White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/186171/

Version: Accepted Version

Article:
Kuhnhäuser, B.G., Bellot, S., Couvreur, T.L.P. et al. (2021) A robust phylogenomic framework for the calamoid palms. Molecular Phylogenetics and Evolution, 157. 107067. ISSN: 1055-7903

https://doi.org/10.1016/j.ympev.2020.107067

Article available under the terms of the CC-BY-NC-ND licence (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Accepted version

A robust phylogenomic framework for the calamoid palms

Molecular Phylogenetics and Evolution, Volume 157, April 2021, 107067

https://doi.org/10.1016/j.ympev.2020.107067

Benedikt G. Kuhnhäuser[a,b], Sidonie Bellot[b], Thomas L. P. Couvreur[c], John Dransfield[b], Andrew Henderson[d], Rowan Schley[b], Guillaume Chomicki[e], Wolf L. Eiserhardt[b,f], Simon J. Hiscock[a], William J. Baker[b,*]

[a] Department of Plant Sciences, University of Oxford, United Kingdom
[b] Royal Botanic Gardens, Kew, Richmond, United Kingdom
[c] IRD, DIADE, Université de Montpellier, Montpellier, France
[d] New York Botanical Garden, The Bronx, United States
[e] Department of Biosciences, Durham University, United Kingdom
[f] Department of Biology, Aarhus University, Denmark

*Corresponding author at: Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, United Kingdom. E-mail address: w.baker@kew.org (W.J. Baker).

Abstract

Well-supported phylogenies are a prerequisite for the study of the evolution and diversity of life on earth. The subfamily Calamoideae accounts for more than one fifth of the palm family (Arecaceae), occurs in tropical rainforests across the world, and supports a billion-dollar industry in rattan products. It contains ca. 550 species in 17 genera, 10 subtribes and three tribes, but their phylogenetic relationships remain insufficiently understood. Here, we sequenced almost one thousand nuclear genomic regions for 75 systematically selected Calamoideae, representing the taxonomic diversity within all calamoid genera. Our phylogenomic analyses resolved a maximally supported phylogenetic backbone for the Calamoideae, including several higher-level relationships not previously inferred. In-depth analysis revealed low gene tree conflict for the backbone but complex deep evolutionary histories within several subtribes. Overall, our phylogenomic framework sheds new light on the evolution of palms and provides a robust foundation for future comparative studies, such as taxonomy, systematics, biogeography, and macroevolutionary research.

1 Introduction

Tropical rainforests are the most diverse terrestrial biome, harboring almost half of all plant species in just 7 % of the Earth's land surface (Eiserhardt et al., 2017). A deeper understanding of their origin, assembly and diversification can be attained by studying phylogenies of important rainforest groups (Pennington et al., 2004; Davis et al., 2005; Couvreur et al., 2011; Eiserhardt et al., 2017; Ramírez-Barahona et al., 2020). The palm family (Arecaceae Bercht. & J.Presl) is an established model group for the study of tropical rainforest evolution (Couvreur and Baker, 2013) and has been used to gain insights into the assembly (Couvreur et al., 2011; Kissling et al., 2012a; Baker and Couvreur, 2013a, b) and distribution (Kissling et al., 2012b; Reichgelt et al., 2018) of tropical rainforests through time. However, phylogenetic relationships in several important groups of palms, such as the diverse subfamily Calamoideae Griff., remain incompletely understood (Baker and Dransfield, 2016).

The Calamoideae is one of five currently recognized subfamilies of the palms (Dransfield et al., 2008) and strongly supported as sister to all other palms (Asmussen and Chase, 2001; Hahn, 2002; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016a; Faurby et al., 2016). It contains ca. 550 species in 17 genera, 10 subtribes and three tribes (Table 1), accounting for one quarter of the palm family. Calamoid palms occur in tropical rainforests across the world and are most diverse in the Asia-Pacific region (Table 1). They have a multitude of uses as food or construction materials, including the utilization of the stems of many climbing species as raw materials for a multi-billion dollar rattan cane furniture industry (Sunderland and Dransfield, 2002). Members of the Calamoideae are easily identified by their fruits with overlapping scales (Fig. 1 a) and the presence of spines (Fig. 1 b). They also exhibit an extraordinary variety of growth forms (Fig. 1 c-e), the climbing habit (Fig. 1 e) being predominant (Kissling et al., 2019).

The monophyly of the Calamoideae is unequivocally supported by phylogenetic evidence (Uhl et al., 1995; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016a) and a large suite of well-defined reproductive and morphological-anatomical synapomorphies (Uhl and Moore Jr., 1971; Baker et al., 1999; Dransfield et al., 2008; Horn et al., 2009; Tomlinson et al., 2011). Within the subfamily, the monophyly of most higher-level taxa has been tested in previous studies, with the majority of genera represented by at least two species in the foundational phylogenetic studies of the Calamoideae (Baker et al., 2000a, b). These studies resulted in a phylogenetic classification of the Calamoideae (Baker et al., 2000a) that is followed in the prevailing classification of palms (Baker and Dransfield, 2016; Dransfield et al., 2008). The only major development since the classification of Baker et al. (2000a) relates to subtribe Calaminae, in which *Calamus sensu* Dransfield et al. (2008) had been inferred to be paraphyletic (Baker et al., 2000c). Five genera nested within *Calamus* have now been synonymized to render the genus monophyletic (*Calospatha* Becc., *Ceratolobus* Blume ex Schult. & Schult.f., *Daemonorops* Blume, *Pogonotium* J.Dransf., and *Retispatha* J.Dransf.; Baker and Dransfield, 2008; Baker, 2015; Henderson and Floda, 2015). However, several knowledge gaps regarding the monophyly of genera and higher taxa remain. Although generally resolved as monophyletic, tribe Lepidocaryeae was inferred as paraphyletic by some analyses (Baker et al., 2009; Faurby et al., 2016). Five of the eighteen sections of *Calamus* (Beccari, 1908), as well as the synonymized genera *Schizospatha* Furtado, *Zalaccella* Becc. (both synonyms to *Calamus*) and *Lophospatha* Burret (synonym to *Salacca*), remain unsampled. Finally, morphologically intermediate

species (e.g. in Salaccinae; Dransfield et al., 2008) call current generic boundaries into question, but were not included in previous phylogenetic studies.

Several previous studies have investigated higher-level phylogenetic relationships within the Calamoideae (Fig. 2). In Lepidocaryeae, one single topology has been resolved by the majority of previous studies with intermediate to maximal support, with Raphiinae placed as sister to Mauritiinae, *Mauritia* as sister to *Mauritiella*, and *Eremospatha* as sister to *Laccosperma* (Baker et al., 2000a, b; Asmussen et al., 2006; Baker et al., 2009). In contrast, recent work placed Ancistrophyllinae as sister to Mauritiinae (Faye et al., 2016). In Calameae, Korthalsiinae has been placed as sister to the other subtribes by most studies (Baker et al., 2000b; Asmussen et al., 2006; Baker et al., 2009; Faurby et al., 2016; Shahimi, 2018). Among the remaining subtribes, Salaccinae has been generally resolved as sister to all other taxa (Baker et al., 2000b; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016a; Faurby et al., 2016). Other relationships of Korthalsiinae and Salaccinae were only rarely resolved, and with low support. Contested relationships remain between the three tribes, and among several subtribes of Calameae. Tribe Calameae was placed by most studies as sister to Lepidocaryeae (Baker et al., 2000a, b; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016a; Shahimi, 2018), but recent analyses resolved Calameae as sister to Eugeissoneae with intermediate support (Barrett et al., 2016a; Faurby et al., 2016). In Calameae, multiple contrasting relationships between the subtribes Calaminae, Metroxylinae, Pigafettinae and Plectocomiinae have been resolved (Baker et al., 2000a, b; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016a; Faurby et al., 2016). In Plectocomiinae, *Plectocomia* has been resolved either as sister to *Plectocomiopsis* (Baker et al., 2000b; Faurby et al., 2016) or as sister to both *Myrialepis* and *Plectocomiopsis* (Baker et al., 2000a; Baker et al., 2009; Shahimi, 2018).

Overall, this lack of resolution in higher-level relationships may be in part due to the limited sampling of molecular markers in most previous phylogenetic studiaes. Rapid technological advances in DNA sequencing (Barrett et al., 2016b; Heather and Chain, 2016) have opened up the possibility of targeted sequencing of hundreds of phylogenetically informative molecular markers (Weitemier et al., 2014; Dodsworth et al., 2019) and allow the revisiting of contentious relationships in the Calamoideae under a new phylogenomic framework. The development of probes specifically designed for targeted sequence capture of nuclear markers that are phylogenetically informative in palms (Heyduk et al., 2015; de La Harpe et al., 2019) has led to substantial advances in the understanding of phylogenetic relationships, for example within the subfamily Arecoideae Burnett (Comer et al., 2016; Loiseau et al., 2019). The PhyloPalm probe kit (Loiseau et al., 2019) has potential to become the new standard for phylogenomic analyses in palms because it combines phylogenetic informativeness with backward-compatibility by targeting both the 795 most informative markers of de La Harpe et al. (2019) and all 176 markers of the foundational target-capture study in palms (Heyduk et al., 2015). The genomic regions of Heyduk et al. (2015) were initially selected for a phylogenetic study of the palm genus *Sabal* Adans. (subfamily Coryphoideae) and were refined by applying a strict filtering strategy to identify putatively orthologous genes in five species representing the palm subfamilies Arecoideae, Coryphoideae and Nypoideae. The additional 795 markers by Loiseau et al. (2019) were selected primarily based on phylogenetic informativeness and putative orthology in 20 species representing the subfamilies Arecoideae, Ceroxyloideae and Coryphoideae from a larger collection of 4,184 genomic regions (de La Harpe et al., 2019). These were identified in the *Elaeis guineensis* Jacq. (Arecoideae) reference genome (Singh et al., 2013) mainly based on low-copy signature in *Geonoma undata* Klotzsch (Arecoideae), known functions,

and variation in rates of molecular evolution, and also include 131 additional loci commonly used in palm phylogenetics, as well as 133 putatively neutral markers (de La Harpe et al., 2019). However, the combined PhyloPalm probes have not yet been tested in a phylogenomic study, and their utility for the Calamoideae remains unclear because the probes were developed mainly for the phylogenetically divergent Arecoideae and Coryphoideae (Heyduk et al., 2015; Loiseau et al., 2019).

Here, we investigated the phylogenetic relationships within the subfamily Calamoideae by generating sequence data from the 971 nuclear loci of the PhyloPalm probe kit. We applied targeted sequencing to the most comprehensive, systematic sampling of Calamoideae yet published with the aims of 1) testing the monophyly of all genera, subtribes and tribes, 2) establishing a robust hypothesis of higher-level phylogenetic relationships, and 3) examining gene tree conflict in relation to systematics. In doing so, we strived to provide a strengthened foundation for future systematic, comparative, and evolutionary studies within the subfamily.

2 Material and methods

2.1 Sampling

We employed a systematic sampling strategy with the objective of achieving a balanced representation of the taxonomic diversity of all calamoid genera. First, we selected at least two species per genus (except for monotypic genera). Second, we chose one species per section within *Calamus* (Beccari, 1908, 1911; Furtado, 1953, 1956; Kramadibrata, 1992), *Korthalsia* Blume (Dransfield, 1981), *Metroxylon* Rottb. (Beccari, 1918), *Plectocomiopsis* Becc. (Dransfield, 1982), *Raphia* P.Beauv. (Otedoh, 1982; Helmstetter et al., 2020), and *Salacca* (Beccari, 1918; Furtado, 1949; Dransfield et al., 2008). Third, we added one species of the synonymized genera *Calospatha*, *Ceratolobus*, *Cornera*, *Daemonorops*, *Pogonotium*, *Retispatha* and *Zalaccella* (all synonyms of *Calamus*), and *Lophospatha* (synonym of *Salacca*). Finally, we added further species in Plectocomiinae, including an undescribed Bornean species of unclear generic affinity, and the larger genera *Calamus*, *Raphia*, *Salacca* based on taxonomic expertise (AJH, BGK, JD, TPLC, WJB). In total, 75 species of the Calamoideae were sampled, plus four outgroup species representing all other palm subfamilies. Detailed information of the selected species, including their classification and voucher information, are provided in Table 2.

2.2 Sequencing

DNA was extracted using a modified cetrimonium bromide (CTAB) protocol (Doyle and Doyle, 1987). We used 10-25 mg tissue from silica-dried leaves or herbarium specimens, ground to fine powder using a 2010 Geno/Grinder (SPEX SamplePrep, Stanmore, UK). Where necessary, extracts were cleaned using AMPure XP magnetic beads (Beckman Coulter, High Wycombe, UK). DNA concentrations were quantified using a Quantus fluorometer (Promega UK Ltd, Southampton, UK). DNA fragment sizes were assessed using gel electrophoresis for all silica-dried samples, but only for a subset of herbarium samples as these had consistently small fragment sizes below 1000 bp.

DNA was prepared for targeted sequence capture using a NEBNext Ultra II library kit (New England BioLabs Ltd, Hitchin, UK). Extracts with DNA fragment sizes larger than 1000 bp were sheared using a Covaris ME220 Focused-ultrasonicator (Covaris Ltd, Brighton, UK) to attain

fragments closer to the size of 300-400 bp targeted during library preparation. Libraries were prepared with (50-)200 ng input DNA, size selection using magnetic beads, dual indexing with NEBNext Multiplex Oligos for Illumina (New England BioLabs Ltd, Hitchin, UK), and 10-12 PCR cycles. Library preparation followed the manufacturer's protocol but was conducted with half volumes (Hale et al., 2020). DNA concentration of prepared libraries was measured using a Quantus fluorometer and distribution of DNA fragment lengths assessed using an Agilent 4200 TapeStation (Agilent Technologies LDA UK Limited, Stockport, UK). Equal amounts of DNA from 16 to 38 indexed libraries of similar DNA fragment length distributions were combined to pools of 300-1500 ng DNA.

The pooled DNA was hybridized for 24 h at 65°C to the PhyloPalm probes (Loiseau et al., 2019) using a myBaits hybridization capture kit (Arbor Biosciences, Ann Arbor, Michigan, USA). The amount of off-target DNA was reduced by four washes whilst the hybridized DNA was bound to magnetic beads, and the hybridized DNA was amplified using 10-12 PCR cycles. Sequencing was conducted at Macrogen Inc. (Seoul, Korea) on an Illumina HiSeq X Ten sequencing platform (Illumina, San Diego, California, USA), generating 2 x 150 bp paired-end reads.

2.3 Phylogeny reconstruction

Raw sequence data were cleaned of adapters and low-quality reads with Trimmomatic v. 0.38 (Bolger et al., 2014) using the MAXINFO algorithm with strictness 0.8 to favor read correctness over read length and a minimum read length of 36 bases to eliminate short reads for which alignment might be problematic. Sequence quality before and after trimming was validated using FastQC v. 0.11.5 (Andrews, 2010). PhyloPalm markers ("exons") were retrieved using a target file of nucleotide sequences matching the PhyloPalm probes in HybPiper v. 1.3.1 (Johnson et al., 2016) using BWA v. 0.7.17-r1188 (Li and Durbin, 2009). Reads matching each exon were independently assembled with SPADES v. 3.11.1 (Bankevich et al., 2012) as implemented in HybPiper, with a lowered coverage cutoff value of 3 to increase exon recovery. Potential paralogs were identified in HybPiper by assessing for each exon whether more than one contig covering at least 85% of the length of the reference sequence was recovered. All exons receiving paralogy warnings were excluded from downstream analyses. In addition to exons, adjacent "splash zones" (Dodsworth et al., 2019) were recovered in HybPiper. The splash zone may contain highly variable intronic regions that are potentially useful to resolve relationships among closely related species (Weitemier et al., 2014). Because splash zones in isolation might be too variable to be meaningfully aligned (Gardner et al., 2019), splash zones and exons were combined in HybPiper into "supercontigs". Downstream analyses were conducted both for exons only and for supercontigs. We hereafter refer to both as "genomic regions".

Genomic regions were aligned individually using MAFFT v. 7.310 (Katoh and Standley, 2013) with local pairwise alignments, 1000 cycles of iterative refinement and optional reverse complementation of sequences. Fragmentary sites in the alignments were removed with trimAl v. 1.2rev59 (Capella-Gutiérrez et al., 2009) using the Automated1 algorithm, combined with a subsequent removal of sites with less than 20 % occupancy, and of sequences covering less than 10 % of the alignment length. These parameters were established by stepwise testing of different settings, with visual inspection of alignment quality by spot checks before and after each step using Geneious Prime v. 2020.0.5 (https://www.geneious.com). The quality of the final trimmed alignments was visually inspected in FluentDNA v. 2.5.3 (Seaman and Buggs, 2020). No alignment errors were found, and no manual alterations were made.

Phylogenies of the individual genomic regions were estimated by maximum likelihood phylogenetic analysis in RAxML v. 8.2.11 (Stamatakis, 2014), using a general time reversible model of nucleotide substitution with gamma distributed rate variation among sites and 1000 rapid bootstrap replicates. In addition, we tested the effect of substitution model testing for each region by performing gene tree estimation in IQ-TREE v. 2.0.6 (Minh et al., 2020) with joint model testing in ModelFinder (Kalyaanamoorthy et al., 2017), using 1000 ultrafast bootstrap replicates. Unexpectedly long branches indicative of possibly erroneous sequence data or alignment were removed using TreeShrink v. 1.3.3 (Mai and Mirarab, 2018) with the false positive error rate set to 0.05. To avoid unsupported topologies influencing subsequent analyses in ASTRAL (Zhang et al., 2018), internal branches with bootstrap support below 10 % were collapsed using Newick utilities v. 1.6 (Junier and Zdobnov, 2010), following optimal thresholds identified by the simulations of Zhang et al. (2018). Four species trees were inferred from all individual exon or supercontig trees with or without model testing using the coalescent-based method ASTRAL v. 5.6.3 (Zhang et al., 2018).

Maximum likelihood analysis of concatenated datasets can estimate species trees more accurately than coalescent-based analyses when the amount of incomplete lineage sorting is low or gene tree error is high (Mirarab and Warnow, 2015). We therefore concatenated all individual exon and supercontig alignments, respectively, using AMAS (Borowiec, 2016) and then estimated two additional species trees for each concatenated alignment using RAxML with the same settings as above, but with an extended majority-rule criterion for stopping bootstrapping after a sufficient number of replicates had been sampled (Pattengale et al., 2010). This threshold was reached after 50 bootstrap replicates for the concatenated exons, and after 100 replicates for the concatenated supercontigs. In addition, we tested the effect of partitioning of the concatenated datasets by selecting partitioning schemes using PartitionFinder v. 2.2.1 (Lanfear et al., 2017) with a general time reversible model of nucleotide substitution with gamma distributed rate variation among sites and the rcluster algorithm (Lanfear et al., 2014), followed by phylogeny inference in RAxML with the same settings as above, but with the addition of the partitioning scheme selected by PartitionFinder. 616 partitions were selected for the concatenated exons, and 576 partitions for the concatenated supercontigs. The threshold for stopping bootstrapping was reached for both analyses after 50 bootstrap replicates.

Species trees were rooted using the outgroup (Table 2) with phyx (Brown et al., 2017). Concordance and conflict between the topologies of the species trees were investigated with a strict consensus analysis in ape v. 5.3 (Paradis et al., 2004) in R v. 3.6.3 (R Core Team, 2020), and by calculating normalized Robinson-Foulds distances (Robinson and Foulds, 1981) in phangorn v. 2.5.5 (Schliep, 2011) in R. Gene tree conflict of selected higher-level relationships was visually investigated for the coalescence supercontig phylogeny without model testing using DiscoVista (Sayyari et al., 2018) after collapsing gene tree branches with bootstrap support below 90% using Newick utilities. We arbitrarily interpreted gene tree conflict as “low” if the proportion of gene trees supporting the species tree topology was at least 30 percentage points higher than for both alternative topologies, and as “substantial” if the difference to one or both alternative topologies was smaller than 20 percentage points. Cases of conflict intermediate between these two categories were not observed for the investigated relationships.

2.4 Data availability

Raw sequence data are deposited in the European Nucleotide Archive of the European Bioinformatics Institute (https://www.ebi.ac.uk/ena) under project number PRJEB40689. Scripts for

all phylogenetic analyses are available on GitHub at https://github.com/BenKuhnhaeuser/PhyloFrame. The PhyloPalm target file, alignments, gene trees, and species trees are deposited on Zenodo at https://doi.org/10.5281/zenodo.4359280 (Kuhnhäuser et al., 2020).

3 Results

3.1 Informativeness of PhyloPalm markers in the Calamoideae

Targeted sequencing of the 971 PhyloPalm markers produced between 394,577 and 13,234,981 trimmed reads per calamoid sample (Table 3). Of these, a median of 22.2 % were on target. After assembly of the mapped reads into contiguous sequences and their alignment to the target reference, a median of 947 exons were retrieved (Table 3), with less than 695 exons for only three species. Two exons were not retrieved for any species. In addition, we excluded four exons that were recovered for less than 10% of species, and 20 exons that received paralogy warnings. The remaining 945 exons were used for downstream analyses either directly or, after retrieval of adjacent splash zones, as part of supercontigs.

The median number of retrieved reads, proportion of targets on read, and recovered proportion of the targeted sequence length were higher in the outgroup than in the calamoid samples (Table 3), but no apparent taxonomic bias existed within the Calamoideae (Supplementary Fig. 1). Across all taxa including the outgroup, lower target length proportions of the Heyduk markers were recovered compared to the other PhyloPalm markers (Supplementary Fig. 1).

The trimmed exon alignments had a total length of over one million base pairs and contained 27.6 % parsimony informative sites (Table 4). Cropping of the outgroup from the concatenated alignment showed that variation within the Calamoideae accounted for 85.4 % of the parsimony informative sites. Concatenated supercontig alignments were only 66 % longer than the concatenated exon alignments but contained over twice as many parsimony informative sites, signifying higher genomic variation within the splash zones than within the exons (Table 4).

3.2 Phylogenetic relationships within the Calamoideae

The eight species trees inferred in this study resolved identical higher-level relationships but differed at shallower levels (Fig. 3, Supplementary Fig. 2 a). Overarching methodologies (coalescence vs. concatenation) and datasets (exons vs. supercontigs) had the strongest effect on differences in tree topology, whereas substitution model testing in the coalescence analyses and partitioning of concatenated analyses had the smallest effects (Supplementary Fig. 2 b). When only considering overarching methodologies and datasets, the coalescence analyses inferred highly similar relationships, whereas larger degrees of conflict existed between the other possible pairwise comparisons of species tree topologies (Supplementary Fig. 2 b).

All species trees inferred in this study consistently supported the monophyly of all currently recognized tribes, subtribes and genera of Calamoideae, and inferred identical relationships among all tribes and subtribes with maximal support (Fig. 3, Supplementary Fig. 2 a). At tribal level, Eugeissoneae was resolved as sister to Calameae. In Lepidocaryeae, Raphiinae was resolved as sister to Ancistrophyllinae, and subtribal relationships in Calameae were resolved as (Korthalsiinae, (Salaccinae, (Metroxylinae, (Pigafettinae, (Plectocomiinae, Calaminae))))). Within all subtribes except Plectocomiinae, identical genus-level relationships were resolved. In Ancistrophyllinae, *Laccosperma*

was resolved as sister to *Oncocalamus*, and in Mauritiinae, *Mauritia* was resolved as sister to *Mauritiella*. In Salaccinae, *Salacca secunda* was resolved as sister to the remaining species of the genus. In *Calamus*, several higher-level clades were consistently resolved with maximal support, with all synonymized genera nested within the genus (Fig. 3 a). All six sections of *Calamus* that were represented by at least two species (sections I, II, V, XII, XV and XVIII; Table 2) were resolved as polyphyletic (Fig. 3 a). Discordance between the different phylogenetic inferences was limited to generic relationships in subtribe Plectocomiinae, all species relationships within *Korthalsia* and *Raphia*, and some species relationships within *Calamus* and *Salacca* (Supplementary Fig. 2 a).

We found low gene tree conflict for all tribal and subtribal relationships, as well as higher-level relationships in Mauritiinae and Salaccinae (Fig. 4 a,b,c,e,f). In contrast, substantial gene tree conflict was identified in Ancistrophyllinae and especially Plectocomiinae (Fig. 4 d,g). In Calaminae, low gene tree conflict existed for the consistent placement of *Calamus castaneus* and *C. zollingeri* subsp. *zollingeri* as sister to all other species of the genus (Fig. 3, Supplementary Fig. 2 a), whereas substantial conflict was inferred for two of the other three investigated relationships in the backbone of *Calamus* (Fig. 4 h).

4 Discussion

4.1 Efficiency of PhyloPalm probes

The PhyloPalm probes were developed primarily for the subfamily Arecoideae and closely related subfamilies (Loiseau et al., 2019). The Calamoideae being sister to all other palms, it is thus not surprising that the proportion of on target reads, as well as the recovered proportion of the targeted markers, were lower in the Calamoideae than in the outgroup representing all other subfamilies (Table 3). Recovery of the included Heyduk markers was lower compared to the remaining PhyloPalm markers for all samples including the outgroup (Supplementary Fig. 1).

Despite the lower recovery rates in the Calamoideae compared to the outgroup, our results show that the retrieved sequence data are phylogenetically highly informative for the subfamily (Table 4) and can be used to resolve relationships both at deep and shallow levels (Figs. 3, 4, Supplementary Fig. 2 a). Splash zones, which were retrieved as flanking regions to the targeted exons and were combined with these to supercontigs, were highly variable and more than doubled the total number of parsimony informative sites (Table 4), substantially increasing the data basis for phylogenetic inferences.

We note that the sampling in this study was limited to only a fraction of the species within most genera, and that most phylogenetic uncertainty in this study pertained to intra-generic relationships. Nevertheless, the usefulness of 795 of the 971 PhyloPalm probes for elucidating species-level relationships has been demonstrated in the palm tribe Geonomateae (subfamily Arecoideae) by resolving most phylogenetic relationships with strong support, and concluding that the remaining poorly supported relationships reflected gene tree conflict rather than lack of data (Loiseau et al., 2019). This is a further indication that the PhyloPalm probes might have high potential for clarifying species-level relationships within the Calamoideae.

4.2 Higher-level relationships within the Calamoideae

Our phylogenetic inferences of higher-level relationships within the Calamoideae establish a robust evolutionary framework (Fig. 3) that is based on genomic-scale sampling of phylogenetically informative markers and accounts for gene tree conflict (Fig. 4). The firm establishment of Eugeissoneae as sister to Calameae (Fig. 4 a) resolves the previously ambiguous relationships among these tribes (Fig. 2). In Lepidocaryeae, our findings (Fig. 3) supersede previous phylogenetic inferences that resolved conflicting relationships among subtribes and within Ancistrophyllinae with intermediate to maximal support (see Introduction) but were based on very limited genomic data. Raphiinae was resolved with maximal support and by a clear majority of gene trees as sister to Ancistrophyllinae (Fig. 4 b), whereas previous analyses predominantly resolved Raphiinae as sister to Mauritiinae, albeit only with weak to intermediate support (Baker et al., 2000a, b; Asmussen et al., 2006; Baker et al., 2009; Faurby et al., 2016). In Ancistrophyllinae, *Oncocalamus* was resolved with maximal support as sister to *Laccosperma*, whereas all previous studies had inferred a sister relationship between *Eremospatha* and *Laccosperma* (Baker et al., 2000a, b; Asmussen et al., 2006; Baker et al., 2009; Faurby et al., 2016; Faye et al., 2016). This can potentially be explained by the relatively strong degree of gene tree conflict in Ancistrophyllinae (Fig. 4 d) and the dramatically increased sampling of genomic regions in the present study compared to previous studies, likely leading to a more representative reconciliation of conflicting gene tree histories. In contrast to these upheavals, the previously consistently inferred sister relationship of *Mauritia* and *Mauritiella* in Mauritiinae (Baker et al., 2000a, b, 2009; Faurby et al., 2016) was validated (Fig. 4 c). In Calameae, all subtribal relationships were resolved with maximal support and low gene tree conflict (Figs. 3, 4 e), elucidating the previously poorly understood relationships among the subtribes Metroxylinae, Pigafettinae, Plectocomiinae and Calaminae (Baker et al., 2000a, b; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016a; Faurby et al., 2016; Shahimi, 2018), and confirming the well-established placements of Korthalsiinae and Salaccinae (Baker et al., 2000b; Asmussen et al., 2006; Baker et al., 2009; Faurby et al., 2016; Shahimi, 2018). In Plectocomiinae, the presence of strong gene tree conflict (Fig. 4 g) mirrors the findings of contrasting relationships among genera in previous studies (Baker et al., 2000a, b; Baker et al., 2009; Faurby et al., 2016; Shahimi, 2018). In Calaminae, the backbone inferred in this study (Fig. 3, Supplementary Fig. 2 a) is congruent with the summary relationships presented in the only previous comprehensive study of this subtribe (Baker et al., 2000c). In-depth analysis of selected higher-level relationships in Calaminae revealed little gene tree conflict for some higher-level relationships, whereas strong conflict existed for others (Fig. 4 h, Supplementary Fig. 2 a).

4.3 Potential causes of gene tree conflict

Substantial gene tree conflict was observed in Ancistrophyllinae, Plectocomiinae and Calaminae (Fig. 4 d,g,h). Gene tree conflict can be caused by many processes, including incomplete lineage sorting, hybridization, and gene duplications (Degnan and Rosenberg, 2009). Incomplete lineage sorting can occur if there is insufficient time for coalescence between species splitting events, which is likely if the number of generations between splitting events is small and / or the population size is large (Naciri and Linder, 2015). Neither factors are well documented in the three subtribes, although the very short branches between genera in Ancistrophyllinae and Plectocomiinae and between several clades in Calaminae (Fig. 3 b) suggest that incomplete lineage sorting might be an important cause of gene tree conflict. Hybridization seems a plausible explanation because distributions of genera are strongly overlapping within the African Ancistrophyllinae and the Asian Plectocomiinae, respectively (Table 1), and because several different species of Calaminae often co-occur in the same habitats

(Petoe et al., 2020). This case seems to be particularly strong for the unplaced species Plectocomiinae sp. nov. (Fig. 4 g), which appears morphologically intermediate between the three genera of Plectocomiinae (BGK, pers. obs.). In addition, all species of Plectocomiinae and Calaminae are dioecious (Dransfield et al., 2008), making outcrossing obligatory and potentially facilitating hybridization. However, more detailed biological interpretations are prevented by the lack of knowledge of reproductive ecology, such as flowering times or pollinators, in all three subtribes. In contrast, gene or genome duplications seem unlikely to have been a major source of gene tree conflict in the investigated relationships because only 20 (2.1 %) of the 969 retrieved exons were identified as potentially paralogous and were removed from the analyses (see 3.1), and because polyploidy is very rare in palms, with no polyploid Calamoideae known to date (Dransfield et al., 2008; Barrett et al., 2019).

4.4 Implications for higher-level classification

Our phylogenomic inferences allow a critical re-examination of the prevailing higher-level classification of the Calamoideae (Baker and Dransfield, 2016). Generally, the current classification is corroborated by the resolution of all currently accepted genera, subtribes and tribes as monophyletic (Fig. 3). In Salaccinae, the morphologically intermediate species *Salacca griffithii* A.J.Hend. and *S. secunda* blur the boundaries between *Eleiodoxa* and *Salacca* (Dransfield et al., 2008). The topology resolved here permits the maintenance of the taxonomic status quo (Fig. 4 f), although an argument could be made on morphological grounds for sinking the monotypic *Eleiodoxa* into *Salacca* (indeed *Eleiodoxa* was originally described in *Salacca* [Griffith, 1844]). However, more complete taxon sampling in *Salacca* should be conducted before making changes to the current classification. In Plectocomiinae, the three currently recognized genera are well-characterized and easily distinguished by vegetative and reproductive characters (Dransfield et al., 2008), but the presence of substantial gene tree conflict (Fig. 4 g) hints to a complex evolutionary history of the subtribe.

In Calaminae, the strongly supported nesting of all synonymized genera within *Calamus* (Fig. 3 a) validates recent systematic decisions to synonymize *Calospatha*, *Ceratolobus*, *Daemonorops*, *Pogonotium* and *Retispatha* (Baker and Dransfield, 2008; Baker, 2015; Henderson and Floda, 2015). Because uncertainty exists for several higher-level relationships within *Calamus* (Fig. 4 h), a broadly circumscribed *Calamus* currently remains the most practical solution. Notwithstanding this, a new infrageneric classification is sorely needed to structure the overwhelming diversity of the over 400 species of *Calamus*, as highlighted by the polyphyly of all sections of Calamus that were represented by at least two species (Table 2, Fig. 3 a).

5 Conclusion

As a pantropical and species-rich lineage, the Calamoideae can contribute to a better understanding of the origin and diversity of tropical rainforests, and this study provides the first steps towards that goal. Our phylogenomic inferences are based on the most extensive gene and taxon sampling for the Calamoideae yet published, covering almost one thousand nuclear markers and a systematic selection of 75 species representing the diversity of all calamoid genera. Here, we shed light on previously poorly understood relationships and confirm the existing classification. We find little gene tree conflict in the backbone of the Calamoideae but reveal complex, deep evolutionary histories in subtribes Ancistrophyllinae, Calaminae and Plectocomiinae. Overall, our study provides a strong

phylogenetic framework for future comparative studies, such as taxonomy, systematics, biogeography, and macroevolutionary research. To pursue these ends, much denser sampling is needed to fully unveil species relationships and the extent and possible origins of gene tree conflict. In this regard, a phylogeny including all ca. 550 calamoid species is now underway, as part of a broader global endeavor to complete the tree of life for all palm species (Bellot et al., 2020).

Author contributions

BGK: Conceptualization, Funding acquisition, Resources, Methodology, Investigation, Formal analysis, Validation, Visualization, Data curation, Project administration, Writing – Original draft; SB: Funding acquisition, Resources, Methodology, Investigation, Supervision; TLPC: Methodology, Resources; JD: Resources, Methodology, Validation; AH: Resources, Methodology, Validation; RS: Methodology, Investigation; GC: Conceptualization, Methodology, Supervision; WLE: Conceptualization, Funding acquisition, Resources, Methodology, Investigation, Supervision; SH: Conceptualization, Methodology, Supervision; WJB: Conceptualization, Funding acquisition, Resources, Methodology, Validation, Project administration, Supervision. All authors: Writing – Review & editing.

Funding

This work was supported by funding of the Natural Environment Research Council (NE/L002612/1), a Clarendon Scholarship by the Clarendon Fund and Brasenose College, a Graduate Student Research Award by the Society of Systematic Biologists, and an Emily Holmes Memorial Scholarship by the Amar-Franses & Foster-Jenkins Trust to BGK; funding by the Garfield Weston Foundation under the Global Tree Seed Bank project to SB; a research grant (00025354) from VILLUM FONDEN to WLE; two Bentham-Moxon Trust grants (BMT1-2017 and BMT36-2018) to BGK and WJB; a grant from IKEA of Sweden AB to WJB; and grants from the Calleva Foundation and the Sackler Trust to the Plant and Fungal Tree of Life Project (PAFTOL) at the Royal Botanic Gardens, Kew. Declarations of interest: none.

Acknowledgements

This research would not have been possible without the generous provision of silica-dried and herbarium material by the Royal Botanic Gardens, Kew (UK), the New York Botanical Garden (USA), Naturalis (Netherlands), the Sarawak Herbarium (Malaysia), the Forest Research Institute Malaysia, the Conservatoire et Jardin botaniques de la Ville de Genève (Switzerland), Fairchild Tropical Botanic Garden (USA), and Ángela Cano (University of Cambridge). In addition, we thank Imalka Kahandawala, Edith Kapinos and Rosie Woods for providing samples from the DNA bank at the Royal Botanic Gardens, Kew. We are grateful for support in conducting DNA extractions by László Csiba, assistance in library preparation by Michelle Siros, and general phylogenomic lab support by Robyn Cowan, Panagiota Malakasi and Dion Devey. Finally, we would like to thank Suzanne Mogue Kamga for advice on sample selection in *Raphia*, Bruce Murphy for support in phylogenomic analyses, and the Royal Botanic Gardens, Kew, for computational support. Two anonymous reviewers provided helpful comments on an earlier version of this manuscript.

References

Adorador, J.T., Fernando, E.S., 2020. Two new species of rattans (*Calamus*, Arecaceae) from forests over limestone in the Philippines. Phytotaxa 447, 88-102. https://doi.org/10.11646/phytotaxa.447.2.2.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed 18 August 2020).

Asmussen, C.B., Chase, M.W., 2001. Coding and noncoding plastid DNA in palm systematics. Am. J. Bot. 88, 1103-1117. https://doi.org/10.2307/2657094.

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. https://doi.org/10.1111/j.1095-8339.2006.00521.x.

Baker, W.J., 2015. A revised delimitation of the rattan genus *Calamus* (Arecaceae). Phytotaxa 197, 139-152. https://doi.org/10.11646/phytotaxa.197.2.7.

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. Biogeog. 40, 274-285. https://doi.org/10.1111/j.1365-2699.2012.02795.x.

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. Biogeog. 40, 286-298. https://doi.org/10.1111/j.1365-2699.2012.02794.x.

Baker, W.J., Dransfield, J., 2008. *Calospatha* subsumed in *Calamus* (Arecaceae: Calamoideae). Kew Bull. 63, 161-162. https://doi.org/10.1007/s12225-007-9007-5.

Baker, W.J., Dransfield, J., 2016. Beyond Genera Palmarum: progress and prospects in palm systematics. Bot. J. Linn. Soc. 182, 207-233. https://doi.org/10.1111/boj.12401.

Baker, W.J., Dransfield, J., Harley, M.M., Bruneau, A., 1999. Morphology and cladistic analysis of subfamily Calamoideae (Palmae). In: Henderson, A., Borchsenius, F. (Eds.), Evolution and classification of palms. The New York Botanical Garden Press, Bronx, NY, pp. 307-324.

Baker, W.J., Dransfield, J., Hedderson, T.A., 2000a. Phylogeny, character evolution, and a new classification of the calamoid palms. Syst. Bot. 25, 297-322. https://doi.org/10.2307/2666644.

Baker, W.J., Hedderson, T.A., Dransfield, J., 2000b. Molecular phylogenetics of subfamily Calamoideae (Palmae) based on nrDNA ITS and cpDNA *rps*16 intron sequence data. Mol. Phylogenet. Evol. 14, 195-217. https://doi.org/10.1006/mpev.1999.0696.

Baker, W.J., Hedderson, T.A., Dransfield, J., 2000c. Molecular phylogenetics of *Calamus* (Palmae) and related rattan genera based on 5S nrDNA spacer sequence data. Mol. Phylogenet. Evol. 14, 218-231. https://doi.org/10.1006/mpev.1999.0697.

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. https://doi.org/10.1093/sysbio/syp021.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455-477. https://doi.org/10.1089/cmb.2012.0021.

Barrett, C.F., Bacon, C.D., Antonelli, A., Cano, Á., Hofmann, T., 2016b. An introduction to plant phylogenomics with a focus on palms. Bot. J. Linn. Soc. 182, 234-255. https://doi.org/10.1111/boj.12399.

Barrett, C.F., Baker, W.J., Comer, J.R., Conran, J.G., Lahmeyer, S.C., Leebens-Mack, J.H., Li, J., Lim, G.S., Mayfield-Jones, D.R., Perez, L., Medina, J., Pires, J.C., Santos, C., Stevenson, D.W., Zomlefer, W.B., Davis, J.I., 2016a. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. New Phytol. 209, 855-870. https://doi.org/10.1111/nph.13617.

Barrett, C.F., McKain, M.R., Sinn, B.T., Ge, X.-J., Zhang, Y., Antonelli, A., Bacon, C.D., 2019. Ancient polyploidy and genome evolution in palms. Genome Biol. Evol. 11, 1501-1511. https://doi.org/10.1093/gbe/evz092.

Beccari, O., 1908. Asiatic palms – Lepidocaryeae. Part I. The species of *Calamus*. Ann. Roy. Bot. Gard. Calcutta 11, 1-518.

Beccari, O., 1911. Asiatic Palms – Lepidocaryeae. Part II. The species of *Daemonorops*. Ann. Roy. Bot. Gard. Calcutta 12, 1-237.

Beccari, O., 1918. Asiatic Palms – Lepidocaryeae. Part III. The species of the genera *Ceratolobus*, *Calospatha*, *Plectocomia*, *Plectocomiopsis*, *Myrialepis*, *Zalacca*, *Pigafetta*, *Korthalsia*, *Metroxylon*, *Eugeissona*. Ann. Roy. Bot. Gard. Calcutta 12, 1-231.

Bellot, S., Odufuwa, P., Dransfield, J., Eiserhardt, W.L., Sang, J., Pérez-Escobar, O.A., Petoe, P., Usher, E., Baker, W., 2020. Why and how to develop DNA barcoding for palms? A case study of *Pinanga*. Palms 64, 109-120.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120. https://doi.org/10.1093/bioinformatics/btu170.

Borowiec, M.L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4, e1660. https://doi.org/10.7717/peerj.1660.

Brown, J.W., Walker, J.F., Smith, S.A., 2017. Phyx: phylogenetic tools for unix. Bioinformatics 33, 1886-1888. https://doi.org/10.1093/bioinformatics/btx063.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972-1973. https://doi.org/10.1093/bioinformatics/btp348.

Comer, J.R., Zomlefer, W.B., Barrett, C.F., Stevenson, D.W., Heyduk, K., Leebens-Mack, J.H., 2016. Nuclear phylogenomics of the palm subfamily Arecoideae (Arecaceae). Mol. Phylogenet. Evol. 97, 32-42. https://doi.org/10.1016/j.ympev.2015.12.015.

Couvreur, T.L.P., Baker, W.J., 2013. Tropical rain forest evolution: palms as a model group. BMC Biol. 11, 48. https://doi.org/10.1186/1741-7007-11-48.

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. BMC Biol. 9, 44. https://doi.org/10.1186/1741-7007-9-44.

Davis, Charles C., Webb, Campbell O., Wurdack, Kenneth J., Jaramillo, Carlos A., Donoghue, Michael J., 2005. Explosive radiation of Malpighiales supports a mid-Cretaceous origin of modern tropical rain forests. Am. Nat. 165, E36-E65. https://doi.org/10.1086/428296.

de La Harpe, M., Hess, J., Loiseau, O., Salamin, N., Lexer, C., Paris, M., 2019. A dedicated target capture approach reveals variable genetic markers across micro- and macro-evolutionary time scales in palms. Mol. Ecol. Resour. 19, 221-234. https://doi.org/10.1111/1755-0998.12945.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332-340. https://doi.org/10.1016/j.tree.2009.01.009.

Dodsworth, S., Pokorny, L., Johnson, M.G., Kim, J.T., Maurin, O., Wickett, N.J., Forest, F., Baker, W.J., 2019. Hyb-Seq for flowering plant systematics. Trends Plant Sci. 24, 887-891. https://doi.org/10.1016/j.tplants.2019.07.011.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11-15.

Dransfield, J., 1981. A synopsis of the genus *Korthalsia* (Palmae: Lepidocaryoideae). Kew Bull. 36, 163-194. https://doi.org/10.2307/4119016.

Dransfield, J., 1982. A reassessment of the genera *Plectocomiopsis*, *Myrialepis* and *Bejaudia* (Palmae: Lepidocaryoideae). Kew Bull. 37, 237-254. https://doi.org/10.2307/4109967.

Dransfield, J., Uhl, N.W., Asmussen, C.B., Baker, W.J., Harley, M.M., Lewis, C.E., 2008. Genera Palmarum – the evolution and classification of palms. Royal Botanic Gardens, Kew. https://doi.org/10.34885/92.

Eiserhardt, W.L., Couvreur, T.L.P., Baker, W.J., 2017. Plant phylogeny as a window on the evolution of hyperdiversity in the tropical rainforest biome. New Phytol. 214, 1408-1422. https://doi.org/10.1111/nph.14516.

Evans, T.D., Sengdala, K., Thammavong, B., Viengkham, O.V., Dransfield, J., 2002. A synopsis of the rattans (Arecaceae: Calamoideae) of Laos and neighbouring parts of Indochina. Kew Bull. 57, 1-84. https://doi.org/10.2307/4110822.

Faurby, S., Eiserhardt, W.L., Baker, W.J., Svenning, J.C., 2016. An all-evidence species-level supertree for the palms (Arecaceae). Mol. Phylogenet. Evol. 100, 57-69. https://doi.org/10.1016/j.ympev.2016.03.002.

Faye, A., Pintaud, J.-C., Baker, W.J., Vigouroux, Y., Sonke, B., Couvreur, T.L.P., 2016. Phylogenetics and diversification history of African rattans (Calamoideae, Ancistrophyllinae). Bot. J. Linn. Soc. 182, 256-271. https://doi.org/10.1111/boj.12454.

Furtado, C.X., 1949. Palmae Malesicae X. The Malayan species of *Salacca*. Gard. Bull. Singapore 12, 378–403.

Furtado, C.X., 1953. The genus *Daemonorops* in Malaya. Gard. Bull. Singapore 14, 49-147.

Furtado, C.X., 1956. Palmae Malesicae XIX. The genus *Calamus* in the Malayan Peninsula. Gard. Bull. Singapore 15, 32-265.

Gardner, E.M., Johnson, M.G., Pereira, J.T., Ahmad Puad, A.S., Arifiani, D., Sahromi, Wickett, N.J., Zerega, N.J.C., 2019. Paralogs and off-target sequences improve phylogenetic resolution in a densely-sampled study of the breadfruit genus (*Artocarpus*, Moraceae). bioRxiv, 854232. https://doi.org/10.1101/854232.

Govaerts, R., Dransfield, J., Zona, S.F., Hodel, D.R., Henderson, A., 2020. World checklist of Arecaceae. Facilitated by the Royal Botanic Gardens, Kew. Available at: http://wcsp.science.kew.org (accessed 18 August 2020).

Griffith, W., 1844. The palms of British East India. Calcutta J. Nat. Hist. 5, 1-103 & 311-355.

Hahn, W.J., 2002. A molecular phylogenetic study of the Palmae (Arecaceae) based on *atp*B, *rbc*L, and 18S nrDNA sequences. Syst. Biol. 51, 92-112. https://doi.org/10.1080/106351502753475899.

Hale, H., Gardner, E.M., Viruel, J., Pokorny, L., Johnson, M.G., 2020. Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. Appl. Plant Sci. 8, e11337. https://doi.org/10.1002/aps3.11337.

Heather, J.M., Chain, B., 2016. The sequence of sequencers: The history of sequencing DNA. Genomics 107, 1-8. https://doi.org/10.1016/j.ygeno.2015.11.003.

Helmstetter, A.J., Mogue Kamga, S., Bethune, K., Lautenschläger, T., Zizka, A., Bacon, C.D., Wieringa, J.J., Stauffer, F., Antonelli, A., Sonké, B., Couvreur, T.L.P., 2020. Unraveling the phylogenomic relationships of the most diverse African palm genus *Raphia* (Calamoideae, Arecaceae). Plants 9, 549. https://doi.org/10.3390/plants9040549.

Henderson, A., 2020. A revision of *Calamus* (Arecaceae, Calamoideae, Calameae, Calaminae). Phytotaxa 445, 1-656. https://doi.org/10.11646/phytotaxa.445.1.1.

Henderson, A., Floda, D., 2015. *Retispatha* subsumed in *Calamus* (Arecaceae). Phytotaxa 192, 58-60. https://doi.org/10.11646/phytotaxa.192.1.8.

Heyduk, K., Trapnell, D.W., Barrett, C.F., Leebens-Mack, J., 2015. Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. Biol. J. Linn. Soc. 117, 106-120. https://doi.org/10.1111/bij.12551.

Horn, J.W., Fisher, J.B., Tomlinson, P.B., Lewis, C.E., Laubengayer, K., 2009. Evolution of lamina anatomy in the palm family (Arecaceae). Am. J. Bot. 96, 1462-1486. https://doi.org/10.3732/ajb.0800396.

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J.C., Wickett, N.J., 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl. Plant Sci. 4, 1600016. https://doi.org/10.3732/apps.1600016.

Junier, T., Zdobnov, E.M., 2010. The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. Bioinformatics 26, 1669-1670. https://doi.org/10.1093/bioinformatics/btq243.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587-589. https://doi.org/10.1038/nmeth.4285.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772-780. https://doi.org/10.1093/molbev/mst010.

Kissling, W.D., Baker, W.J., Balslev, H., Barfod, A.S., Borchsenius, F., Dransfield, J., Govaerts, R., Svenning, J.C., 2012b. Quaternary and pre-Quaternary historical legacies in the global distribution of a major tropical plant lineage. Glob. Ecol. Biogeogr. 21, 909-921. https://doi.org/10.1111/j.1466-8238.2011.00728.x.

Kissling, W.D., Balslev, H., Baker, W.J., Dransfield, J., Göldel, B., Lim, J.Y., Onstein, R.E., Svenning, J.-C., 2019. PalmTraits 1.0, a species-level functional trait database of palms worldwide. Sci. Data 6, 178. https://doi.org/10.1038/s41597-019-0189-0.

Kissling, W.D., Eiserhardt, W.L., Baker, W.J., Borchsenius, F., Couvreur, T.L.P., Balslev, H., Svenning, J.C., 2012a. Cenozoic imprints on the phylogenetic structure of palm species assemblages worldwide. Proc. Natl. Acad. Sci. U.S.A. 109, 7379-7384. https://doi.org/10.1073/pnas.1120467109.

Kramadibrata, P., 1992. A revision of the genus *Calamus* (Palmae) section *Macropodus* sensu Furtado. Ph.D. thesis, University of Reading.

[dataset] Kuhnhäuser, B.G., Bellot, S., Couvreur, T.L.P., Dransfield, J., Henderson, A., Schley, R., Chomicki, G., Eiserhardt, W.L., Hiscock, S.J., Baker, W.J., 2020. Trees and alignments for: A robust phylogenomic framework for the calamoid palms. Zenodo. Version 1. https://doi.org/10.5281/zenodo.4359280.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14, 82. https://doi.org/10.1186/1471-2148-14-82.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772-773. https://doi.org/10.1093/molbev/msw260.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754-1760. https://doi.org/10.1093/bioinformatics/btp324.

Loiseau, O., Olivares, I., Paris, M., de La Harpe, M., Weigand, A., Koubínová, D., Rolland, J., Bacon, C.D., Balslev, H., Borchsenius, F., Cano, A., Couvreur, T.L.P., Delnatte, C., Fardin, F., Gayot, M., Mejía, F., Mota-Machado, T., Perret, M., Roncal, J., Sanin, M.J., Stauffer, F., Lexer, C., Kessler, M., Salamin, N., 2019. Targeted capture of hundreds of nuclear genes unravels phylogenetic relationships of the diverse Neotropical palm tribe Geonomateae. Front. Plant Sci. 10, 864. https://doi.org/10.3389/fpls.2019.00864.

Mai, U., Mirarab, S., 2018. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. BMC Genomics 19, 272. https://doi.org/10.1186/s12864-018-4620-2.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530-1534. https://doi.org/10.1093/molbev/msaa015.

Mirarab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, i44-i52. https://doi.org/10.1093/bioinformatics/btv234.

Mogue Kamga, S., Brokamp, G., Cosiaux, A., Awono, A., Fürniss, S., Barfod, A.S., Muafor, F.J., Le Gall, P., Sonké, B., Couvreur, T.L.P., 2020. Use and cultural significance of *Raphia* palms. Econ. Bot. 74, 207-225. https://doi.org/10.1007/s12231-020-09487-z.

Mondal, S., Basu, S.K., Chowdhury, M., 2020. A new species of *Calamus* (Arecaceae) from Eastern Himalaya, India. Turczaninowia 23, 92-98. https://doi.org/10.14258/turczaninowia.23.3.9.

Naciri, Y., Linder, H.P., 2015. Species delimitation and relationships: the dance of the seven veils. Taxon 64, 3-16. https://doi.org/10.12705/641.24.

Otedoh, M., 1982. A revision of the genus *Raphia* Beauv. (Palmae). J. Niger. Inst. Oil Palm Res. 6, 145-189.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289-290. https://doi.org/10.1093/bioinformatics/btg412.

Pattengale, N.D., Alipour, M., Bininda-Emonds, O.R., Moret, B.M., Stamatakis, A., 2010. How many bootstrap replicates are necessary? J. Comput. Biol. 17, 337-354. https://doi.org/10.1089/cmb.2009.0179.

Pennington, R.T., Cronk, Q.C.B., Richardson, J.A., 2004. Introduction and synthesis: plant phylogeny and the origin of major biomes. Philos. Trans. Roy. Soc. B. 359, 1455-1464. https://doi.org/10.1098/rstb.2004.1539.

Petoe, P., Kuhnhäuser, B.G., Geri, C., Baker, W.J., 2020. Lanjak Entimau Wildlife Santuary – a palm hotspot in the heart of Borneo. Palms 64, 57-84.

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available at: https://www.R-project.org/ (accessed 18 August 2020).

Ramírez-Barahona, S., Sauquet, H., Magallón, S., 2020. The delayed and geographically heterogeneous diversification of flowering plant families. Nat. Ecol. Evol. https://doi.org/10.1038/s41559-020-1241-3.

Reichgelt, T., West, C.K., Greenwood, D.R., 2018. The relation between global palm distribution and climate. Sci. Rep. 8, 4721. https://doi.org/10.1038/s41598-018-23147-2.

Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. Math. Biosci. 53, 131-147. https://doi.org/10.1016/0025-5564(81)90043-2.

Sayyari, E., Whitfield, J.B., Mirarab, S., 2018. DiscoVista: Interpretable visualizations of gene tree discordance. Mol. Phylogenet. Evol. 122, 110-115. https://doi.org/10.1016/j.ympev.2018.01.019.

Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27, 592-593. https://doi.org/10.1093/bioinformatics/btq706.

Seaman, J., Buggs, R.J.A., 2020. FluentDNA: nucleotide visualization of whole genomes, annotations, and alignments. Front. Genet. 11, 292. https://doi.org/10.3389/fgene.2020.00292.

Shahimi, S., 2018. Systematics and evolution of the rattan genus *Korthalsia* Bl. (Arecaceae) with special reference to domatia. Ph.D. thesis, University of Reading.

Singh, R., Ong-Abdullah, M., Low, E.-T.L., Manaf, M.A.A., Rosli, R., Nookiah, R., Ooi, L.C.-L., Ooi, S.E., Chan, K.-L., Halim, M.A., Azizi, N., Nagappan, J., Bacher, B., Lakey, N., Smith, S.W., He, D., Hogan, M., Budiman, M.A., Lee, E.K., DeSalle, R., Kudrna, D., Goicoechea, J.L., Wing, R.A., Wilson, R.K., Fulton, R.S., Ordway, J.M., Martienssen, R.A., Sambanthamurthi, R., 2013. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. Nature 500, 335-339. https://doi.org/10.1038/nature12309.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313. https://doi.org/10.1093/bioinformatics/btu033.

Sunderland, T.C.H., Dransfield, J., 2002. Species profiles rattans (Palmae: Calamoideae). In: Dransfield, J., Tesoro, F.O., Manokaran, N. (Eds.), Rattan: current research issues and prospects for conservation and sustainable development. Food and Agriculture Organisation of the United Nations, Rome, pp. 9-22.

Thiers, B., 2020. Index Herbariorum: a global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. Available at: http://sweetgum.nybg.org/science/ih/ (accessed 18 August 2020).

Tomlinson, P.B., Horn, J.W., Fisher, J.B., 2011. The anatomy of palms: Arecaceae - Palmae. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199558926.001.0001.

Uhl, N.W., Dransfield, J., Davis, J.I., Luckow, M.A., Hansen, K.S., Doyle, J.J., 1995. Phylogenetic relationships among palms: cladistic analyses of morphological and chloroplast DNA restriction site variation. In: Rudall, P.J., Cribb, P.J., Cutler, D.F., Humphries, C.J. (Eds.), Monocotyledons: systematics and evolution. Royal Botanic Gardens, Kew, Richmond, UK, pp. 623–661.

Uhl, N.W., Moore Jr., H.E., 1971. The palm gynoecium. Am. J. Bot. 58, 945-992. https://doi.org/10.1002/j.1537-2197.1971.tb10050.x.

Weitemier, K., Straub, S.C.K., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A., Liston, A., 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. Appl. Plant Sci. 2. https://doi.org/10.3732/apps.1400042.

Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19, 153. https://doi.org/10.1186/s12859-018-2129-y.

Table 1. Classification, diversity, and distribution of calamoid genera. Classification follows Baker and Dransfield (2016). Species numbers according to Govaerts et al. (2020) but updated for recent and upcoming publications of *Raphia* (Helmstetter et al., 2020; Mogue Kamga et al., 2020), *Mauritiella* (Torres Jiménez et al., in prep.) and *Calamus* (Adorador and Fernando, 2020; Henderson, 2020; Mondal et al., 2020). Distributions according to Dransfield et al. (2008). [1]One species of *Raphia* occurs in South and Central America. [2]One species of *Calamus* occurs in Africa.

TRIBE			
Subtribe			
Genus	Authority	Species	Distribution
EUGEISSONEAE	W.J. Baker & J. Dransf.		
Eugeissona	Griff.	6	Malay Peninsula, Borneo
LEPIDOCARYEAE	Mart.		
Ancistrophyllinae	Becc.		
Oncocalamus	(G.Mann & H.Wendl.) H.Wendl	4	West and Central Africa
Eremospatha	(G.Mann & H.Wendl.) Schaedtler	11	West and Central Africa
Laccosperma	(G.Mann & H.Wendl.) Drude	7	West and Central Africa
Raphiinae	H. Wendl.		
Raphia	P.Beauv.	21	Africa, Madagascar[1]
Mauritiinae	Meisn.		
Lepidocaryum	Mart.	1	Northern South America
Mauritia	L.f.	2	Northern South America
Mauritiella	Burret	5	Northern South America
CALAMEAE	Kunth		
Korthalsiinae	Becc.		
Korthalsia	Blume	28	Mainland Southeast Asia to New Guinea
Salaccinae	Becc.		
Eleiodoxa	(Becc.) Burret	1	Malay Peninsula to Borneo
Salacca	Reinw.	23	Mainland Southeast Asia to Borneo
Metroxylinae	Blume		
Metroxylon	Rottb.	7	Moluccas to Samoa
Pigafettinae	J. Dransf. & N.W. Uhl		
Pigafetta	(Blume) Becc.	2	Sulawesi to New Guinea
Plectocomiinae	J. Dransf. & N.W. Uhl		
Plectocomia	Mart. & Blume	15	Himalayas to Borneo
Myrialepis	Becc.	1	Mainland Southeast Asia to Sumatra
Plectocomiopsis	Becc.	6	Mainland Southeast Asia to Borneo
Calaminae	Meisn.		
Calamus	L.	415	India to Fiji[2]

Table 2. Sampling. Classification to genus level follows Baker and Dransfield (2016). Authorities of infrageneric sections are indicated with superscript letters as follows. [A]Otedoh (1982), [B]Helmstetter et al. (2020), [C]Dransfield (1981), [D]Beccari (1918), [E]Dransfield et al. (2008), [F]Furtado (1949), [G]Dransfield (1982), [H]Beccari (1908), [I]Kramadibrata (1992), [J]Furtado (1956). Synonymized genera are given in the same column as sections but are italicized. Species of generic types are marked with *. For *Calamus*, the type *C. rotang* is represented by the closely related *C. godefroyi* (Evans et al., 2002). Herbarium codes in the voucher information follow Thiers (2020).

TRIBE **Subtribe** *Genus* Section	Species	Voucher
EUGEISSONEAE		
Eugeissona	*E. tristis* Griff.*	Baker 501 (K)
	E. utilis Becc.	Baker 712 (SAR)
LEPIDOCARYEAE		
Ancistrophyllinae		
Oncocalamus	*O. mannii* (H.Wendl.) H.Wendl.*	Sunderland 1768 (K)
	O. tuleyi Sunderl.	Dransfield JD7007 (K)
Eremospatha	*E. laurentii* De Wild.	1984-1058 (K)
	E. wendlandiana Dammer ex Becc.	Dransfield JD7004 (K)
Laccosperma	*L. opacum* Drude*	Sunderland 1750 (K)
	L. secundiflorum (P.Beauv.) Kuntze	Sunderland 1763 (K)
Raphiinae		
Raphia		
Flabellatae[A,B]	*R. farinifera* (Gaertn.) Hyl.	Dransfield JD7516 (K)
Monfiliformes[A]	*R. regalis* Becc.	Couvreur 753 (WAG)
Moniliformes[A,B]	*R. textilis* Welw.	Couvreur 1075 (WAG)
Obclavatae[A,B]	*R. sudanica* A.Chev.	Michon 56 (G)
Raphia[A,B]	*R. monbuttorum* Drude	Mogue 4 (WAG)
Temulentae[A,B]	*R. hookeri* G.Mann & H.Wendl.	Couvreur 984 (WAG)
Mauritiinae		
Lepidocaryum	*L. tenue* Mart.*	Dransfield JD7012 (K)
Mauritia	*M. carana* Wallace	Traill 1103 (K)
	M. flexuosa L.f.*	Ely 17 (K)
Mauritiella	*M. aculeata* (Kunth) Burret*	1988-4331 (K)
	M. armata (Mart.) Burret	Couvreur 257 (K)
CALAMEAE		
Korthalsiinae		
Korthalsia		
I[C]	*K. robusta* Blume	Baker 552 (K)
II[C]	*K. jala* J.Dransf.	Baker 558 (K)
III[C]	*K. rostrata* Blume	Kuhnhäuser 34 (K)
IV[C]	*K. rigida* Blume*	Baker 498 (KEP)
Salaccinae		
Eleiodoxa	*E. conferta* (Griff.) Burret*	Dransfield JD6514 (K)
Salacca		
Salacca[D,E]	*S. secunda* Griff.	Henderson 3176 (K)

TRIBE Subtribe *Genus* Section	Species	Voucher
Salacca[D,F]	*S. zalacca* (Gaertn.) Voss*	1984-3376 (K)
Leiosalacca[D,F]	*S. affinis* Griff.	Baker 708 (SAR)
Lophospatha	*S. lophospatha* J.Dransf. & Mogea	Clemens 26380 (K)
Metroxylinae		
Metroxylon		
Coelococcus[D]	*M. salomonense* (Warb.) Becc.	Zona 651 (FTG)
Metroxylon[D]	*M. sagu* Rottb.*	Baker 550 (K)
Pigafettinae		
Pigafetta	*P. elata* (Mart.) H.Wendl.	Baker 508 (K)
	P. filaris (Giseke) Becc.*	Dransfield JD7610 (K)
Plectocomiinae		
Plectocomia	*P. elongata* Mart. ex Blume*	Dransfield JD6200 (K)
	P. himalayana Griff.	Baker 815 (K)
Myrialepis	*M. paradoxa* (Kurz) J.Dransf.*	Baker 491 (K)
Plectocomiopsis		
I[G]	*P. geminiflora* (Griff.) Becc.*	Baker 492 (K)
II[G]	*P. mira* J.Dransf.	Kuhnhäuser 44 (K)
Unplaced	Plectocomiinae sp. nov.	Kuhnhäuser 35 (K)
Calaminae		
Calamus		
I[H,I]	*C. erectus* Roxb.	Baker 814 (K)
I(?)[I]	*C. acanthophyllus* Becc.	Khamphone 141 (K)
II[H,I]/Platyspathus[J]	*C. arborescens* Griff.	Henderson 3182 (NY)
II[H,I]/Podocephalus[J]	*C. castaneus* Griff.	Baker 507 (K)
II(?)[I]	*C. thysanolepis* Hance	Shui Ying Hu 12415 (K)
III[H,I]	*C. deerratus* G.Mann & H.Wendl.	Sunderland 1754 (K)
IV[H,I]	*C. radiatus* Thwaites	de Zoysa 3 (K)
V[H,I]	*C. godefroyi* Becc.*	Evans 153 (K)
V[H,I]	*C. usitatus* Blanco	Baker 564 (K)
V[H,I]	*C.* aff. *usitatus* Blanco	Henderson 4020 (K)
V(?)[I]/*Zalaccella*	*C. harmandii* Pierre ex Becc.	Khamphone 398 (K)
VI[I]	*C. acanthochlamys* J.Dransf.	1989-2701 (K)
VII[H,I]	*C. rheedei* Griff.	Unknown 291 (K)
VIII[I]/Macropodus[J]	*C. peregrinus* Furtado	Henderson 3959 (NY)
IX[I]	*C. rhabdocladus* Burret	Henderson 3763 (NY)
X[H,I]/Macropodus[J]	*C. ciliaris* Blume	1992-2048 (K)
XI[H,I]/Rhombocalamus[J]	*C. rhomboideus* Blume	Baker 565 (K)
XII[H,I]	*C. symphysipus* Mart. ex Walp.	Henderson 4234 (NY)
XII[H,I]	*C. vitiensis* Warb. ex Becc.	Baker 568 (K)
XIII[H,I]/Coleospathus[J]	*C. ornatus* Blume	Dransfield JD7628 (KEP)
XIV[H,I]	*C. zollingeri* subsp. *zollingeri* Becc.	Henderson 4318 (NY)
XV[H,I]	*C. aruensis* Becc.	Dransfield JD7571 (K)
XV[I]	*C. compsostachys* Burret	RITF s.n. (K)

TRIBE		
Subtribe		
Genus		
Section	Species	Voucher
XV[H,I]/Phyllanthectus[J]	*C. oxleyanus* Teijsm. & Binn. ex Miq.	Rajasegar 20 (K)
XV[I]	*C. moseleyanus* Beccari	Baker 545 (K)
XVI[I]/*Cornera*	*C. conirostris* Becc.	Baker 532 (K)
XVII[I]	*C. pogonacanthus* Becc. ex H.J.P.Winkl.	Baker 522 (K)
XVIII[I]	*C. koordersianus* Becc.	Kramadibrata s.n. (K)
XVIII[I]	*C. pedicellatus* Becc. ex K.Heyne	Henderson 4361 (NY)
Calospatha	*C. calospathus* (Ridl.) W.J.Baker & J.Dransf.	1990-2783 (K)
Ceratolobus	*C. pseudoconcolor* (J.Dransf.) W.J.Baker	1975-3398 (K)
Daemonorops	*C. melanochaetes* (Blume) Miq.	Henderson 3992 (NY)
Daemonorops	*C. crinitus* subsp. *sabut* (Beccari) Henderson	Rajasegar 8 (K)
Pogonotium	*C. ursinus* (Becc.) W.J.Baker	Baker 517 (K)
Retispatha	*C. dumetosus* (J.Dransf.) A.J.Hend. & Floda	Baker 530 (K)
Unplaced[H,I]	*C. discolor* Mart.	Henderson 3996 (K)
Unplaced[I]	*C. inermis* T.Anderson	Oulathong 213 (K)
OUTGROUP		
Nypa	*N. fruticans* Wurmb.*	Chase 12603 (K)
Kerriodoxa	*K. elegans* J.Dransf.*	Baker 1987-2685 (K)
Ceroxylon	*C. quindiuense* (H.Karst) H.Wendl.	Baker 1976-1160 (K)
Asterogyne	*A. martiana* H.Wendl. ex Hemsl.*	Cano ACS361 (G)

Table 3. Sequence retrieval of PhyloPalm markers. Summary statistics from targeted sequencing of 971 PhyloPalm exons for 75 species of the Calamoideae and four outgroup species representing all other palm subfamilies. Q1 = 25 % quartile, Q3 = 75 % quartile.

		Reads total	Reads on target		Exons with minimum reference length			
		#	#	%	> 0 %	25 %	50 %	75 %
Calamoideae	Minimum	394577	61164	6.1	133	11	0	0
	Q1	2299154	399717	15.7	923	874	607	176
	Median	3343687	713749	22.2	947	912	722	307
	Q3	5290890	1061112	27.1	952	926	769	357
	Maximum	13234981	3858999	39.8	959	941	823	463
Outgroup	Median	5675654	1340924	31.6	963	955	917.5	698

Table 4. Alignment statistics. Summary statistics for trimmed alignments of 945 exons, supercontigs and splash zones analyzed in this study. Missing = missing or unidentified bases, Pars. inf. sites = parsimony informative sites. GC = guanine and cytosine bases. Statistics for concatenated splash zones were inferred indirectly from differences between concatenated exons and supercontigs in alignment length, number of missing, variable and parsimony informative sites, and bases.

				Missing	Variable sites		Pars. inf. sites		GC
		Taxa	Length	%	#	%	#	%	%
Exons	Minimum	13	69	0.0	18	13.7	11	2.9	35.4
	Median	75	1139	10.7	547	47.7	298	26.1	44.8
	Maximum	79	3810	62.4	2287	77.0	1426	50.2	66.6
	Concatenated	79	1167751	18.3	576659	49.4	322370	27.6	46.2
Supercontigs	Minimum	10	88	0.0	42	21.3	22	4.0	30.3
	Median	69	1604	10.2	933	59.2	487	30.8	40.3
	Maximum	77	10121	57.6	7359	87.3	4749	69.0	63.0
	Concatenated	79	1943739	28.0	1202869	61.9	664320	34.2	41.2
Splash zones	Concatenated	79	775988	42.6	626210	80.7	341950	44.1	29.4

Fig. 1. Diversity of calamoid palms. a, **b**, Typical characters. **a**, Fruit of *Raphia vinifera* P.Beauv with overlapping scales. **b**, Spines of *Calamus leloi* J.Dransf. **c**-**e**, Variety of growth forms. **c**, Arborescent habit of *Mauritiella aculeata* (Kunth) Burret. **d**, Acaulescent habit of *Salacca sarawakensis* Mogea. **e**, Climbing habit of *Calamus scipionum* Lour. Photographs BGK.

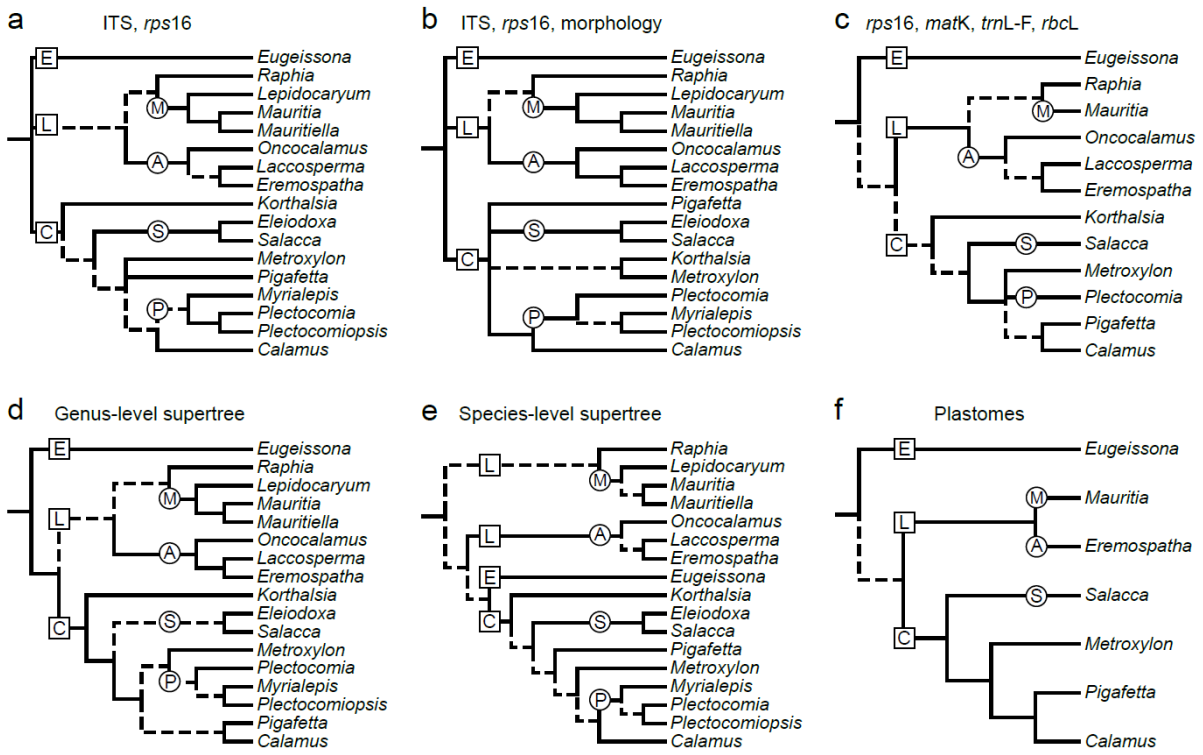

Fig. 2. Previous phylogenetic inferences of the Calamoideae. Selected previous inferences of relationships among calamoid tribes, subtribes and genera. Most topologies resolve the same generic relationships within the Lepidocaryeae, as well as identical placements of *Korthalsia* and the Salaccinae within the Calameae. In contrast, inferred relationships among the remaining genera of the Calameae and among the three tribes are often differing and poorly supported. **a**, Fig. 7 in Baker et al. (2000c). **b**, Fig. 4 in Baker et al. (2000a). **c**, Strict consensus (Fig. 1) in Asmussen et al. (2006). **d**, Most congruent supertree (Fig. 3) in Baker et al. (2009). **e**, Fig. 2 in Faurby et al. (2016). **f**, Fig. 3 in Barrett et al. (2016b). **a**-**c**, **f**, Jackknife or bootstrap branch support of 90 % or higher is indicated by solid lines, support between 50 % and 90 % is indicated by dashed lines, branches with support below 50 % are collapsed. **d**, Branches supported by 5 or more trees are indicated by solid lines, branches supported by less than 5 trees are indicated by dashed lines. **e**, Posterior branch support of 0.9 or higher is indicated by solid lines, branches with support between 0.5 and 0.9 and branches without support values provided in Table S1 (Faurby et al., 2016) are indicated by dashed lines. Data used for phylogenetic inferences are indicated above each subfigure. ITS is a nuclear ribosomal region, *rps*16, *mat*K, *trn*L-F and *rbc*L are plastid regions. Squares indicate tribes: C=Calameae, E=Eugeissoneae, L=Lepidocaryeae, and circles indicate selected subtribes: A=Ancistrophyllinae, M=Mauritiinae, P=Plectocomiinae, S=Salaccinae.

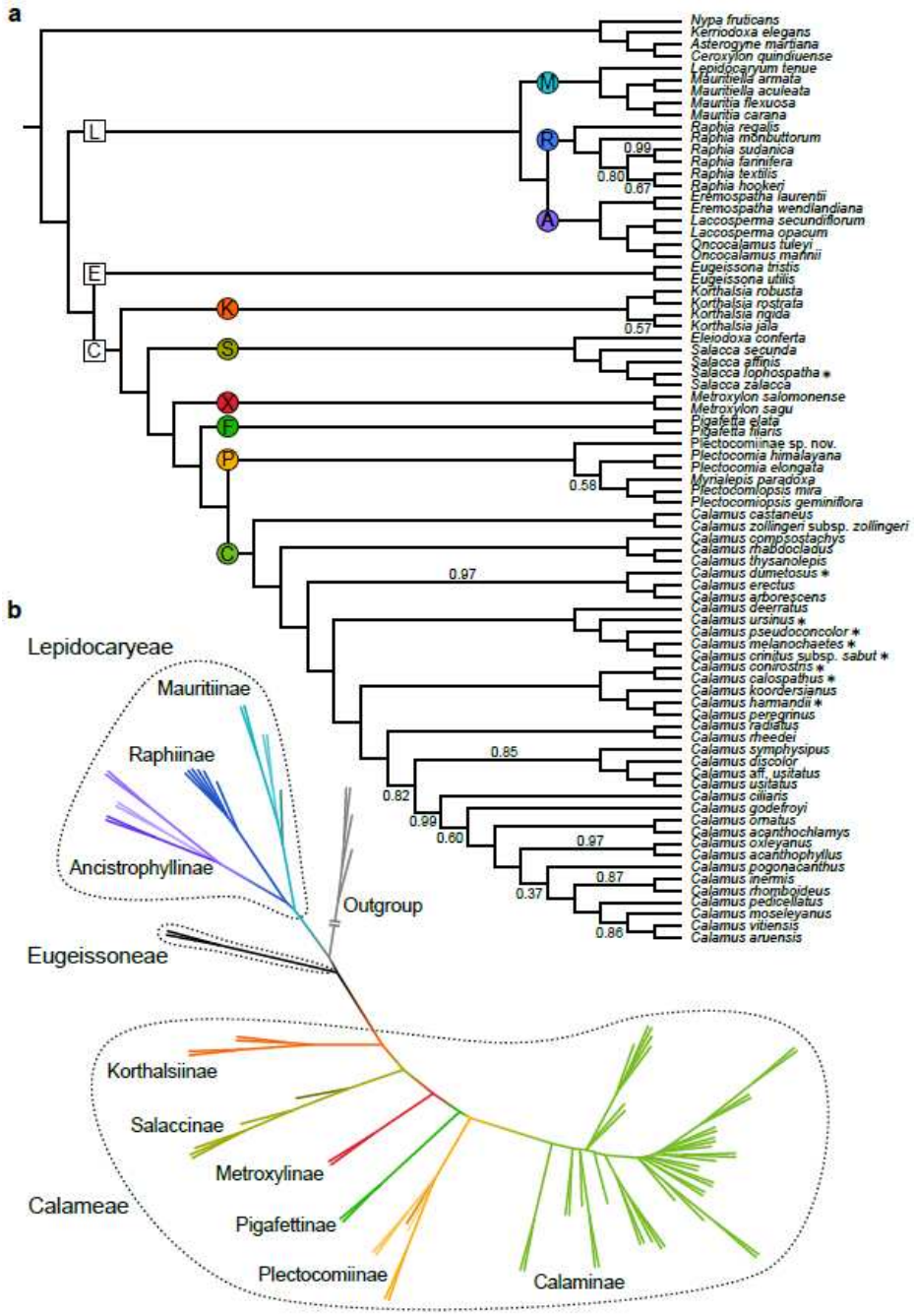

Fig. 3. Coalescence-based phylogenetic inferences of the Calamoideae. Phylogenetic trees inferred using coalescence analysis of 945 supercontigs and 75 calamoid species, plus four outgroup species representing the remaining four palm subfamilies. Supercontig trees were estimated using a general time reversible model of nucleotide substitution with gamma distributed rate variation among sites. Identical tribal, subtribal and generic relationships (except within subtribe Plectocomiinae) were inferred with maximal support by all phylogenomic analyses conducted in this study (Supplementary Fig. 2a). **a**, Cladogram showing relationships among species, with local posterior probabilities indicated only if below 1. Species representing synonymized genera are indicated by *. Squares indicate tribes: C=Calameae, E=Eugeissoneae, L=Lepidocaryeae, and circles indicate subtribes: A=Ancistrophyllinae, C=Calaminae, F=Pigafettinae, K=Korthalsiinae, M=Mauritiinae, P=Plectocomiinae, R=Raphiinae, S=Salaccinae, X=Metroxylinae. **b**, Phylogram with internal branch lengths proportional to coalescent units between branching events, as estimated by ASTRAL. Subtribes are colored as in subfigure **a**, but with the different genera in the Ancistrophyllinae, Mauritiinae, Salaccinae and Plectocomiinae marked by different shades. Terminal branch lengths are not estimated by ASTRAL and were set to a constant, arbitrary length.

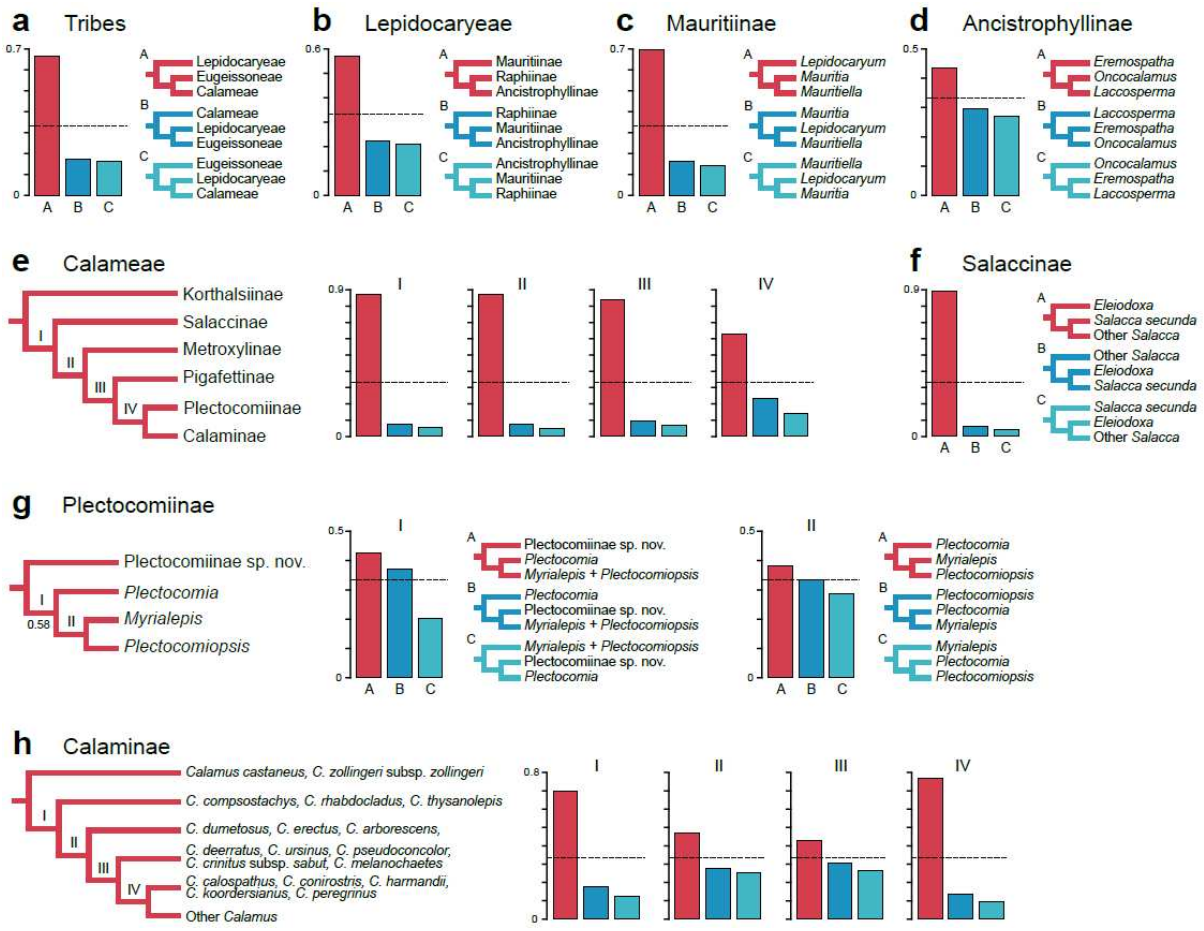

Fig. 4. Gene-tree conflict in higher-level relationships of the Calamoideae. Gene tree quartet frequencies for alternative topologies of selected higher-level relationships, based on coalescence analysis of 945 supercontigs (Fig. 3). Tribal and subtribal relationships, as well as relationships within Mauritiinae and Salaccinae, are supported by a clear majority of gene trees, whereas substantial conflict exists within Ancistrophyllinae, Plectocomiinae and partly Calaminae. Quartet frequencies are represented as bar graphs, with red bars (left) representing the main topology resolved by the coalescence supercontig analysis, and blue and turquoise bars (middle and right) representing alternative topologies. Small differences between the main topology and one or both alternative topologies indicate the presence of substantial conflicting phylogenetic signal. Dashed horizontal lines mark the expectation for equal frequencies of the three possible topologies (Y=0.333), i.e. maximal gene tree conflict. Branch support in local posterior probabilities and shown only when below 1. **a**, Tribes. **b**, Lepidocaryeae. **c**, Mauritiinae. **d**, Ancistrophyllinae. **e**, Calameae. **f**, Salaccinae. **g**, Plectocomiinae. **h**, Calaminae. In panels in **e** and **h**, only the main topology is shown.