


RESEARCH ARTICLE

Open Access



A plastid phylogenomic framework for the palm family (Arecaceae)

Gang Yao¹, Yu-Qu Zhang^{2,3}, Craig Barrett⁴, Bine Xue⁵, Sidonie Bellot⁶, William J. Baker⁶ and Xue-Jun Ge^{2,7*} 

Abstract

Background Over the past decade, phylogenomics has greatly advanced our knowledge of angiosperm evolution. However, phylogenomic studies of large angiosperm families with complete species or genus-level sampling are still lacking. The palms, Arecaceae, are a large family with ca. 181 genera and 2600 species and are important components of tropical rainforests bearing great cultural and economic significance. Taxonomy and phylogeny of the family have been extensively investigated by a series of molecular phylogenetic studies in the last two decades. Nevertheless, some phylogenetic relationships within the family are not yet well-resolved, especially at the tribal and generic levels, with consequent impacts for downstream research.

Results Plastomes of 182 palm species representing 111 genera were newly sequenced. Combining these with previously published plastid DNA data, we were able to sample 98% of palm genera and conduct a plastid phylogenomic investigation of the family. Maximum likelihood analyses yielded a robustly supported phylogenetic hypothesis. Phylogenetic relationships among all five palm subfamilies and 28 tribes were well-resolved, and most inter-generic phylogenetic relationships were also resolved with strong support.

Conclusions The inclusion of nearly complete generic-level sampling coupled with nearly complete plastid genomes strengthened our understanding of plastid-based relationships of the palms. This comprehensive plastid genome dataset complements a growing body of nuclear genomic data. Together, these datasets form a novel phylogenomic baseline for the palms and an increasingly robust framework for future comparative biological studies of this exceptionally important plant family.

Keywords Arecaceae, Nuclear-plastid discordance, Palmae, Plastome, Phylogenomics

*Correspondence:

Xue-Jun Ge

xjge@scbg.ac.cn

¹ College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou 510642, China

² Key Laboratory of Plant Resources Conservation and Sustainable Utilization, and Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

³ Present Address: College of Pharmacy, Shaanxi University of Chinese Medicine, Xi'an, China

⁴ Department of Biology, West Virginia University, Morgantown, WV, USA

⁵ College of Horticulture and Landscape Architecture, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

⁶ Royal Botanic Gardens, Kew TW9 3AB, Surrey, UK

⁷ Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Guangzhou, China

Background

Over the past decade, enormous progress in phylogenomics has been made in clarifying many recalcitrant relationships among angiosperms [1–5]. However, disentangling the relationships among angiosperm lineages that have undergone rapid radiations remains a major challenge [1, 4–7] and is often complicated by the existence of long branches [8], hybridization [9], incomplete lineage sorting (ILS) [4, 10], polyploidization [9, 11], and lateral gene transfer [12]. The importance of broad taxon sampling in improving the accuracy of phylogenetic inference has been highlighted by many authors [13–16], since including more taxa can improve the detection of multiple substitutions [17, 18] and thus



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

may be helpful to alleviate the influence of systematic error, such as “long-branch attraction” (LBA) [14, 19]. On the other hand, including more genomic data is also known to be helpful in improving phylogenetic inference, because this increases phylogenetic signal and may also reveal conflicting gene histories that can shed light on evolutionary history. For instance, comparing nuclear and plastid trees can help detect cases of past or recent hybridization [20, 21]. Previous phylogenetic studies of large angiosperm families based on the analysis of one or a few short genetic loci have included comprehensive or even complete generic sampling, but some key relationships remained unresolved or weakly supported, such as in the grass family Poaceae [22], the mint family Lamiaceae [23], and the palm family Arecaceae [24, 25]. An increasing number of phylogenomic studies (i.e., usually based on hundreds of generic regions) have been conducted recently to improve phylogenetic resolution in large angiosperm families, such as in Apiaceae [26], Apocynaceae [27], Asteraceae [28], Brassicaceae [29], Cucurbitaceae [30], Fabaceae [31, 32], Gesneriaceae [33], Lamiaceae [34], Orchidaceae [35], Poaceae [36], and Rosaceae [37, 38]. However, in these studies, sampling was mostly focused on the subfamily and/or tribal levels. Phylogenomic analyses with a comprehensive generic-level sampling of large angiosperm families remain rare and thus urgently needed to better understand the angiosperm tree of life.

Arecaceae, the palms, are an iconic large family of flowering plants [25, 39]. They represent one of the most diverse monocot families with ca. 181 genera and 2600 species [40–42], which are classified currently into five subfamilies and 28 tribes based on the results of extensive molecular phylogenetic studies [40]. Members of the palm family can be readily identified by their “woody” type growth, with primary thickening of their vascular tissues, plicate leaves with unique development, and inflorescences subtended by adaxial two-keeled bracts [43]. The palms exhibit a remarkable degree of morphological variation and over 90% of palm species are restricted to tropical rainforests, thus making Arecaceae an important model system to study the evolutionary history of tropical biodiversity [44, 45]. Additionally, the palm family includes many economically important species as hundreds of species are used by communities throughout the world for food, materials, medicine, and other uses [46]. Examples include the betelnut (*Areca catechu* L.), coconut (*Cocos nucifera* L.), date palm (*Phoenix dactylifera* L.), oil palm (*Elaeis guineensis* Jacq.), sago palm (*Metroxylon* Rottb. spp.), and wax palm (*Copernicia cerifera* (Arruda) Mart.). Furthermore, many palm species are economically important in the horticulture industry,

being widely used in gardens and as ornamental trees and shrubs in landscaping [43].

As an ecologically and economically important angiosperm family, the palms have attracted the attention of many botanists and evolutionary biologists [25, 39, 41–51]. Molecular phylogenetic studies have greatly advanced our knowledge of the taxonomy and phylogeny of palms. A series of phylogenetic studies strongly supported the monophyly of Arecaceae and its affiliation within the commelinid monocots [6, 47, 52, 53]. However, as recovered in previous phylogenetic studies [25, 39, 47], diversification along the backbone of Arecaceae is complicated and disentangling palm tribal-level relationships is challenging. Early phylogenetic studies based on limited taxon sampling and a few plastid and nuclear DNA regions recovered several major clades within Arecaceae, but relationships among these clades were incompletely resolved [54–56]. Asmussen et al. [57] reconstructed the phylogenetic relationships of 161 palm genera based on the analysis of four plastid DNA regions (*matK*, *rbcl*, *rps16* intron, *trnL-F*) and recognized five subfamilies, viz., Arecoideae, Calamoideae, Ceroxyloideae, Coryphoideae, and Nypoideae. The first complete generic-level phylogenetic study of Arecaceae was conducted based on analyses of multiple plastid and nuclear loci, morphology, and restriction profiling [24]. In this study, which was central to the establishment of the formal phylogenetic classification [24], the monophyly of all the five subfamilies was strongly supported with increases in support values for many nodes compared with previous studies. Nevertheless, relationships among a number of tribes and genera were still weakly supported and some of the tribes and genera were even clustered in polytomies, especially among members of the subfamily Arecoideae, which represents over 60% of the generic diversity and nearly 55% of the species diversity in the family [24]. In a phylogenetic analysis of the palms using an all-evidence super-tree approach based on ca. 35.5% species-level taxon sampling and 13 DNA markers (9 plastid and 4 nuclear regions), low support was obtained for relationships among some tribes and subtribes as well [25]. Seventy-four plastid genomes of 54 palm genera representing five subfamilies and 19 tribes were analyzed by Chen et al. [58], and relationships among all of them were mostly resolved with high support, but multiple tribes within the largest subfamily Arecoideae were not included. In a recent phylogenomic study of 2333 angiosperm genera based on an analysis of 353 nuclear genes, 67 genera representing 26 palm tribes and all the five palm subfamilies were included, the subfamily-level relationships and most tribal-level relationships were highly supported except some tribal-level relationships within Arecoideae were resolved with weak support [59]. An updated version of

this phylogenetic tree is now available online via the Kew Tree of Life Explorer (KTLE) [60] and includes 179 palm genera. However, some tribal-level relationships in Arecoideae, Ceroxyloideae, and Coryphoideae were poorly resolved in this phylogenetic tree. Multiple phylogenomic studies based on large-scale nuclear or plastid genomic data were also conducted recently in different palm lineages, such as in Arecoideae [39, 48], Dypsidinae [41], Calamoideae [49], Lepidocaryeae [61], Phytelephea [62], Geonomateae [63], *Brahea* Mart. [64], and *Raphia* P. Beauv. [65]. All these studies continuously improved our understanding of the evolution of the family. Nevertheless, in terms of phylogenetic relationships among tribes, subtribes, and genera, many nodes are still unresolved or weakly supported [24, 25, 39, 48, 57, 59]. Especially along the backbone of the largest subfamily Arecoideae, phylogenetic relationships among the six tribes within the core arecoids (Areceae, Euterpeae, Geonomateae, Leopoldinieae, Manicarieae and Pelagodoxeae [66]) need to be further clarified. Additionally, about 17 palm genera from the two largest subfamilies Arecoideae and Coryphoideae remain unplaced at the subtribal level due to their unresolved phylogenetic placements [40].

Plastid genomes provide a large number of nucleotide sites for phylogenetic inference and can be sequenced from many samples at a lower cost than nuclear genome regions [5, 13, 34]. A plastome tree can be compared to a nuclear tree to provide complementary insights on lineage history, for instance by revealing cases of past hybridization only detectable through chloroplast capture [20, 21], or aspects related to maternal inheritance [67]. Thus, plastid phylogenomics have been adopted extensively in recent studies of plant lineages both at deep and shallow taxonomic levels, such as in liverworts [68], ferns [69, 70], gymnosperms [71], and angiosperms [5, 10, 13, 31, 34, 37, 72–74], providing critical insights into the recalcitrant phylogenetic relationships within these groups. In contrast with the continuously increasing body of nuclear-based palm phylogenomic studies [39, 41, 49, 59–63], a comprehensive generic-level plastid phylogenomic study of the family is still lacking [47, 58]. Here, we conducted a plastid phylogenomic analysis of palms based on a nearly complete generic-level sampling and aimed at (1) shedding new light on the relationships among tribes, subtribes, and genera of palms and (2) evaluating how plastome and nuclear histories differ across the family.

Results

Plastome and dataset characteristics

Plastomes of 210 individuals representing 182 palm species and 111 genera were newly sequenced in the present study (Additional file 1: Table S1). The average

sequencing coverage of these newly assembled plastomes ranged from $117\times$ to $2343\times$, while plastome length ranged from 153,584 bp (*Chuniophoenix suoitienensis* A.J. Hend.) to 160,194 bp (*Borassus madagascariensis* (Jum. & H. Perrier) Jum. & H. Perrier). However, the species *Tahina spectabilis* J. Dransf. & Rakotoarivivo, the whole plastome, which was sequenced by Barrett et al. [47], has the smallest size (126,251 bp) among all the palm plastomes analyzed in the present study due to the loss of one copy of the inverted repeat. For the 49 accessions that have limited plastid sequences included in the incomplete-105 regions matrix (including 105 plastid regions from 349 accessions, among which 49 do not have complete or nearly complete plastome data), 381 sequences of plastid genes and intergenic regions were obtained in total from the National Center for Biotechnology Information (NCBI) database (Additional file 2: Table S2). The complete-coding (including 83 plastid coding regions of 300 accessions that have complete or nearly complete plastome data), complete-105 regions (including 105 plastid regions of 300 accessions that have complete or nearly complete plastome data), and incomplete-105 regions matrices had aligned lengths of 77,140 bp, 101,842 bp, and 101,931 bp, respectively, and with 1.34%, 1.28%, and 15.32% missing data, respectively.

Phylogenetic relationships obtained from different data matrices

Within Areceae, phylogenetic relationships among the five subfamilies and 28 tribes were consistent and mostly strongly supported (or highly supported defined here as Bootstrap (BS) $\geq 85\%$) in the trees obtained from all three matrices (Fig. 1; Additional files 4, 5: Figs. S1, S2), except several tribal-level nodes moderately (defined here as $70\% \leq \text{BS} < 85\%$) or weakly (defined here as $\text{BS} < 70\%$) supported (Fig. 1; Additional files 4, 5: Fig. S1, S2), such as the sister relationships between the tribes Calameae and Lepidocaryeae (BS = 84%) within Calamoideae, between the RRC clade [66] and the core arecoids (BS = 84%), between (Areceae, Euterpeae) and the core core Geonomateae (BS < 50%), and between ((*Pholidostachys* H. Wendl. ex Hook. f., *Welfia* H. Wendl.), Manicarieae) and ((Areceae, Euterpeae), core Geonomateae) (BS = 84%) within Arecoideae in the incomplete-105 regions tree (Fig. 1), the crown of the RRC clade (BS = 82%) in the complete-coding tree (Additional file 4: Fig. S1), the sister relationship between (Areceae, Euterpeae) and the core core Geonomateae (BS = 81%) in the complete-105 regions tree (Additional file 5: Fig. S2), and the sister relationship between Areceae and Euterpeae (BSs = 71%, 57%, 73% in Fig. 1, Additional files 4, 5: Fig. S1, S2) in all three trees. Phylogenetic relationships at the subtribal level

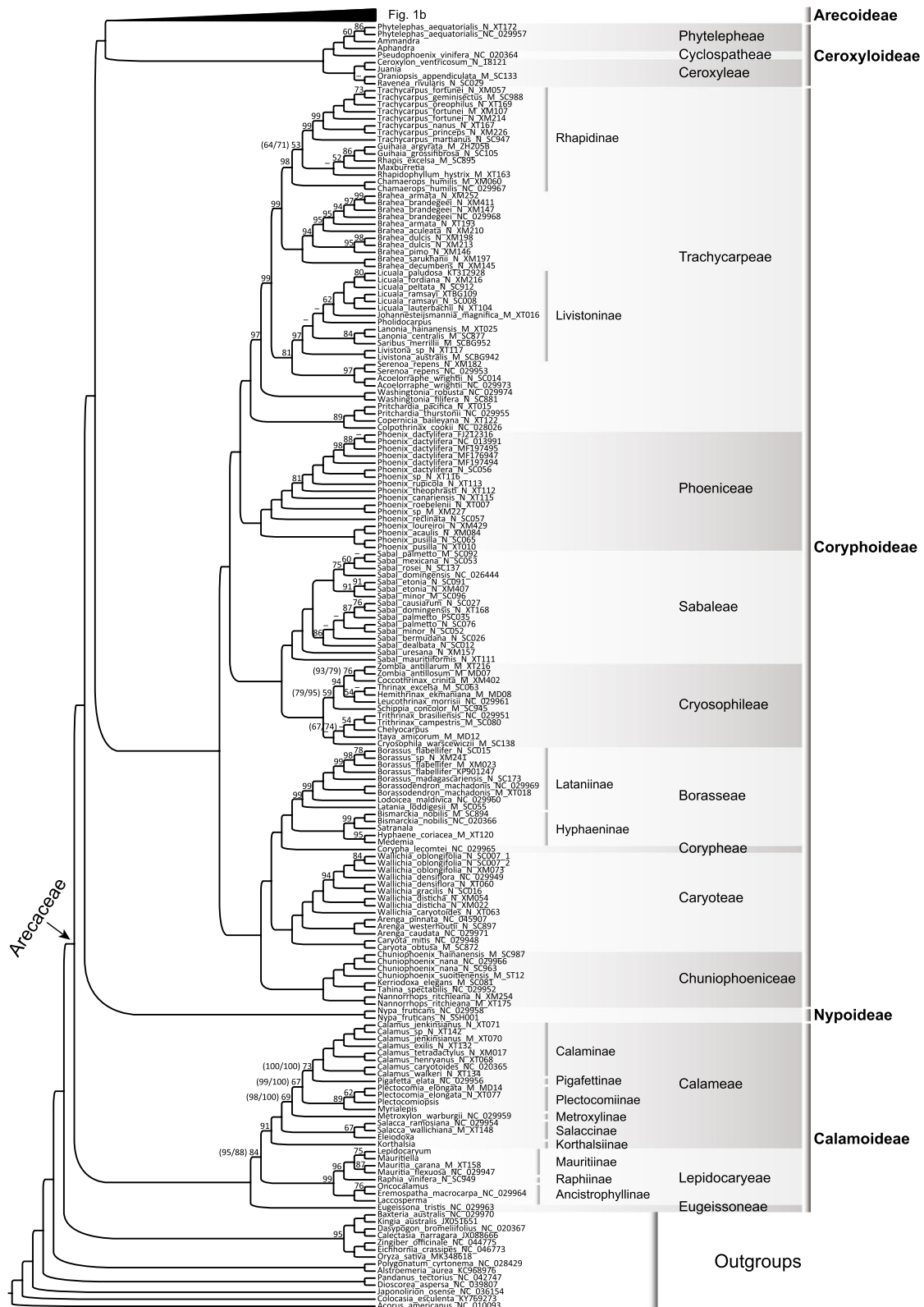


Fig. 1 Maximum likelihood (ML) phylogenetic tree of Areaceae inferred from the incomplete-105 regions matrix. Bootstrap support (BS) values inferior to 100% are shown, with dashes denoting a support inferior to 50%. Numbers in brackets indicate the BS values obtained from the analyses of the complete-coding/complete-105 matrices. Taxon names are followed by their GenBank accession numbers (sequences downloaded from NCBI) or voucher specimen information (plastomes newly obtained). For the 49 genera with a large proportion of missing data and for which individual plastid regions had to be retrieved from NCBI (see the “Methods” section), only the genus name is provided, but detailed sequence accession numbers can be found in Additional file 2: Table S2

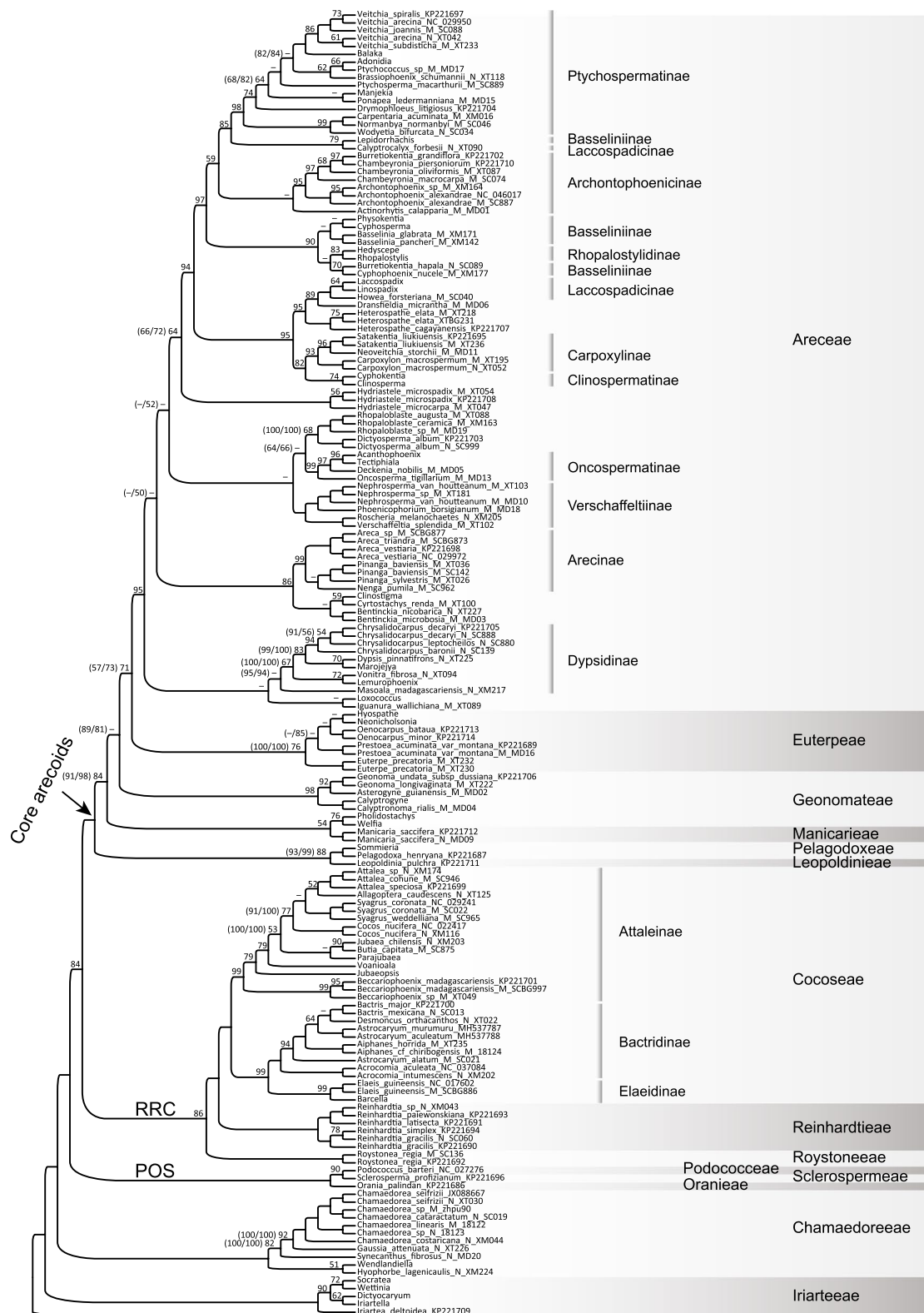


Fig. 1 continued

were also consistent and mostly strongly supported in all the trees (Fig. 1; Additional files 4, 5: S1, S2), except that several nodes within the tribe Calameae were weakly or moderately supported in the incomplete-105 regions tree (Fig. 1), and several nodes within Areceae were weakly or moderately supported in all the trees (Fig. 1b; Additional files 4, 5: Figs. S1, S2). The generic-level relationships within the family were also largely resolved with moderate to strong support (Fig. 1; Additional files 4, 5: Figs. S1, S2). Several generic-level nodes within the tribe Areceae showed conflicted topologies but were weakly supported in all the trees (Fig. 1b; Additional files 4, 5: Figs. S1, S2).

In the maximum likelihood (ML) tree based on the complete-105 regions matrix (Additional file 5: Fig. S2), stronger support values were recovered at most nodes compared with those from the complete-coding matrix (Additional file 4: Fig. S1). This was especially the case for the sister relationships between Mauritiinae and Raphiinae (from 80 to 99%; Additional files 4, 5: Figs. S1, S2), between ((Areceae, Euterpeae), Geonomateae) and Manicariaceae (from 91 to 98%; Additional files 4, 5: Figs. S1, S2), and between Areceae and Euterpeae (from 57 to 73%; Additional files 4, 5: Figs. S1, S2). Tribal- and subtribal-level placements of the 49 genera with a large proportion of missing data were mostly resolved with strong support in the analysis based on the incomplete-105 regions matrix (Fig. 1). Among them, the stem nodes of 29 genera had bootstrap support values $\geq 70\%$ (Fig. 1; Additional file 3: Table S3), including values $\geq 95\%$ for *Acanthophoenix* H. Wendl., *Aphandra* Barfod, *Balaka* Becc., *Barcella* Trail ex Drude, *Calyptrigyne* H. Wendl., *Juania* Drude, *Laccosperma* Drude, *Medemia* Wurttemb. ex H. Wendl., *Satranala* J. Dransf. & Beentje, *Sommieria* Becc., and *Tectiphiala* H.E. Moore, which all have the proportion of missing data over 89% in the incomplete-105 regions matrix (Additional file 3: Table S3). However, some deeper nodes associated with the stem nodes of these 49 genera that have a large proportion of missing data obtained lower supports from the incomplete-105 regions matrix (Fig. 1) than from the complete-105 regions matrix (Additional file 5: Fig. S2), such as the sister relationships between the core Geonomateae and (Areceae, Euterpeae) (81% vs. <50%), between subtribes Calaminae and Pigafettinae (100% vs. 73%), between Plectocomiinae and (Calaminae, Pigafettinae) (100% vs. 67%), and between Metroxylinae and ((Calaminae, Pigafettinae), Plectocomiinae) (100% vs. 69%).

Phylogenetic results based on the incomplete-105 regions matrix

Considering the highly congruent topologies obtained across all the three ML analyses, we focus here on

describing phylogenetic relationships mainly derived from the incomplete-105 regions matrix (Fig. 1), because this matrix included a more comprehensive taxon sampling compared with the other two matrices.

Arecaceae were placed within the commelinid monocots, and the monophyly of the family and the commelinid clade was both strongly supported (BS = 100%; Fig. 1). Within Arecaceae, the monophyly of all the five subfamilies was strongly supported (BSs = 100%), and phylogenetic relationships among these subfamilies were well-resolved with strong support values (BSs = 100%; Fig. 1). The subfamily Calamoideae was placed as sister to the remaining subfamilies and then followed successively by the monotypic subfamily Nypoideae and the second-largest subfamily Coryphoideae, which was placed as sister to the (Arecoideae, Ceroxyloideae) clade (Fig. 1).

Within Calamoideae, phylogenetic relationships among the three tribes (viz. Calameae, Eugeissonaeae, and Lepidocaryeae) were well-resolved and the sister relationship between Calameae and Lepidocaryeae was moderately supported (BS = 84%, Fig. 1). Within the tribe Lepidocaryeae, the subtribe Ancistrophyllinae was placed as sister to the (Mauritiinae, Raphiinae) clade, and relationships among the three subtribes were strongly supported (BSs $\geq 96\%$; Fig. 1). Within Calameae, the subtribe Korthalsiinae was placed as sister to the remaining members of the tribe with high support (BS = 91%; Fig. 1), and then followed by the subtribe Salaccinae (BS = 100%; Fig. 1), and Metroxylinae was placed as sister to the ((Calaminae, Pigafettinae), Plectocomiinae) clade. Relationships among the latter four subtribes were weakly or moderately supported.

Two major clades were recovered in the subfamily Coryphoideae: ((Phoeniceae, Trachycarpeae), (Cryosophileae, Sabaleae)) and (((Borasseae, Corypheeae), Caryoteae), Chuniophoeniceae), with relationships at the tribal level all strongly supported (BSs = 100%; Fig. 1). The monophyly of each of these tribes was also strongly supported (BSs = 100%), except for Corypheeae, represented here by only one accession of *Corypha lecomtei* Becc. ex Lecomte. Within Trachycarpeae, the phylogenetic positions of the seven genera (viz. *Acoelorrhaphe* H. Wendl., *Serenoa* Hook.f., *Brahea* Mart., *Colpothrinax* Schaedler, *Copernicia* Mart. ex Endl., *Pritchardia* Seem. & H. Wendl. ex H. Wendl., and *Washingtonia* H. Wendl.) that were not classified at subtribal level [40] were well-resolved with strong support. The clade ((*Copernicia*, *Pritchardia*), *Colpothrinax*) (BS = 89%; Fig. 1) was placed as sister to all the other members of Trachycarpeae with strong support (BS = 100%), then followed by the genus *Washingtonia*, also with strong support (BS = 97%). In addition, within this tribe, the genus *Brahea* was sister to

the subtribe Rhipidinae with strong support (BS=99%), and the (*Acoelorrhapha*, *Serenoa*) subclade (BS=97%) was sister to the subtribe Livistoninae (BS=81%). Monophyly of the two subtribes Hyphaeninae and Lataniinae, in the tribe Borasseae, was strongly supported (BSs \geq 99%).

Within the subfamily Ceroxyloideae, relationships among all the three tribes were well-resolved with strong support (BSs=100%; Fig. 1), with the tribe Ceroxyleae placed as sister to (Cyclospatheae, Phytelepheae). Furthermore, relationships among all eight genera of the subfamily were most strongly supported except for two nodes: (*Ammandra* O.F. Cook, *Phytelephas* Ruiz & Pav.) (BS=60%) and (*Oraniopsis* (Becc.) J. Dransf., A.K. Irvine & N.W. Uhl, *Ravenia* C.D. Bouché) (BS < 50%).

In the largest subfamily Arecoideae (Fig. 1b), the tribe Iriarteae was placed as sister to the rest of the subfamily, followed successively by Chamaedoreae and ((Podococceae, Sclerospermeae), Oranieae), i.e., the POS clade [66], and these relationships were all strongly supported (BSs=100%). Relationships among the three tribes within the POS clade were also well resolved with strong support (BSs \geq 90%). The remaining members of Arecoideae formed two major clades: the ((Cocoseae, Reinhardtiae), Roystoneae) (RRC [66]) clade and the core arecoids, the monophyly of which was strongly supported (BS=86% and 100%, respectively). Within the RRC clade, the sister relationship between Cocoseae and Reinhardtiae, and the relationships among the three subtribes ((Bactridinae, Elaeidinae), Attaleinae) of Cocoseae, all obtained strong support (BSs \geq 99%). Within the core arecoids, the two tribes Leopoldinieae and Pelagodoxeae formed a well-supported clade (BS=88%) and were placed as sister to the rest of the core arecoids with high support (BS=100%). The tribe Manicariae formed a weakly supported clade (BS=54%) with two genera (*Pholidostachys* and *Welfia*) of Geonomateae, while the core Geonomateae was sister to (Areceae, Euterpeae) with no support (BS < 50%). Additionally, the sister relationship between Areceae and Euterpeae was moderately supported (BS=71%).

Within the largest tribe Areceae, the monophyly of some subtribes, such as Arecinae (BS=99%), Carpoxylinae (93%), Oncospermatinae (99%), Ptychospermatinae (98%), and Verschaffeltiinae (100%) received strong support. Some lineages also received strong support, such as ((*Calypstrocalyx* Blume, *Lepidorrhachis* O.F. Cook), Ptychospermatinae) (BS=85%), (Basseliniinae (except *Lepidorrhachis*), Rhopalostylidinae) (BS=90%), (((Laccospadicinae (except *Calypstrocalyx*), *Dransfieldia* W.J. Baker & Zona), *Heterospatha* Scheff.), (Carpoxylinae, Clinospermatinae)) (BS=95%), and (((*Clinostigma* H. Wendl., *Cyrtostachys* Blume), *Bentinckia* Berry ex Roxb.), Arecinae) (BS=86%). However, the monophyly of both

Basseliniinae and Laccospadicinae was not supported here, because the genera *Calypstrocalyx* of Laccospadicinae and *Lepidorrhachis* of Basseliniinae formed a moderately supported clade (BS=79%) that was sister to Ptychospermatinae with high support (BS=85%), and was distantly related to the other members of either Basseliniinae or Laccospadicinae. Additionally, the monophyletic subtribe Rhopalostylidinae (BS=83%) was nested deeply within a clade composed of the members of Basseliniinae except *Lepidorrhachis*, although the placement of Rhopalostylidinae within this clade was weakly supported. Among the ten genera that were not classified at the subtribal level, the two genera *Dransfieldia* and *Heterospatha* formed a highly supported clade (BS=95%), which was sister to (Carpoxylinae, Clinospermatinae) with high support (BS=95%). The genus *Hydriastele* H. Wendl. & Drude represents an independent lineage within Areceae and was weakly supported as the sister of a large clade comprised by Archontophoenicinae, Basseliniinae, Carpoxylinae, Clinospermatinae, *Dransfieldia*, *Heterospatha*, Laccospadicinae, Rhopalostylidinae, Ptychospermatinae (BS=64%). The two genera *Dictyosperma* H. Wendl. & Drude and *Rhopaloblaste* Scheff. formed a weakly supported clade (BS=69%) that was sister to Oncospermatinae (BS < 50%). The three genera *Bentinckia*, *Clinostigma*, and *Cyrtostachys* formed a clade, but with no support (BS < 50%), that was sister to Arecinae with high support (BS=86%). Finally, the genus *Iguanura* Blume was recovered as sister to *Loxococcus* H. Wendl. & Drude, and then they were collectively sister to the subtribe Dypsidinae, but relevant nodes obtained no support (BS < 50%).

Discussion

Insights into the plastid phylogenomic resolution of the palms based on a nearly complete generic-level sampling

Our study represents the first phylogenomic analysis of palms based on both a comprehensive generic sampling and a large number of plastid genes. In the present study, several nodes weakly or moderately supported in the analysis of the incomplete-105 regions matrix obtained strong support in analyses of the other two matrices, such as the sister relationships between the tribes Calameae and Lepidocaryeae within Calamoideae (BS=84% in the incomplete-105 regions tree, Fig. 1; BSs=95%, 88%, in the complete-coding tree and complete-105 regions tree, Additional files 4, 5: Figs. S1, S2), and between (Areceae, Euterpeae) and the core Geonomateae within Arecoideae (BS < 50% in the incomplete-105 regions tree, Fig. 1; BS=89% in the complete-coding tree, Additional file 4: Fig. S1). Our results are largely congruent with previous plastid phylogenetic analyses [47, 48, 57, 58, 64]

but provide higher support for some tribal and generic relationships, especially those among Geonomateae, Leopoldinieae, Manicarieae, and Pelagodoxeae (Fig. 1; Additional files 4, 5: Figs. S1, S2).

The relationships among all the palm subfamilies and tribes were resolved with high support in the present study, except the sister relationship between Areceae and Euterpeae, which obtained moderate support (Fig. 1; Additional file 5: Fig. S2). However, this sister relationship was strongly supported in previous studies based on both nuclear genomic data [39] and plastome data [48, 58], as well as in a recent study based on a combined analysis of three nuclear regions (*RPB2*, *CISP4*, *WRKY6*) and one plastid region (*trnT-trnD*) [75]. Relationships among subtribes and genera of the family were also mostly resolved with moderate or high support (Fig. 1; Additional files 4, 5: Figs. S1, S2). However, clarifying some subtribal and generic relationships within the largest tribe Areceae is still a great challenge. Subtribes and genera within this tribe may have undergone rapid radiations, and perhaps including more species combined with the inclusion of more non-coding sequences (non-CDS) into further plastid phylogenomic analyses focused on this tribe would help improve the resolution of nodes that remain unclear, for instance, the placements of the genera *Actinorhytis*, *Iguanura*, *Loxococcus*, *Ponapea* and *Ptychosperma*, and subtribes Oncospermatinae, Verschaffeltiinae.

Previous theoretical [19, 76] and empirical studies [13, 14] have demonstrated that increased taxon sampling can improve resolution and support in phylogenetic analyses, even if including taxa with relatively slow substitution rates and large proportion of missing data. This was further supported by the present study. Our results from the analysis of the incomplete-105 regions matrix showed that moderate or strong support was obtained for the phylogenetic placements of nearly 60% of the 49 genera that have a large proportion of missing data (approximately up to 90% or even higher; Fig. 1; Additional file 3: Table S3). The well-supported phylogenetic positions of these genera indicate that their limited sequence data may be enough to clarify their placements on the phylogenetic tree. This is supported by the fact that, for many of these genera, the same phylogenetic position was recovered with high support in the KTLE nuclear tree (PP=1.00) [60]. This was notably the case for *Acanthophoenix*, *Balaka*, *Barcella*, *Calyptrogynne*, *Juania*, *Korthalsia* Blume, *Medemia*, *Myrialepis* Becc., *Satranala*, *Sommieria*, and *Tectiphiala*. Thus, the scaffold approach [14] of adding taxa with a large proportion of missing data into the analysis seems to be a good method to overcome the drawback of incomplete taxon sampling in phylogenomic studies.

Comparison between plastid and nuclear phylogenomic trees

The current version of the KTLE tree [60] has not yet been used as the basis for an in-depth discussion of palm relationships. Nevertheless, it provides important insight into the generic-level relationships of palms and is largely congruent with other published nuclear trees, such as the generic-level relationships within the subfamily Calamoideae [49] and the tribe Geonomateae [63], the tribal-level relationships within Calamoideae [49] and Ceroxyloideae [66], and the relationships among major clades of the subfamily Arecoideae [39] and among the five palm subfamilies [39, 49]. The plastome tree inferred here is complementary to these nuclear trees and brings some advantages because it is less likely to be biased by very high rates of substitution, paralogy, or incomplete lineage sorting, which can result in phylogenetic inference errors or conflicts among gene trees. On the other hand, comparing plastome and nuclear trees can reveal traces of past hybridization events that have disappeared from nuclear genomes, such as those involved in chloroplast capture [21, 77]. Considering these nuclear and plastid trees together thus can strengthen our understanding of the palm phylogeny and evolutionary history.

Phylogenetic topologies among all the five subfamilies, among major clades within Arecoideae, among most tribes within Coryphoideae, and among most subtribes and genera in the family were consistent in analyses based on both plastid (Fig. 1) [47, 48, 58] and nuclear data [39, 49, 59, 60, 63, 66, 78] and also mostly obtained strong supported. Additionally, although the sister relationships between Phoeniceae and Trachycarpeae within Coryphoideae recovered in the plastome tree (Fig. 1) were not derived from the analysis of 353 nuclear genes [59, 60], but it was highly supported in Faurby et al.'s [25] supertree analysis as well as Cano et al.'s [79] study based on analysis of four nuclear regions (*CISP4*, *CISP5*, *PRK*, *RPB2*) and one plastid region (*matK*).

Widespread nuclear-plastid discordance within each palm subfamily (excluding the monotypic subfamily Nypoideae) was observed at different taxonomic levels between our plastome tree (Fig. 1; Additional files 4, 5: Figs. S1, S2) and previously published nuclear trees [39, 49, 59, 60, 63, 66]. Detailed information about the discordances regarding relationships among tribes and subtribes is summarized in Fig. 2. For some of these discordances, the nuclear topology was most often better supported by morphological evidence than the plastid topology. For example, Calaminae and Plectocomiinae, which were placed as sisters in the nuclear trees [49, 59, 60] but not in our plastid trees (Fig. 2d), are mostly climbing palms that have bi-symmetric and aperturate pollen grains, while members of Pigafettinae are massive

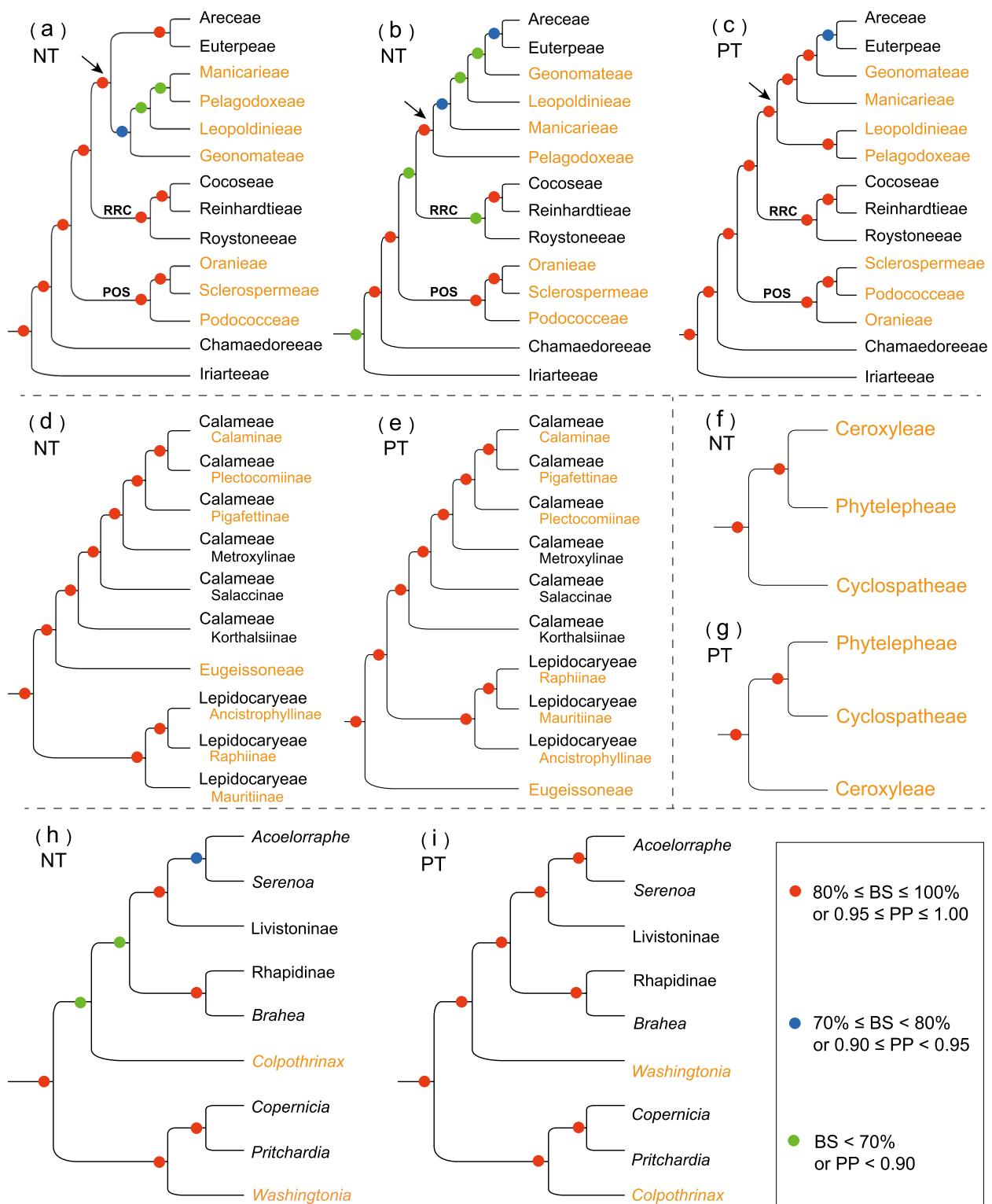


Fig. 2 Comparison between the palm topologies derived from nuclear and plastome data. **a–c** Tribal relationships in Arecoideae, the arrowhead indicates the crown of the core arecoids; **d, e** tribal and subtribal relationships in Calamoideae; **f, g** tribal relationships in Ceroxyloideae; **h, i** subtribal relationships in Trachycarpeae (Coryphoideae); NT nuclear topology, PT plastome topology; NT were obtained **a** from the maximum likelihood analysis of Comer et al. [39]; **b, h** from the multispecies coalescent analysis of KTLE [60] (KTLE tree); **d** from the multispecies coalescent analysis of Kuhnhauser et al. [49]; and **f** from the maximum parsimony analysis of Baker et al. [66]; PT of **c, e, g,** and **i** were all obtained from the present study

trees with sub-actinomorphic and inaperturate pollen grains [43]. However, for other discordances, the plastid topology is better supported by morphological evidence than the nuclear topology. This is the case for Calameae and Lepidocaryeae, which are sisters in our plastid topology (Fig. 1; Fig. 2e) but not in previous nuclear topologies (Fig. 2d; [49, 59, 60]). These tribes are characterized by flowers not borne in a cupule of bracts, less than 20 stamens and fruits lacking an endocarp, while the tribe Eugeissoneae is characterized by its large flowers that are born in a cupule of bracts, more than 20 stamens and fruits with a thick endocarp [43]. Finally, it is interesting to note that the plastid topology ((Podococceae, Sclerospermeae), Oranieae) is well supported by the geographic distribution of the tribes, as Podococceae and Sclerospermeae are endemic to western Africa, while Oranieae are mainly distributed from southeast Asia to New Guinea, with only three species in Madagascar [43].

Nuclear-plastid discordance has been reported widely in other angiosperm lineages both at deep and shallow nodes, such as the asterids [20], Caryophyllales [13, 80], Fagales [21], Fabaceae [31, 32], Orchidaceae [35], and *Magnolia* L. (Magnoliaceae) [81], and the discordance is mostly interpreted as resulting from hybridization and/or ILS [20, 21, 77, 81]. In the case of palms, our study suggests that the conflicting nuclear and plastid topologies could be mainly due to incomplete lineage sorting during rapid radiations, because the nodes involved tend to be associated with very short internal branches and long external branches (Additional file 6: Fig. S3). This is notably the case for the diversifications among major lineages within the core arecoids and the POS clade in Arecoideae, among tribes in Calamoideae, and along the backbone of Trachycarpeae in Coryphoideae. Incomplete lineage sorting is expected to make ancestral genetic polymorphisms persist during evolutionary radiations and could therefore have induced the observed phylogenetic incongruences [82]. However, such patterns of short internal branch coupled with long external branches resulting from fast diversification combined with a lack of species-level sampling can make phylogenetic inference prone to errors, and incongruences around these nodes may thus also be due to tree estimation errors. Comprehensive species-level sampling will be instrumental to avoid this pitfall in further studies. On the other hand, hybridization is another phenomenon that can lead to phylogenetic discordance, and it has been shown that palms can hybridize at least in gardens and have done so in the past [83]. Accordingly, some palm lineages involved in nuclear-plastid discordance have overlapping or closely adjacent distribution ranges, which may facilitate hybridization between them. This is for instance the case of Geonomateae, Leopoldinieae, and Manicarieae

(core arecoids), whose members often co-occur in Amazonia [43]. Hybridization therefore seems another plausible explanation for the palm nuclear-plastid discordances observed in the present study. Moreover, some of the conflicted topologies are highly supported in both the plastome tree (Fig. 1) and nuclear trees [39, 49, 59, 60, 66], suggesting possible hybridization events involved in chloroplast capture may have occurred in relevant nodes, such as those observed among tribes within the two subfamilies Calamoideae (Fig. 2d, e) and Ceroxyloideae (Fig. 2f, g), and among the three tribes within the POS clade (Fig. 2a–c). Further phylogenetic studies and gene tree frequency analyses will be necessary to enable the accurate detection and quantification of ILS and hybridization across the palms and to achieve higher phylogenetic resolution.

Implications for palm taxonomy

The palm family has benefited from considerable systematic studies over recent decades [24, 39–43, 48–50, 55–57, 61]. This rich history has led to continual improvements in our understanding of palm evolution and relationships, which underpins the current taxonomic classification. As circumscribed in the latest taxonomic treatment of the palms provided by Baker and Dransfield [40], members of the family are classified into five subfamilies and 28 tribes, and most genera within large tribes are also classified at the subtribal level.

Monophyly of all the five subfamilies and tribes circumscribed in Baker and Dransfield [40] was mostly highly supported here by plastome data. An exception is the tribe Geonomateae in the subfamily Arecoideae (Fig. 1), for which the two genera *Pholidostachys* and *Welfia* did not fall within the core Geonomateae but were sister to the tribe Manicarieae with weak support (BS=54%; Fig. 1). The sister relationship between (*Pholidostachys*, *Welfia*) and Manicarieae was also reported previously in an analysis based on four plastid regions and extensive generic sampling, with higher support (BS=89%) [57]. However, the monophyly of Geonomateae (including *Welfia* and *Pholidostachys*) was highly supported in nuclear phylogenomic studies [60, 63], with *Welfia* and *Pholidostachys* successively sister to the remaining members of Geonomateae with weak support in the KTLE tree [60] and high support in an analysis based on 795 nuclear genes [63].

The taxonomic circumscriptions of subtribes recognized within some large palm tribes by Baker and Dransfield [40] were also supported in the present plastome study (Fig. 1), with the exception of subtribes Basselininae and Laccospadicinae in the largest tribe Areceae. Similar results were also recovered in the KTLE nuclear tree [60]. Our results, combined with those from the KTLE

topology [60] and other recent phylogenetic studies of the palms [25, 64, 66, 79], indicate the need for some updates to the taxonomic classification of the family.

- (1) The highly supported placement of the genus *Calypstrocalyx* distant from the remaining members of the subtribe Laccospadicinae recovered in the plastome analyses (Fig. 1b) was also supported in the analysis of Baker et al. [66] based on two nuclear regions (*PRK*, *RPB2*) and the KTLE nuclear tree [60]. Thus, the circumscription of Laccospadicinae should be revised and the subtribal placement of *Calypstrocalyx* should be reassessed.
- (2) The monophyletic subtribe Rhopalostylidinae was nested deeply within the Basseliniinae with high support as recovered in the present plastome tree (Fig. 1b). A similar result for the placement of Rhopalostylidinae within Basseliniinae was also highly supported in Faurby et al. [25], and the placement of the genus *Rhopalostylis* (Rhopalostylidinae) within Basseliniinae was also highly supported in the KTLE nuclear tree [60]. We recommend that the Basseliniinae should be expanded to include Rhopalostylidinae.
- (3) The phylogenetic positions of multiple palm genera which had not been classified to subtribes before have been clarified in our study with moderate to strong support (BSs $\geq 70\%$; Fig. 1; Additional files 4, 5: Figs. S1, S2). These include *Bentinckia*, *Cyrtostachys*, *Dictyosperma*, *Dransfieldia*, *Heterospatha*, *Hydriastele*, and *Rhopaloblaste* in the tribe Areceae and *Acoelorrhapha*, *Brahea*, *Colpothrinax*, *Copernicia*, *Pritchardia*, *Serenoa*, and *Washingtonia* in the tribe Trachycarpeae. The isolated phylogenetic positions of these genera out of all currently well-defined palm subtribes were also recovered in Faurby et al. [25] and in the KTLE nuclear tree [60], and similar results for the placements of some of these genera were also supported in recent studies focused on different palm lineages [25, 64, 66, 79]. Improvements to the subtribal classification of palms may be achievable in light of these new results, especially in concert with the growing body of nuclear phylogenomic evidence.

Conclusions

Our study is the most comprehensive plastome-based phylogenomic analysis yet conducted for palms. Our results improve upon previous plastid-based studies that were limited either in taxon or in locus sampling, providing resolution and support throughout the palm family,

with most tribe- and genus-level nodes strongly supported. The plastome data provided here complement previously reported nuclear trees and our understanding of the evolutionary history of the palms, thus strengthening our understanding of palm relationships. The robust phylogenetic hypothesis of the palms reconstructed here will support future studies on the biogeography, classification, diversification, and evolution of the family. The present phylogenomic investigation of the palms provides a case study of how to use phylogenomics to disentangle relationships among lineages of a large angiosperm family. Comprehensive, species-level, nuclear phylogenomic studies of palms are also nearing completion [84] and will provide important opportunities for the evaluation of nuclear-plastid discordance.

Methods

Taxon sampling

Plastomes of 210 accessions representing 182 species and 111 genera of Areceaceae were newly sequenced for this project, and detailed information about these accessions and voucher specimens is provided in Additional file 1 (Table S1). We further added complete or nearly complete plastid genome sequences from NCBI (<https://www.ncbi.nlm.nih.gov/>), corresponding to 76 accessions representing 71 species and 63 genera of the palm family (45 of these genera were duplicate with those newly sequenced here) (Additional file 1: Table S1). Most of these were derived from the phylogenomic studies conducted by Barrett et al. [47] and Comer et al. [48]. In addition, another 49 palm accessions representing 49 genera that have at least five plastid DNA regions in NCBI (www.ncbi.nlm.nih.gov) were also included (Additional file 2: Table S2), most of which were from Baker et al. [24]. In total, the final taxon sampling included 335 palm accessions representing 276 species and 178 genera, accounting for 98.3% of all currently circumscribed palm genera, and representing all the subtribes, tribes, and subfamilies [40]. Three palm genera, viz. *Jailoloa* Heatubun & W.J. Baker, *Sabina-ria* R. Bernal & Galeano, and *Wallaceodoxa* Heatubun & W.J. Baker, which are recently described genera [40], were not sampled here because of the lack of DNA material or published plastid sequence data, but their phylogenetic positions within the family were resolved previously in analyses based mainly on nuclear data [60, 79, 85, 86]. Additionally, complete plastome sequences of four genera of the family Dasypogonaceae and ten genera representing the other ten monocot orders (Additional file 1: Table S1) were selected as outgroups based on the phylogenetic framework provided by Givnish et al. [87] and Li et al. [52].

Plastome DNA extraction, sequencing, assembly, and annotation

Total genomic DNAs were extracted from silica-dried leaves following the CTAB protocol of Doyle and Doyle [88]. DNAs were sheared to approximately 500-bp fragments through ultrasonic treatment and used to construct short-insert libraries following the manufacturer's protocol (NEBNext[®] Ultra II[™] DNA Library Prep Kit for Illumina[®]) and sequenced from both ends on the Illumina HiSeq 2500 platform at Beijing Genomics Institute (BGI, Shenzhen, China) to generate 2 × 150-bp sequencing reads. Approximately 3 GB of raw data was generated for each sample. Plastid reads were assembled using the software GetOrganelle [89] with parameter settings as follows: “-t 30 -R 15 -k 75, 85, 95, 105 -F embplant_pt,” using the plastid genomes of *Nypa fruticans* Wurmb (GenBank accession number: NC_029958) and *Veitchia arecina* Becc. (NC_029950) as references. All the plastid genes were then annotated using the software PGA [90], with the annotated plastome of *Amborella trichopoda* Baill. (NC_005086) as a reference, following the recommendation of Qu et al. [90]. GenBank accession numbers of the complete or nearly complete plastome newly sequenced here as well as those obtained from NCBI are listed in Additional file 1 (Table S1), and the GenBank accession numbers corresponding to the sequences from the 49 accessions with few plastid DNA regions available are listed in Additional file 2 (Table S2).

Phylogenetic dataset construction

Coding regions of 79 protein-coding genes, four ribosomal RNA genes, six transfer RNA genes with sequence lengths above 200 bp (viz. *trnA-UGC*, *trnG-UCU*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, and *trnV-UAC*), and intron regions of all 12 intron-containing coding genes (viz. *atpF*, *clpP*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl12*, *rpl16*, *rpoC1*, *rps12*, *rps16*, *ycf3*) were extracted from the plastomes as conducted in Geneious 11.1.5 [91]. Additionally, to coordinate with the 49 accessions for which at least five plastid regions were available, sequences of four intergenic spacer regions (viz. *rps15-ycf1*, *trnD-trnT*, *trnL-F*, and *trnQ-rps16*) were also extracted. Sequence matrices corresponding to each region were aligned independently using the plugin of MAFFT [92] in Geneious with the default settings. The removal of ambiguously aligned sites in 17 matrices (including that of the *trnK-UUU* region, all the 12 intron-containing and four intergenic regions) was conducted in Gblocks 0.91b [93], with the option “Allowed Gap Positions” set as “All.”

Three different data sets were constructed: (1) a complete-coding matrix, including all 83 coding regions (viz. 79 protein-coding genes and four ribosomal RNA genes) of 300 accessions (including 129 palm genera and 28 palm tribes) that possessed complete or nearly complete plastome data;

(2) a complete-105 regions matrix (including 105 plastid regions of 300 accessions), constructed by adding the complete-coding matrix to the other 22 plastid regions (i.e., six rRNA regions, 12 intron-containing regions, and four intergenic spacer regions); and (3) an incomplete-105 regions matrix (including 105 plastid regions of 349 accessions, among which 49 accessions had limited plastome data), constructed based on the complete-105 regions matrix, with the additional inclusion of the 49 accessions (representing 49 genera) with at least five plastid regions available in NCBI. Information about the proportion of missing data for each of the 49 genera included in the incomplete-105 regions matrix is presented in Additional file 3 (Table S3).

Phylogenetic analyses

The plastome has long been considered to comprise a single linkage group [94, 95], and thus, plastid genes are usually concatenated in order to maximize the overall phylogenetic signal [5, 13, 34]. In the present study, all three matrices were analyzed under an unpartitioned scheme. We inferred phylogenetic trees from the three matrices, using the maximum likelihood (ML) approach implemented in RAxML-HPC2 (8.1.24) [96] on the CIPRES cluster [97], employing the GTR + Γ model with the default number of rate categories ($C = 25$). We conducted a rapid bootstrap (BS) analysis with 1000 pseudoreplicates. The trees obtained were visualized and edited using FigTree v.1.4.4 [98]. The alignments and ML trees are available from figshare [99].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-023-01544-y>.

Additional file 1: Table S1. List of accessions sampled, with voucher specimens, GenBank accession numbers and the length of plastomes newly sequenced. Sequences marked with * were obtained from NCBI (<https://www.ncbi.nlm.nih.gov/>).

Additional file 2: Table S2. Accessions of the 49 genera with few plastid DNA regions available and therefore a large proportion of missing data in the incomplete-105 regions matrix, with GenBank accession numbers of sequences used in the present study. The character ‘—’ indicates that the sequence was unavailable from NCBI (<https://www.ncbi.nlm.nih.gov/>) and treated as missing data in the matrix.

Additional file 3: Table S3. Proportion of missing data for the 49 accessions (genera) with few DNA sequences available, compared to the total length of the incomplete-105 regions matrix, and bootstrap value for the stem nodes of these accessions (genera) obtained from maximum likelihood (ML) analysis of the incomplete-105 regions matrix (Fig. 1).

Additional file 4: Fig. S1. Maximum likelihood phylogenetic tree of Areaceae inferred from the complete-coding matrix. Bootstrap values inferior to 100% are shown, with dashes denoting a support inferior to 50%.

Additional file 5: Fig. S2. Maximum likelihood phylogenetic tree of Areaceae inferred from the complete-105 regions matrix. Bootstrap values inferior to 100% are shown, with dashes denoting a support inferior to 50%.

Additional file 6: Fig. S3. Maximum likelihood phylogenetic tree (including branch lengths) of Areaceae inferred from the incomplete-105 regions matrix.

Acknowledgements

The authors thank Yu-Ying Zhou for her kind help in molecular experiments and Duc-Thanh Le, Li-Xiu Guo, and Zhi-Ping Ruan for field collection. We thank South China Botanical Garden, Xiamen Botanical Garden, and Xishuangbanna Tropical Botanical Garden for permission to collect material. We thank Patrick Griffith, Larry Noblick, Joanna Tucker Lima, Chad Husby, Brett Jestrow, and the staff at both Montgomery Botanical Center and the Fairchild Tropical Botanical Garden for permission to collect material and accommodations for CFB.

Authors' contributions

This study was conceived by X.-J.G., C.B., and W.J.B. Analyses were performed by X.-J.G. and G.Y. The manuscript was written by G.Y. with contributions from C.B., B.X., S.B., W.J.B., and X.-J.G. The samples were collected by X.-J.G., Y.Q.Z., and C.B. The authors read and approved the final manuscript.

Funding

This work was financially supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB31000000), the International Partnership Program of the Chinese Academy of Science (grant number: 151853KYSB20190027), and the Biological Resources Program of the Chinese Academy of Sciences (ZSSD-009) through X.-J.G. We thank the West Virginia University Department of Biology and the International Palm Society for providing funding to C.F.B. W.J.B. and S.B. are supported by funding from the Calvea Foundation to the Royal Botanic Gardens Kew (Completing the Plant Tree of Life).

Availability of data and materials

All data generated in this study are available on the NCBI. The GenBank accession numbers are listed in Additional file 1: Table S1. All plant material collected for this study came from plants grown in a botanical garden. Sequence alignments underlying analyses and phylogenetic trees are available from figshare (<https://doi.org/10.6084/m9.figshare.20489916>) [99].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 18 August 2022 Accepted: 14 February 2023

Published online: 08 March 2023

References

- Zeng LP, Zhang Q, Sun R, Kong H, Zhang N, Ma H. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat Commun*. 2014;5(1):4956.
- One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*. 2019;574:679–85.
- Zhang C, Zhang T, Luebert F, Xiang Y, Huang C-H, Hu Y, et al. Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Mol Biol Evol*. 2020;37(11):3188–210.
- Guo X, Fang D, Sahu SK, Yang S, Guang X, Folk R, et al. *Chloranthus* genome provides insights into the early diversification of angiosperms. *Nat Commun*. 2021;12:6930.
- Li H-T, Luo Y, Gan L, Ma P-F, Gao L-M, Yang J-B, et al. Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biol*. 2021;19(1):232.
- Barrett CF, Davis JJ, Leebens-Mack J, Conran JG, Stevenson DW. Plastid genomes and deep relationships among the commelinid monocot angiosperms. *Cladistics*. 2013;29(1):65–87.
- Guo C, Luo Y, Gao L-M, Yi T-S, Li H-T, Yang J-B, et al. Phylogenomics and the flowering plant tree of life. *J Integr Plant Biol*. 2022; published online: doi: <https://doi.org/10.1111/jipb.13415>
- Susko E, Roger AJ. Long branch attraction biases in phylogenetics. *Syst Biol*. 2021;70(4):838–43.
- Chalopin D, Clark LG, Wysocki WP, Park M, Duvall MR, Bennetzen JL. Integrated genomic analyses from low-depth sequencing help resolve phylogenetic incongruence in the Bamboos (Poaceae: Bambusoideae). *Front Plant Sci*. 2021;12: 725728.
- Ma P-F, Zhang Y-X, Zeng C-X, Guo Z-H, Li D-Z. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). *Syst Biol*. 2014;63(6):933–50.
- Rothfels CJ. Polyploid phylogenetics. *New Phytol*. 2021;230(1):66–72.
- Galtier N, Daubin V. Dealing with incongruence in phylogenomic analyses. *Phil Trans R Soc B*. 2008;363:4023–9.
- Yao G, Jin J-J, Li H-T, Yang J-B, Mandala VS, Croley M, et al. Plastid phylogenomic insights into the evolution of Caryophyllales. *Mol Phylogenet Evol*. 2019;134:74–86.
- Xi Z, Ruhfel BR, Schaefer H, Amorim AM, Sugumaran M, Wurdack KJ, et al. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proc Nat Acad Sci USA*. 2012;109(43):17519–24.
- Jiang W, Chen S-Y, Wang H, Li D-Z, Wiens JJ. Should genes with missing data be excluded from phylogenetic analyses? *Mol Phylogenet Evol*. 2014;80:308–18.
- Pick KS, Philippe H, Schreiber F, Erpenbeck D, Jackson DJ, Wrede P, et al. Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. *Mol Biol Evol*. 2010;27(9):1983–7.
- Jeffroy O, Brinkmann H, Delsuc F, Philippe H. Phylogenomics: the beginning of incongruence? *Trends Genet*. 2006;22(4):225–31.
- Dávalos LM, Perkins SL. Saturation and base composition bias explain phylogenomic conflict in Plasmodium. *Genomics*. 2008;91(5):433–42.
- Wiens JJ. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Syst Biol*. 2005;54(5):731–42.
- Stull GW, Soltis PS, Soltis DE, Gitzendanner MA, Smith SA. Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. *Am J Bot*. 2020;107(5):790–805.
- Yang Y-Y, Qu X-J, Zhang R, Stull GW, Yi T-S. Plastid phylogenomic analyses of Fagales reveal signatures of conflict and ancient chloroplast capture. *Mol Phylogenet Evol*. 2021;163: 107232.
- Gallagher TJ, Peterson PM, Soreng RJ, Zuloaga FO, Li DZ, Clark LG, et al. Grasses through space and time: an overview of the biogeographical and macroevolutionary history of Poaceae. *J Syst Evol*. 2022;60(3):522–69.
- Li B, Cantino PD, Olmstead RG, Bramley GL, Xiang C-L, Ma Z-H, et al. A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification. *Sci Rep*. 2016;6(1):1–18.
- Baker WJ, Savolainen V, Asmussen-Lange CB, Chase MW, Dransfield J, Forest F, et al. Complete generic-level phylogenetic analyses of palms (Arecaceae) with comparisons of supertree and supermatrix approaches. *Syst Biol*. 2009;58(2):240–56.
- Faurby S, Eiserhardt WL, Baker WJ, Svenning J-C. An all-evidence species-level supertree for the palms (Arecaceae). *Mol Phylogenet Evol*. 2016;100:57–69.
- Zhao L, Yang Y-Y, Qu X-J, Ma H, Hu Y, Li H-T, et al. Phylotranscriptomic analyses reveal multiple whole-genome duplication events, the history of diversification and adaptations in the Araceae. *Ann Bot*. 2023;131(1):199–214.
- Fishbein M, Livshultz T, Straub SC, Simões AO, Boutte J, McDonnell A, et al. Evolution on the backbone: Apocynaceae phylogenomics and new perspectives on growth forms, flowers, and fruits. *Am J Bot*. 2018;105(3):495–513.
- Huang C-H, Zhang C, Liu M, Hu Y, Gao T, Qi J, et al. Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. *Mol Biol Evol*. 2016;33(11):2820–35.
- Liu LM, Du XY, Guo C, Li DZ. Resolving robust phylogenetic relationships of core Brassicaceae using genome skimming data. *J Syst Evol*. 2021;59(3):442–53.
- Guo J, Xu W, Hu Y, Huang J, Zhao Y, Zhang L, et al. Phylotranscriptomics in Cucurbitaceae reveal multiple whole-genome duplications

- and key morphological and molecular innovations. *Mol Plant*. 2020;13(8):1117–33.
31. Zhang R, Wang Y-H, Jin J-J, Stull GW, Bruneau A, Cardoso D, et al. Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. *Syst Biol*. 2020;69(4):613–22.
 32. Zhao Y, Zhang R, Jiang K-W, Qi J, Hu Y, Guo J, et al. Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. *Mol Plant*. 2021;14(5):748–73.
 33. Ogutcen E, Christe C, Nishii K, Salamin N, Möller M, Perret M. Phylogenomics of Gesneriaceae using targeted capture of nuclear genes. *Mol Phylogenet Evol*. 2021;157: 107068.
 34. Zhao F, Chen Y-P, Salmaki Y, Drew BT, Wilson TC, Scheen A-C, et al. An updated tribal classification of Lamiaceae based on plastome phylogenomics. *BMC Biol*. 2021;19(1):1–27.
 35. Pérez-Escobar OA, Dodsworth S, Bogarín D, Bellot S, Balbuena JA, Schley RJ, et al. Hundreds of nuclear and plastid loci yield novel insights into orchid relationships. *Am J Bot*. 2021;108(7):1166–80.
 36. Huang W, Zhang L, Columbus JT, Hu Y, Zhao Y, Tang L, et al. A well-supported nuclear phylogeny of Poaceae and implications for the evolution of C₄ photosynthesis. *Mol Plant*. 2022;15(4):755–77.
 37. Zhang SD, Jin JJ, Chen SY, Chase MW, Soltis DE, Li HT, et al. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. *New Phytol*. 2017;214(3):1355–67.
 38. Xiang Y, Huang C-H, Hu Y, Wen J, Li S, Yi T, et al. Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol Biol Evol*. 2017;34(2):262–81.
 39. Comer JR, Zomlefer WB, Barrett CF, Stevenson DW, Heyduk K, Leebens-Mack JH, et al. Nuclear phylogenomics of the palm subfamily Arecoideae (Arecaceae). *Mol Phylogenet Evol*. 2016;97:32–42.
 40. Baker WJ, Dransfield J. Beyond Genera Palmarum: progress and prospects in palm systematics. *Bot J Linn Soc*. 2016;182(2):207–33.
 41. Eiserhardt WL, Bellot S, Cowan RS, Dransfield J, Hansen LESF, Heyduk K, et al. Phylogenomics and generic limits of Dyspidinae (Arecaceae), the largest palm radiation in Madagascar. *Taxon*. 2022;71(6):1170–95.
 42. Hodel DR, Baker WJ, Bellot S, Pérez-Calle V, Cumberledge A, Barrett CF. Reassessment of the Archontophoenicinae of New Caledonia and description of a new species. *Palms*. 2021;65(3):109–31.
 43. Dransfield J, Uhl NW, Asmussen CB, Baker WJ, Harley MM, Lewis CE. Genera Palmarum—the evolution and classification of palms. Richmond (UK): Royal Botanic Gardens, Kew. 2008;732 pp.
 44. Couvreur TL, Forest F, Baker WJ. Origin and global diversification patterns of tropical rain forests: inferences from a complete genus-level phylogeny of palms. *BMC Biol*. 2011;9:44.
 45. Lim JY, Huang H, Farnsworth A, Lunt DJ, Baker WJ, Morley RJ, et al. The Cenozoic history of palms: global diversification, biogeography and the decline of megathermal forests. *Glob Ecol Biogeogr*. 2022;31(3):425–39.
 46. Cámara-Leret R, Faurby S, Macía MJ, Balslev H, Gödel B, Svenning J-C, et al. Fundamental species traits explain provisioning services of tropical American palms. *Nat Plants*. 2017;3(2):1–7.
 47. Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, et al. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytol*. 2016;209(2):855–70.
 48. Comer JR, Zomlefer WB, Barrett CF, Davis JI, Stevenson DW, Heyduk K, et al. Resolving relationships within the palm subfamily Arecoideae (Arecaceae) using plastid sequences derived from next-generation sequencing. *Am J Bot*. 2015;102(6):888–99.
 49. Kuhnhäuser BG, Bellot S, Couvreur TL, Dransfield J, Henderson A, Schley R, et al. A robust phylogenomic framework for the calamoid palms. *Mol Phylogenet Evol*. 2021;157: 107067.
 50. Dransfield J, Uhl NW, Asmussen CB, Baker WJ, Harley MM, Lewis CE. A new phylogenetic classification of the palm family. *Arecaceae Kew Bull*. 2005;60:559–69.
 51. Schley RJ, Pellicer J, Ge X-J, Barrett C, Bellot S, Guignard MS, Novák P, Suda J, Fraser D, Baker WJ, et al. The ecology of palm genomes: repeat-associated genome size expansion is constrained by aridity. *New Phytol*. 2022;236:433–46.
 52. Li H-T, Yi T-S, Gao L-M, Ma P-F, Zhang T, Yang J-B, et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants*. 2019;5(5):461–70.
 53. Timilsena PR, Wafula EK, Barrett CF, Ayyampalayam S, McNeal JR, Rentsch JD, et al. Phylogenomic resolution of order- and family-level monocot relationships using 602 single-copy nuclear genes and 1375 BUSCO genes. *Front Plant Sci*. 2022;13: 876779.
 54. Baker WJ, Asmussen CB, Barrow SC, Dransfield J, Hedderson TA. A phylogenetic study of the palm family (Palmae) based on chloroplast DNA sequences from the *trnL-trnF* region. *Plant Syst Evol*. 1999;219(1):111–26.
 55. Asmussen CB, Chase MW. Coding and noncoding plastid DNA in palm systematics. *Am J Bot*. 2001;88(6):1103–17.
 56. Hahn WJ. A molecular phylogenetic study of the Palmae (Arecaceae) based on *atpB*, *rbcL*, and 18S nrDNA sequences. *Syst Biol*. 2002;51(1):92–112.
 57. Asmussen CB, Dransfield J, Deickmann V, Barford AS, Pintaud J-C, Baker WJ. A new subfamily classification of the palm family (Arecaceae): evidence from plastid DNA phylogeny. *Bot J Linn Soc*. 2006;151(1):15–38.
 58. Chen DJ, Landis JB, Wang HX, Sun QH, Wang Q, Wang HF. Plastome structure, phylogenomic analyses and molecular dating of Arecaceae. *Front Plant Sci*. 2022;13: 960588.
 59. Baker WJ, Bailey P, Barber V, Barker A, Bellot S, Bishop D, et al. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Syst Biol*. 2022;71(2):301–19.
 60. The Kew Tree of Life Explorer (KTLE). 2022 continuously updated. <https://treeoflife.kew.org/tree-of-life>. (Data release 2.0, January 2022; Accessed 1 June 2022).
 61. Jiménez MFT, Prata E, Zizka A, Cohn-Haft M, de Oliveira AV, Emilio T, et al. Phylogenomics of the palm tribe Lepidocaryeae (Calamoideae: Arecaceae) and description of a new species of *Mauritiella*. *Syst Bot*. 2021;46(3):863–74.
 62. Escobar S, Helmstetter AJ, Montúfar R, Couvreur TL, Balslev H. Phylogenomic relationships and historical biogeography in the South American vegetable ivory palms (Phytelephea). *Mol Phylogenet Evol*. 2022;166: 107314.
 63. Loiseau O, Olivares I, Paris M, de La Harpe M, Weigand A, Koubínová D, et al. Targeted capture of hundreds of nuclear genes unravels phylogenetic relationships of the diverse Neotropical palm tribe Geonomateae. *Front Plant Sci*. 2019;10:864.
 64. Barrett CF, Sinn BT, King LT, Medina JC, Bacon CD, Lahmeyer SC, et al. Phylogenomics, biogeography and evolution in the American genus *Brahea* (Arecaceae). *Bot J Linn Soc*. 2019;190(3):242–59.
 65. Helmstetter AJ, Kamga SM, Bethune K, Lautenschläger T, Zizka A, Bacon CD, et al. Unraveling the phylogenomic relationships of the most diverse African palm genus *Raphia* (Calamoideae, Arecaceae). *J Plants*. 2020;9(4):549.
 66. Baker WJ, Norup MV, Clarkson JJ, Couvreur TL, Dowe JL, Lewis CE, et al. Phylogenetic relationships among arecoid palms (Arecaceae: Arecaceae). *Ann Bot*. 2011;108(8):1417–32.
 67. Moner AM, Furtado A, Henry RJ. Chloroplast phylogeography of AA genome rice species. *Mol Phylogenet Evol*. 2018;127:475–87.
 68. Yu Y, Yang JB, Ma WZ, Pressel S, Liu HM, Wu YH, et al. Chloroplast phylogenomics of liverworts: a reappraisal of the backbone phylogeny of liverworts with emphasis on Ptilidiales. *Cladistics*. 2020;36(2):184–93.
 69. Du X-Y, Lu J-M, Zhang L-B, Wen J, Kuo L-Y, Mynssen CM, et al. Simultaneous diversification of Polypodiaceae and angiosperms in the Mesozoic. *Cladistics*. 2021;37(5):518–39.
 70. Wei R, Yang J, He LJ, Liu HM, Hu JY, Liang SQ, et al. Plastid phylogenomics provides novel insights into the infrafamilial relationship of Polypodiaceae. *Cladistics*. 2021;37(6):717–27.
 71. Qu X-J, Wu C-S, Chaw S-M, Yi T-S. Insights into the existence of isomeric plastomes in Cupressaceae (Cupressaceae). *Genome Biol Evol*. 2017;9(4):1110–9.
 72. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc Natl Acad Sci USA*. 2010;107(10):4623–8.
 73. Gitzendanner MA, Soltis PS, Wong GKS, Ruhfel BR, Soltis DE. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *Am J Bot*. 2018;105(3):291–301.
 74. Xu L-S, Herrando-Moraira S, Susanna A, Galbany-Casals M, Chen Y-S. Phylogeny, origin and dispersal of *Saussurea* (Asteraceae) based on chloroplast genome data. *Mol Phylogenet Evol*. 2019;141: 106613.
 75. Pichardo-Marciano FJ, Nieto-Blázquez ME, MacDonald AN, Galeano G, Roncal J. Phylogeny, historical biogeography and diversification rates in

- an economically important group of Neotropical palms: Tribe Euterpeae. *Mol Phylogenet Evol.* 2019;133:67–81.
76. Wiens JJ. Missing data, incomplete taxa, and phylogenetic accuracy. *Syst Biol.* 2003;52(4):528–38.
 77. Liu B-B, Ren C, Kwak M, Hodel RGJ, Xu C, He J, et al. Phylogenomic conflict analyses in the apple genus *Malus* s.l. reveal widespread hybridization and allopolyploidy driving diversification, with insights into the complex biogeographic history in the Northern Hemisphere. *J Integr Plant Biol.* 2022;64:1020–1043.
 78. Osborne OG, Ciezarek A, Wilson T, Crayn D, Hutton I, Baker WJ, et al. Speciation in *Howea* palms occurred in sympatry, was preceded by ancestral admixture, and was associated with edaphic and phenological adaptation. *Mol Biol Evol.* 2019;36(12):2682–97.
 79. Cano Á, Bacon CD, Stauffer FW, Antonelli A, Serrano-Serrano ML, Perret M. The roles of dispersal and mass extinction in shaping palm diversity across the Caribbean. *J Biogeogr.* 2018;45(6):1432–43.
 80. Walker JF, Yang Y, Feng T, Timoneda A, Mikenas J, Hutchison V, et al. From cacti to carnivores: improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. *Am J Bot.* 2018;105(3):446–62.
 81. Dong SS, Wang YL, Xia NH, Liu Y, Liu M, Lian L, et al. Plastid and nuclear phylogenomic incongruences and biogeographic implications of *Magnolia* s.l. (Magnoliaceae). *J Syst Evol.* 2022;60(1):1–15.
 82. Feng S, Bai M, Rivas-González I, Li C, Liu S, Tong Y, et al. Incomplete lineage sorting and phenotypic evolution in marsupials. *Cell.* 2022;185(10):1646–60.
 83. Flowers JM, Hazzouri KM, Gros-Balthazard M, Mo Z, Koutroumpa K, Perakis A, et al. Cross-species hybridization and the origin of North African date palms. *Proc Natl Acad Sci USA.* 2019;116(5):1651–1658.
 84. Bellot S, Odufuwa P, Dransfield J, Eiserhardt WL, Perez-Escobar OA, Petoe P, et al. Why and how to develop DNA barcoding for Palms? A case study of Pinanga Palms. 2020;64(3):109–20.
 85. Alapetite E, Baker WJ, Nadot S. Evolution of stamen number in Ptychospermatinae (Arecaceae): insights from a new molecular phylogeny of the subtribe. *Mol Phylogenet Evol.* 2014;76:227–40.
 86. Heatubun CD, Zona S, Baker WJ. Three new genera of arecoid palm (Arecaceae) from eastern Malesia. *Kew Bull.* 2014;69(3):1–18.
 87. Givnish TJ, Zuluaga A, Spalink D, Soto Gomez M, Lam VK, Saarela JM, et al. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *Am J Bot.* 2018;105(11):1888–910.
 88. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 1987;19:11–5.
 89. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 2020;21(1):1–31.
 90. Qu X-J, Moore MJ, Li D-Z, Yi T-S. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods.* 2019;15:50.
 91. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Syst Biol.* 2012;28(12):1647–9.
 92. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
 93. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 2000;17(4):540–52.
 94. Vogl C, Badger J, Kearney P, Li M, Clegg M, Jiang T. Probabilistic analysis indicates discordant gene trees in chloroplast evolution. *J Mol Evol.* 2003;56(3):330–40.
 95. Birky CW Jr. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc Natl Acad Sci USA.* 1995;92(25):11331–8.
 96. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006;22(21):2688–90.
 97. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetics trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans, LA. 2010; pp.1–8.
 98. Rambaut A. FigTree, v.1.4.4. 2018. <http://tree.bio.ed.ac.uk/software/figtree/>. Accessed 16 Jan. 2022.
 99. Yao G, Zhang YQ, Barrett C, Xue B, Bellot S, Baker WJ, Ge XJ. A plastid phylogenomic framework for the palm family (Arecaceae). 2023. Figshare Dataset. <https://doi.org/10.6084/m9.figshare.20489916>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

