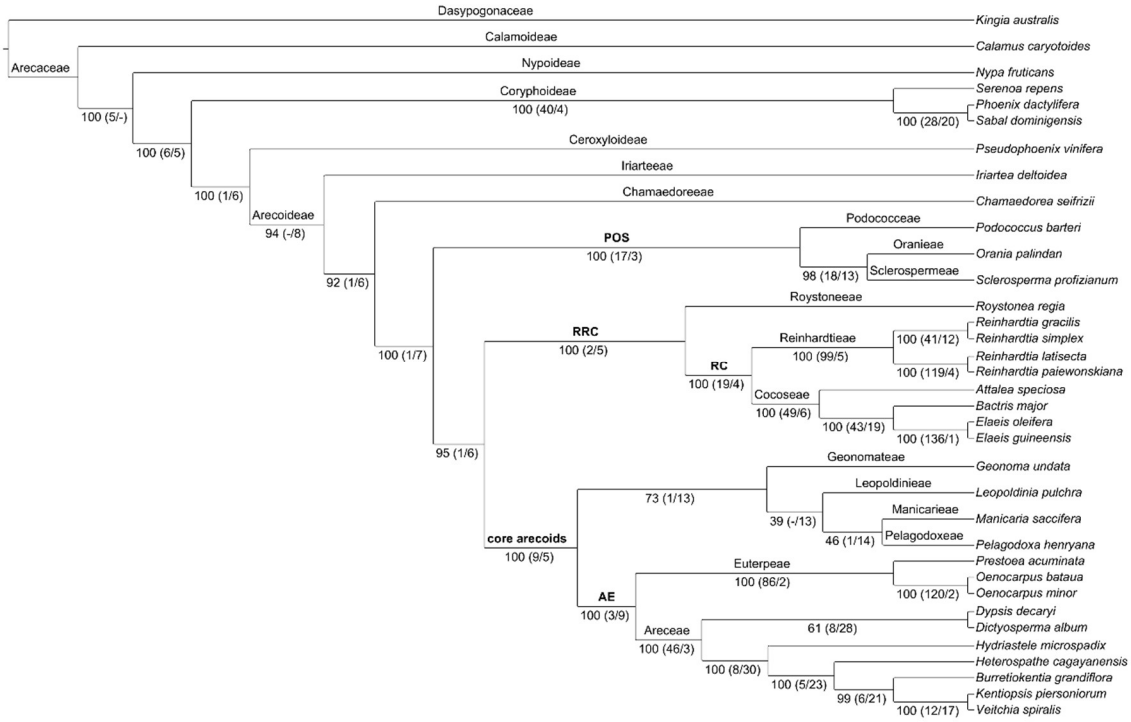



**Figure 3.4.** Species tree from the ML concatenated analysis of the 168 nuclear genes. The Bayesian analysis had identical topology for the tribal relationships (posterior probabilities > 0.95). Labels above the branches = family, subfamily, tribe, and major clade (bold). Labels below branches = bootstrap support. Numbers in parentheses = gene trees supporting (monophyletic) or rejecting (polyphyletic) the clade with a bootstrap value ≥ 75; a dash (–) indicates no genes trees with a bootstrap value of ≥ 75. Major clades: AE (Areceae + Euterpeae), core arecoids (Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae), POS (Podococceae, Oranieae, and Sclerospermeae), RC (Reinhardtieae + Cocoseae), and RRC (Roystoneeae, Reinhardtieae, and Cocoseae).

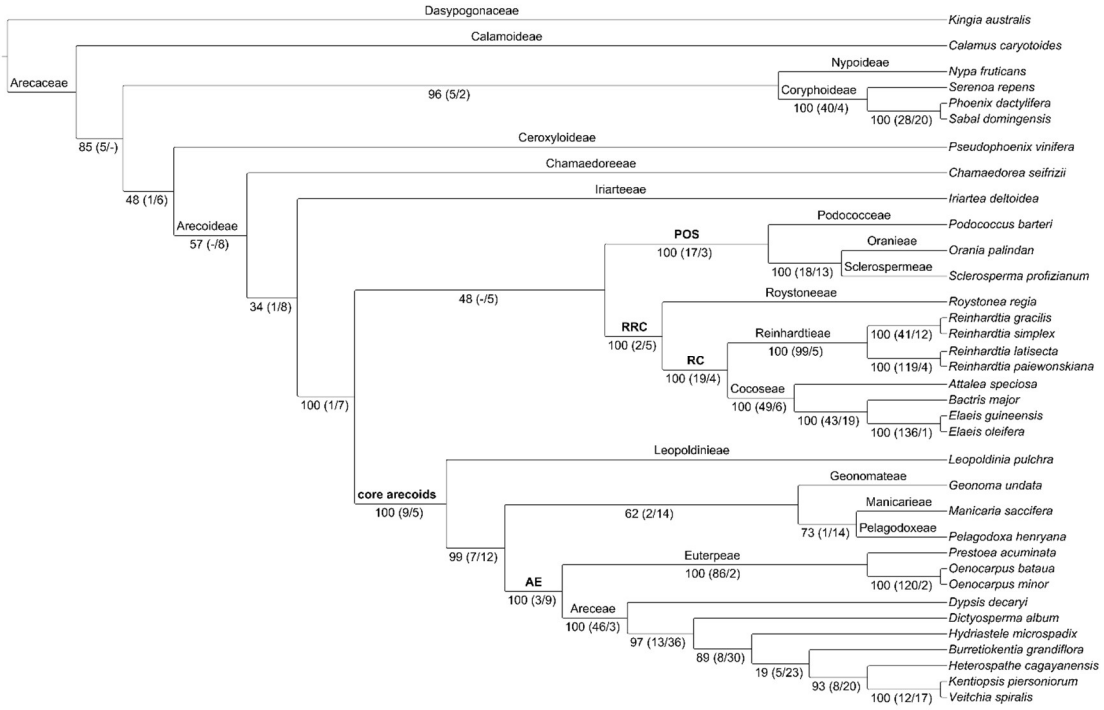



**Figure 3.5.** Species tree from the ASTRAL analysis of the best gene trees of the 168 nuclear genes. Labels above the branches = family, subfamily, tribe, and major clade (boldface font); labels below branches = bootstrap support; numbers in parentheses = gene trees supporting (monophyletic) or rejecting (polyphyletic) the clade with a bootstrap value ≥ 75; a dash (–) indicates no genes trees with a bootstrap value of ≥ 75. Major clades: AE (Areceae + Euterpeae), core arecoids (Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae), POS (Podococceae, Oranieae, and Sclerospermeae), RC (Reinhardtieae + Cocoseae), and RRC (Roystoneeae, Reinhardtieae, and Cocoseae).

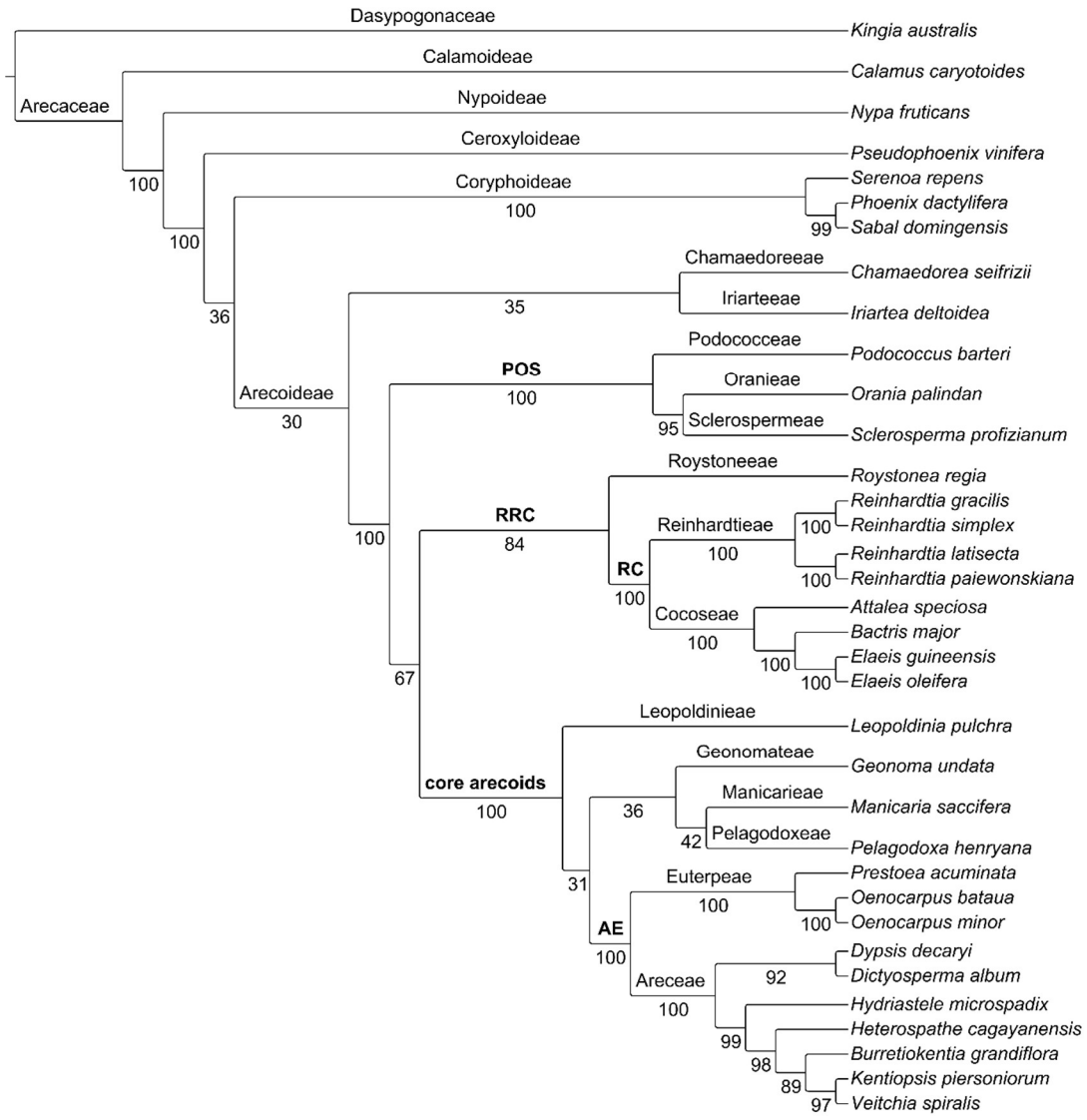



**Figure 3.6.** Species tree (most parsimonious) from the MP concatenated analysis of the 168 nuclear genes. Labels above the branches = family, subfamily, tribe, and major clade (boldface font); labels below branches = bootstrap support. Major clades: AE (Areceae + Euterpeae), core arecoids (Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae), POS (Podococceae, Oranieae, and Sclerospermeae), RC (Reinhardtieae + Cocoseae), and RRC (Roystoneeae, Reinhardtieae, and Cocoseae).

# CHAPTER 4

# SYSTEMATICS OF THE PALM SUBFAMILY ARECOIDEAE (ARECACEAE): PHYLOGENOMICS AND IMPLICATIONS FOR HISTORICAL BIOGEOGRAPHY[3]

[3] Comer JR, Zomlefer WB, Barrett CF, Davis JI, Stevenson DW, Heyduk K, Leebens-Mack J. To be submitted to *Taxon*.

**Abstract**

Arecaceae, the palm family, are distributed throughout the tropics and include species of significant economic value (coconut, date palms, and oil palms). Recent phylogenomic studies have provided new insights into the evolution of subfamily Arecoideae. To evaluate proposed historical biogeographic hypotheses, two phylogenies—one based on chloroplast (114 genes) data, and one on nuclear (168 genes) data—were dated using a penalized likelihood method. The resulting chronograms were used for ancestral area analyses implementing the dispersal-extinction-cladogenesis model. Age estimates were consistent with previous reports. Two radiation events were inferred within subfamily Arecoideae: the first during the initial diversification of the subfamily, and the second within the core arecoids (tribes Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae). North America was inferred as the ancestral area of subfamily Arecoideae and the area of most of the early diversification. Ancestral area analyses provided support for the dispersal of tribe Oranieae from Madagascar to the Indopacific by rafting on India. Stochastic mapping simulations provided some examples of range shifts occurring within time frames that were consistent with this "out of India" hypothesis.

**Introduction**

Members of the palm family Arecaceae are easily recognized by their "woody" growth (from cell expansion and secondary wall thickening), plicate leaves, and a two-keeled bract (prophyll) enclosing a developing inflorescence (Dransfield & al., 2008; Moore, 1973; Uhl & Dransfield, 1987). Arecaceae comprise 183 genera (ca. 2600 species) in five subfamilies primarily distributed throughout the tropics to sub-tropics

(Balick, 1988; Dransfield, 2008; Fadini & al., 2009; Peters & al., 2004; Trias-Blasi & al., 2015). Arecoideae is the largest and most diverse subfamily (Dransfield & al., 2008) and comprises 109 genera (1300 species) in 14 tribes, including coconut (*Cocos nucifera*) and oil palm (*Elaeis guineensis*). The subfamily is distributed throughout the tropics with two primary centers of diversity (Fig. 4.1 map inset), the Americas and the Indopacific (Dransfield & al., 2008; Moore, 1973). Members of Arecoideae are distinguished by reduplicately pinnate leaves and unisexual flowers arranged in triads (clusters of two staminate and one pistillate flowers), acervuli (lines of flowers), or derivatives (Dransfield & al., 2008; Moore, 1973; Uhl & Dransfield, 1987).

***Palm systematics and subfamily Arecoideae*—**Moore (1973) outlined a classification scheme for the palms based on inferred evolutionary lines, without assigning formal ranks. The arecoid line was the largest group. Uhl & Dransfield (1987) provided a formal classification for the Arecaceae in *Genera Palmarum.* The arecoid line was split into three subfamilies (Arecoideae, Ceroxyloideae, and Phytelephantoideae) with the caryotoid line submerged into Arecoideae. Arecoideae comprised six tribes (Areceae, Caryoteae, Cocoeae, Geonomeae, Iriarteeae, and Podococceae) characterized by floral clusters in triads or derivatives triads (Uhl & Dransfield, 1987).

The first molecular phylogenetic study of Arecaceae (Uhl & al., 1995) utilized random fragment length polymorphism (RFLP) data from the chloroplast genome. Subsequent studies used one to a few (one to five) molecular markers from both chloroplast (e.g., *rbcL*) and nuclear genomes (e.g., *PRK*; Asmussen & Chase, 2001; Asmussen & al., 2006; Asmussen & al., 2000; Baker & al., 1999; Baker & al., 2011; Cuenca & al., 2008; Cuenca & Dransfield, 2009; Eiserhardt & al., 2011; Hahn, 2002a,

2002b; Lewis & Doyle, 2002; Meerow & al., 2009). These studies provided insights into the relationships within the palms and were the foundation for a revised classification of the palms, and for subfamily Arecoideae in particular (See Table 4.1; Dransfield & al., 2008; Dransfield & al., 2005). Tribe Caryoteae was placed in subfamily Coryphoideae, and tribe Chamaedoreeae (Hyophorbeae *sensu* Uhl & Dransfield [1987]) was transferred from subfamily Ceroxyloideae to Arecoideae. Fourteen tribes were recognized within Arecoideae: Areceae, Chamaedoreeae, Cocoseae, Euterpeae, Geonomateae, Iriarteeae, Leopoldinieae, Manicarieae, Oranieae, Pelagodoxeae, Podococceae, Reinhardtieae, Roystoneeae, and Sclerospermeae.

A common feature of these studies was low support (bsv ≤ 60) due to low phylogenetic signal, particularly at the deeper nodes, making inferences about tribal relationships difficult. However, three trends were supported: 1) subfamily Arecoideae was monophyletic; 2) tribe Chamaedoreeae, Iriarteeae, or a Chamaedoreeae + Iriarteeae clade were often recovered as the earliest diverging lineages within the subfamily; and 3) three major clades were frequently recovered (Podococceae-Oranieae-Sclerospermeae [POS], Roystoneeae-Reinhardtieae-Cocoseae [RRC], and Areceae-Euterpeae-Geonomateae-Leopoldinieae-Manicarieae-Pelagodoxeae [core arecoids]).

The phylogenetic study by Baker & al. (2009) included at least one taxon from every palm genus (except *Tahina*) and all of the available molecular data (nine chloroplast and six nuclear markers). Tribe Iriarteeae was the earliest diverging lineage within subfamily Arecoideae (Fig. 4.1A). The three major clades (core arecoids, POS, and RRC) were recovered as monophyletic, with the POS and core arecoids as sister clades. Roystoneeae was the basal lineage within the RRC clade, and Podococceae was

the basal lineage of the POS clade. Within the core arecoids two clades were recovered: Geonomateae-Leopoldinieae-Manicarieae (Leopoldinieae as the earliest diverging lineage) and Areceae-Euterpeae-Pelagodoxeae (Euterpeae as the basal lineage).

The most recent phylogenetic studies of Arecoideae by Comer & al. (2015) and Comer & al. (in review) included data from 114 chloroplast genes and 168 nuclear genes from representatives for the 14 tribes of Arecoideae (Figs. 4.1B and 4.2B). The results of these studies were generally congruent with Baker & al. (2009; Fig. 4.1A). The plastid phylogeny recovered Chamaedoreeae as the earliest diverging lineage within the subfamily, and the POS clade was sister to the RRC and core arecoid clades. Tribe Oranieae was sister to Podococceae and Sclerospermeae. Relationships within the core arecoids were unresolved except for the Areceae + Euterpeae (AE clade). The nuclear data (Comer & al., in review) did not resolve the relationships between Chamaedoreeae, Iriarteeae, and those within the core arecoids (except the AE clade). Podococceae was recovered as the basal lineage of the POS clade, and Leopoldinieae was the basal lineage of the core arecoids. The AE clade was recovered, as well as a clade comprising of Manicarieae and Pelagodoxeae.

***Biogeography and ancestral areas of Arecoideae tribes*—**The geographic distributions of palms have been of interest due to their tropical–subtropical range and their key role in the ecosystems in these areas (Corner, 1966; Dransfield & al., 2008; Fadini & al., 2009; Gentry, 1988; Moore, 1973; Peters & al., 2004; Tomlinson, 2006). These features, along with a relatively well-documented fossil record, have made the Arecaceae a model family for examining the origin and evolution of tropical rainforests (Baker & Couvreur, 2013a, 2013b; Couvreur & Baker, 2013; Couvreur & al., 2011;

Harley & Baker, 2001; Harley, 2006). In addition, some palms, particularly some taxa in subfamily Arecoideae, have current geographic distributions that are likely the result of long-distance and anthropogenic dispersal (Dransfield & al., 2008; Moore, 1973). For example, tribe Chamaedoreeae is distributed in the Americas and the Mascarene Islands (*Hyophorbe*); *Elaeis oleifera* is native to South America and *E. guineensis* (oil palm) is native to Africa; and coconut (*Cocos nucifera*) has an extant distribution throughout the tropics (Dransfield & al., 2008; Moore, 1973). Within the POS clade, tribes Podococceae and Sclerospermeae are endemic to western equatorial Africa, while tribe Oranieae is primarily distributed in the Malesian area with three species endemic to Madagascar (Fig. 4.1 map inset; Dransfield & al., 2008; Moore, 1973).

Couvreur & al. (2011) and Baker & Couvreur (2013a, 2013b) used the palm phylogeny of Baker & al. (2009) to estimate divergence times and reconstruct the ancestral areas of the family. These studies inferred a Laurasian origin for the palms during the mid-Cretaceous (ca. 100 My) with subfamilies Arecoideae and Ceroxyloideae diverging in North America ca. 78 My (Fig. 4.1A). Arecoideae later dispersed into South America where early tribal lineages diversified. The most recent common ancestor (MRA) of the POS and core arecoids clades was inferred to have originated in Eurasia with subsequent dispersals into Africa, the Indopacific, and South America.

Comer & al. (2015) used their chloroplast phylogeny to test these hypotheses of Baker & Couvreur (2013a). Ancestral area reconstruction yielded alternative hypotheses (the Americas) for the origin of the POS and the core arecoids clades (Fig. 4.1B). Progenitors of the POS clade were inferred to have dispersed from the Americas into Africa and later into the Indopacific (tribe Oranieae). The core arecoids were inferred to

have diversified in South America with possible independent dispersal into the Indopacific by tribes Areceae and Pelagodoxeae. A phylogeny based on nuclear data (Comer & al., in review) generally supported these hypotheses and suggested that the dispersal pattern of tribe Oranieae may be explained by the "out of India" hypothesis, whereby the tribe dispersed from Africa/Madagascar to India and then "rafted" to the Indopacific (see Conti & al., 2002 and Morley, 2003).

***Study objectives*—**The main objective of this study was to test the hypotheses (Fig. 4.1) proposed by Baker & Couvreur (2013a) versus Comer & al. (2015; in review) concerning ancestral areas of subfamily Arecoideae using phylogenies generated from chloroplast (Comer& al., 2015) and nuclear (Comer & al., in review) data. Penalized likelihood analyses were conducted for the chloroplast and nuclear phylogenies to obtain estimates of divergence times. These dated phylogenies were then used for ancestral area reconstruction analyses using the dispersal-extinction-cladogenesis model (DEC; Ree & al., 2005; Ree & Smith, 2008).

**Materials and methods**

***Taxon sampling*—**Forty-five palm species and one outgroup were included (see Table 4.1). Sampling for the chloroplast analysis was increased to include representatives of all five Arecaceae subfamilies in addition to the 14 Arecoideae tribes. Appendix 4.1 lists the vouchers and original publication information.

***Phylogenetic reconstruction*—**The chloroplast data matrix, expanded from Comer & al. (2015), was constructed by concatenating 114 individually aligned genes. RAxML version 8.1.11 (Stamatakis, 2006, 2014) was used with the GTRGAMMA substitution model to search for the best scoring maximum likelihood (ML) tree (rapid

hill-climbing algorithm; Stamatakis & al., 2007) and to conduct 1000 rapid bootstrap replicates ("–f a" option). The ML tree from supermatrix analysis, generated from a 168 nuclear gene data set (Comer & al., in review), was used in the following analyses.

***Fossil calibrations*—**Three fossils were selected for calibration (Baker & Couvreur, 2013a; Couvreur & al., 2011), with minimum (min.) and maximum (max.) ages assigned following Magallón & al. (2015). The oldest known monocot fossil, dated between 110−120 My (Friis & al., 2004) was used to calibrate the root node (Dasypogonaceae/Arecaceae split; min. 100 My; max. 120 My). The oldest unequivocal palm fossil, *Sabalites carolinensis* (Coryphoideae, Berry, 1914), dated to the Late Coniacian/early Santonian (ca. 83.5 My), was used to calibrate the stem node of subfamily Coryphoideae (min. 83.5 My and max. 91.85 My). Srivastava & Srivastava (2014) documented the oldest fossil of tribe Cocoseae (Arecoideae), *Cocos binoriensis*, dating from the Maastrichtian–Danian (65.5−61.7 My; min. 61.7 My; max. 67.87 My). This fossil was used to calibrate the crown node of Cocoseae subtribe Attaleinea (*Attalea, Beccariophoenix*, and *Cocos*) in the chloroplast phylogeny and the crown node of Cocoseae for the nuclear phylogeny, since subtribe Attalinea was represented by *Attalea* alone.

***Divergence time estimation*—**Divergence times were estimated using a semi-parametric penalized likelihood (PL) as implemented in the R (version 3.2.2) package, "ape" (Analyses of Phylogenetics and Evolution; version 3.3) (Paradis & al., 2004; Sanderson, 2002; R Core Team, 2015). This method utilizes a rate-smoothing parameter, lambda (λ), allowing for variation between a fully parametric model (each branch with its own substitution rate; $\lambda = 0$, not clock-like) and a fully nonparametric model (rate

variability minimized between adjacent branches; λ = 1, clock-like). Optimal λ values were estimated for both phylogenies using the cross-validation method (chronoPL) for a series of λ values ranging from -1–1. Full PL analyses were implemented in ape (chronos) to obtain age point estimates for the ML trees with the fossil calibrations. Minimum and maximum age ranges were estimated from 100 bootstrap replicates for each phylogeny with the same fossil calibrations and optimal λ values used for the age point estimates (Magallón & al., 2015). Branch lengths were not modified as the ancestral area reconstruction implemented required fully bifurcating trees.

***Ancestral area reconstruction*—**The chronograms from the PL analyses were used for ancestral area reconstruction under the dispersal-extinction-cladogenesis model (DEC) as implemented by Lagrange version 20130526 (Ree & al., 2005; Ree & Smith, 2008) and the R package BioGeoBears (BioGeography with Bayesian [and Likelihood] Evolutionary Analysis in R Scripts) version 0.2.1 (Matzke, 2013; Matzke, 2014). Geographic coding is included in the Lagrange M1 input/output file in the supplemental material. Seven geographic ranges were designated (see Fig. 4.1) based on Couvreur & al. (2011), Baker & Couvreur (2013a), and Comer & al. (2015), and taxa were coded according to their current geographic ranges (Dransfield & al., 2008). Representatives of subfamilies Calamoideae, Ceroxyloideae, and Coryphoideae were coded based on the inferred ancestral areas of Baker & Couvreur (2013a). Maximum concurrent range size was restricted to two geographic areas for Lagrange analyses (Baker & Couvreur, 2013a), and three concurrent ranges for BioGeoBears analyses (smallest allowable size based on geographic coding for taxa; i.e., some taxa concurrently occupy three geographic ranges). Two models were tested in both Lagrange and BioGeoBears M0 (DEC with equal

dispersal probabilities between all areas) and M1 (DEC with time-stratified dispersal probabilities for six discrete time frames). Five of the time periods, between 100 My to the present, followed Couvreur & al. (2011) and Baker & Couvreur (2013a). A sixth time period was added to accommodate results of the PL analyses covering 120−100 My. Dispersal probabilities were based on Morley (2003) and Scotese (1991) (see supplemental material Lagrange M1 input/output for dispersal matrices). Three additional nested models of the time-stratified model were tested in BioGeoBears: M1J (M1 model with found-event speciation (parameter "j"), i.e. long distance dispersal, M1A (M1 model with variable range-switching rates (parameter "a"), allowing for lineages to switch ranges without leaving descendants in the ancestral area), and M1AJ (M1 model with founder-event speciation and variable range switching rates). BioGeoBears was used to run 100 stochastic mapping simulations under the M1A model.

**Results**

Results are summarized in Tables 4.2 and 4.3, Figs. 4.2–4.7, and the supplemental materials deposited to DRYAD (DOI: to be deposited upon manuscript acceptance). The expanded chloroplast data set a matrix (71,861 bp) and the nuclear data set (Comer & al., in review) are also available from DRYAD (DOI: *** and DOI: ***, respectively; to be deposited upon manuscript acceptance).

Topologies for the two phylogenies (Figs. 4.2) are largely congruent, differing in the relationships between the core arecoid tribes, except for Areceae and Euterpeae (AE clade) and the basal lineage of the POS clade. The ML recovered by RAxML for the chloroplast data set placed Calamoideae as the basal lineage within the Arecaceae and strongly supported (bsv 100) Nypoideae as sister to Arecoideae-Ceroxyloideae-

Coryphoideae. Coryphoideae was sister to an Arecoideae + Ceroxyloideae clade. The four palm subfamilies represented by multiple taxa were all strongly supported as monophyletic (bsv 100). Within subfamily Arecoideae, tribe Iriarteeae was supported as the basal lineage, and Chamaedoreeae was moderately supported (bsv 76) as sister to the three major clades (core arecoids, POS, and RRC). The monophyletic POS clade (bsv 88) was strongly supported as sister to a core arecoid + RRC clade. Podococceae and Sclerospermeae were supported as sister tribes (bsv 71). Tribes Reinhardtieae and Cocoseae were strongly supported as monophyletic (bsv 100 and 99 respectively). The sister relationship of tribes Reinhardtieae and Cocoseae was strongly supported (bsv 100) but monophyly of the RRC clade was not supported (bsv 49). The core arecoid clade was strongly supported as monophyletic, as were tribes Areceae and Euterpeae. Relationships between the core arecoid tribes were not well supported, with the exception of the AE clade (bsv 91).

Chronograms from the penalized likelihood (PL) analyses are shown in Figs. 4.3 with age ranges from the bootstrap replicates analyses, and Table 4.2 provides a comparison of PL point estimates of node ages for the chloroplast and nuclear phylogenies. In general, divergence time range estimates for shared clades overlapped in the two phylogenies. Stem node point estimates for the Arecaceae were ca. 116 My (chloroplast) and 106 My (nuclear) but age range estimates overlapped. The estimated times of divergence for subfamilies Ceroxyloideae and Arecoideae were ca. 83 My (chloroplast) and 89 My (nuclear), and the crown node age of the Arecoideae was dated at ca. 81 My chloroplast) and 86 My (nuclear).

Table 4.3 includes summary statistics from the ancestral areas reconstruction (including likelihood values). For both Lagrange and BioGeoBears analyses, the time-stratified dispersal model (M1) was favored over the null (M0) model with all dispersal probabilities equal. Inferred ancestral areas with relative probabilities ≥ 0.10 were mapped onto the chloroplast and nuclear phylogenies (Figs. 4.4 and 4.5). The M1J model conferred a significantly better ($P < 0.5$) likelihood score to the chloroplast phylogeny. The M1A model conferred a significantly better ($P < 0.5$) likelihood score to both phylogenies (Figs. 4.4 and 4.5). Inferred ancestral ranges within two likelihood scores are provided in the Lagrange input/output files (supplemental material) along with the relative probability of the inferred ancestral areas.

North America was inferred as the ancestral area of subfamily Arecoideae with all analyses and was generally favored as the ancestral range for the most recent common ancestor of the POS, core arecoid, and RRC clades. The ancestral area of the POS clade inferred by Lagrange was North America but South America also had a relative probability greater than 0.1 for the chloroplast phylogeny. BioGeoBears favored Eurasia (with Africa having the second highest relative probability; see Fig. 4.4) for the chloroplast phylogeny and Africa for the nuclear phylogeny (Fig. 4.5). The core arecoid ancestral area was inferred as North America or North and South America in the Lagrange analyses; BioGeoBears favored North America and Eurasia for the chloroplast phylogeny and North America, South America, and the Pacific for the nuclear phylogeny. Figure 4.6 shows examples of the stochastic mapping and along with the remaining 99 replicates (DRYAD DOI: to be deposited upon manuscript acceptance) represent possible dispersal histories.

## Discussion

***Implications for systematics*—**The ML chloroplast phylogeny (Fig. 4.2A) generated for this study is largely congruent with the chloroplast ML tree from Comer & al. (2015), strongly supporting a monophyletic subfamily Arecoideae. The main topological difference between the two chloroplast phylogenies was the recovered basal lineage. Comer & al. (2015) recovered Chamaedoreeae as the basal lineage and Iriarteeae as sister to the remaining tribes (bsv 100). The expanded chloroplast phylogeny presented in this study (Fig. 4.2A) recovered Iriarteeae as the basal lineage and Chamaedoreeae as sister to the rest of the subfamily (bsv 76).

The PL analyses (Table 4.2; Fig. 4.3) of both the nuclear and chloroplast phylogenies support rapid divergences among the deeper nodes of subfamily Arecoideae (and Ceroxyloideae), as shown also by the age estimates of Baker & Couvreur (2013a; see Table 4.2). Following the split of subfamily Coryphoideae from the MRCA of Arecoideae and Ceroxyloideae, 2 to 3 My passed before each of the following divergence events: 1) subfamilies Arecoideae and Ceroxyloideae; 2) tribe Iriarteeae from the rest Arecoideae; 3) Chamaedoreeae from the three major clades (POS, RRC, core arecoid) 4) POS clade from the RRC + core arecoid clades; and 5) RRC and core arecoid clades. Age estimates for these nodes are also characterized by a wider range inferred from the bootstrap replicate trees (blue bars in Figs. 4.3). While the time intervals between divergence events within the core arecoid clade are not as rapid as the earliest divergences within the subfamily, they exhibit relatively wide age ranges, particularly among the earliest nodes of the clade (chloroplast ± 20 My; nuclear ± 10 My). Support for most of these nodes was low (bsv < 70) in both phylogenies.

While age estimates were relatively consistent, some nodes had larger differences in the estimated ages between the chloroplast and nuclear phylogenies (Table 4.2). A leading contributor to these inconsistencies is the varying placement of the fossil calibrations (Magallón & al., 2013; Magallón, 2014).The number of substitutions per site did not vary significantly for any lineages between the phylogenies, and although the overall longer branch lengths of the nuclear phylogeny were proportional to the branch lengths of the chloroplast phylogeny (Fig. 4.7). The fossil calibration used within tribe Cocoseae did not result in large age differences (ca. ± 3 My, Table 4.2) estimated for Cocoseae and the other members of the RRC clade. Relative topological placement of taxa within the phylogeny altered age estimates. For example, Podococceae and Oranieae (POS clade) differed in placement and estimated age between phylogenies but the relative position of Sclerospermeae and age estimate were consistent between phylogenies. The nodes with the largest differences (> 5 My) in age estimates were characterized by relatively long branches dividing into shorter branches (Fig. 4.7). This type of node has been documented as particularly sensitive to different model specifications and distance from the fossil calibration, resulting in large age estimation ranges (Magallón & al., 2013; Magallón, 2014).

***Historical biogeography*—**Diversification of the earliest lineages of Arecaceae was inferred as dating from the mid-Cretaceous (105−101 My) in Laurasia. The MRA of subfamilies Arecoideae and Ceroxyloideae was estimated to have diverged from Coryphoideae ca. 85−92 My in North America (BioGeoBears) or Laurasia (Lagrange). Subsequently Arecoideae and Ceroxyloideae diverged in North America ca. 83−89 My. These results are consistent with Baker & Couvreur (2013a) and Comer & al. (2015).

Diversification of subfamily Arecoideae was estimated to have originated with Iriarteeae ca. 81–86 My, followed by Chamaedoreeae ca. 79–83 My. Results of the ancestral area reconstruction favored North America as the center for the early diversification of Arecoideae. Progenitors of the POS clade likely diverged from the MRCA of the RRC and core arecoids clades in North America 77–81 My, with the RRC and core arecoids diverging ca. 75–78 My. These results suggest a more restricted ancestral area as the early center of diversification for subfamily Arecoideae in the Americas than suggested by Comer & al. (2015).

Tribe Cocoseae likely began diversifying in North America during the Late Cretaceous (73–74 My), with subsequent dispersals into South America and later to Africa and Madagascar. Diversification of the core arecoids began during the early Eocene (50–56 My) in North America, North America–Eurasia, or in South America and the Pacific Ocean Islands. Tribes Pelagodoxeae and Areceae were inferred to have dispersed independently into the Indopacific.

The POS clade was estimated to have begun diversifying during the mid-Eocene (38–41 My) but the analyses resulted in different inferences for dating the origin for the POS clade. Previous ancestral area analysis (Comer & al., 2015) suggested that the POS clade began diversifying in Africa, with tribe Oranieae later dispersing into the Indopacific. Comer & al. (in review) proposed that following the dispersal into Africa, Oranieae may have dispersed to its current range in Madagascar and the Indopacific via the “out of India” hypothesis. Support for this hypothesis would require the presence of the POS clade in Africa prior to India splitting from Africa and Madagascar ca. 72–66 My (Morley, 2003). These age estimates for the stem lineage of the POS clade predate

the split splitting of India from Africa and Madagascar, and the BioGeoBears analysis of the nuclear phylogeny (Figs. 4.5) suggests the presence of the POS clade in Africa prior to diversification. While Eurasia had the highest relative probability in the BioGeoBears analysis of the chloroplast data (Fig. 4.4), Africa also had a relatively high probability. Although the Lagrange analyses suggest that the POS clade began diversifying prior to dispersal from North America (relative probability > 0.10), there were inferences of Africa as the ancestral area within two log-likelihood values (see Lagrange files in the supplemental materials). Differences in inferred ancestral areas are likely due, at least in part, to topological differences between phylogenies since the ranges of the daughter lineages are factored into the probability for the ancestral area (e.g. placement of Oranieae or Podococceae; see Ree & al., 2005; Ree & Smith, 2008; Matzke, 2014).

Although ancestral area reconstructions (Figs. 4.4 and 4.5) did not elucidate the path or timing of dispersal events, stochastic mapping simulations provided some insight (simulations available at DOI: ***; to be deposited upon acceptance of manuscript). Stochastic simulations showed the POS clade reaching Africa and Madagascar prior to the Maastrichtian 71–66 My (Fig. 4.6). The earliest fossil evidence for the POS clade, triporate pollen from India dating from the Eocene (56–34 My), has been compared to pollen of extant *Sclerosperma* (Sclerospermeae) (Misra & al., 1996). However, Harley (2006) questions the assignment of this pollen to *Sclerosperma*. Fossilized triporate pollen from Senegal, Africa (late Eocene to early Miocene, 38–23 My) comprises unequivocal evidence of the POS clade (Harley and Baker, 2001; Harley, 2006; Medus, 1975). A *Sclerosperma* leaf compression dating from the mid-Miocene (ca. 13 My) was described from central Africa, suggesting that ancestors of the clade may have occupied a

broader range in Africa (Lakhanpal, 1966; Pan & al., 2006). Some simulations (e.g., Fig. 4.6) also showed range shifts from Africa and/or Madagascar to include Eurasia around the time the POS began diversifying (mid-Eocene, 41–38 My), when the Indian and Asian plates collided (50–39 My), making dispersals from India to Southeast Asia possible (Morley, 2003).

***Conclusions*—**Both the chloroplast and nuclear phylogenies yielded similar age estimates to those of Baker & Couvreur (2013a). Our results suggest two main radiation events within the history of subfamily Arecoideae. The first occurred with the diversification of the Arecoideae during the Upper Cretaceous (86–80 My), and the second within the core arecoids in the early Eocene (56–50 My). Results of the ancestral area analyses provided further support for Arecoideae and Ceroxyloideae diverging in North America. North America, rather than the Americas (Comer & al., 2015) or South America (Baker & Couvreur, 2013a), was also inferred as the center for much of the early diversification within the Arecoideae. Ancestral area analyses supported the out of India hypothesis for tribe Oranieae, and stochastic mapping simulations returned scenarios that fit the timing for this proposed dispersal pathway (Fig. 4.6).

**Acknowledgements**

The authors thank Larry Noblick and Patrick Griffith (Montgomery Botanical Center) and Brett Jestrow (Fairchild Tropical Botanic Garden) for assistance collecting many of the palms used in this study. They are also grateful to Anders Lindstrom and Kampon Tansacha (Nong Nooch Tropical Botanical Garden) for hosting J.R.C. and allowing him to sample the palm collection. Several species of *Reinhardtia* were kindly provided by Jeff Marcus (Floribunda Palms and Exotics), and Thomas Couvreur supplied

tissue for *Podococcus* and *Sclerosperma*. Funding was provided by the National Science Foundation (DEB-083009, J. H. Leebens-Mack, PI and W. B. Zomlefer, co-PI). Additional travel funds were provided to J.R.C. by the Department of Plant Biology, University of Georgia (Palfrey Grant for Graduate Student Research).

**Literature cited**

**Asmussen, C.B., & Chase, M.W.** 2001. Coding and noncoding plastid DNA in palm systematics. *Amer. J. Bot.* 88: 1103–1117.

**Asmussen, C.B., Dransfield, J., Deickmann, V., Barfod, A.S., Pintaud, J.-C., & Baker, W.J.** 2006. A new subfamily classification of the palm family (Arecaceae): Evidence from plastid DNA phylogeny. *Bot. J. Linn. Soc.* 151: 15–38.

**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. Pp. 525–537 in: Wilson, K.L., & Morrison, D.A., (eds), *Monocots: Systematics and Evolution*. Collingwood VIC : CSIRO Publishing.

**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 chloroplast DNA sequences from the *trnL-trnF* region. *Pl. Syst. Evol.* 219: 111–126.

**Baker, W.J., & Couvreur, T.L.P.** 2013a. Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical biogeography. *J. Biogeogr.* 40: 274–285.

**Baker, W.J., & Couvreur, T.L.P.** 2013b. Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. II. Diversification history and origin of regional assemblages. *J. Biogeogr.* 40: 286–298.

**Baker, W.J., Norup, M.V., Clarkson, J.J., Couvreur, T.L.P., Dowe, J.L., Lewis, C.E., Pintaud, J.-C., Savolainen, V., Wilmot, T., & Chase, M.W.** 2011. Phylogenetic relationships among arecoid palms (Arecaceae: Arecoideae). *Ann. Bot.* 108: 1417–1432.

**Baker, W.J., Savolainen, V., Asmussen-Lange, C.B., Chase, M.W., Dransfield, J., Forest, F., Harley, M.M., Uhl, N.W., & Wilkinson, M.** 2009. Complete generic-level phylogenetic analyses of palms (Arecaceae) with comparisons of supertree and supermatrix approaches. *Syst. Biol.* 58: 240–-256.

**Balick, M.J.** 1988. *Jessenia* and *Oenocarpus* : Neotropical oil palms worthy of domestication. *F. A. O. Pl. Prod. Protect. Pap.* 88: 1–-191.

**Berry, E.W.** 1914. The Upper Cretaceous and Eocene floras of South Carolina and Georgia. *Proff. Pap. U.S. Geol. Surv.* 84: 1–-200.

**Comer, J.R., Zomlefer, W.B., Barrett, C.F., Davis, J.I., Stevenson, D.W., Heyduk, K., & Leebens-Mack, J.H.** 2015. Resolving relationships within the palm subfamily Arecoideae (Arecaceae) using plastid sequences derived from next-generation sequencing. *Amer. J. Bot.* 102: 888–-899.

**Conti, E., Eriksson, T., Schönenberger, J., Sytsma, K.J., & Baum, D.A.** 2002. Early Tertiary out-of-India dispersal of Crypteroniaceae: evidence from phylogeny and molecular dating. *Evolution* 56: 1931–1942.

**Corner, E.J.H.** 1966. *The natural history of Palms.* London: Weidenfeld and Nicolson

**Couvreur, T.L.P., & Baker, W.J.** 2013. Tropical rain forest evolution: palms as a model group. *B. M. C Biol.* 11: 48.

**Couvreur, T.L.P., Forest, F., & Baker, W.J.** 2011. Origin and global diversification patterns of tropical rain forests: Inferences from a complete genus-level phylogeny of palms. *B. M. C. Biol.* 9: 44.

**Cuenca, A., Asmussen-Lange, C.B., & Borchsenius, F.** 2008. A dated phylogeny of the palm tribe Chamaedoreeae supports Eocene dispersal between Africa, North and South America. *Molec. Phylogen. Evol.* 46: 760–775.

**Cuenca, A., & Dransfield, J.B.** 2009. Phylogeny and evolution of morphological characters in tribe Chamaedoreeae (Arecaceae). *Taxon* 58: 1092–1110.

**Dransfield, J., Uhl, N.W., Asmussen, C.B., Baker, W.J., Harley, M.M., & Lewis, C.** 2008. *Genera Palmarum: The evolution and classification of palms*. Richmond: Royal Botanical Gardens, Kew.

**Dransfield, J., Uhl, N.W., Asmussen, C.B., Baker, W.J., Harley, M.M., & Lewis, C.E.** 2005. A new phylogenetic classification of the palm family, Arecaceae. *Kew Bull.* 60: 559–569.

**Eiserhardt, W.L., Pintaud, J.-C., Asmussen-Lange, C., Hahn, W.J., Bernal, R., Balslev, H., & Borchsenius, F.** 2011. Phytogeny and divergence times of Bactridinae (Arecaceae, Palmae) based on plastid and nuclear DNA sequences. *Taxon* 60: 485–498.

**Fadini, R.F., Fleury, M., Donatti, C.I., & Galetti, M.** 2009. Effects of frugivore impoverishment and seed predators on the recruitment of a keystone palm. *Acta Oecol.* 35: 188–196.

**Friis, E.M., Pedersen, K.R., & Crane, P.R.** 2004. Araceae from the Early Cretaceous of Portugal: Evidence on the emergence of monocotyledons. *Proc. Natl. Acad. Sci. U.S.A.* 101: 16565–16570.

**Gentry, A.H.** 1988. Changes in plant community diversity and floristic composition on environmental and geographical gradients. *Ann. Missouri Bot. Gard.* 75: 1–34.

**Hahn, W.J.** 2002a. A molecular phylogenetic study of the Palmae (Arecaceae) based on *atpB*, *rbcL*, and *18S* nrDNA sequences. *Syst. Biol.* 51: 92–112.

**Hahn, W.J.** 2002b. A phylogenetic analysis of the Arecoid Line of palms based on plastid DNA sequence data. *Molec. Phylogen. Evol.* 23: 189–204.

**Harley, M.M., & Baker, W.J.** 2001. Pollen aperture morphology in (Arecaceae): Application within phylogenetic analyses, and a summary of the fossil record of palm-like pollen. *Grana* 40: 45–77.

**Harley, M.M.** 2006. A summary of fossil records for Arecaceae. *Bot. J. Linn. Soc.* 151: 39–67.

**Lakhanpal, R.** 1966. Some middle Tertiary plant remains from South Kivu, Congo. *Ann. Mus. Roy. Afrique Centr. Ser. 8* 52: 21–30.

**Lewis, C.E., & Doyle, J.J.** 2002. A phylogenetic analysis of tribe Areceae (Arecaceae) using two low-copy nuclear genes. *Pl. Syst. Evol.* 236: 1–17.

**Magallón, S., Hilu, K.W., & Quant, D.** 2013. Land plant evolutionary timeline: Gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *Amer. J. Bot.* 100: 556–573.

**Magallón, S.** 2014. A review of the effect of relaxed clock method, long branches, genes, and calibrations in estimation of angiosperm age. *Bot. Sci.* 92: 1–22.

**Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., & Hernández-Hernández, T.** 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* 207: 437–453.

**Matzke, N.J.** 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers Biogeog.* 5: 242–248.

**Matzke, N.J.** 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Syst. Biol*. 63: 951–970.

**Medus, J.** 1975. Palynologie de sédiments Tertiaires du Sénégal méridional. *Pollen & Spores* 17: 545–601.

**Meerow, A.W., Noblick, L., Borrone, J.W., Couvreur, T.L.P., Mauro-Herrera, M., Hahn, W.J., Kuhn, D.N., Nakamura, K., Oleas, N.H., & Schnell, R.J.** 2009. Phylogenetic analysis of seven WRKY genes across the palm subtribe Attaleinae (Arecaceae) Identifies *Syagrus* as Sister Group of the Coconut. *PLoS ONE* 4: e7353.

**Misra, B.K., Singh, A., & Ramanujam, C.G.K.** 1996. Trilatiporate pollen from Indian Palaeogene and Neogene sequences: evolution, migration and continental drift. *Rev. Palaeobot. Palynol.* 91: 331–352.

**Moore, H.E.** 1973. The major groups of palms and their distribution. *Gentes Herbarum* 11: 27–141.

**Morley, R.J.** 2003. Interplate dispersal paths for megathermal angiosperms. *Perspect. Pl. Ecol. Evol. Syst.* 6: 5–20.

**Pan, A.D., Jacobs, B.F., Dransfield, J., & Baker, W.J.** 2006. The fossil history of palms (Arecaceae) in Africa and new records from the Late Oligocene (28–27 Mya) of north-western Ethiopia. *Bot. J. Linn. Soc.* 151: 69–81.

**Paradis, E., Claude, J., & Strimmer, K.** 2004. APE: Analyses of phylogenetics and volution in R language. *Bioinformatics* 20: 289–290.

**Peters, H.A., Pauw, A., Silman, M.R., & Terborgh, J.W.** 2004. Falling palm fronds structure Amazonian rainforest sapling communities. *Proc. Roy. Soc. London Ser B, Biol. Sci.* 271: S367–S369.

**Ree, R.H., Moore, B.R., Webb, C.O., & Donoghue, M.J.** 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 59: 2299–2311.

**Ree, R.H., & Smith, S.A.** 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* 57: 4–14.

**Sanderson, M.J.** 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Molec. Biol. Evol.* 19: 101–109.

**Scotese, C.R.** 1991. Palaeogeography and paleoceanography of Tethys Jurassic and Cretaceous plate tectonic reconstructions. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 87: 493–501.

**Srivastava, R., & Srivastava, G.** 2014. Fossil fruit of Cocos L. (Arecaceae) from Maastrichtian-Danian sediments of central India and its phytogeographical significance. *Acta Palaeobot.* 54: 67–75.

**Stamatakis, A.** 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.

**Stamatakis, A.** 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

**Stamatakis, A., Blagojevic, F., Nikolopoulos, D.S., & Antonopoulos, C.D.** 2007. Exploring New Search Algorithms and Hardware for Phylogenetics: RAxML meets the IBM cell. *J. VLSI Signal Processing Syst. Signal Image Video Technol.* 48: 271–286.

**R Core Team.** 2015. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

**Tomlinson, P.B.** 2006. The uniqueness of palms. *Bot. J. Linn. Soc.* 151: 5–14.

**Trias-Blasi, A., Baker, W.J., Haigh, A.L., Simpson, D.A., Weber, O., & Wilkin, P.** 2015. A genus-level phylogenetic linear sequence of monocots. *Taxon* 27: 30.

**Uhl, N.W., & Dransfield, J.** 1987. *Genera palmarum: a classification of palms based on the work of Harold E.Moore, Jr.* Lawrence: Allen Press.

**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. Pp. 623–662 in: P.J. Rudall, P.J. Cribb, D.F. Cutler, & C.J. Humphries, (ed), *Monocotyledons: Systematics and Evolution*. Kent: Whitstable Litho Printers.

**Appendix 4.1.** Taxa included in this study with voucher information, in the following order: subfamily (tribe); ***species***; *voucher specimen* (herbarium); GenBank accession number (chloroplast). Chloroplast data from Comer & al. (2015) with GenBank accession numbers; nuclear data from Comer & al. (in review), Sequence Read Archive study accession: SRP061467, unless otherwise indicated. For taxa with different vouchers for chloroplast and nuclear data, information for nuclear data is presented first. * = Nuclear data from Heyduk & al. (2015).

---

Arecoideae (Areceae); ***Areca vestiaria*** Giseke; *Zomlefer 2310* (FTG, NY); KP221698. ***Burretiokentia grandiflora*** Pintaud & Hodel*; Comer 297* (BKF); KP221702. ***Dictyosperma album**** (Bory) H. L. Wendl. & Drude ex Scheff.*; Noblick 5069* (FTG); KP221703. ***Dypsis decaryi*** (Jum.) Beentje & J. Dransf.*; Noblick 5056* (FTG); KP221705. ***Drymophloeus litigiosus*** (Becc.) H. E. Moore; *Comer 299* (BKF); KP221704. ***Heterospathe cagayanensis*** Becc.*; Kyburz s.n.* [31 May 1995] (FTG); KP221707. ***Hydriastele microspadix*** (Warb. ex K. Schum. & Lauterb.) Burret*; Noblick 5667* (FTG); KP221708. ***Kentiopsis piersoniorum*** Pintaud & Hodel*; Comer 274* (GA); KP221710. ***Satakentia liukiuensis*** (Hatus.) H. E. Moore; *Comer 275* (GA); KP221695. ***Veitchia spiralis*** H. Wendl.*; Zona 724* (FTG); KP221697.

Arecoideae (Chamaedoreeae); ***Chamaedorea seifrizii*** Burret*; Zomlefer 2358* (FTG, GA, NY); Givnish & al. (2010).

Arecoideae (Cocoseae); ***Attalea speciosa*** Mart. ex Spreng.*; Noblick 4950* (FTG). ***Bactris major**** Jacq.*; Noblick 5467* (FTG); KP221699. ***Beccariophoenix madagascariensis*** Jum. & H. Perrier; *Jestrow 2014-FTG-022* (FTG); KP221701. ***Cocos nucifera*** L.; Huang & al. (2013), NC_022417. ***Elaeis guineensis*** Jacq.*;* Bourgis & al. (2011);

Uthaipaisanwong & al. (2012), NC_017602. ***E. oleifera*** (Kunth) Cortés*;* Singh & al. (2013); Jansen & al. (2005), EU016883--EU016962.

Arecoideae (Euterpeae); ***Oenocarpus bataua*** Mart.*; Comer 294* (BKF); KP221713. ***O. minor*** Mart.*; Comer 300* (BKF); KP221714. ***Prestoea acuminata*** (Willd.) H. E. Moore var. ***montana*** (Graham) A. J. Hend. & Galeano; *Comer 317* (GA); KP221689.

Arecoideae (Geonomateae); ***Geonoma undata*** Klotzsch subsp. ***dussiana*** (Becc.) A. J. Hend.*; Roncal 025* (FTG); KP221706.

Arecoideae (Iriarteeae); ***Iriartea deltoidea*** Ruiz & Pav.*; Stevenson s.n.* [July 2009] (GA); KP221709.

Arecoideae (Leopoldinieae); ***Leopoldinia pulchra*** Mart.*; Comer 325* (GA); KP221711.

Arecoideae (Manicarieae); ***Manicaria saccifera*** Gaertn.*; Noblick 5482* (FTG); KP221712.

Arecoideae (Oranieae); ***Orania palindan*** (Blanco) Merr.*; Horn 4981* (FTG); KP221686.

Arecoideae (Pelagodoxeae); ***Pelagodoxa henryana*** Becc.*; Comer 276* (GA); KP221687.

Arecoideae (Podococceae); ***Podococcus barteri*** Mann & H. Wendl.*; Sunderland 1803* (K); KP221688.

Arecoideae (Reinhardtieae); ***Reinhardtia gracilis*** (H. Wendl.) Drude ex Dammer*; Comer 295* (BKF); KP221690. ***R. latisecta*** (H. Wendl.) Burret*; Comer 323* (GA); KP221691. ***R. paiewonskiana*** Read, Zanoni & M. Mejía*; Comer 324* (GA); KP221693. ***R. simplex*** (H. Wendl.) Drude ex Dammer*; Comer 320* (GA) KP221694.

Arecoideae (Roystoneeae); ***Roystonea regia*** (Kunth) O. F. Cook*; Noblick 5248* (GA); KP221692.

Arecoideae (Sclerospermeae); ***Sclerosperma profizianum*** Valk. & Sunderl.*; Stauffer & Ouattara 5-010* (G); KP221696.

Calamoideae (Calameae); ***Calamus caryotoides*** A. Cunn. ex Mart.*; Perry s.n.* [14 July 1997] (FTG); Barret & al. (2013), NC_020365.

Calamoideae (Lepidocaryeae); ***Mauritia flexuosa*** L.f.; *Zomlefer 2333* (FTG; NY); Barrett & al. (2015), KT312914.

Coryphoideae (Borasseae); ***Bismarckia nobilis*** Hildebrandt & H. Wendl.; *Noblick 5054* (FTG); Barrett & al. (2013), NC_020366.

Coryphoideae (Caryoteae); ***Caryota mitis*** Lour.; *Zona, Lewis, & Roncal 920* (FTG); Barrett & al. (2015), KT312915.

Coryphoideae (Chuniophoeniceae); ***Chuniophoenix nana*** Burret; *931085-C;* Barrett & al. (2015), KT312934.

Coryphoideae (Cryosophileae); ***Trithrinax brasiliensis*** Mart.; *Noblick 5282* (FTG)*;* Barrett & al. (2015), KT312918.

Coryphoideae (Phoeniceae); ***Phoenix dactylifera*** L.*;* Al-Dous & al. (2011); Yang & al. (2010), GU811709.

Coryphoideae (Sabaleae); ***Sabal domingensis*** Becc.; *Jestrow 2012-207* (FTG); Heyduk & al. (2015).

Coryphoideae (Trachycarpeae); ***Serenoa repens*** (W. Bartram) Small; *Soltis & Miles 2935* (FLAS); Johnson & al. (2012); *Zomlefer 2334* (FTG; NY); Barrett & al. (2015), KT312920.

Ceroxyloideae (Cyclospatheae); ***Pseudophoenix vinifera*** (Mart.) Becc.*; Zomlefer 2355* (FTG); Barrett & al. (2013), NC_020364.

Ceroxyloideae (Ceroxyleae); ***Ravenea hildebrandtii*** C. D. Bouché.; *Zomlefer 2357* (FTG); Givnish & al. (2010).

Nypoideae; ***Nypa fruticans*** Wurmb; *Chase 34461* (K); Johnson & al. (2012); *Cuenca NAC34* (FTG); Barrett & al. (2015), KT312925.

Dasypogonaceae (Outgroup); ***Kingia australis*** R.Br.; *Thiele 3703* (PERTH); Givnish & al. (2010), JX051651.

**Table 4.1.** Subfamilial circumscription of Arecaceae and tribes of subfamily Arecoideae (Dransfield & al., 2008, Dransfield & al., 2009 and Trias-Blasi, 2015), with species sampled in this study. All species included in chloroplast data set; boldfaced species included in nuclear data set.

| Subfamily | Tribe | Genera | Species sampled |
|---|---|---|---|
| Calamoideae | | 18 (600) | ***Calamus caryotoides***; *Mauritia flexuosa* |
| Ceroxyloideae | | 8 (40) | ***Pseudophoenix vinifera*** |
| Coryphoideae | | 47 (450) | *Caryota mitis; Chuniophoenix nana; Bismarkia nobilis;* ***Phoenix dactylifera***; ***Sabal domingensis***; ***Serenoa repens***; *Trithrinax brasiliensis* |
| Nypoideae | | 1 (1) | ***Nypa fruticans*** |
| Arecoideae | | 109 (1300) | |
| | Areceae | 61 (630) | *Areca vestiaria;* ***Burretiokentia grandiflora***; ***Dictyosperma album***; *Drymophloeus litigiosus;* ***Dypsis decaryi***; ***Heterospathe cagayanensis***; ***Hydriastele microspadix***; ***Kentiopsis piersoniorum***; *Satakentia liukiuensis;* ***Veitchia spiralis*** |
| | Chamaedoreeae | 5 (120) | ***Chamaedorea seifrizii*** |
| | Cocoseae | 18 (360) | ***Attalea speciosa***; ***Bactris major***; *Beccariophoenix madagascariensis*; *Cocos nucifera*; ***Elaeis guineensis***; ***E. oleifera*** |
| | Euterpeae | 5 (30) | ***Oenocarpus bataua*****;** ***O. minor*****;** ***Prestoea acuminata*** **var.** ***montana*** |
| | Geonomateae | 6 (80) | ***Geonoma undata*** **subsp.** ***Dussiana*** |
| | Iriarteeae | 5 (30) | ***Iriartea deltoidea*** |
| | Leopoldinieae | 1 (3) | ***Leopoldinia pulchra*** |
| | Manicarieae | 1 (1) | ***Manicaria saccifera*** |
| | Oranieae | 1 (25) | ***Orania palindan*** |
| | Pelagodoxeae | 2 (2) | ***Pelagodoxa henryana*** |
| | Podococceae | 1 (2) | ***Podococcus barteri*** |

| | | |
|---|---|---|
| Reinhardtieae | 1 (6) | ***Reinhardtia gracilis*; *R. latisecta*; *R. paiewonskiana*; *R. simplex*** |
| Roystoneeae | 1 (10) | ***Roystonea regia*** |
| Sclerospermeae | 1 (3) | ***Sclerosperma profizianum*** |

**Table 4.2.** Comparison of age estimates from Baker & Couvreur (2013a) and the penalized likelihood analyses for the chloroplast and nuclear data sets from this study.

| | Chloroplast | | Nuclear | | Baker & Couvreur (2013a) | |
|---|---|---|---|---|---|---|
| Clade | Stem | Crown | Stem | Crown | Stem | Crown |
| Arecaceae | 116.21 | 105.65 | 105.65 | 101.02 | 117.86 | 100.14 |
| Calamoideae | 105.65 | 52.82 | 101.02 | - | 100.14 | 80.21 |
| Nypoideae | 95.09 | - | 96.41 | - | 93.5 | - |
| Coryphoideae | 84.53 | 63.39 | 91.85 | 47.59 | 86.62 | 66.02 |
| Ceroxyloideae | 82.56 | 41.28 | 89.04 | - | 78.29 | 52.17 |
| Arecoideae | 82.56 | 80.6 | 89.04 | 86.29 | 78.29 | 73.63 |
| Iriarteeae | 80.6 | - | 86.29 | - | 73.63 | 26.84 |
| Chamaedoreeae | 78.64 | - | 83.23 | - | 70.53 | 40.63 |
| POS | 76.68 | 38.34 | 80.54 | 40.76 | 57 | 43 |
| Oranieae | 38.34 | - | 18.98 | - | 33.26 | - |
| Podococceae | 19.17 | - | 40.76 | - | 42.8 | - |
| Sclerospermeae | 19.17 | - | 18.98 | - | 33.26 | - |
| RRC | 74.72 | 72.75 | 77.71 | 74.24 | 67 | 63.59 |
| Roystoneeae | 72.75 | - | 74.24 | - | 63.59 | - |
| Reinhardtieae | 70.79 | 47.19 | 71.11 | 49.58 | 59.43 | - |
| Cocoseae | 70.79 | 68.83 | 71.11 | 67.87 | 59.43 | 55.77 |
| core arecoids | 74.72 | 49.81 | 77.71 | 56.24 | 57 | 44 |
| Geonomateae | 16.6 | - | 24.93 | - | 39.98 | 28.52 |
| Leopoldinieae | 24.91 | - | 12.51 | - | 41.02 | - |
| Manicarieae | 16.6 | - | 6.08 | - | 39.98 | - |
| Pelagodoxeae | 24.91 | - | 6.08 | - | 41.38 | 17.98 |
| AE | 33.21 | 22.14 | 56.24 | 37.54 | NA | NA |
| Areceae | 22.14 | 14.76 | 37.54 | 27.65 | 41.38 | 34.11 |
| Euterpeae | 22.14 | 11.07 | 37.54 | 17.71 | 42.62 | 31.54 |

**Table 4.3.** Summary statistics from the dispersal-extinction cladogenesis models as implemented in BioGeoBears and Lagrange using the results of the penalized likelihood analysis of the chloroplast and nuclear data sets including: log-likelihood (LnL), number of free parameters (Parameters) and their estimates (dispersal [d], extinction [e], rate of range-switching [a], and founder-event speciation [j]), as well as Akaike information criterion (AIC), including correction for sample size (AICc). Results of the likelihood ratio test comparing models implemented in BioGeoBears are provided with the alternative (Alt) and null (Null) models, degrees of freedom (df), chi-square value ($\chi^2$; one tailed), and the p value (*P)*. Significance ($P < 0.05$) = *

| | | | | | | | | | | Likelihood ratio test | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chloroplast | Model | LnL | Parameters | d | e | a | j | AIC | AICc | Alt | Null | df | $\chi^2$ | *P* |
| BioGeoBears[1] | M0 | -130.8 | 2 | 0.0024 | 0.0016 | Fixed[3] | Fixed[3] | 265.6 | 265.9 | M1+J | M1 | 1 | 4.8 | 0.028* |
| | M1 | -120.0 | 2 | 0.0192 | 0.0033 | Fixed[3] | Fixed[3] | 243.9 | 244.2 | M1+A | M1 | 1 | 9 | 0.0027* |
| | M1+J | -117.6 | 3 | 0.0148 | 0.0016 | Fixed[3] | 0.1232 | 241.1 | 241.7 | M1+A+J | M1+A | 1 | 0.52 | 0.47 |
| | M1+A | -115.5 | 3 | 0.0124 | 0.0000 | 0.0101 | Fixed[3] | 236.9 | 237.5 | | | | | |
| | M1+A+J | -115.2 | 4 | 0.0124 | 0.0000 | 0.0067 | 0.0555 | 238.4 | 239.4 | | | | | |
| Lagrange[2] | M0 | -122.9 | 2 | 0.0040 | 0.0020 | Fixed[3] | Fixed[3] | 249.8 | 250.1 | | | | | |
| | M1 | -115.2 | 2 | 0.0283 | 0.0036 | Fixed[3] | Fixed[3] | 234.4 | 234.7 | | | | | |

| | | | | | | | | | | Likelihood ratio test | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nuclear | Model | LnL | Parameters | d | e | a | j | AIC | AICc | Alt | Null | df | $\chi^2$ | *P* |
| BioGeoBears[1] | M0 | -102.1 | 2 | 0.0023 | 0.0006 | Fixed[3] | Fixed[3] | 208.3 | 208.7 | M1+J | M1 | 1 | 1.78 | 0.18 |
| | M1 | -93.2 | 2 | 0.0191 | 0.0020 | Fixed[3] | Fixed[3] | 190.5 | 190.9 | M1+A | M1 | 1 | 4.51 | 0.034* |
| | M1+J | -92.4 | 3 | 0.0148 | 0.0000 | Fixed[3] | 0.0931 | 190.7 | 191.5 | M1+A+J | M1+A | 1 | 0.0008 | 0.98 |
| | M1+A | -91.0 | 3 | 0.0148 | 0.0000 | 0.0062 | Fixed[3] | 188.0 | 188.7 | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | M1+A+J | -91.0 | 4 | 0.0148 | 0.0000 | 0.0062 | 0.0011 | 190.0 | 191.3 |
| Lagrange[2] | M0 | -95.8 | 2 | 0.0049 | 0.0023 | Fixed[3] | Fixed[3] | 195.6 | 196.0 |
| | M1 | -84.8 | 2 | 0.0417 | 0.0041 | Fixed[3] | Fixed[3] | 173.6 | 174.0 |

Notes: [1]The maximum number of ranges used for BioGeoBears analyses were three. [2]The maximum number of ranges used for Lagrange analyses were two.

[3]Fixed parameters were set to zero.

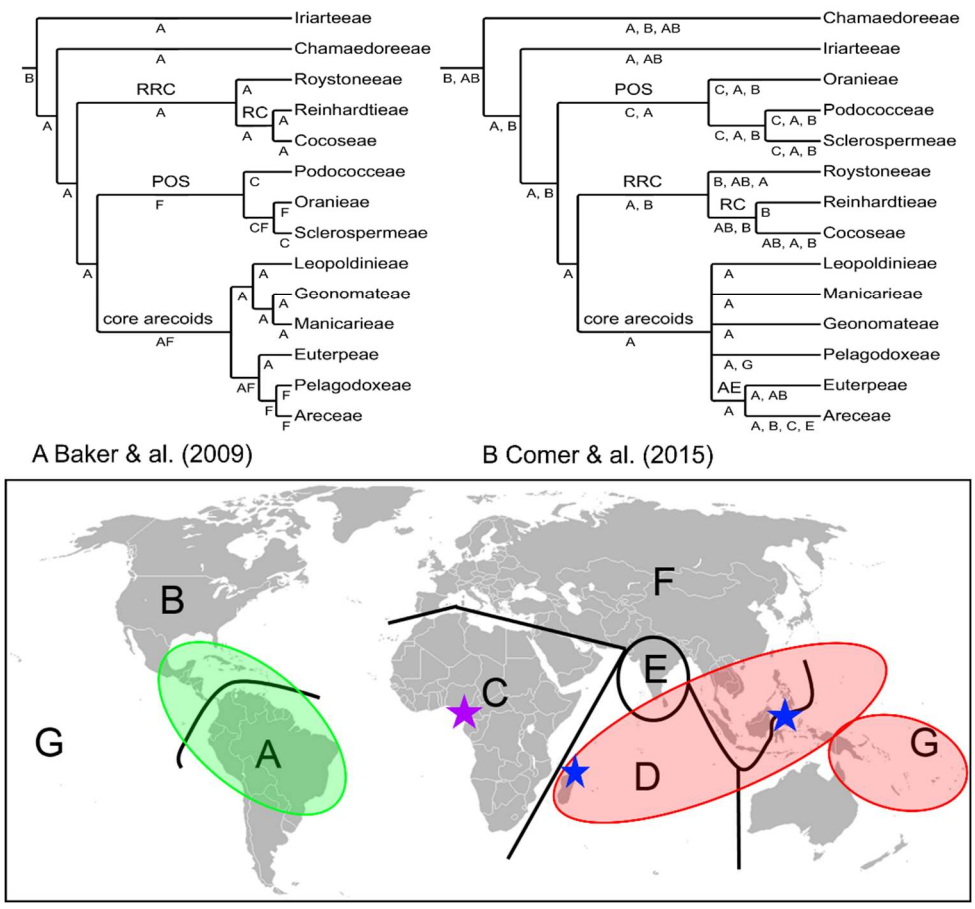
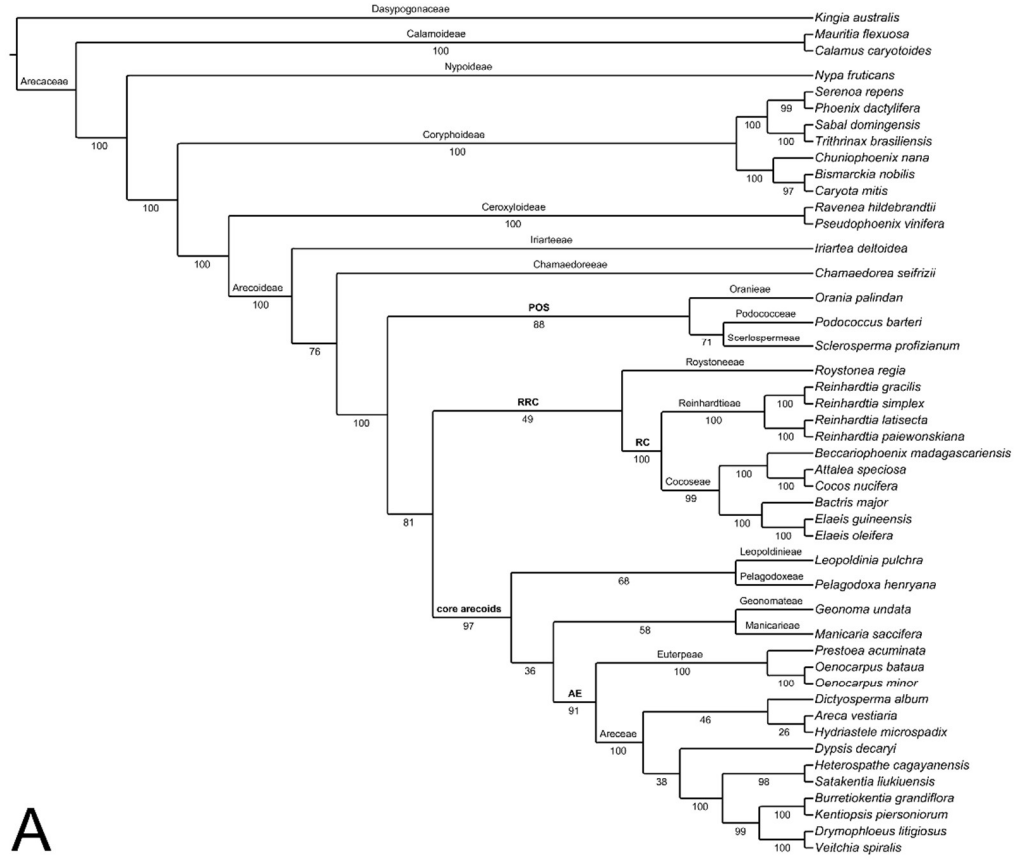


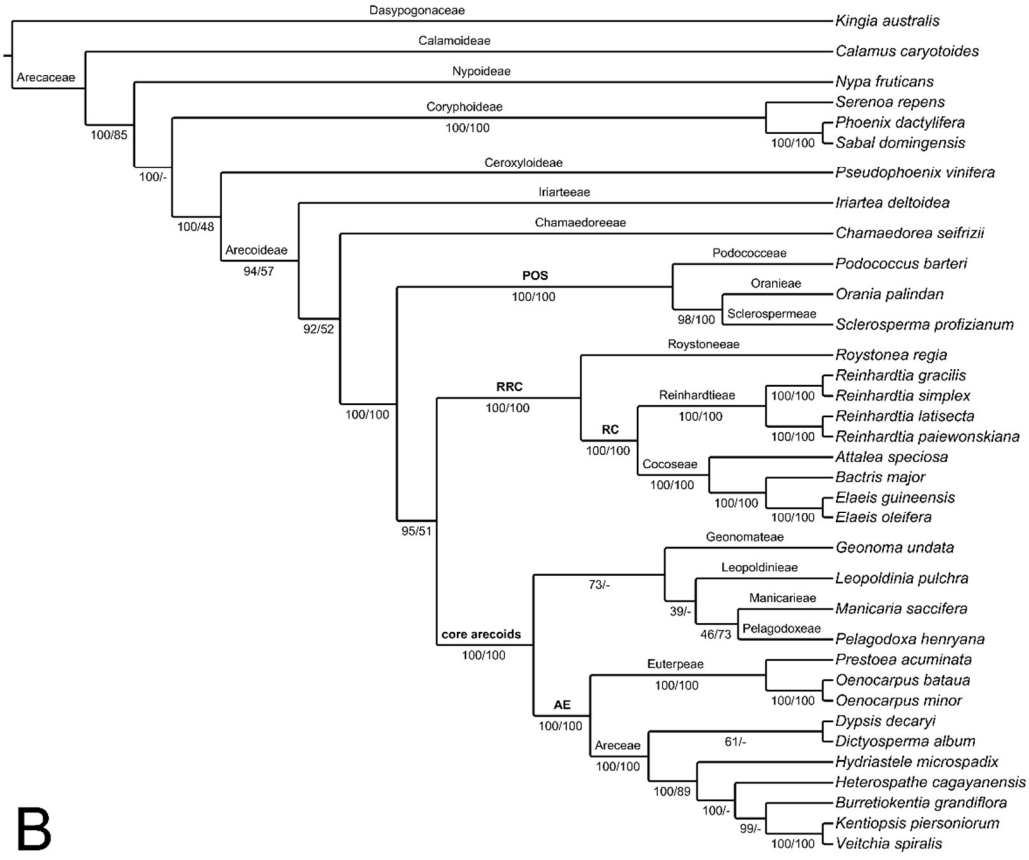

**Figure 4.1.** Tribal phylogenies of subfamily Arecoideae from two recent studies: A) most congruent supertree modified from Fig. 3 in Baker & al. (2009, all branches supported by > 1 input tree), with the inferred ancestral areas of Baker & Couvreur (Fig. 1, 2013a) below branches, B) chloroplast summary tree modified from Fig. 4 in Comer & al. (2015, branches supported by BS ≥ 79 and PP ≥ 0.95) and inferred ancestral areas below branches. Map inset of the geographic areas used for ancestral area analyses (Baker and Couvreur, 2013a and Couvreur & al., 2011): A = South America; B = North America, Central America, and the Caribbean; C = Africa and Arabia; D = Indian Ocean Islands and Madagascar; E = India and Sri Lanka; F = Eurasia to Wallace's line; G = Australia and Pacific east of Wallace's line. Labels above branches are major clades: AE (Areceae + Euterpeae); core arecoids (Areceae, Euterpeae, Geonomateae, Leopoldinieae,

Manicarieae, and Pelagodoxeae); POS (Podococceae, Oranieae, and Sclerospermeae); RC (Reinhardtieae + Cocoseae); and RRC (Roystoneeae, Reinhardtieae, and Cocoseae). Green and red ovals = centers of Arecoideae diversity; purple star = current range for Podococceae and Sclerospermeae; blue stars = current range of Oranieae.

**Figure 4.2. (Next page)** Chloroplast and nuclear phylogenies used for divergence time estimation and ancestral area analyses: A) The ML tree from the analysis of the 114 chloroplast genes, B) the ML tree from the supermatrix analysis 168 nuclear genes (modified from Comer & al., in review). Labels above the branches = family, subfamily, tribe, and major clade (bold). Labels below branches = bootstrap support (nuclear phylogeny shows ML/ASTRAL bsv). Major clades: AE (Areceae + Euterpeae), core arecoids (Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae), POS (Podococceae, Oranieae, and Sclerospermeae), RC (Reinhardtieae + Cocoseae), and RRC (Roystoneeae, Reinhardtieae, and Cocoseae).


Dasypogonaceae
Kingia australis
Calamoideae
Mauritia flexuosa
Calamus caryotoides
Nypoideae
Arecaceae
Nypa fruticans
Serenoa repens
Phoenix dactylifera
Coryphoideae
Sabal domingensis
Trithrinax brasiliensis
Chuniophoenix nana
Bismarckia nobilis
Caryota mitis
Ceroxyloideae
Ravenea hildebrandtii
Pseudophoenix vinifera
Iriarteeae
Iriartea deltoidea
Chamaedoreeae
Chamaedorea seifrizii
Arecoideae
Oranieae
Orania palindan
POS
Podococceae
Podococcus barteri
Sclerospermeae
Sclerosperma profizianum
Roystoneeae
Roystonea regia
RRC
Reinhardtieae
Reinhardtia gracilis
Reinhardtia simplex
Reinhardtia latisecta
Reinhardtia paiewonskiana
RC
Beccariophoenix madagascariensis
Cocoseae
Attalea speciosa
Cocos nucifera
Bactris major
Elaeis guineensis
Elaeis oleifera
Leopoldinieae
Leopoldinia pulchra
Pelagodoxeae
Pelagodoxa henryana
core arecoids
Geonomateae
Geonoma undata
Manicarieae
Manicaria saccifera
Euterpeae
Prestoea acuminata
Oenocarpus bataua
Oenocarpus minor
AE
Dictyosperma album
Areceae
Areca vestiaria
Hydriastele microspadix
Dypsis decaryi
Heterospathe cagayanensis
Satakentia liukiuensis
Burretiokentia grandiflora
Kentiopsis piersoniorum
Drymophloeus litigiosus
Veitchia spiralis

A


Dasypogonaceae
Kingia australis
Calamoideae
Calamus caryotoides
Arecaceae
Nypoideae
Nypa fruticans
Coryphoideae
Serenoa repens
Phoenix dactylifera
Sabal domingensis
Ceroxyloideae
Pseudophoenix vinifera
Iriarteeae
Iriartea deltoidea
Chamaedoreeae
Chamaedorea seifrizii
Arecoideae
Podococceae
Podococcus barteri
POS
Oranieae
Orania palindan
Sclerospermeae
Sclerosperma profizianum
Roystoneeae
Roystonea regia
RRC
Reinhardtia gracilis
Reinhardtieae
Reinhardtia simplex
Reinhardtia latisecta
RC
Reinhardtia paiewonskiana
Attalea speciosa
Cocoseae
Bactris major
Elaeis guineensis
Elaeis oleifera
Geonomateae
Geonoma undata
Leopoldinieae
Leopoldinia pulchra
Manicarieae
Manicaria saccifera
Pelagodoxeae
Pelagodoxa henryana
core arecoids
Euterpeae
Prestoea acuminata
Oenocarpus bataua
Oenocarpus minor
AE
Dypsis decaryi
Areceae
Dictyosperma album
Hydriastele microspadix
Heterospathe cagayanensis
Burretiokentia grandiflora
Kentiopsis piersoniorum
Veitchia spiralis

B

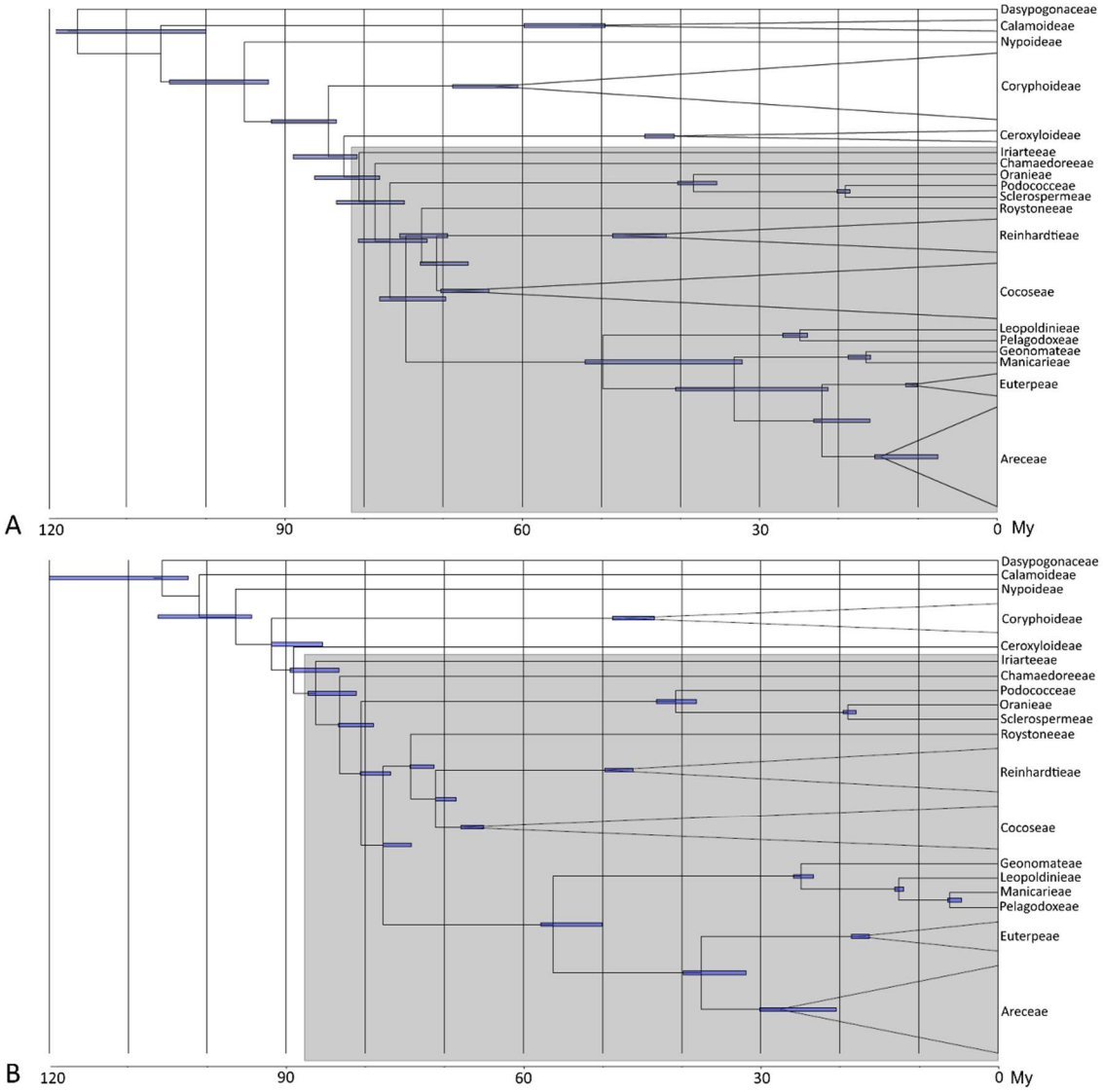



**Figure 4.3.** Chronograms from the penalized likelihood analyses: A) chloroplast phylogeny, B) nuclear phylogeny. Blue bars indicate the age range from the analysis of the bootstrap replicates. Subfamily Arecoideae is shaded grey.

**Figure 4.4. (Next page)** Chloroplast phylogeny mapped with inferred ancestral areas with relative probabilities > 0.1. Labels above branches = BioGeoBears results. Labels below branches = Lagrange results. * = relative probability > 0.5.

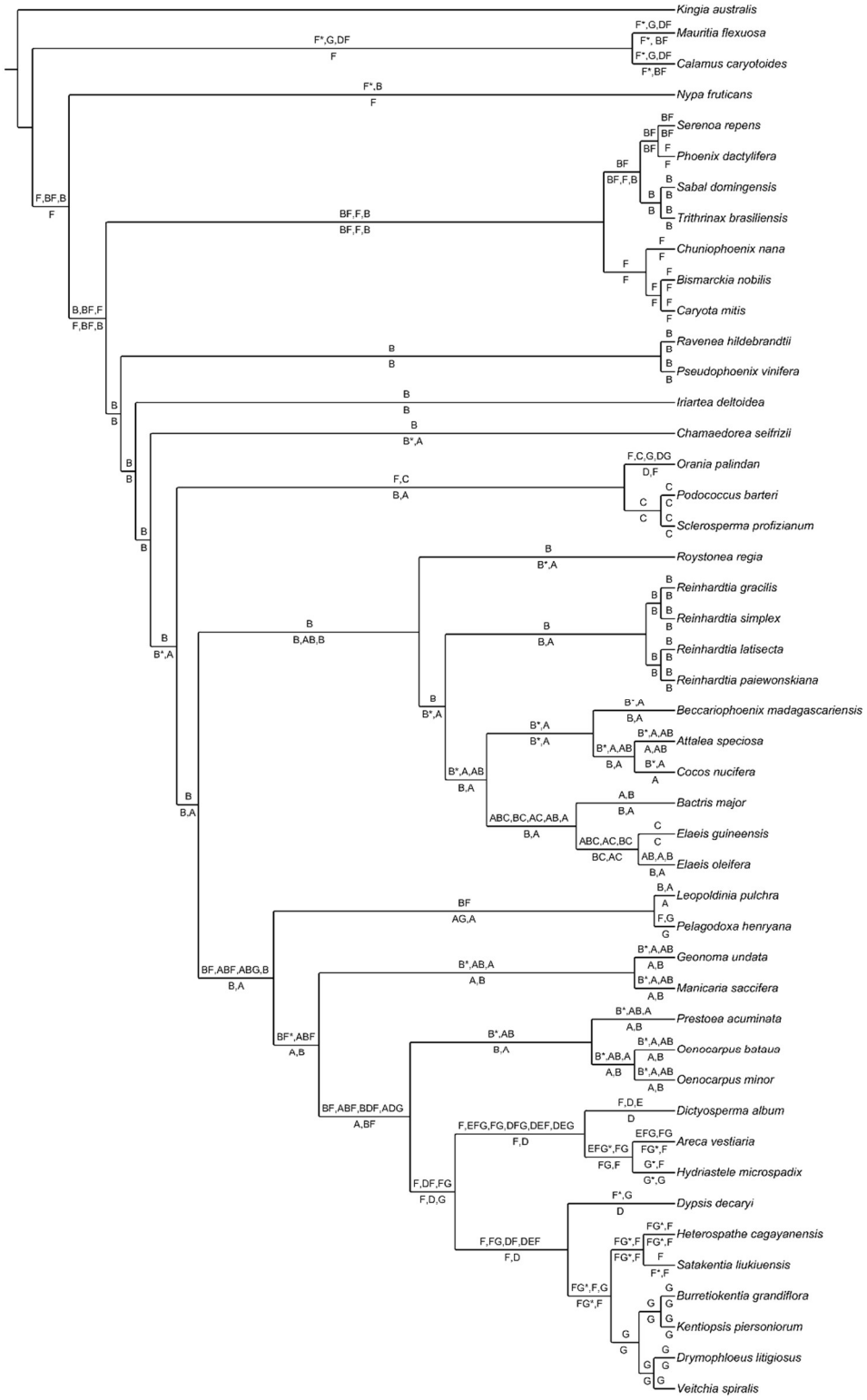
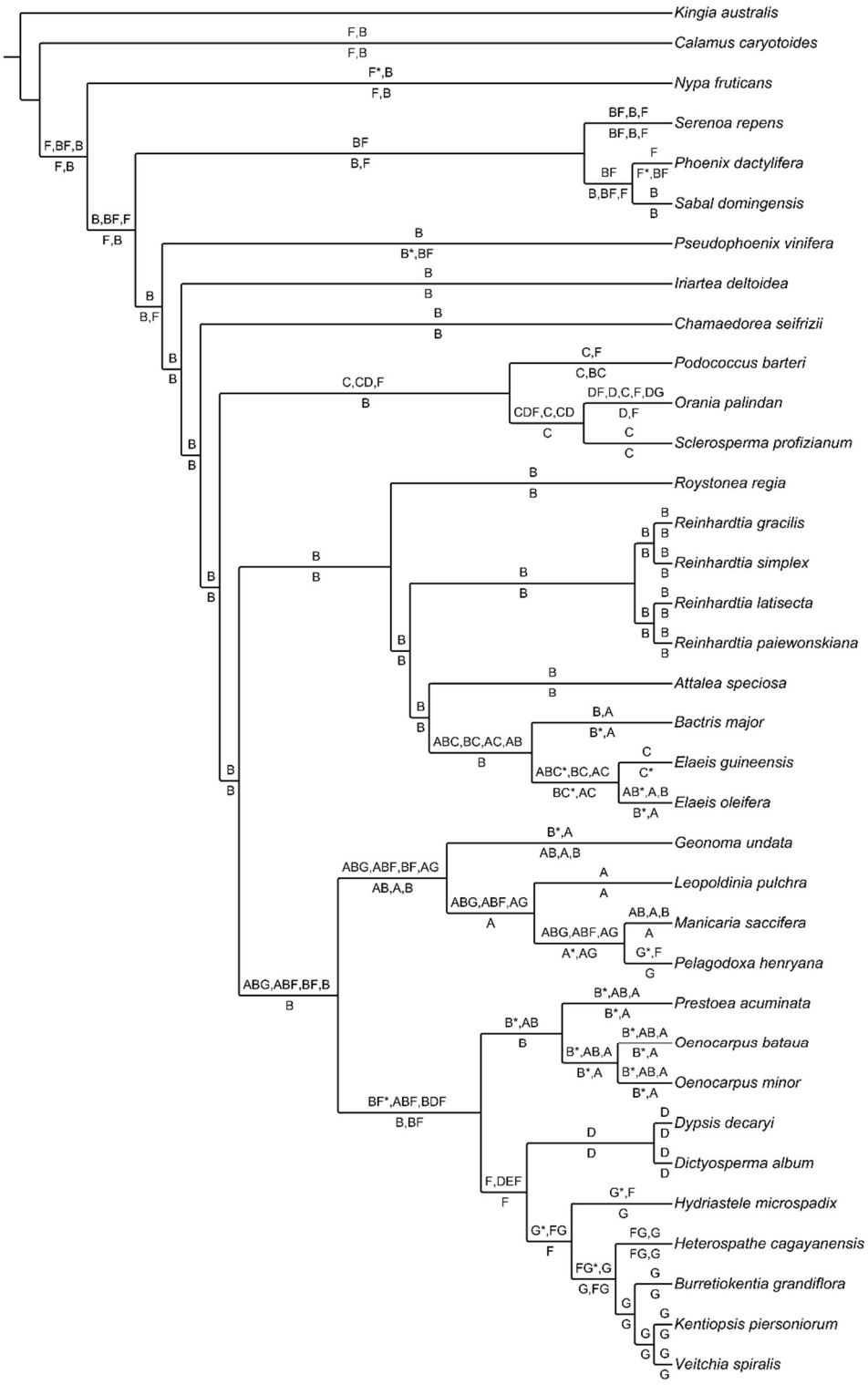

Kingia australis
Mauritia flexuosa
Calamus caryotoides
Nypa fruticans
Serenoa repens
Phoenix dactylifera
Sabal domingensis
Trithrinax brasiliensis
Chuniophoenix nana
Bismarckia nobilis
Caryota mitis
Ravenea hildebrandtii
Pseudophoenix vinifera
Iriartea deltoidea
Chamaedorea seifrizii
Orania palindan
Podococcus barteri
Sclerosperma profizianum
Roystonea regia
Reinhardtia gracilis
Reinhardtia simplex
Reinhardtia latisecta
Reinhardtia paiewonskiana
Beccariophoenix madagascariensis
Attalea speciosa
Cocos nucifera
Bactris major
Elaeis guineensis
Elaeis oleifera
Leopoldinia pulchra
Pelagodoxa henryana
Geonoma undata
Manicaria saccifera
Prestoea acuminata
Oenocarpus bataua
Oenocarpus minor
Dictyosperma album
Areca vestiaria
Hydriastele microspadix
Dypsis decaryi
Heterospathe cagayanensis
Satakentia liukiuensis
Burretiokentia grandiflora
Kentiopsis piersoniorum
Drymophloeus litigiosus
Veitchia spiralis

**Figure 4.5. (Next page)** Nuclear phylogeny mapped with inferred ancestral areas with relative probabilities > 0.1. Labels above branches = BioGeoBears results. Labels below branches = Lagrange results. * = relative probability > 0.5.


Kingia australis
Calamus caryotoides
Nypa fruticans
Serenoa repens
Phoenix dactylifera
Sabal domingensis
Pseudophoenix vinifera
Iriartea deltoidea
Chamaedorea seifrizii
Podococcus barteri
Orania palindan
Sclerosperma profizianum
Roystonea regia
Reinhardtia gracilis
Reinhardtia simplex
Reinhardtia latisecta
Reinhardtia paiewonskiana
Attalea speciosa
Bactris major
Elaeis guineensis
Elaeis oleifera
Geonoma undata
Leopoldinia pulchra
Manicaria saccifera
Pelagodoxa henryana
Prestoea acuminata
Oenocarpus bataua
Oenocarpus minor
Dypsis decaryi
Dictyosperma album
Hydriastele microspadix
Heterospathe cagayanensis
Burretiokentia grandiflora
Kentiopsis piersoniorum
Veitchia spiralis

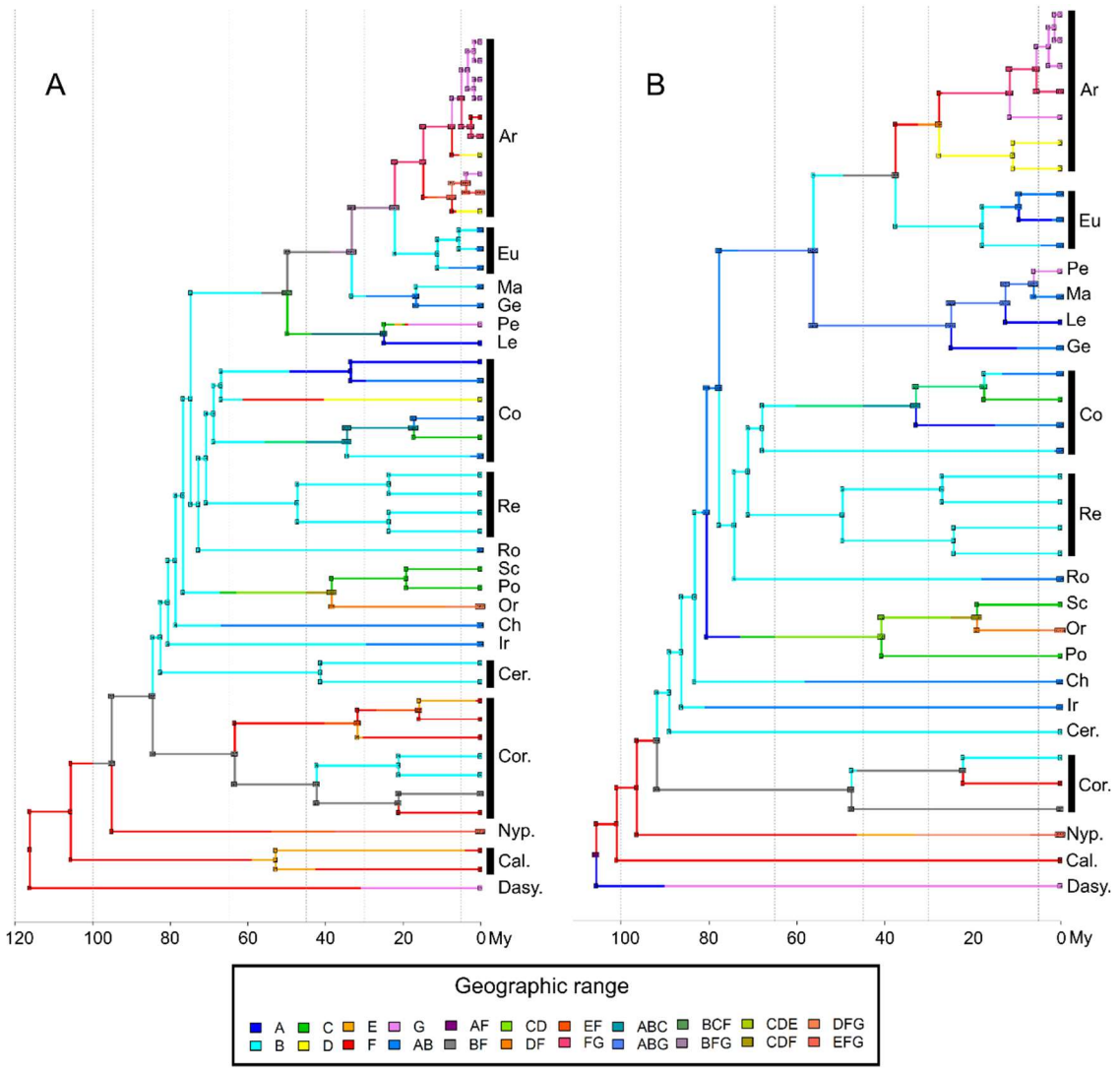



**Figure 4.6.** Examples from the stochastic mapping simulations representing possible dispersal pathways: A) chloroplast phylogeny, B) nuclear phylogeny. Families: Dasy. = Dasypogonaceae; Arecaceae (subfamilies): Cal. = Calamoideae. Cer. = Ceroxyloideae. Cory. = Coryphoideae. Nyp. = Nypoideae. Arecoideae tribes: Ar = Areceae, Ch = Chamaedoreeae, Co = Cocoseae, Eu = Euterpeae, Ge = Geonomateae, Ir = Iriarteeae, Le = Leopoldinieae, Ma = Manicarieae, Or = Oranieae, Pe = Pelagodoxeae, Po = Podococceae, Re = Reinhardtieae, Ro = Roystoneeae, Sc = Sclerospermeae.

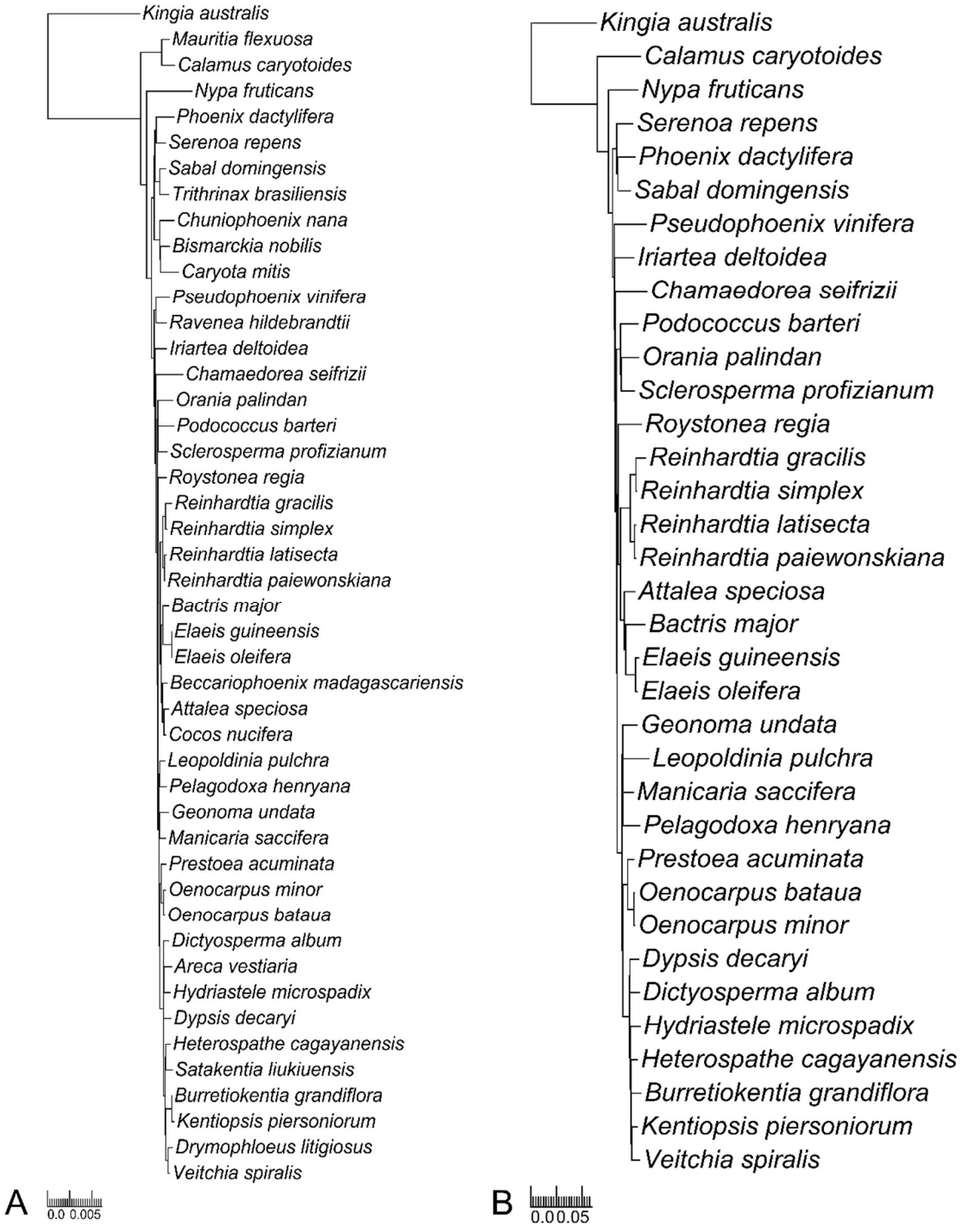



**Figure 4.7.** Phylograms of: A) the chloroplast phylogeny, B) nuclear phylogeny.

# CHAPTER 5

# CONCLUSIONS

Since the foundation of modern palm systematics based on Moore (1973) and Uhl and Dransfield (1987), understanding of the evolutionary history of the Arecaceae has progressed with molecular data (Uhl et al., 1995; Baker et al., 1999). Prior to the study presented here, the largest data set (Baker et al. 2009) had nearly full generic coverage of the Arecaceae and all molecular data available at that time (nine plastid and six nuclear markers). While providing new insights into the relationships of subfamily Arecoideae, deeper relationships were not well resolved. Next generation sequencing techniques allow relatively rapid generation of large sequence data sets for a large number of taxa (Jansen et al., 2007; Givnish et al., 2010; Steele et al., 2012). This dissertation utilized a phylogenomics approach incorporating next generation sequencing technology to generate the largest sequence data set (114 plastid and 168 nuclear genes) to date for subfamily Arecoideae (Comer et al., 2015; Comer et al., in review).

With these data, the plastid (Comer et al., 2015) and the nuclear phylogenies (Comer et al., in review) were largely congruent and consistent with previous studies in recovering a monophyletic subfamily Arecoideae and three major clades: core arecoid clade (tribes Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae), POS clade (Podococceae, Oranieae, and Sclerospermeae), and RRC clade (Roystoneeae, Reinhardtieae, and Cocoseae). The AE clade (tribes Areceae and Euterpeae) was also consistently and strongly supported. Phylogenies differed in their

placements of the tribes within the core arecoids clade, which had weak to no support. The POS clade was now placed sister to a core arecoid/RCC clade.

Two radiation events were inferred for subfamily Arecoideae. The first radiation was at the deepest nodes of the Arecoideae (i.e. the successive divergences of tribes Iriarteeae and Chamaedoreeae). The plastid phylogeny recovered Iriarteeae as the earliest diverging lineage with Chamaedoreeae only moderately supported (bsv 76) as sister to the rest of the Arecoideae, while placements of Chamaedoreeae and Iriarteeae were unresolved in the nuclear phylogeny. Age estimates from penalized likelihood analyses also suggest these earlier lineages underwent rapid diversification, with less than three million years separating divergence events. The second inferred radiation event occurred within the core arecoids clade. With the exception of the AE clade, relationships among the major core arecoid lineages were unresolved with both data sets. Age estimates within the core arecoids clade often had wide and overlapping ranges, suggesting uncertainty in these results and/or possible rapid diversification.

The results of ancestral area analyses were generally consistent with previous studies (Baker et al., 2011; Baker and Couvreur, 2013a, b) in supporting North America as the area of origin for subfamily Arecoideae. North America was also inferred as the center for much of the early diversification within the Arecoideae, with subsequent dispersals into South America, Africa, and the Indopacific. The current geographic distribution of tribe Oranieae may be explained by the “out of India” hypothesis. Following the dispersal of the early progenitors of the POS clade into Africa, tribe Oranieae may have dispersed from Africa and Madagascar to India prior to India splitting

away from Africa/Madagascar ca. 66 million years ago. Oranieae may then have rafted on India to the Indopacific, where the tribe dispersed and diversified.

This dissertation sets the foundation for future phylogenomic work for subfamily Arecoideae and also the Arecaceae. Representatives of all five palm subfamilies were samples for this study, with focus on subfamily Arecoideae. The thesis provide a methodological standard and phylogenetic framework for future systematic studies of the palm family.

**Literature cited**

BAKER, W. J., AND T. L. P. COUVREUR. 2013a. Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical biogeography. *Journal of Biogeography* 40: 274–285.

BAKER, W. J., AND T. L. P. COUVREUR. 2013b. Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. II. Diversification history and origin of regional assemblages. *Journal of Biogeography* 40: 286–298.

BAKER, W. J., C. B. ASMUSSEN, S. C. BARROW, J. DRANSFIELD, AND T. A. HEDDERSON. 1999. A phylogenetic study of the palm family (Palmae) based on chloroplast DNA sequences from the *trnL - trnF* region. *Plant Systematics and Evolution* 219: 111–126.

BAKER, W. J., V. SAVOLAINEN, C. B. ASMUSSEN-LANGE, M. W. CHASE, J. DRANSFIELD, F. FOREST, M. M. HARLEY, et al. 2009. Complete generic-level phylogenetic analyses of palms (Arecaceae) with comparisons of supertree and supermatrix approaches. *Systematic Biology* 58: 240–256.

BAKER, W. J., M. V. NORUP, J. J. CLARKSON, T. L. P. COUVREUR, J. L. DOWE, C. E. LEWIS, J.-C. PINTAUD, et al. 2011. Phylogenetic relationships among arecoid palms (Arecaceae: Arecoideae). *Annals of Botany* 108: 1417–1432.

COMER, J.R., ZOMLEFER, W.B., BARRETT, C.F., DAVIS, J.I., STEVENSON, D.W., HEYDUK, K., & LEEBENS-MACK, J.H. 2015. Resolving relationships within the palm subfamily Arecoideae (Arecaceae) using plastid sequences derived from next-generation sequencing. *American Journal of Botany* 102: 888–899.

COMER, J.R., ZOMLEFER, W.B., BARRETT, C.F., STEVENSON, D.W., HEYDUK, K., & LEEBENS-MACK, J.H. In review. Nuclear phylogenomics of the palm subfamily Arecoideae (Arecaceae). *Molecular Phylogenetics and Evolution*.

GIVNISH, T. J., M. AMES, J. R. MCNEAL, M. R. MCKAIN, P. R. STEELE, C. W. DEPAMPHILIS, S. W. GRAHAM, et al. 2010. Assembling the Tree of the Monocotyledons: Plastome Sequence Phylogeny and Evolution of Poales1. *Annals of the Missouri Botanical Garden* 97: 584–616.

JANSEN, R. K., Z. CAI, L. A. RAUBESON, H. DANIELL, C. W. DEPAMPHILIS, J. LEEBENS-MACK, K. F. MÜLLER, et al. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences* 104: 19369–19374.

MOORE, H. E. 1973. The major groups of palms and their distribution. *Gentes Herbarum* 11: 27–141.

STEELE, P. R., K. L. HERTWECK, D. MAYFIELD, M. R. MCKAIN, J. LEEBENS-MACK, AND J. C. PIRES. 2012. Quality and quantity of data recovered from massively parallel

sequencing: Examples in Asparagales and Poaceae. *American Journal of Botany* 99: 330–348.

UHL, N. W., AND J. DRANSFIELD 1987. Genera palmarum: a classification of palms based on the work of Harold E. Moore, Jr. Allen Press, Lawrence, Kansas, USA.

UHL, N. W., J. DRANSFIELD, J. I. DAVIS, M. A. LUCKOW, K. S. HANSEN, AND J. J. DOYLE. 1995. Phylogenetic relationships among palms: cladistic analyses of morphological and chloroplast DNA restriction site variation. . *In* P. J. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries [eds.], Monocotyledons: Systematics and evolution, vol. 2, 623–662. Whitstable Litho Printers, Kent, England.