# Phylogenomics and Biogeography of the Calamoid Palms

Thesis submitted by

**Benedikt Georg Kuhnhäuser**

for the award of Doctor of Philosophy

University of Oxford

Department of Plant Sciences

Brasenose College

©Benedikt Georg Kuhnhäuser, 2021

First supervisor: Prof Simon J Hiscock, University of Oxford

Second supervisors: Dr William J Baker, Royal Botanic Gardens Kew

Dr Guillaume Chomicki, University of Sheffield

Dr Wolf L Eiserhardt, Aarhus University & Royal Botanic Gardens Kew

Dr Sidonie Bellot, Royal Botanic Gardens Kew

*To Metterling*

## Acknowledgements

First and foremost, I would like to thank my love Diem-Tu for her continuous support, open ear, and eye for details, throughout the four years during which this thesis was written. We have walked the ways of our doctoral theses together. I would also like to thank my parents and grandmother for their unlimited support throughout my life.

I am more than grateful for the outstanding support by my large supervisory team that assembled over the years and provided a continuous stream of simultaneously critical and constructive feedback and encouragement, allowing me to steadily improve my abilities as an aspiring scientist. In particular, I thank Bill, who has been not only a supervisor to me but also a highly valued mentor.

My first encounter with the fantastic diversity of calamoid palms was in the tropical rainforests of Borneo in Sarawak, Malaysia. I thank my friend Peter Petoe for being my field companion during these seven weeks where we encountered more than a hundred palm species, “stingless” bees, and countless leeches. My supervisors Bill and Wolf, as well as palm authority John Dransfield, were kind enough to join parts of this field trip and introduced me to the intricacies of identifying and collecting palms. I am also grateful for the incredible hospitality of our local counterparts Connie Geri and Sirukit anak Dubod.

Finally, I am grateful for the invaluable contributions that a large number of other people and institutions have provided to my research activities. They are acknowledged in the results chapters of this thesis.

# Contents

**Acknowledgements** 3

**General abstract** 8

**1 General introduction** 10

1.1 Tropical rainforests 10

1.1.1 Diversification 11

1.1.2 Context dependence 12

1.1.3 Palms as a model group 13

1.2 The palm subfamily Calamoideae 15

1.2.1 Systematics 19

1.2.2 Morphology 20

1.2.3 Biogeography 28

1.3 Scope, aims, and research questions 30

1.4 Thesis structure 31

**2 A robust phylogenomic framework for the calamoid palms** 33

CRediT authorship contribution statement 33

Abstract 34

2.1 Introduction 34

2.2 Material and methods 43

2.2.1 Sampling 43

2.2.2 Sequencing 47

2.2.3 Phylogeny reconstruction 48

2.2.4 Data availability 51

2.3 Results 52

2.3.1 Informativeness of PhyloPalm markers in the Calamoideae 52

2.3.2 Phylogenetic relationships within the Calamoideae 54

2.4 Discussion 59

2.4.1 Efficiency of PhyloPalm probes 59

2.4.2 Higher-level relationships within the Calamoideae 60
2.4.3 Potential causes of gene tree conflict 61
2.4.4 Implications for higher-level classification 62
2.5 Conclusion 63
Funding 64
Declaration of Competing Interest 65
Acknowledgements 65
Supplementary material 66
**3 A dated species-level phylogeny of the calamoid palms 69**
Abstract 69
3.1 Introduction 70
3.2 Material and Methods 72
3.2.1 Sampling 72
3.2.2 Sequencing and phylogeny reconstruction 73
3.2.3 Fossil calibration 77
3.2.4 Divergence time estimation 83
3.3 Results 85
3.3.1 Retrieval of orthologous loci 85
3.3.2 Effects of alignment filtering 88
3.3.3 Phylogenetic placement of data-deficient samples 89
3.3.4 Species-level phylogenetic relationships of the Calamoideae 91
3.3.5 A time-calibrated phylogeny of the Calamoideae 93
3.4 Discussion 106
3.4.1 Towards a complete species-level phylogeny 106
3.4.2 Paralogy in the PhyloPalm loci 107
3.4.3 Systematic implications 109
3.4.4 Considerably older divergence time estimates compared to previous studies 112
3.4.5 Long lag times between lineage originations and build-up of extant diversity 114
3.4.6 Break-up of Gondwana linked to early phylogenetic history 116
3.5 Conclusion 119

Funding 120

Acknowledgements 120

Supplementary material 122

**4 Historical biogeography of the Calamoideae in the Asia-Pacific 165**

Abstract 165

4.1 Introduction 167

4.2 Material and Methods 174

4.2.1 Dated phylogeny 174

4.2.2 Biogeographic analyses 174

4.3 Results 179

4.3.1 DEC+J best model under all scenarios incorporating distance or connectivity 179

4.3.2 Ancestral ranges and regional radiations 182

4.3.3 Dispersals patterns across the Asia-Pacific 191

4.3.4 Differing roles of biogeographic regions in shaping the distribution of the Calamoideae in the Asia-Pacific 194

4.4 Discussion 197

4.4.1 Biogeographic history of the Asian-Pacific Calamoideae 197

4.4.2 Strong biogeographic significance of Indian and Australian collisions with Asia 199

4.4.3 Differing roles of land connectivity within Sundaland and Sahul 199

4.4.4 van Welzen's line demarcates contrasting dispersal frequencies in the Asia-Pacific 201

4.4.5 Prevalence of stepping-stone dispersal 202

4.4.6 Predominance of dispersals from Borneo to the periphery of the Asia-Pacific 202

4.4.7 Distribution better explained by present than past geology 203

4.4.8 Different roles of regions in establishing extant diversity 204

4.5 Conclusion 209

Funding 211

Acknowledgements 211

Supplementary Material 212

**5 General Discussion** **222**

5.1 The first robust phylogenomic framework of the Calamoideae 223

5.2 The most detailed dated phylogeny of a tropical rainforest lineage to date 225

5.3 Unparalleled insights into the biogeography of the Asia-Pacific region 228

5.4 New views on tropical rainforest evolution 231

5.5 Conclusion 236

**References** **237**

**Appendix** **267**

Appendix 1. Lanjak Entimau Wildlife Sanctuary – a palm hotspot in the heart of Borneo 267

Appendix 2. A robust phylogenomic framework for the calamoid palms 296

## General abstract

Tropical rainforests are the most diverse terrestrial biome, but their origins remain incompletely understood. Detailed analyses of the temporal and spatial context of the evolution of old and species-rich tropical rainforest lineages have potential to yield novel insights. However, only few such analyses are currently available.

The palm subfamily Calamoideae is proposed here as an appropriate study system to address this knowledge gap. It has a rich fossil record, comprises 555 species, and has a pantropical distribution with the centre of diversity in the Asia-Pacific region. There is a general understanding of its taxonomy, phylogeny, and distribution, but species-level relationships, divergence times, and biogeographic history of the subfamily are known only in vague outlines.

Here, three steps were undertaken to realise the full potential of the Calamoideae for providing new insights into the origins of tropical rainforests. First, targeted sequencing of 971 genes of 75 species was conducted, resulting in a robust phylogenomic framework of the Calamoideae. Second, sequencing was expanded to 653 calamoid specimens representing 502 species (90%) of the Calamoideae, resulting in 381 species (68%) in the final species-level phylogeny after data filtering. Divergence times were estimated based on four dating strategies, indicating an Early Cretaceous age of crown Calamoideae but an Oligocene and Miocene origin of most extant species. Finally, ancestral ranges and dispersal patterns were inferred for the Calamoideae in the Asia-Pacific region following the testing of 198 biogeographic models, showing that the subfamily reached its broad distribution during the Oligocene and Miocene convergence of Australia and Asia, and that migration pathways

closely followed a stepping-stone model with dispersals out of the centre and into the periphery of the Asia-Pacific region.

This work represents the largest densely sampled and dated molecular phylogeny of any major pantropical rainforest lineage to date and elucidates the biogeographic history of a characteristic component of the tropical rainforests in the Asia-Pacific region in unprecedented detail. It is hoped that it will serve as a detailed framework for future comparative analyses.

# 1 General introduction

## 1.1 Tropical rainforests

The exceptional diversity of tropical rainforests, housing almost half of all plant species in less than 7% of earth's land surface, has attracted the attention of biologists for over a century (Wallace, 1878; Eiserhardt et al., 2017). Tropical rainforests are defined as tall, dense, evergreen forests characterised by a wet, hot climate, the absence of pronounced dry seasons, and little seasonal variation (Corlett and Primack, 2011). At present, they form a belt around the equator and are broadly delimited by the name giving Tropics of Cancer and Capricorn at 23° latitude North and South, respectively. Tropical rainforests are commonly characterised by minimum annual rainfalls of 2000 mm and a coldest month mean temperature of at least 18°C (Whitmore, 1998; Morley, 2000; Hill and Hill, 2001). Definitions of rainforests based on vegetation physiognomy include a high proportion of leaves that are large, a predominance of entire-margined leaves, and the presence of leaf drip tips (Upchurch and Wolfe, 1987; Burnham and Johnson, 2004). Whilst trees form the main structure of tropical rainforests, shrubs, ground herbs, epiphytes and lianas form an integral part of the biome, accounting for up to two thirds of its species diversity (Gentry, 1992; Balslev et al., 1998). Depending on which definitions are used, the estimated extent and distribution of tropical rainforests on a global scale can differ considerably (Eiserhardt et al., 2017).

Fossil evidence indicates that angiosperm-dominated, multi-layered, and closed-canopy biomes resembling present-day tropical rainforests have existed since the Cretaceous-Paleogene boundary at 66 million years ago (Morley, 2011; Carvalho et al., 2021). However, leaf fossils consistent with tropical rainforest physiognomy (Upchurch and

Wolfe, 1987) and dated molecular phylogenies of tropical rainforest lineages, such as Annonaceae, Arecaceae, Malpighiales and Menispermaceae (Davis et al., 2005; Couvreur et al., 2011a; Couvreur et al., 2011b; Wang et al., 2012), point to a considerably earlier origin of several tropical rainforest lineages at least 100 million years ago.

### 1.1.1 Diversification

The high species diversity of tropical rainforests has been attributed to environmental stability over large geological timescales resulting in low extinction rates (Stebbins, 1974; Davis et al., 2005; Couvreur et al., 2011a), the recent stimulation of high speciation rates (Richardson et al., 2001; Erkens et al., 2007), or the combination of both (Koenen et al., 2015; Pennington et al., 2015). It must be noted, however, that the dissection of net diversification rates into speciation and extinction rates is methodologically challenging (Purvis, 2008; Rabosky, 2010; Beaulieu and O'Meara, 2015; Rabosky, 2016; Louca and Pennell, 2020, 2021).

Immigration of species from other biomes may have added to the diversity of tropical rainforests. Although little evidence is currently available for this scenario (Donoghue and Edwards, 2014), future species-level phylogenies of lineages with mixed distributions inside and outside of tropical rainforests are expected to contribute to a better understanding of this factor (Eiserhardt et al., 2017).

Assuming overall positive net diversification rates, the old age of tropical rainforests and their lineages could explain their richness even if net diversification rates were low (Wiens, 2011; Stadler et al., 2014). Phylogenetic evidence from tropical rainforest taxa

overall supports this view (Eiserhardt et al., 2017), suggesting that the age of tropical rainforests at least partly explains their high diversity.

Environmental factors such as current and past climates, habitat heterogeneity, area or insularity have been shown to impact the distribution and diversification of tropical rainforest lineages (Kissling et al., 2012a; Kissling et al., 2012b). More hotly debated is the role of ecological limits, which are hypothesised to result in a negative feedback loop on species diversity once certain carrying capacities are reached (Rabosky, 2009; Wiens, 2011; Donoghue and Edwards, 2014; Harmon and Harrison, 2015; Rabosky and Hurlbert, 2015).

### 1.1.2 Context dependence

The largest tropical rainforests currently occur in the Americas, the Asia-Pacific, and Africa. While seeking overarching patterns explaining the origins of tropical rainforest diversity, it should not be forgotten that their evolution is context dependent, with each region shaped by different geological histories, climatic peculiarities, and unique assemblages of plant and animal lineages (Corlett and Primack, 2011; Hagen et al., 2021).

The geology of the Americas was shaped by the uplift of the Andes in the Neogene and the closure of the Isthmus of Panama between North and South America in the Pliocene (Hoorn et al., 2010). At present, the South American Amazon basin houses the most extensive continuous tropical rainforests on earth (Eiserhardt et al., 2017). In Africa, continental uplift and aridification since the Late Eocene have resulted in a dramatic contraction of tropical rainforest area (Kissling et al., 2012b; Couvreur et al., 2021). The geological history of the Asia-Pacific, comprising the world's largest tropical archipelago, is highly complex. The assembly and distribution of the tropical rainforests of the Asia-Pacific

were shaped by the subsequent collisions of India and Australia with Asia, the submergence and re-emergence of landmasses (such as Sumatra or Java) in the recent geological past, and a fluctuating connectivity of landmasses within Sundaland (Thai-Malay Peninsula, Sumatra, Java and Borneo) and Sahul (New Guinea and Australia), respectively, during strongly alternating sea levels in the Quaternary (Hall, 1998, 2011, 2012a, 2017).

Whereas the geological events in the Americas and the Asia-Pacific facilitated migration and provided new habitats and niches that might have stimulated diversification (Hughes et al., 2013; Kooyman et al., 2019), the loss of tropical rainforest area in Africa was likely associated with highly increased extinction rates (Couvreur et al., 2021; Hagen et al., 2021). This interpretation is supported by the similarly high richness of tropical rainforest plants in the Americas and the Asia-Pacific (Slik et al., 2015; Joyce et al., 2020a), whereas Africa has a remarkably lower number of tropical rainforest species and has been labelled accordingly the 'odd man out' (Richards, 1973; Couvreur, 2015).

### 1.1.3 Palms as a model group

The palm family Arecaceae, one of the most emblematic tropical plant lineages (Tomlinson, 2006; Corlett and Primack, 2011), is an established model group to study tropical rainforest evolution (Couvreur et al., 2011a; Couvreur and Baker, 2013). Indeed, physiological constraints render most palms intolerant to freezing (Tomlinson, 2006) and restrict their distribution to tropical and subtropical climates (Kissling et al., 2012a), with over 90% of species occurring in tropical rainforests (Couvreur et al., 2011a).

Palms have an abundant and well-documented fossil record going back to the Late Cretaceous (Harley, 2006) and have existed throughout the recorded history of tropical

rainforests (Baker and Couvreur, 2013a). They were the most abundant angiosperms in some biota, such as disturbed environments in North America, from the Late Cretaceous onwards (Wing et al., 1993; Wing and Boucher, 1998; Morley, 2000), and continue to dominate habitats and play key roles in all tropical rainforests of the world (Dransfield et al., 2008; Eiserhardt et al., 2011).

The palm family has elicited interest of researchers for centuries (Martius et al., 1823-50; Beccari, 1908, 1911, 1918; Uhl and Dransfield, 1987), culminating in the availability of outstanding fundamental data resources, such as a comprehensive family monograph on genus level (Dransfield et al., 2008; Baker and Dransfield, 2016) and a complete and continuously updated checklist of all palm species including distribution data (Govaerts and Dransfield, 2005; Govaerts et al., 2021).

The Arecaceae consists of ca. 2500 species in 181 genera and 5 subfamilies (Baker and Dransfield, 2016). The monophyly of the family is unequivocally supported by molecular studies (Chase et al., 1993; Chase et al., 1995; Chase et al., 2000; Lewis and Doyle, 2001; Asmussen et al., 2006; Chase et al., 2006; Baker et al., 2009; Barrett et al., 2016b), and morphologically by the synapomorphy of a unique leaf developmental mode in which the folded lamina expands and splits on predetermined lines (Dransfield et al., 2008).

A good general understanding of the infra-familiar relationships of palms is available from a large number of phylogenetic studies (Uhl et al., 1995; Baker et al., 1999a; Asmussen et al., 2000; Baker et al., 2000a; Baker et al., 2000c; Hahn, 2002b, a; Asmussen et al., 2006; Barrett et al., 2016b), synthesised in a complete and dated genus-level phylogeny of palms (Baker et al., 2009; Couvreur et al., 2011a; Baker and Couvreur, 2013a), as well as an all-evidence species-level phylogeny (Faurby et al., 2016), providing frameworks to identify and

address unresolved questions in the evolution of palms. However, substantial uncertainty remains for many critical relationships, especially in the subfamilies Arecoideae and Calamoideae (Baker and Dransfield, 2016).

## 1.2 The palm subfamily Calamoideae

With 555 species in 17 genera, 10 subtribes and 3 tribes (Fig. 1.1), the Calamoideae is the second largest palm subfamily and accounts for more than one fifth of the palms (Dransfield et al., 2008; Baker and Dransfield, 2016). Excepting the monotypic subfamily Nypoideae, the Calamoideae is the only subfamily that is almost completely restricted to wet tropical and subtropical climates (Couvreur et al., 2011a; Reichgelt et al., 2018). Calamoid palms occur in the tropical rainforests of the Asia-Pacific, Africa, and the Americas. Members of the subfamily display a broad variety of growth forms, but their climbing members, the rattans, are probably best known. The predominantly climbing genus *Calamus* contains over 400 species, making it the by far most species-rich genus of the palm family. Most species of the Calamoideae occur in the rainforests of the Asia-Pacific (Baker and Couvreur, 2012).




**Figure 1.1. Diversity of the Calamoideae. a**, *Mauritiella armata* (Mauritiinae). **b**, *Raphia farinifera* (Raphiinae). **c**, *Laccosperma acutiflorum* (Ancistrophyllinae). **d**, *Eugeissona minor* (Eugeissoneae). Photos: **a**-**b**, William J. Baker. **c**, John Dransfield. **d**, Benedikt G. Kuhnhäuser.




**Figure 1.1** (*continued*)**. e**, *Korthalsia jala* (Korthalsiinae). **f**, *Salacca rupicola* (Salaccinae). **g**, *Metroxylon sagu* (Metroxylinae). **h**, *Pigafetta filaris* (Pigafettinae). Photos: **e**-**f**, Benedikt G. Kuhnhäuser. **g**-**h**, William J. Baker.




**Figure 1.1** (*continued*)**. i**, *Plectocomiopsis geminiflora* (Plectocomiinae). **j**, *Calamus formicarius* (Calaminae). **k**, *Calamus ingens* (Calaminae). **l**, *Calamus pygmaeus* (Calaminae). Photos: Benedikt G. Kuhnhäuser.

### 1.2.1 Systematics

The Calamoideae is strongly supported as sister to all other palms by recent phylogenetic studies (Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016b; Faurby et al., 2016). The oldest fossil records of the Calamoideae are *Dicolpopollis* pollen from the Cenomanian (100.5-93.9 million years ago) of Australia (Burger, 1990; Macphail and Jordan, 2015). Fossil-constrained molecular age estimates date the stem node age to 100 million years and the crown node age to 80 million years (Baker and Couvreur, 2013a), although these age estimates did not use such early fossils.

The monophyly of the subfamily is unequivocally supported by phylogenetic evidence (Uhl et al., 1995; Baker et al., 1999a; Asmussen et al., 2000; Baker et al., 2000c; Lewis and Doyle, 2001; Hahn, 2002a; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016b; Faurby et al., 2016) and a large suite of well-defined reproductive and anatomical synapomorphies: 1) overlapping, reflexed scales covering ovary and fruit, 2) anatropous ovules with the micropyle facing the centre of the gynoecium, 3) carpels with open ventral sutures, 4) organisation of flowers in sympodial pairs (dyads) or derivates thereof, 5) transverse veins sheathed by many layers of fibres, 6) adaxial nonvascular fibres with subepidermal distribution, 7) inflated protoxylem parenchyma in petiolar vascular bundles, and 8) phloem of stems and leaves with simple sieve plates (Uhl and Moore Jr., 1971; Baker et al., 1999b; Dransfield et al., 2008; Horn et al., 2009; Tomlinson et al., 2011).

All subtribes, as well as tribes Calameae and Eugeissoneae are strongly supported as monophyletic (Baker et al., 2000a; Baker et al., 2000c; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016b; Faurby et al., 2016; Faye et al., 2016). However, uncertainty exists regarding the monophyly of tribe Lepidocaryeae, the relationships between the three

calamoid tribes, and relationships between the subtribes within tribes Calameae and Lepidocaryeae, respectively (Asmussen et al., 2000; Baker et al., 2000a; Baker et al., 2000c; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016b; Faurby et al., 2016; Faye et al., 2016).

Whereas all-evidence "supertree" approaches have been utilised to close existing knowledge gaps by combining all available genetic and morphological data (Baker et al., 2009; Faurby et al., 2016), their power is limited by the complete lack of molecular phylogenetic data for over 80% of the Calamoideae (Faurby et al., 2016).

### 1.2.2 Morphology

Calamoid palms are most easily identified by their fruits with reflexed overlapping scales and the presence of spines (Figs. 1.2 - 1.3), and yet exhibit an extraordinary variety of growth forms, life history strategies, flowers and inflorescences, sexual systems, and leaf forms (Figs. 1.1, 1.4 - 1.6). Their leaves have reduplicate leaflets (Λ-shaped in section) and are usually pinnate (feather-shaped). Subtribe Mauritiinae (Lepidocaryeae), however, is highly distinctive by having palmate (fan-shaped) leaves (Dransfield et al., 2008). The Calamoideae includes tall trees and stemless undergrowth plants but is most notable for the climbing habit in 90% of its species (Couvreur et al., 2015). Nine genera are solely composed of trees and shrubs, seven genera only contain climbers, and the genus *Calamus* consists mainly of climbers, but has a proportion of ca. 10% non-climbing species (Couvreur et al., 2015).




**Figure 1.2. Diversity of spines in the Calamoideae. a**, *Plectocomia mulleri*. **b**, *Calamus exilis*. Note the short, bristly spines. **c**, *Calamus leloi*. **d**, *Calamus ursinus*. Photos: Benedikt G. Kuhnhäuser.

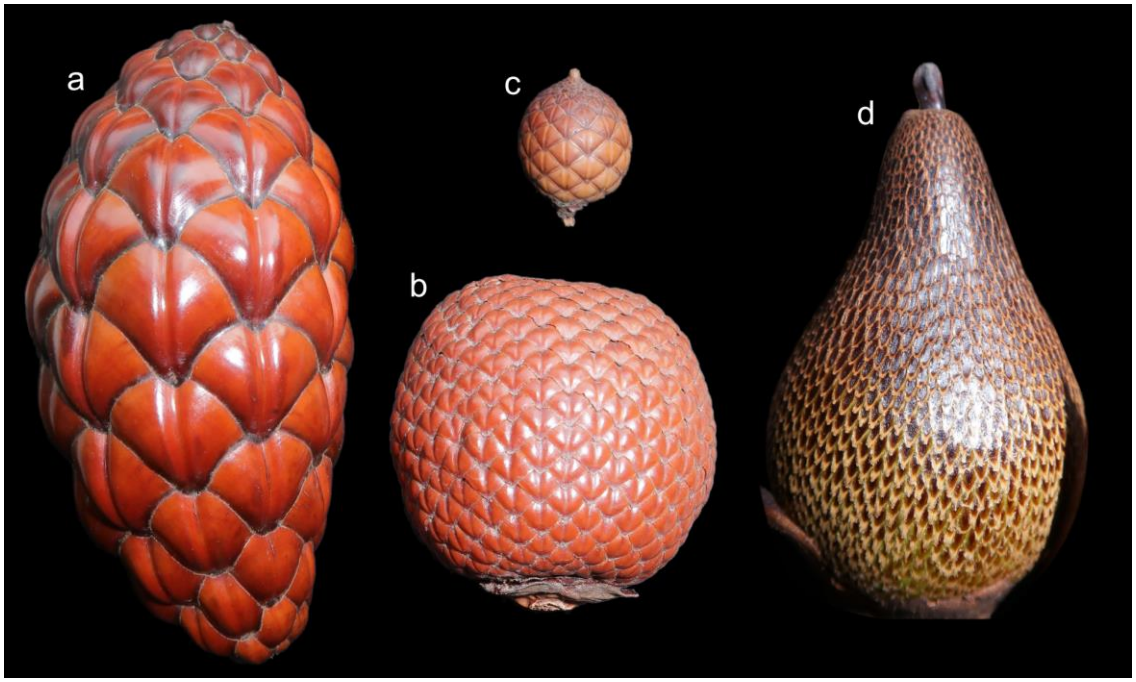



**Figure 1.3. Fruits of the Calamoideae.** Note reflexed scales covering the fruits. **a**, *Raphia vinifera*. **b**, *Mauritia flexuosa*. **c**, *Calamus* sp. **d**, *Eugeissona minor*. Photos: Benedikt G. Kuhnhäuser.




**Figure 1.4. Life history strategies in the Calamoideae. a**, Hapaxanthy in *Metroxylon salomonense*, superficially resembling a terminal inflorescence. **b**, Pleonanthy in *Pigafetta filaris*, with inflorescences in leaf axils. Photos: William J. Baker.




**Figure 1.5. Flowers and inflorescences in the Calamoideae. a**, Infrutescence of *Eugeissona minor*. Note persistent lignified tepals. **b**, Female (pistillate) inflorescence of *Calamus longipes*. Note the pairs (dyads) of fertile female flowers and dysfunctional male (staminate) flowers. **c**, Male (staminate) inflorescence of *Calamus ruptiloides*. **d**, Inflorescence of *Calamus hallierianus*. The entire inflorescence is enclosed in a persistent bract. Photos: Benedikt G. Kuhnhäuser.




**Figure 1.6. Leaf forms in the Calamoideae. a**, *Mauritiella armata* with palmate leaves. **b**, *Korthalsia jala* with diamond-shaped leaflets. **c**, *Salacca sarawakensis* with entire bifid leaves. **d**, *Calamus ruptiloides* with pinnate leaves and lanceolate leaflets. Photos: Benedikt G. Kuhnhäuser.

### *Reproductive systems*

Reproductive systems in the Calamoideae are highly diverse (Fig. 1.4). Nine calamoid genera have a hapaxanthic reproductive strategy, where a prolonged vegetative phase is followed by a compressed reproductive phase that exhausts the shoot and results in its death (Fig. 1.4a). In two additional genera, only single species are hapaxanthic. The flowering structure superficially appears to be terminal but is morphologically identical to pleonanthy (Fig. 1.4b), occurring in the remaining eight genera, where the reproductive phase is more integrated in the vegetative growth and principally occurs indeterminately. Similarly, breeding systems are also highly variable between genera. By the variation of sexes within the two flowers of the dyad, a broad bandwidth of sexual systems exists in different genera, ranging from hermaphroditism to polygamy (mixture of hermaphroditic and unisexual flowers), monoecy and dioecy (Baker et al., 1999b). Five genera are hermaphroditic, two are monoecious, and ten are dioecious. In contrast to the highly diverse life history strategies and sexual systems of calamoid palms, low diversity in chromosome numbers has been observed. Diploid chromosome numbers range between 2n=26 and 2n=32, and no polyploid species are known (Dransfield et al., 2008).

### *Pollen*

In addition to its distinct macromorphological characters, the Calamoideae has been described as the palynologically most distinctive palm subfamily (Harley and Morley, 1995). Pollen grains of the subfamily can be highly characteristic and often allow the allocation to tribe or genus level, such as of the equatorially disulcate pollen genus *Dicolpopollis* to the tribes Metroxylinae, Plectocomiinae and Calaminae, the spiny *Mauritiidites* to subtribe

Mauritiinae, and the large and extended monosulcate *Quilonipollenites* to *Eugeissona* (Harley, 2006).

***Climbing***

The climbing habit characterises the subtribes Ancistrophyllinae, Korthalsiinae, Plectocomiinae and Calaminae, which have several morphological adaptations and innovations to climbing (Fig. 1.7). A high ratio of internode length to stem diameter can be used reliably to identify the climbing habit in Calamoideae (Baker et al., 2000a). All rattans possess recurved, hook-like spines that act as climbing aids. In subtribes Korthalsiinae, Plectocomiinae and Calaminae, these spines are usually organised in whorls ("grapnel spines"), improving the climbing efficiency compared to genera in subtribe Ancistrophyllinae, which have dispersed spines (Baker et al., 1999b). Only the Ancistrophyllinae also have acanthophylls, i.e. reflexed, shortened and stiff leaflets acting as climbing aids. Most rattans possess specialised climbing organs that are stiff but flexible, armed with grapnel spines or acanthophylls, and can be more than 10m long (Isnard and Rowe, 2008). The cirrus is a leafless elongation of the leaf rachis occurring in all rattan genera. In some species of *Calamus*, an alternative climbing organ, the flagellum, exists. It is a modified, sterile inflorescence converted into a climbing organ (Dransfield et al., 2008). Functional cirri and flagella never co-occur in the same species, although some cirrate rattans also have a short vestigial (dysfunctional) flagellum, and some flagellate rattans have so-called subcirrate leaf tips with distant, highly reduced leaflets. In addition to spines and climbing organs, most species of *Calamus* also have a knee-like swelling of the leaf sheath below the insertion of the petiole. The function of the "knee" has not been investigated

systematically, but it has been hypothesised that it contributes to climbing efficiency (Isnard and Rowe, 2008).

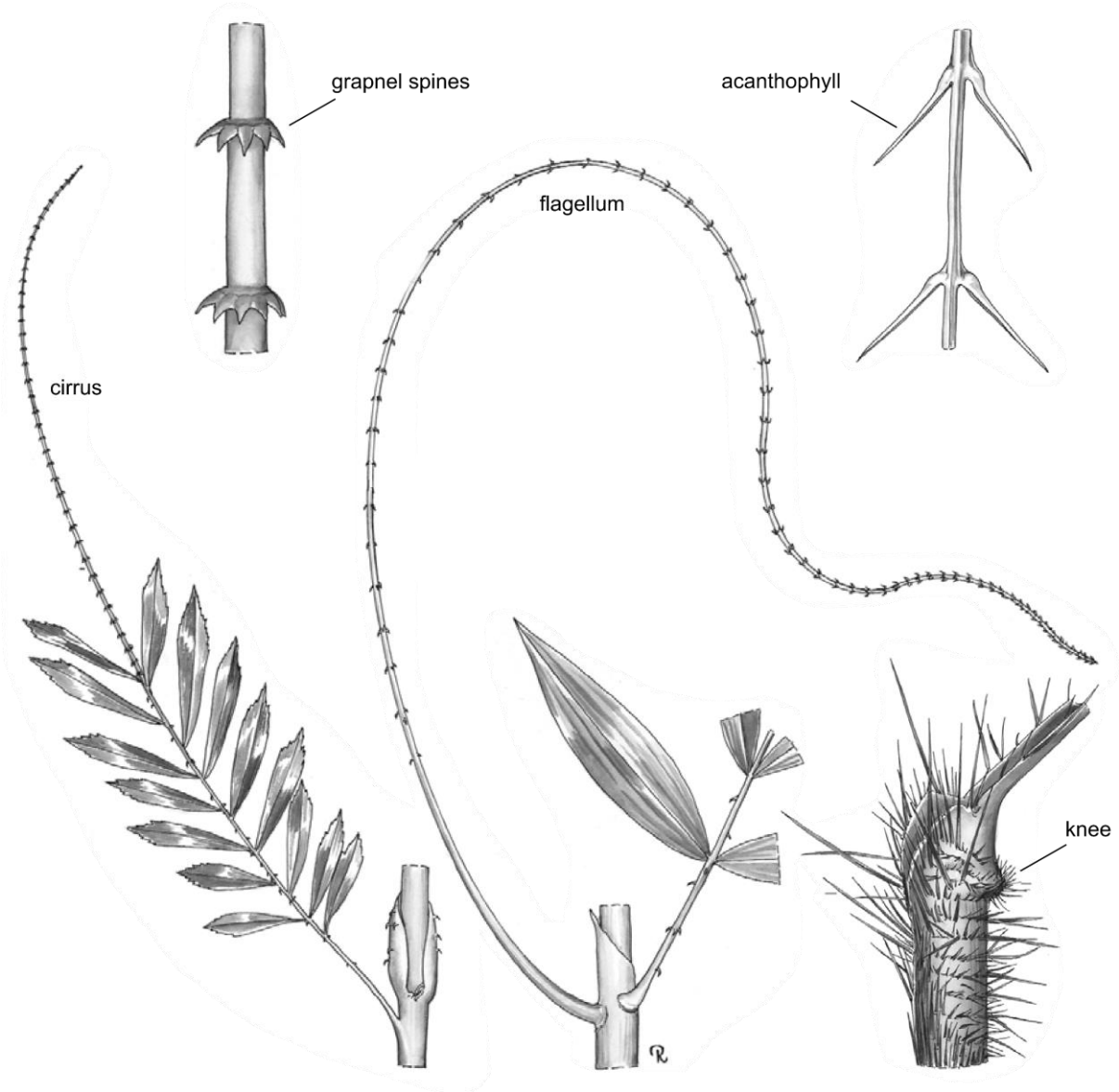



**Figure 1.7. Climbing adaptations in the Calamoideae.** Modified from Dransfield et al. (2008).

Rattans employ a climbing strategy that can be considered fundamentally different from other climbing plants (Putz and Chai, 1987; Putz, 1990). Whereas most climbers actively attach to support structures by mechanisms of twining or with adventitious roots, rattans climb passively by attaching to neighbouring plants with their hook-like spines or acanthophylls borne on cirri or flagella. This enables them to attach to trees with a larger stem diameter than possible for most other climbers. Rows of spines along the length of the climbing organ create a ratchet-like mechanism leading to closer attachment to supports upon movements of either the rattan or the support. The climbing organs of rattans are more short-lived than in other climbing plants and need to be continuously replaced. Upon establishment in the canopy, rattans maintain their position by the combination of decreased longitudinal growth and downward slippage of the stem when attachment organs are replaced.

### 1.2.3 Biogeography

With ca. 500 species in ten genera, the Asia-Pacific accounts for over 90% of the Calamoideae. Six genera are confined to the West of Wallace's line (a biogeographic boundary running between Bali and Lombok, Borneo and Sulawesi, and east of the Philippines), two to the East, and two are spread across the whole region (Baker and Couvreur, 2012). *Calamus* has the broadest distribution, reaching from India to Fiji, and having one disjunct representative in Africa. Malesia, the region comprising the Thai-Malay Peninsula, Sumatra, Java, Borneo, the Philippines, Sulawesi, the Moluccas, the Lesser Sunda Islands and New Guinea, is the centre of calamoid species richness (Baker and Couvreur, 2012). Palms in this region show high degrees of endemism on genus and species level

(Sterling et al., 2007; Baker and Couvreur, 2012), demonstrating the importance of geographic isolation in this region.

Asian calamoid palms are thought to have colonised the Malesian archipelago from West to East, with several subsequent back-dispersals in the opposite direction (Baker and Couvreur, 2013b). However, the biogeographic history of the Calamoideae in the Asia-Pacific is currently known only in its rough outlines owing to the lack of a detailed understanding of their phylogenetic relationships.

Consistent with general patterns of plant and palm diversity, Africa has a much lower diversity of calamoid palms than Asia, containing 40 species in five genera (Kissling et al., 2012b; Couvreur, 2015). A recent study of the African subtribe Ancistrophyllinae, accounting with 21 species in three genera for half of the African Calamoideae and one third of all African palms, traces the low number of species back to one or several extinction events in the late Eocene or early Oligocene, and to constant diversification rates since then (Faye et al., 2016).

In the Americas, only nine calamoid species occur. Eight of these belong to the three genera of tribe Mauritiinae, which is thought to have expanded from Africa into America ca. 70 million years ago (Baker and Couvreur, 2013a). One further species belongs to the African genus *Raphia*, potentially presenting a Holocene long-distance dispersal event (Helmstetter et al., 2020). The causes for the paucity of the Calamoideae in the Americas compared to Africa and Asia are not clear.

## 1.3 Scope, aims, and research questions

Dated molecular phylogenies can provide deep insights into the temporal and spatial origins of tropical rainforests (Pennington et al., 2004; Pennington et al., 2006; Richardson and Pennington, 2016; Eiserhardt et al., 2017). Most dated molecular phylogenies of tropical rainforest lineages that are available to date are either based on dense sampling of relatively small lineages (Grudinski et al., 2014; Faye et al., 2016; Strijk et al., 2020) or sparse sampling of large lineages (Couvreur et al., 2011a; Couvreur et al., 2011b; Ramírez-Barahona et al., 2020), but a notable lack of densely sampled studies of large lineages exist (Eiserhardt et al., 2017).

The aim of the present study is to address this knowledge gap by 1) generating a densely sampled phylogeny of the species-rich palm subfamily Calamoideae, 2) estimating divergence times, incorporating its rich fossil record, and 3) conducting biogeographic analyses of the subfamily with a focus on the geologically complex Asia-Pacific region.

These objectives are associated with the following research questions: 1) What are the higher-level and species-level relationships of the Calamoideae? 2) When did the subfamily originate, and what are divergence times of extant calamoid lineages? 3) Where did the Calamoideae originate, and how did it attain its high diversity across the tropical rainforests of the Asia-Pacific region? 4) What are implications for tropical rainforest evolution?

## 1.4 Thesis structure

The structure of this work is closely aligned with these overarching research aims and questions. In the beginning of the project, seven weeks of field work in Borneo were conducted to collect high-quality plant material for phylogenomic analyses. During this time, I encountered 85 calamoid palm species and collected samples of 71 different calamoid taxa. Most of my photographs in the introduction (Chapter 1) are based on this work. The findings of an expedition to the remote primary rainforests of Lanjak Entimau Wildlife Sanctuary during my field work in Borneo have been published in the journal Palms (Appendix 1). To achieve near-complete species-level coverage of the Calamoideae, the sampling was then expanded to 653 specimens representing 502 species (90%) of the Calamoideae based on an extensive survey of the collections held by 24 herbaria worldwide.

Chapter 2 investigates higher-level relationships of the Calamoideae based on targeted sequencing of 971 genes and 75 calamoid species representing all tribes, subtribes, and genera, as well as the majority of infrageneric sections. To test whether the results are dependent on the methods employed, eight different approaches of phylogeny reconstruction are compared. Gene-tree conflict in higher-level relationships is analysed to allow the differentiation of highly robust relationships that are supported by a strong majority of gene trees from ambiguous relationships with high conflict between the phylogenies of individual gene trees. This chapter has already been published in the journal Molecular Phylogenetics and Evolution (Appendix 2).

In Chapter 3, this phylogenomic framework is substantially expanded by sampling 653 calamoid specimens representing 502 (90%) of the 555 currently recognised species of the Calamoideae. Several filtering steps are implemented to exclude data-deficient samples,

resulting in 381 species (68% of the Calamoideae) in the final species-level phylogeny. Divergence times are estimated based on four dating strategies incorporating different fossil placements and maximum age constraints as well as recently discovered fossils of the Calamoideae that are the oldest known palm fossils that can be assigned to a clade nested within the palm family. Implications for the age of the palms and the temporal and spatial context of the origin of the crown Calamoideae are discussed.

Chapter 4 utilises the dated species-level phylogeny to estimate ancestral areas of the Calamoideae in the Asia-Pacific region. For 12 biogeographic regions, 198 biogeographic models are compared varying from unconstrained models to models considering distance between regions, land connectivity, and geological changes through time. For the model best explaining the data, dispersal events between regions are quantified using biogeographic stochastic mapping. On this basis, proportions of emigration and in-situ diversification are inferred for each biogeographic region, revealing contrasting roles of different regions in contributing to the high diversity of the Calamoideae in the tropical rainforests of the Asia-Pacific.

In Chapter 5, the findings of this thesis are summarised. Implications of this work for the understanding of the origins of tropical rainforests are discussed, and future directions of research on the Calamoideae that are enabled by this work and might further increase the understanding of the evolution of tropical rainforests are proposed.

# 2 A robust phylogenomic framework for the calamoid palms

## CRediT authorship contribution statement

**Benedikt G. Kuhnhäuser:** Conceptualization, Funding acquisition, Resources, Methodology, Investigation, Formal analysis, Validation, Visualization, Data curation, Project administration, Writing - original draft, Writing - review & editing. **Sidonie Bellot:** Funding acquisition, Resources, Methodology, Investigation, Supervision, Writing - review & editing. **Thomas L.P. Couvreur:** Resources, Methodology, Writing - review & editing. **John Dransfield:** Resources, Methodology, Validation, Writing - review & editing. **Andrew Henderson:** Resources, Methodology, Validation, Writing - review & editing. **Rowan Schley:** Methodology, Investigation, Writing - review & editing. **Guillaume Chomicki:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Wolf L. Eiserhardt:** Conceptualization, Funding acquisition, Resources, Methodology, Investigation, Supervision, Writing - review & editing. **Simon J. Hiscock:** Conceptualization, Methodology, Supervision, Writing - review & editing. **William J. Baker:** Conceptualization, Funding acquisition, Resources, Methodology, Validation, Project administration, Supervision, Writing - review & editing.

## 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.

## 2.1 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., 2011a; 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., 2011a; Kissling et al., 2012b; Baker and Couvreur, 2013b, a) and distribution (Kissling et al., 2012a; 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, 2002a; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016b; Faurby et al., 2016). It contains ca. 550 species in 17 genera, 10 subtribes and three tribes (Table 2.1), accounting for more than one fifth of the palm family. Calamoid palms occur in tropical rainforests across the world and are most diverse in the Asia-Pacific region (Table 2.1). 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. 2.1a) and the presence of spines (Fig. 2.1b). They also exhibit an extraordinary variety of growth forms (Fig. 2.1c-e), the climbing habit (Fig. 2.1e) being predominant (Kissling et al., 2019).




**Figure 2.1. Diversity of calamoid palms. a**, **b**, Synapomorphies of the Calamoideae. **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.

**Table 2.1. Classification, diversity, and distribution of the calamoid genera.** Classification follows Baker and Dransfield (2016). Species numbers according to Govaerts et al. (2021) 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**<br>**Subtribe**<br>*Genus* | Authority | Species | Distribution |
|---|---|---|---|
| **EUGEISSONEAE** | W.J. Baker & J. Dransf. | | |
| *Eugeissona* | Griff. | 6 | Malay Peninsula and Borneo |
| **LEPIDOCARYEAE** | Mart. | | |
| **Ancistrophyllinae** | Becc. | | |
| *Oncocalamus* | (G.Mann & H.Wendl.) H.Wendl | 4 | Africa |
| *Eremospatha* | (G.Mann & H.Wendl.) Schaedtler | 11 | Africa |
| *Laccosperma* | (G.Mann & H.Wendl.) Drude | 7 | Africa |
| **Raphiinae** | H. Wendl. | | |
| *Raphia* | P.Beauv. | 21 | Africa and Madagascar[1] |
| **Mauritiinae** | Meisn. | | |
| *Lepidocaryum* | Mart. | 1 | South America |
| *Mauritia* | L.f. | 2 | South America |
| *Mauritiella* | Burret | 4 | South America |
| **CALAMEAE** | Kunth | | |
| **Korthalsiinae** | Becc. | | |
| *Korthalsia* | Blume | 28 | Myanmar to New Guinea |
| **Salaccinae** | Becc. | | |
| *Eleiodoxa* | (Becc.) Burret | 1 | Thailand to Borneo |
| *Salacca* | Reinw. | 23 | Myanmar 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 | Myanmar to Sumatra |
| *Plectocomiopsis* | Becc. | 6 | Myanmar to Borneo |
| **Calaminae** | Meisn. | | |
| *Calamus* | L. | 414 | India to Fiji[2] |

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., 2016b) and a large suite of well-defined morphological-anatomical synapomorphies (Uhl and Moore, 1971; Baker et al., 1999b; 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; Baker et al., 2000c). These studies resulted in a phylogenetic classification of the Calamoideae (Baker et al., 2000a) that is followed in the prevailing classification of palms (Dransfield et al., 2008; Baker and Dransfield, 2016). 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., 2000b). Five genera nested within *Calamus* (*Calospatha* Becc., *Ceratolobus* Blume ex Schult. & Schult.f., *Daemonorops* Blume, *Pogonotium* J.Dransf., and *Retispatha* J. Dransf.) have now been synonymized to render the genus monophyletic (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. 2.2). 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; Baker et al., 2000c; 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., 2000c; 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., 2000c; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016b; 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; Baker et al., 2000c; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016b; Shahimi, 2018), but recent analyses resolved Calameae as sister to Eugeissoneae with intermediate support (Barrett et al., 2016b; Faurby et al., 2016). In Calameae, multiple contrasting relationships between the subtribes Calaminae, Metroxylinae, Pigafettinae and Plectocomiinae have been resolved (Baker et al., 2000a; Baker et al., 2000c; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016b; Faurby et al., 2016). In Plectocomiinae, *Plectocomia* has been resolved either as sister to *Plectocomiopsis* (Baker et al., 2000c; Faurby et al., 2016) or as sister to both *Myrialepis* and *Plectocomiopsis* (Baker et al., 2000a; Baker et al., 2009; Shahimi, 2018).

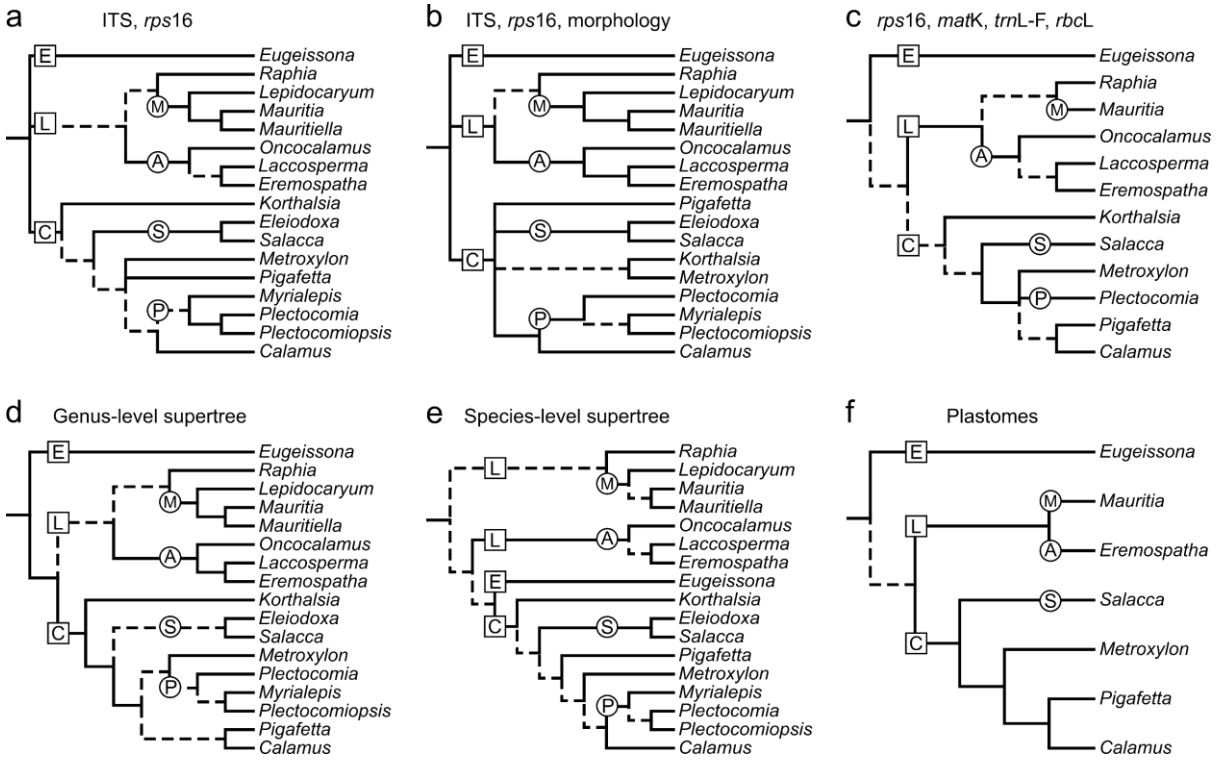



**Figure 2.2. Previous phylogenetic inferences of the Calamoideae.** Selected previous inferences of relationships among calamoid tribes, subtribes and genera. Most topologies resolve the same generic relationships within the Lepidocaryeae, as well as identical placements of *Korthalsia* and the Salaccinae within the Calameae. In contrast, inferred relationships among the remaining genera of the Calameae and among the three tribes are often differing and generally 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.

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 studies. Rapid technological advances in DNA sequencing (Barrett et al., 2016a; 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 123 of the 176 loci of Heyduk et al. (2015), 8 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 primarily for the phylogenetically divergent Arecoideae and Coryphoideae (Heyduk et al., 2015; de La Harpe 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 tribes, subtribes and genera, 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.

## 2.2 Material and methods

### 2.2.1 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, 1981b), *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.

**Table 2.2. 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 (1981b), [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**<br>**Subtribe**<br>***Genus***<br>**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) |

**Table 2.2. Sampling** (*continued*)

| **TRIBE** | | |
|---|---|---|
| **Subtribe** | | |
| ***Genus*** | | |
| **Section** | **Species** | **Voucher** |
| 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) |
| 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) |

**Table 2.2. Sampling** (*continued*)

| **TRIBE** | | |
|---|---|---|
| **Subtribe** | | |
| ***Genus*** | | |
| **Section** | **Species** | **Voucher** |
| 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) |
| 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) |

### 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 × 150 bp paired-end reads.

### 2.2.3 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., 2020), 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 2.2) 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, 2021), 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 coalescent 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.

### 2.2.4 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).

## 2.3 Results

### 2.3.1 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 2.3). 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 2.3), 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.

**Table 2.3. 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 % |
| | **Minimum** | 394577 | 61164 | 6.1 | 133 | 11 | 0 | 0 |
| | **Q1** | 2299154 | 399717 | 15.7 | 923 | 874 | 607 | 176 |
| **Calamoideae** | **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 |

The median number of retrieved reads, proportion of reads on target, and recovered proportion of the targeted sequence length were higher in the outgroup than in the calamoid samples (Table 2.3), but no apparent taxonomic bias existed within the Calamoideae (Supplementary Fig. 2.1). Across all taxa including the outgroup, lower target length proportions of the Heyduk markers were recovered compared to the other PhyloPalm markers (Supplementary Fig. 2.1).

The trimmed exon alignments had a total length of over one million base pairs and contained 27.6% parsimony informative sites (Table 2.4). 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 2.4).

**Table 2.4. 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 |

### 2.3.2 Phylogenetic relationships within the Calamoideae

The eight species trees inferred in this study resolved identical higher-level relationships but differed at shallower levels (Fig. 2.3, Supplementary Fig. 2.2a). Overarching methodologies (coalescent vs. concatenation) and datasets (exons vs. supercontigs) had the strongest effect on differences in tree topology, whereas substitution model testing in the coalescent analyses and partitioning of concatenated analyses had the smallest effects (Supplementary Fig. 2.2b). When only considering overarching methodologies and datasets, the coalescent analyses inferred highly similar relationships, whereas larger degrees of conflict existed between the other possible pairwise comparisons of species tree topologies (Supplementary Fig. 2.2b).

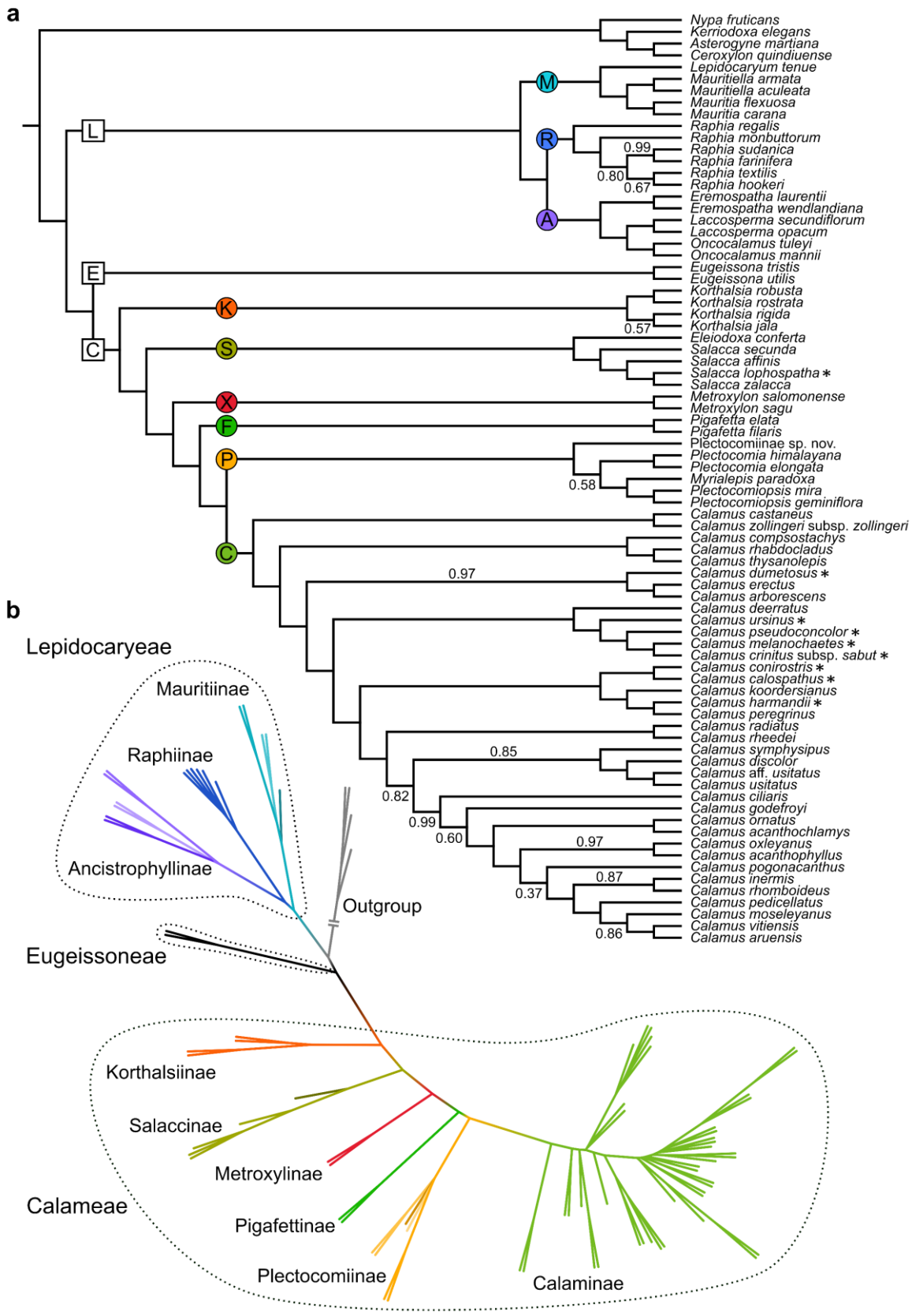



**Figure 2.3. Phylogenomic inferences of the Calamoideae based on coalescent analysis of 945 supercontigs.**

**Figure 2.3** (*continued*). Phylogenetic trees of 75 species representing the diversity of the Calamoideae, plus four outgroup species (*Nypa fruticans*, *Kerriodoxa elegans*, *Asterogyne martiana*, *Ceroxylon quindiuense*) representing the remaining four palm subfamilies. Individual 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 by all phylogenomic analyses conducted in this study (Supplementary Fig. 2.2a). **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.

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. 2.3, Supplementary Fig. 2.2a). 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. 2.3a). All six sections of *Calamus* that

were represented by at least two species (sections I, II, V, XII, XV and XVIII; Table 2.2) were resolved as polyphyletic (Fig. 2.3a). 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. 2.2a).

We found low gene tree conflict for all tribal and subtribal relationships, as well as higher-level relationships in Mauritiinae and Salaccinae (Fig. 2.4a,b,c,e,f). In contrast, substantial gene tree conflict was identified in Ancistrophyllinae and especially Plectocomiinae (Fig. 2.4d, 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. 2.3, Supplementary Fig. 2.2a), whereas substantial conflict was inferred for two of the other three investigated relationships in the backbone of *Calamus* (Fig. 2.4h).

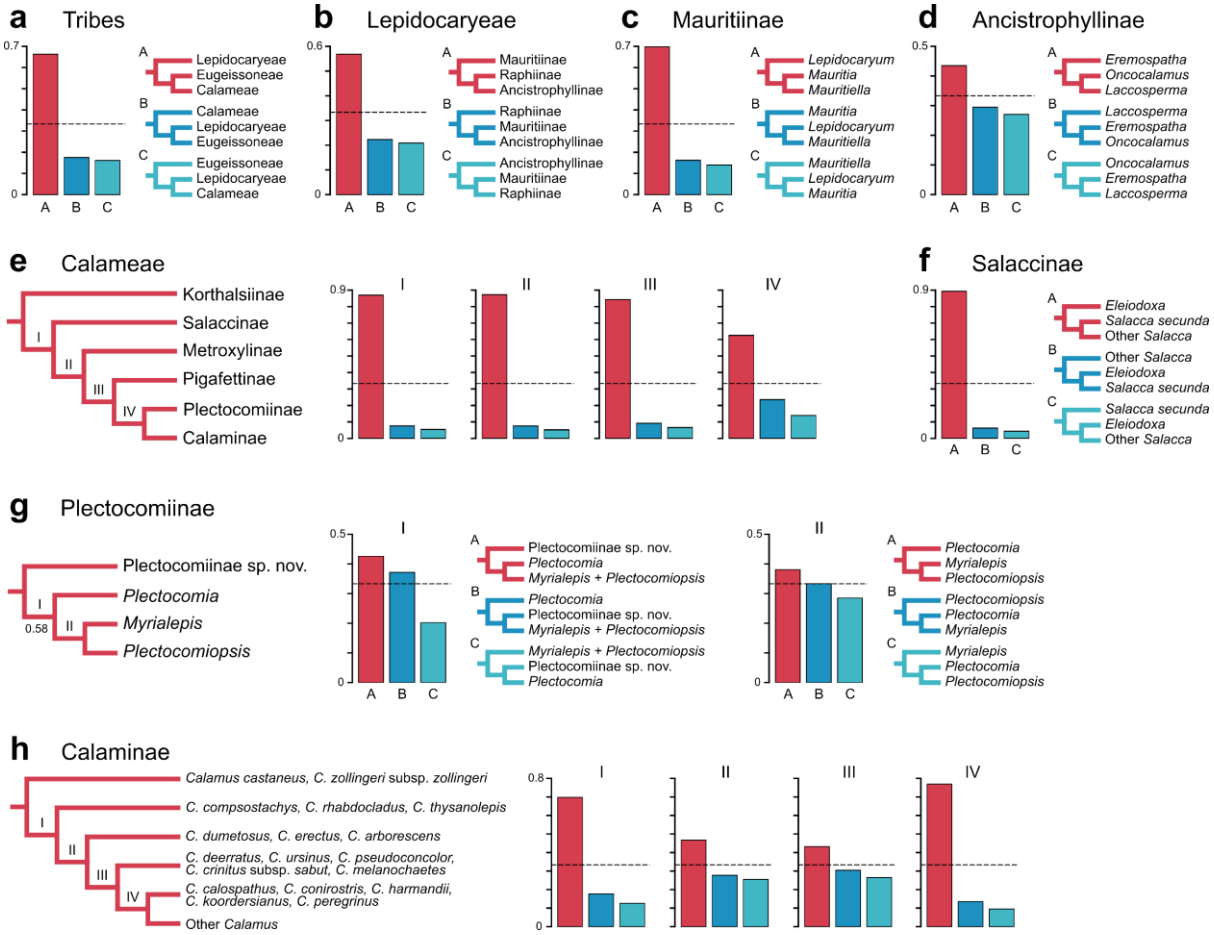



**Figure 2.4. Gene-tree conflict in higher-level relationships of the Calamoideae.** Gene tree quartet frequencies for alternative topologies of selected higher-level relationships, based on coalescent analysis of 945 supercontigs (Fig. 2.3). Tribal and subtribal relationships, as well as relationships 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 coalescent 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.

## 2.4 Discussion

### 2.4.1 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 2.3). Recovery of the included Heyduk markers was lower compared to the remaining PhyloPalm markers for all samples including the outgroup (Supplementary Fig. 2.1).

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 2.4) and can be used to resolve relationships both at deep and shallow levels (Figs. 3, 4, Supplementary Fig. 2.2a). 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 2.4), 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.

### 2.4.2 Higher-level relationships within the Calamoideae

Our phylogenetic inferences of higher-level relationships within the Calamoideae establish a robust evolutionary framework (Fig. 2.3) that is based on genomic-scale sampling of phylogenetically informative markers and accounts for gene tree conflict (Fig. 2.4). The firm establishment of Eugeissoneae as sister to Calameae (Fig. 2.4a) resolves the previously ambiguous relationships among these tribes (Fig. 2.2). In Lepidocaryeae, our findings (Fig. 2.3) 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. 2.4b), whereas previous analyses predominantly resolved Raphiinae as sister to Mauritiinae, albeit only with weak to intermediate support (Baker et al., 2000a; Baker et al., 2000c; Asmussen et al., 2006; Baker et al., 2009; Faye 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; Baker et al., 2000c; Asmussen et al., 2006; Baker et al., 2009; Faurby et al., 2016; Faye et al., 2016). This can potentially be explained by the relatively high degree of gene tree conflict in Ancistrophyllinae (Fig. 2.4d) 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; Baker et al., 2000c; Faurby et al., 2016) was validated (Fig. 2.4c). In Calameae, all subtribal relationships were resolved with maximal support and low gene tree conflict (Figs. 3, 4e), elucidating the previously poorly understood relationships among the subtribes Metroxylinae, Pigafettinae, Plectocomiinae and Calaminae (Baker et al., 2000a; Baker et al., 2000c; Asmussen et al., 2006; Baker et al., 2009; Barrett et al., 2016b; Faurby et al., 2016; Shahimi, 2018), and confirming the well-established placements of Korthalsiinae and Salaccinae (Baker et al., 2000c; Asmussen et al., 2006; Baker et al., 2009; Faurby et al., 2016; Shahimi, 2018). In Plectocomiinae, the presence of strong gene tree conflict (Fig. 2.4g) mirrors the inference of contrasting relationships among genera in previous studies (Baker et al., 2000a; Baker et al., 2000c; Baker et al., 2009; Faurby et al., 2016; Shahimi, 2018). In Calaminae, the backbone inferred in this study (Fig. 2.3, Supplementary Fig. 2.2a) is congruent with the summary relationships presented in the only previous comprehensive study of this subtribe (Baker et al., 2000b). 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. 2.4h, Supplementary Fig. 2.2a).

### 2.4.3 Potential causes of gene tree conflict

Substantial gene tree conflict was observed in Ancistrophyllinae, Plectocomiinae and Calaminae (Fig. 2.4d,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. Hybridization seems a plausible explanation because distributions of genera are strongly overlapping within the African Ancistrophyllinae and the Asian Plectocomiinae, respectively (Table 2.1), 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. 2.4g), which appears morphologically intermediate between the three genera of Plectocomiinae. 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).

### 2.4.4 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. 2.3). 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. 2.4f), 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. 2.4g) hints to a complex evolutionary history of the subtribe. In Calaminae, the strongly supported nesting of all synonymized genera within *Calamus* (Fig. 2.3a) 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. 2.4h), 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 2.2, Fig. 2.3a).

## 2.5 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).

## 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 NERC Independent Research Fellowship (NE/ S014470/1) to GC; 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.

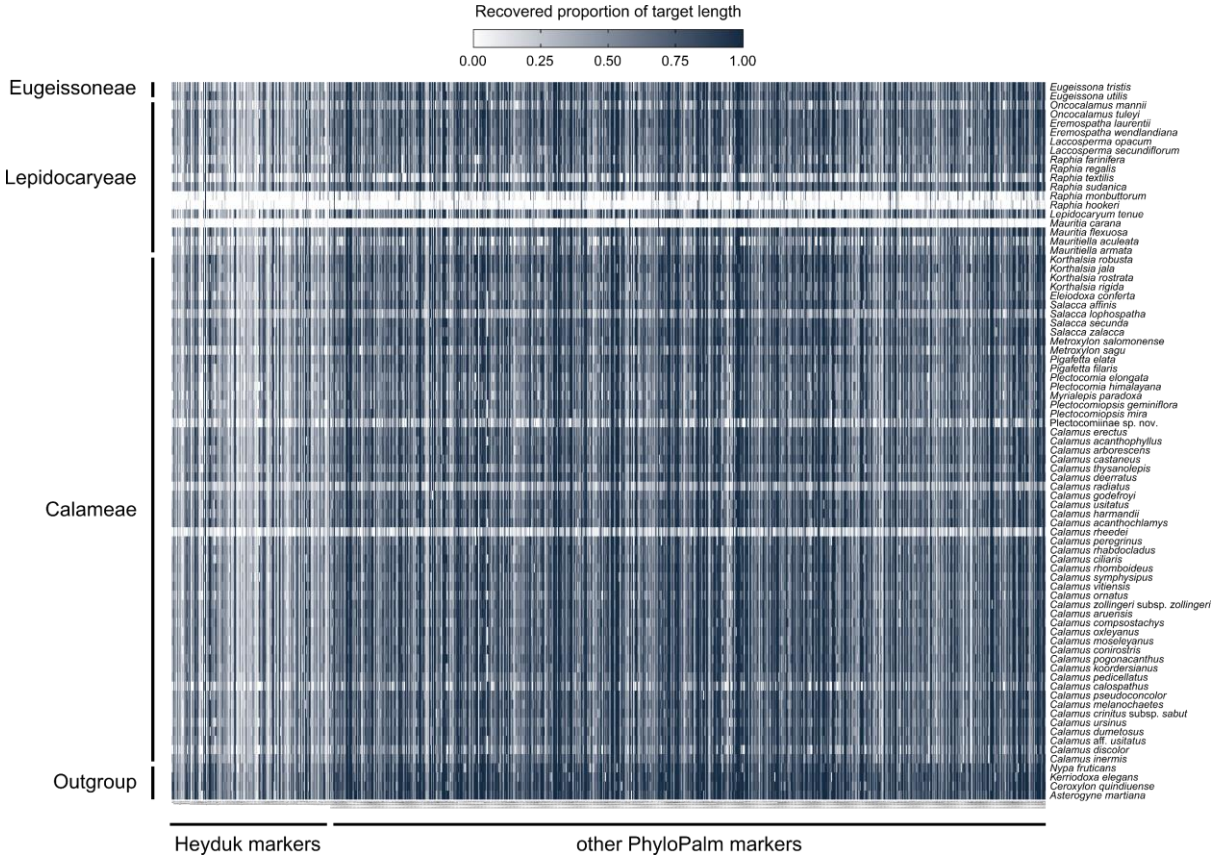
## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

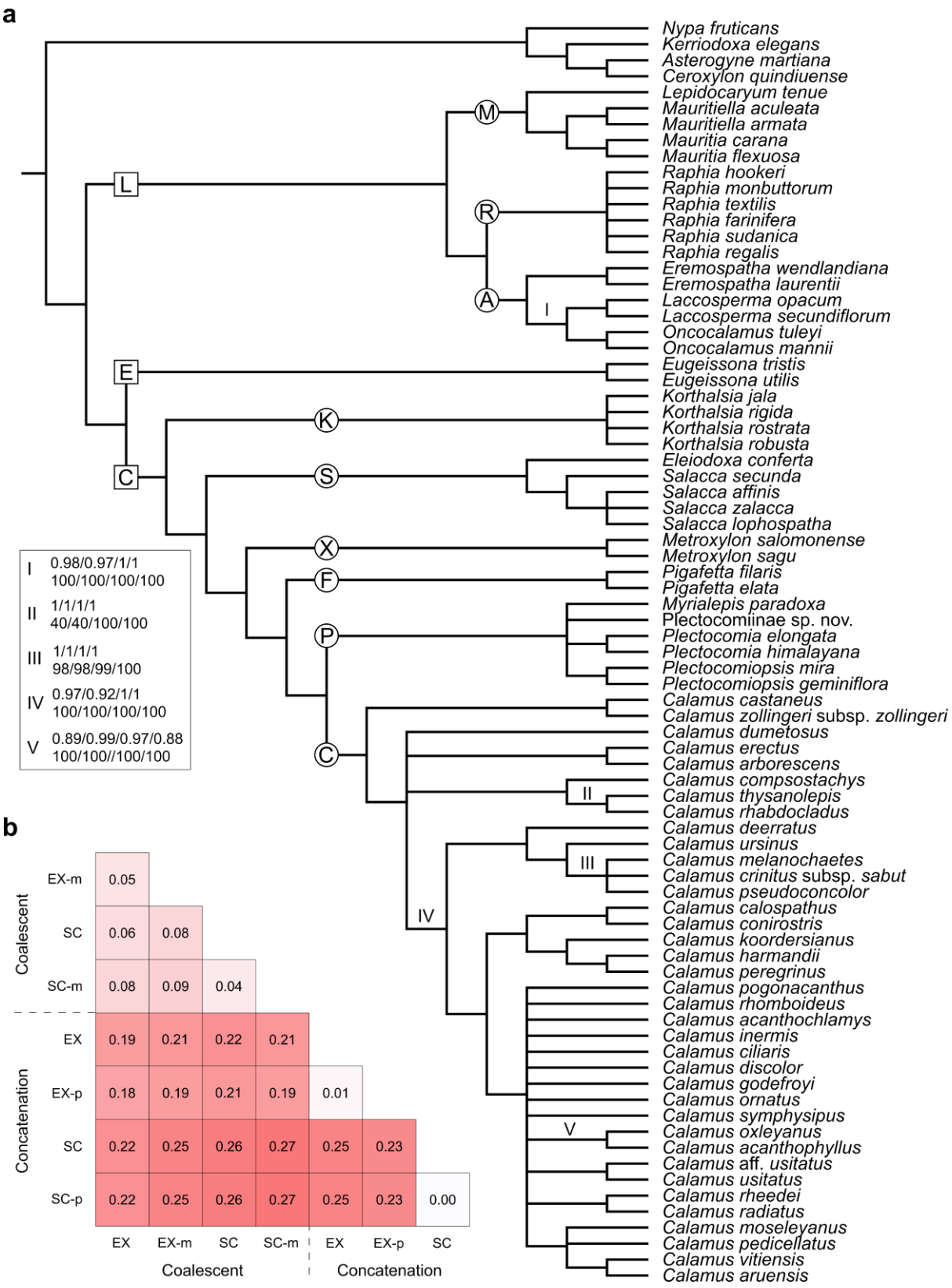
## Acknowledgements

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.

## Supplementary material




**Supplementary Figure 2.1. Retrieval of PhyloPalm markers in the Calamoideae and outgroups.** Data matrix showing recovery of the 971 PhyloPalm markers (Loiseau et al., 2019), which include all 176 markers by Heyduk et al. (2015), relative to the length of the target reference. Retrieved markers that were longer than the target reference are represented as maximal recovery. Darker colors indicate higher recovery.




**Supplementary Figure 2.2. Comparison of phylogenomic inferences of the Calamoideae in this study.**

**Supplementary Figure 2.2** (*continued*). **a**, Strict consensus of four coalescent-based and four concatenation-based species trees of 75 calamoid species and four outgroup species, with conflicting topologies collapsed to polytomies. Un-collapsed branches without support values were maximally supported by all analyses. Branches without maximal support values in all analyses are indicated by roman numerals. For these, the support values are given in the inset, with the upper row giving local posterior probabilities of coalescent analyses (in the order exons / exons with substitution model testing / supercontigs / supercontigs with substitution model testing), and the lower row giving bootstrap percentages of concatenation analyses (in the order exons / exons with partitioning / supercontigs / supercontigs with partitioning). 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**, Matrix of normalized Robinson-Foulds distances between the eight species trees inferred in this study. EX = exons, SC = supercontigs, m = with substitution model testing, p = with partitioning. Smaller distances and lighter colors indicate more similar tree topologies.

# 3 A dated species-level phylogeny of the calamoid palms


## Abstract

Dated and densely sampled phylogenetic trees of major plant groups have great potential to inform the long-standing controversy on the temporal and spatial origins of tropical rainforests, but only few such trees currently exist. In this chapter, the pantropical, old, and species-rich palm subfamily Calamoideae was used to address this issue. Based on the sampling of 653 specimens representing 502 (90%) of the 555 species of the Calamoideae and targeted sequencing of 971 nuclear genes, several filtering steps were implemented to exclude potentially paralogous genes and data-deficient samples, resulting in a highly resolved phylogeny comprising 354 genes and 381 species (68%) of the Calamoideae. Divergence times were inferred using four calibration strategies, incorporating recently discovered *Dicolpopollis* pollen fossils from the Cenomanian (93.9 – 100.5 Ma) of Australia that represent the oldest currently known palm fossils and can be assigned with high confidence to a clade nested in tribe Calameae. All relationships at generic level were fully resolved and strongly supported, and all genera were strongly supported as monophyletic. Species-level relationships were generally highly resolved and strongly supported, largely validating existing infrageneric classifications but also identifying some sections as paraphyletic or polyphyletic. Divergence time estimates were substantially older compared to previous studies, with the estimated minimum crown node age of the Calamoideae of 122.3 Ma predating previously established maximum age estimates for the crown nodes of both the Calamoideae and the palm family Arecaceae by over 20 and 10 million years, respectively. The inferred higher-level relationships and divergence times, as well as the locations of the earliest calamoid fossils, are fully reconcilable with the sequence and timing

of the break-up of Gondwana, suggesting that the pantropical distribution of the Calamoideae may be explained by a Gondwanan origin and geographical disjunction caused by the separation of its constituent landmasses. Despite this Early Cretaceous age of the crown Calamoideae, ages inferred for the vast majority of extant calamoid lineages substantially post-date the Cretaceous, revealing that the Calamoideae are an ancient palm subfamily almost exclusively constituted by young lineages. Overall, this study represents the largest densely sampled, dated species-level phylogeny of any major tropical rainforest lineage to date. It sheds new light on the temporal origin of palms and provides a highly detailed framework for future comparative studies.

## 3.1 Introduction

Dated and densely sampled phylogenetic trees of major plant groups have great potential to inform the long-standing controversy on the temporal and spatial origins of tropical rainforests (Pennington et al., 2004; Eiserhardt et al., 2017). However, only few such trees currently exist, especially in the Asia-Pacific region (Eiserhardt et al., 2017). The rapid advance of DNA sequencing technologies in the last decade, enabling the routinely retrieval of hundreds of genes even from ancient or highly processed materials (Heyduk et al., 2015; Loiseau et al., 2019; Baker et al., 2021; Cowell et al., 2021; Pérez-Escobar et al., 2021), can be utilised to address this knowledge gap.

The palm family is an established model group for tropical rainforest evolution (Morley, 2000; Couvreur et al., 2011a; Baker and Couvreur, 2013b, a; Couvreur and Baker, 2013; Couvreur et al., 2015; Reichgelt et al., 2018; Hansen et al., 2021). Although this potential builds on a large body of knowledge of the taxonomy, systematics, phylogeny,

morphology, distribution, and fossil record of the family (Harley, 2006; Dransfield et al., 2008; Baker et al., 2009; Tomlinson et al., 2011; Iles et al., 2015; Kissling et al., 2019; Govaerts et al., 2021), most of the phylogeny of the ca. 2500 extant palm species is currently only known well at the genus level (Baker et al., 1999a; Baker et al., 2009; Barrett et al., 2016b; Faurby et al., 2016). The inference of species-level relationships and divergence times of large palm lineages promises to provide an even deeper understanding of tropical rainforest evolution.

The monophyletic palm subfamily Calamoideae plays a key role in the understanding of the age and geographic origin of the palms due to its phylogenetic position as sister to all other palms (Baker et al., 2009; Faurby et al., 2016), and because the oldest unambiguous palm fossils that can be attributed to a lineage within the family, the pollen taxon *Dicolpopollis* (Burger, 1990; Macphail and Jordan, 2015), are placed in the Calamoideae (Harley and Morley, 1995; Morley, 2000; Harley, 2006; Iles et al., 2015). The Calamoideae contains 555 species in 17 genera and three tribes (Kuhnhäuser et al., 2021), most of which occur in the Asia-Pacific region, and has an exceptionally distinctive fossil record for multiple lineages (Morley, 2000; Harley, 2006; Iles et al., 2015). Previous studies have laid a strong foundation for a comprehensive dated phylogeny of the Calamoideae, including taxonomic revisions at the genus level (Uhl and Dransfield, 1987; Baker and Dransfield, 2008; Dransfield et al., 2008; Baker, 2015; Henderson and Floda, 2015) and species level (Beccari, 1908, 1911, 1918; Dransfield, 1981b; Madulid, 1981; Dransfield, 1982; Otedoh, 1982; McClatchey, 1998; Sunderland, 2012; Henderson, 2020), increasingly robust phylogenetic frameworks for the subfamily (Baker et al., 2000a; Baker et al., 2000c, b; Baker et al., 2009; Barrett et al., 2016b; Faurby et al., 2016; Kuhnhäuser et al., 2021), several dated phylogenies that provide crown and stem age estimates for the Calamoideae (Couvreur et al., 2011a; Baker and Couvreur,

2013a; Faurby et al., 2016; Antonelli et al., 2017; Givnish et al., 2018; Ramírez-Barahona et al., 2020), and comprehensive reviews of the fossil record (Harley and Morley, 1995; Harley, 2006; Iles et al., 2015; Bogotá-Ángel et al., 2021). However, less than one fifth of the calamoid species have been sampled at a phylogenomic scale (Kuhnhäuser et al., 2021), and molecular sequence data is lacking for most of the remaining species (Faurby et al., 2016). Given the rich fossil record, sister relationship to all other palms, high species diversity, and centre of diversity in the Asia-Pacific region, the inference of species-level relationships and divergence times for the Calamoideae has great potential to yield novel insights into the evolution of palms, and therefore tropical rainforests.

Here, I present a densely sampled, dated species-level phylogeny of the Calamoideae. Based on 381 (68%) of its 555 species, 354 nuclear genomic loci, and the implementation of recently discovered fossils, this dated phylogeny presents the most densely sampled dated species-level phylogeny of any major tropical rainforest plant lineage to date. On this basis, the following two fundamental research questions are addressed: 1) What are the phylogenetic relationships among calamoid species, and 2) When did extant calamoid lineages diverge? In this light, implications for the temporal and spatial origins of palms are discussed.

## 3.2 Material and Methods

### 3.2.1 Sampling

Sampling in this study comprised 653 specimens representing 502 species (90%) of the Calamoideae, including 471 species (95%) of tribe Calameae, six species (100%) of tribe

Eugeissoneae, and 25 species (49%) of tribe Lepidocaryeae (Supplementary Table 3.1). The taxonomy for palms presented in the World checklist of Arecaceae (Govaerts et al., 2021) was followed here, except for the case of *Calamus*, for which the taxonomy of a recent monograph (Henderson, 2020) was accepted. In addition, recently published species in *Raphia* (Helmstetter et al., 2020; Mogue Kamga et al., 2020), *Mauritiella* (Torres Jiménez et al., in press) and *Calamus* (Adorador and Fernando, 2020; Mondal et al., 2020) were recognised. This combined taxonomy recognises 555 accepted species of the Calamoideae, of which 51 occur in tribe Lepidocaryeae, six in Eugeissoneae, and 498 in Calameae (Kuhnhäuser et al., 2021). Of note, an earlier species list of *Calamus* (Vorontsova et al., 2016a) accepted 117 additional species that were sunk in a revision of the genus (Henderson, 2020). To validate these taxonomic decisions and account for different species delimitations, the sampling was augmented by 82 (70%) of the 117 recently synonymised species of *Calamus* (Henderson, 2020).

### 3.2.2 Sequencing and phylogeny reconstruction

DNA extraction and sequencing followed the methods described in Kuhnhäuser et al. (2021). In brief, DNA was extracted from silica-dried leaves or herbarium specimens using a modified CTAB protocol (Doyle and Doyle, 1987) and prepared for targeted sequence capture using a NEBNext Ultra II library kit (New England BioLabs, UK). The libraries were hybridized to the PhyloPalm probe kit (Loiseau et al., 2019) using a myBaits hybridization capture kit (Arbor Biosciences, USA), and sequencing was conducted at Macrogen Inc. (South Korea) on an Illumina HiSeq X Ten sequencing platform (Illumina, USA).

Raw sequence data were cleaned with Trimmomatic v. 0.38 (Bolger et al., 2014) using the MAXINFO algorithm with removal of the 3 leading and trailing bases to eliminate sequencing artefacts, 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). The targeted loci ("exons") were recovered in HybPiper v. 1.3.1 (Johnson et al., 2016). In addition, genomic regions adjacent to the on-target regions ("splash zones") were recovered using the development version of the Hybpiper script "intronerate.py" (https://github.com/mossmatters/HybPiper/blob/develop/intronerate.py; commit of 3 December 2018) because of their high variability (Kuhnhäuser et al., 2021) and potential to increase phylogenetic signal in resolving species-level relationships (Weitemier et al., 2014; Dodsworth et al., 2019). Exons and their respective splash zones were then combined to "genes" using the HybPiper script "retrieve_sequences.py".

Sequences were aligned per gene with MAFFT v. 7.471 (Katoh and Standley, 2013) using the memory-efficient method FFT-NS-2 because analyses using the more computationally demanding methods L-INS-i, G-INS-i, E-INS-i and FFT-NS-i failed due to memory limitations. Based on the stepwise testing of different filtering thresholds to balance the removal of missing data and maintenance of phylogenetically informative data, sites with less than 30% occupancy were removed with trimAl v. 1.4.rev15 (Capella-Gutiérrez et al., 2009).

All alignments were visually inspected using Geneious Prime v. 2020.2.5 (https://www.geneious.com), indicating that multiple alignments contained non-orthologous sequences or chimeric sequences containing segments of different genes. A

conservative filtering approach was therefore applied to validate that only single exons were targeted by the PhyloPalm probe kit, and to identify and exclude potentially paralogous loci. First, all exons in the PhyloPalm target file were mapped against the coding sequences of the genomes of the arecoid palm *Elaeis guineensis* (Chan et al., 2017) and the calamoid palm *Calamus inermis* (Zhao et al., 2018) using BLAST v. 2.10.0 (Altschul et al., 1997). The *Elaeis* genome was used because the PhyloPalm target file is based on the exons of this genome, and mapping the target file back to its original source provides a direct test of whether the targeted loci can be assigned to single-copy loci. This approach is regarded here as a reliable test of paralogy because no other factors such as biological variation or technical differences in sequence data generation are involved. The additional mapping against the *Calamus* genome was performed because of the early divergence of the calamoid palms from all other palm subfamilies and the possibility that gene or genome duplications might be unique to calamoid or non-calamoid palms. Only strictly matching hits (i.e. with an e-value of 0) were retained, and all PhyloPalm loci mapped to two or more different coding sequences in one or both of the two genomes were excluded as putatively paralogous. Second, using the script “paralog_investigator.py” in HybPiper, all exons for which more than one contig with at least 85% of the reference sequence length was recovered, were excluded.

Because of the high potential impact of sequencing and alignment errors on phylogeny reconstruction and molecular dating (Bromham, 2019), all remaining alignments were manually cleaned by deleting unaligned hypervariable sequence segments in Geneious Prime. Sequences shorter than 20% of the alignment length were excluded from the respective alignments to avoid their erroneous placement in gene trees due to the presence of few parsimony informative sites. Subsequently, samples represented by less than 100 alignments were excluded using AMAS (Borowiec, 2016) because preliminary analyses

indicated that a large proportion of taxa represented in fewer gene trees had implausible phylogenetic placements. This may be linked to an observed positive correlation between the number and length of genes that were retrieved, i.e. samples for which only fewer genes were recovered also generally had shorter sequence lengths and thus fewer phylogenetically informative sites. All alignments were then re-aligned in MAFFT using the method L-INS-i.

Gene trees were estimated in IQ-TREE v. 2.0.6 (Minh et al., 2020) using joint testing of nucleotide substitution models in ModelFinder (Kalyaanamoorthy et al., 2017) and 1000 ultrafast bootstrap replicates. Gene tree branches with bootstrap support below 10% were collapsed using Newick utilities v. 1.6 (Junier and Zdobnov, 2010) to avoid unsupported topologies influencing subsequent species tree reconstruction. A coalescent-based species tree was inferred from the individual gene trees using ASTRAL v. 5.7.1 (Zhang et al., 2018) and rooted with phyx (Brown et al., 2017) using as an outgroup the tribe Lepidocaryeae, which is a monophyletic group that is sister to the clade composed of Eugeissoneae and Calameae (Kuhnhäuser et al., 2021). For samples with improbable phylogenetic placements, voucher specimens were compared both to reference specimens of the originally assigned species as well as the closest relatives as indicated by the phylogenetic placement. On this basis, 13 species identifications of specimens were corrected (Supplementary Table 3.1).

Sequencing statistics were computed using the scripts “get_seq_lengths.py” and “hybpiper_stats.py” implemented in HybPiper, alignment summary statistics were calculated using the “summary” function in AMAS, and gene and species tree statistics were computed using the functions “get_var_length.py” and “get_bp_genetrees.py” in SortaDate (Smith et al., 2018) and the function “bss” in PhyKIT v. 1.2.1 (Steenwyk et al., 2021).

### 3.2.3 Fossil calibration

The Calamoideae have an exceptionally rich and distinctive microfossil record that can be assigned with high confidence to extant lineages in the subfamily (Harley and Morley, 1995; Morley, 2000; Harley, 2006; Iles et al., 2015; Bogotá-Ángel et al., 2021). In particular, the fossil pollen form genera *Mauritiidites* van Hoeken-Klinkenberg and *Dicolpopollis* Pflanzl are relevant for providing fossil calibrations because of their old age and clear taxonomic affinities.

***Mauritiidites***

*Mauritiidites* has been assigned to subtribe Mauritiinae based on its monosulcate pollen with spine-like structures on the pollen surface (ektexine) that exhibit inward bulging at the base (Bogotá-Ángel et al., 2021). The earliest well-substantiated records, i.e. having pollen micrographs validated by Bogotá-Ángel et al. (2021), are *M. crassibaculatus* and *M. lehmanii* from the Campanian of West Central and North East Africa (Salami, 1990; Salard-Cheboldaeff, 1990; Edet and Nyong, 1993; Eisawi and Schrank, 2008). These fossils were conservatively assigned to the stem node of subtribe Mauritiinae. The minimum age was set to 72.1 Ma (million years ago), corresponding to the younger bound of the Campanian (83.6-72.1 Ma; Walker et al., 2018). Following the approach of Magallón et al. (2015), this fossil age was implemented as a lognormal prior distribution with offset 72.1, a logarithmic mean of 79.3 (offset plus 10%), and a standard deviation of 1. The resulting soft upper bound (95% quantile) was 94.6 Ma, coinciding with the oldest (but less certain; Bogotá-Ángel et al., 2021) records of Mauritiinae pollen from the Cenomanian (100.5-93.9 Ma; Walker et al., 2018).

***Dicolpopollis***

*Dicolpopollis* is “highly comparable” (Harley and Morley, 1995) to extant equatorially disulcate pollen in subtribes Metroxylinae, Plectocomiinae, and Calaminae. Earlier studies very conservatively assigned *Dicolpopollis* to the stem node of Calameae (Couvreur et al., 2011a; Iles et al., 2015), partly because of previously uncertain higher-level relationships in the Calamoideae (Kuhnhäuser et al., 2021). However, recent evidence strongly supports a clade formed by subtribes Metroxylinae, Plectocomiinae, Calaminae, and Pigafettinae that is nested within Calameae (Kuhnhäuser et al., 2021). Therefore, *Dicolpopollis* was attributed here to the clade formed by these subtribes. Of note, the spheroidal, inaperturate pollen of Pigafettinae (Dransfield et al., 2008) is highly unusual in palms and distinct from the disulcate pollen in the other members of this clade, likely representing an autapomorphy. Further, pollen with only a single opening is found in the former genus *Pogonotium* (Dransfield et al., 2008), a small clade nested within the otherwise disulcate *Calamus* (Kuhnhäuser et al., 2021).

The oldest records of *Dicolpopollis* are *D. kalewensis* Potonié (synonym *D. malesianus* Muller; Huang et al., 2020) from the Cenomanian of Australia (Burger, 1990; Totterdell and Mitchell, 2009; Macphail and Jordan, 2015), with identification and age interpreted as highly reliable (R. Morley, pers. comm.). *D. kalewensis* has been assigned to subtribe Calaminae based on the net-like pollen surface structure (reticulate to perforate-semireticulate ektexine) and thin inner pollen wall (endexine) (Harley and Morley, 1995). However, it has to be noted that highly similar pollen occurs in the sister taxon Plectocomiinae (Dransfield et al., 2008). A second form of *Dicolpopollis*, *D. metroxylonoides* Khan, is similar to *D. kalewensis* in its equatorially disulcate pollen but can be distinguished by its more robust

ektexine and highly thickened endexine (Khan, 1976). In extant calamoid taxa, this form only occurs in the genus *Metroxylon*. Its oldest records are from the Paleocene of Tasmania (Macphail and Jordan, 2015).

Overall, an assignment of *Dicolpopollis kalewensis* fossils to the stem node of the clade defined by Metroxylinae, Pigafettinae, Plectocomiinae and Calaminae can be considered most conservative based on the equatorially disulcate pollen macrostructure alone, whereas a narrower assignment to stem node of the clade defined by Plectocomiinae and Calaminae would reflect the close affinity of the pollen microstructures (reticulate ektexine, thin endexine) (Figure 3.1). Both placements of the fossil were applied here in separate analyses. The minimum age for this fossil calibration was set to 93.9 Ma, corresponding to the younger bound of the Cenomanian (100.5-93.9 Ma; Walker et al., 2018). Following the approach of Magallón et al. (2015), this calibration was implemented as a lognormal prior distribution with offset 93.9, a logarithmic mean of 103.3 (offset plus 10%), and a standard deviation of 1. The resulting soft upper bound (95% quantile) was 122 Ma.

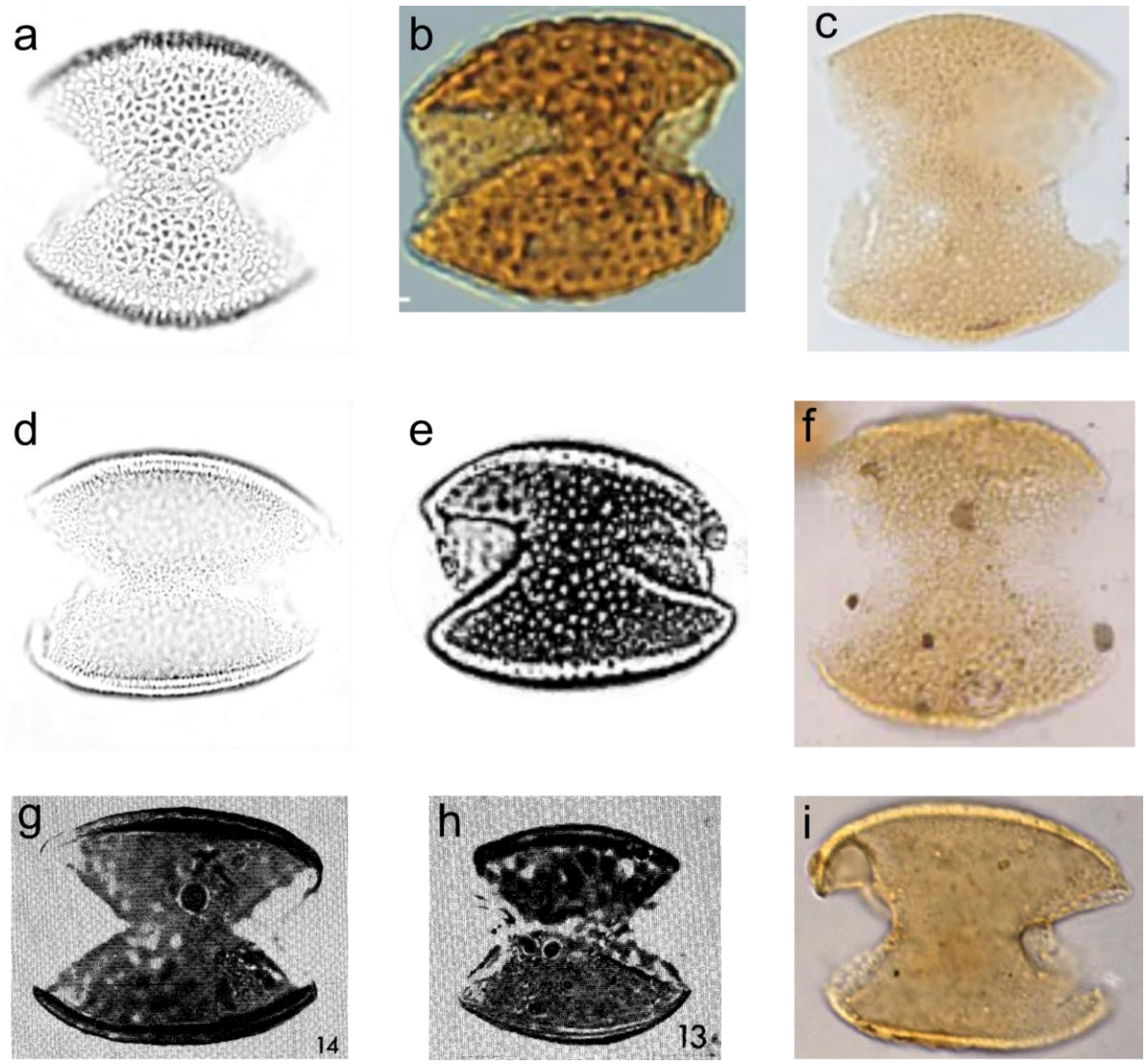



**Figure 3.1. Comparison of extant disulcate calamoid pollen and fossil *Dicolpopollis*.** **a**-**f**, Extant and fossil pollen assignable to subtribes Calaminae and Plectocomiinae / *Dicolpopollis kalewensis* (= *Dicolpopollis malesianus*). **a**, *Calamus formicarius*, extant (Dransfield et al., 2008, p. 203, Fig. i). **b**, *Dicolpopollis* sp., Eocene of Myanmar (Huang et al., 2020, Fig. 7E). **c**, *Dicolpopollis* sp., Cenomanian of Australia (Macphail, unpublished). **d**, *Plectocomia elongata*, extant (Dransfield et al., 2008, p. 185, Fig. d). **e**, *Dicolpopollis malesianus*, Eocene of Java (Harley and Morley, 1995, Plate VII-1). **f**, *Dicolpopollis* sp., Cenomanian of Australia (Macphail and Jordan, 2015, Plate 1c). **g**-**i**, Extant and fossil pollen comparable to subtribe Metroxylinae / *Dicolpopollis metroxylonoides*. **g**, *Metroxylon americanum* [sic.] (= *M. amicarum*), extant (Khan, 1976, Fig. 14). **h**, *Dicolpopollis metroxylonoides*, Miocene of New Guinea (Khan, 1976, Fig. 13). **i**, *Dicolpopollis sp.*, Paleocene of Tasmania (Macphail and Jordan, 2015, Plate 1a). Copyright: **a**, **d**, **e**: M. Harley, RBG Kew. **b**: H. Huang, University of Amsterdam. **c**, **f**, **i**: M.K. Macphail, Australian National University. **g**, **h**: A.M. Khan, Robertson Research (North America), Canada.

***Crown Calamoideae***

To constrain the maximum crown node age of the Calamoideae (i.e. the root of the phylogeny), a secondary calibration was used based on dated phylogenies that include calamoid palms (Couvreur et al., 2011a; Faurby et al., 2016; Antonelli et al., 2017; Givnish et al., 2018; Ramírez-Barahona et al., 2020). Estimates for the stem node age of the Calamoideae range from 28.4 to 127.4 Ma, and for the crown node age from 7.5 to 104.1 Ma (Table 3.1). However, a recent study based on 75 calamoid species and 971 genes (Kuhnhäuser et al., 2021) inferred highly supported tribal and subtribal relationships that are different from the topologies used in most of the above studies, affecting the placement of fossils and potentially suggesting that stem and crown ages of the Calamoideae may have been underestimated. In addition, sparse sampling of the Calamoideae in some studies (Ramírez-Barahona et al., 2020) likely led to an underestimation of the crown age of the Calamoideae because the sampled species did not define the most recent common ancestor of the Calamoideae.

**Table 3.1. Previous age estimates for the stem and crown node of the Calamoideae.** Ma = million years ago, min = lower end of 95% highest posterior density interval (HPD), max = upper end of 95% HPD. UC = unconstrained calibration, RC = relaxed calibration, CC = constrained calibration, cons = conservative fossil set, comp = complete fossil set.

| Reference | Details | Stem (Ma) | | | Crown (Ma) | | |
|---|---|---|---|---|---|---|---|
| | | min | mean | max | min | mean | max |
| Ramírez-Barahona et al., 2020 | UC cons | 35.0 | 57.3 | 81.5 | 8.3 | 30.9 | 52.5 |
| | UC comp | 83.9 | 107.0 | 127.4 | 66.0 | 79.9 | 104.1 |
| | RC cons | 28.4 | 47.9 | 67.6 | 11.5 | 27.0 | 47.0 |
| | RC comp | 78.2 | 94.0 | 112.0 | 66.0 | 75.1 | 88.3 |
| | CC cons | 32.1 | 48.5 | 69.4 | 7.5 | 25.2 | 43.2 |
| | CC comp | 79.1 | 95.6 | 112.5 | 66.0 | 76.3 | 92.1 |
| Givnish et al., 2018 | Fig. S13 | | 84.5 | | | 12.3 | |
| Antonelli et al., 2017 | Table S2 | | 85.8 | | | 65.0 | |
| Faurby et al., 2016 | Fig. S2 | | 104.9 | | | 70.6 | |
| Couvreur et al., 2011 | Table 1 | 92.0 | 100.1 | 108.7 | 70.3 | 80.2 | 90.3 |

Overall, almost all previous crown age estimates of the subfamily are younger than the Cenomanian fossils of *Dicolpopollis* (Burger, 1990; Totterdell and Mitchell, 2009; Macphail and Jordan, 2015), which were established here as nested within the Calamoideae (see discussion above). The omission of the Cenomanian *Dicolpopollis* from previous dated phylogenies can potentially be attributed to the publication of important fossils (Macphail and Jordan, 2015) only after widely used reviews of the palm fossil record had been published (Harley, 2006; Iles et al., 2015). To account for this general underestimation of the crown node age of the Calamoideae in previous studies, the maximum previous age estimate for the stem node of the Calamoideae, 127.4 Ma (Table 3.1), was used as a secondary constraint for the maximum crown age for the Calamoideae. This information was implemented as a uniform prior distribution ranging from 93.9 Ma, i.e. the minimum age of *Dicolpopollis*, to 127.4 Ma. This interval includes the estimated age of the oldest monocot fossil *Liliacidites* (125-113 Ma; Ramírez-Barahona et al., 2020) but is younger than the

earliest angiosperm-like pollen grains (minimum age 132.1 Ma; Ramírez-Barahona et al., 2020). Given the ongoing debate on the age of the angiosperms (Budd et al., 2021; Sauquet et al., 2021; Silvestro et al., 2021), additional analyses without imposing a maximum age constraint for the crown node of Calamoideae were conducted. The four different fossil calibration strategies are summarised in Table 3.2.

**Table 3.2. Calibration strategies.** Age constraints and fossil placements implemented in the four dating analyses in this study. Ma = million years ago. stem MPPC = stem node of Metroxylinae + Pigafettinae + Plectocomiinae + Calaminae, stem PC = stem node of Plectocomiinae + Calaminae, stem Mau = stem node of Mauritiinae.

| **Strategy** | **Calamoideae** | ***Dicolpopollis*** | | ***Mauritiidites*** | |
|---|---|---|---|---|---|
| | **Max. age (Ma)** | **Placement** | **Min. age (Ma)** | **Placement** | **Min. age (Ma)** |
| **MetCal 127.4** | 127.4 | stem MPPC | 93.9 | stem Mau | 72.1 |
| **PleCal 127.4** | | stem PC | | | |
| **MetCal UC** | unconstrained | stem MPPC | | | |
| **PleCal UC** | | stem PC | | | |

### 3.2.4 Divergence time estimation

For molecular dating analyses, only one sample per species or subspecies was retained based on gene tree occupancy (Supplementary Table 3.2) because sampling biases can impact divergence time estimates (Bromham et al., 2018). The alignments were then re-aligned with MAFFT using L-INS-i, and gene trees as well as an ASTRAL species tree were re-estimated as described above.

Because divergence time estimation based on hundreds of genes and taxa was infeasible (Smith et al., 2018), a subset of gene trees was selected as follows: using SortaDate, topological similarity to the species tree ("bipartition"), clock-likeness ("root-to-

tip variance"), and information content ("tree length") were estimated for all gene trees. Because the ASTRAL species tree was assumed to represent the best available estimate of phylogenetic relationships of the Calamoideae, only the ten gene trees with the highest similarity to the ASTRAL species tree were retained. Tree length and clock-likeness of the selected gene trees were then examined, validating that they had sufficient information content (defined here as a tree length of at least one expected substitution per site across all branches) and no extreme root-to-tip variance (defined here as variance of root-to-tip paths below 0.001). To confirm that the phylogenetic information in this data subset was compatible with the all-genes species tree, a phylogeny was estimated from the concatenated alignments of the ten genes using IQ-TREE and ModelFinder with settings as described above and compared to the ASTRAL species-level phylogeny based on all gene trees using a co-phylogenetic plot generated with phytools v. 0.7.70 (Revell, 2012) in R v. 4.0.4 (R Core Team, 2021).

Divergence times were estimated using BEAST v. 2.6.3 (Bouckaert et al., 2019) with parallelisation of analyses using BEAGLE v. 3.2.0 (Ayres et al., 2019). Based on the best-fitting DNA substitution rate models identified by ModelFinder, the HKY model (unequal transition and transversion rates, unequal base frequencies) was selected (Hasegawa et al., 1985). Given that the ASTRAL species tree was assumed to represent the best available estimate of phylogenetic relationships, it was used as a topological constraint for the BEAST analysis. Because branch lengths of the preliminary phylogeny estimated from the ten selected gene trees indicated that rates of molecular evolution varied between genera (and to a lesser extent between species within genera), a relaxed lognormal clock (Drummond et al., 2006) with the Calibrated Yule Model tree prior (Heled and Drummond, 2011) was used. Following recommendations in Drummond and Bouckaert (2015), the speciation rate (birthRateY) was

implemented as a gamma prior with Alpha 0.001 and Beta 1000, and the clock rate standard deviation was implemented as an exponential distribution with mean 0.33333. The mean clock rate prior ("ucldMean") was set to a log normal distribution with M parameter (in real space) 0.002 and S parameter 2, resulting in a distribution with a 2.5% quantile of 0.000005 and a 97.5% quantile of 0.0136 that reflected previous rate estimates in the palm family ranging from 0.00001 to 0.01 (Gaut et al., 1996). The priors of different fossil calibration strategies are outlined above (section 3.2.3 Fossil calibration). For each calibration strategy, three independent runs with 300 million generations (400 million generations for strategy MetCal 127.4) were performed and combined using LogCombiner after discarding the first 10% generations of each run and visually assessing using Tracer v. 1.7.1 (Rambaut et al., 2018) that the independent runs had converged to the same values. For each combined analysis, effective sample sizes were above 100 for all parameters, and above 200 for most parameters, indicating that all analyses had reached convergence. To assess to which degree the implemented priors informed or constrained the divergence time estimates, one run per calibration strategy was performed with 200 million generations that were sampled only from the prior. Effective sampling sizes of all of these runs were above 200.

## 3.3 Results

### 3.3.1 Retrieval of orthologous loci

In total, 655 libraries representing 637 calamoid palm specimens were sequenced, targeting 971 putatively orthologous nuclear genomic "PhyloPalm" loci (Loiseau et al., 2019). A median of 4.4 million reads were retrieved per library, of which a median of 22.1% were on target (Table 3.3). One locus could not be retrieved.

**Table 3.3. Sequencing statistics.** Retrieved sequence reads for 653 libraries of calamoid palms in total, reads on target across all PhyloPalm loci, and reads on target across the subset of 357 orthologous loci that were identified here.

| Statistic | Reads | Reads on target (all loci) | | Reads on target (orthologs) | |
|---|---|---|---|---|---|
| | # | # | % | # | % |
| **Minimum** | 4,653 | 1,005 | 1.6 | 606 | 0.7 |
| **25% quartile** | 2,850,492 | 451,249 | 14.1 | 179,631 | 5.2 |
| **Median** | 4,408,290 | 812,523 | 22.1 | 355,939 | 9.1 |
| **75% quartile** | 5,808,155 | 1,299,166 | 27.8 | 631,652 | 13.6 |
| **Maximum** | 18,316,777 | 9,318,056 | 63.0 | 8,326,457 | 53.0 |

Because visual assessment of multiple sequence alignments indicated the presence of paralogous sequences, additional paralogy tests were conducted. Whereas only 28 putative paralogs were identified using HybPiper, mapping of the PhyloPalm loci, which are based on coding sequences of the *Elaeis guineensis* genome (Chan et al., 2017), back against the coding sequences of the same genome, resulted in the discovery of 455 paralogous loci, i.e. the target file loci mapped to two or more different coding sequences in one or both genomes. In addition, the PhyloPalm loci were mapped against the genome of the phylogenetically divergent *Calamus inermis* (Zhao et al., 2018), revealing 487 putatively paralogous loci. There was substantial overlap between the paralogous loci identified by mapping against the two different genomes, with 338 paralogs identified in both genomes. Out of the 28 putative paralogs identified using Hybpiper, 19 were also identified with the BLAST search. In total, 613 (63%) of the 971 PhyloPalm loci were identified as paralogous (Figure 3.2, Supplementary List 3.1). The PhyloPalm loci consist of 176 loci from the first target-capture study of palms (Heyduk et al., 2015), of which 23 (13%) were inferred here as paralogous, and 795 additional loci (Loiseau et al., 2019), of which 590 (74%) were inferred as paralogous.

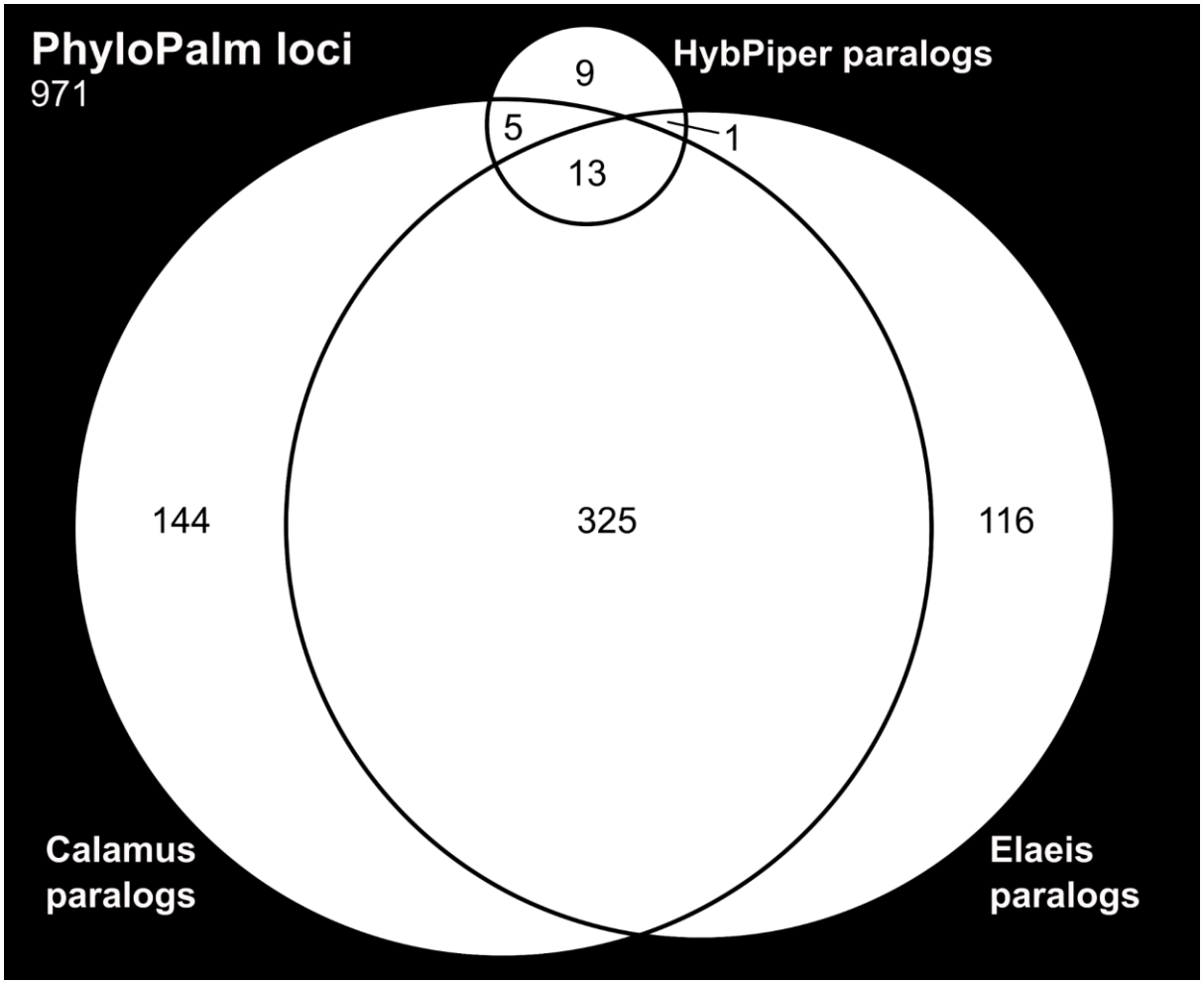



**Figure 3.2. Overlap of different approaches in inferring paralogy in the PhyloPalm loci.** The area of the rectangle corresponds to the 971 PhyloPalm loci. The areas of the circles correspond to the number of paralogs identified by mapping the PhyloPalm targetfile against coding sequences of the genomes of *Calamus inermis* and *Elaeis guineensis*, respectively, as well as using HybPiper. The total number of paralogs is 613. The difference between the rectangle area and the circle areas corresponds to the number of orthologs. The total number of orthologs is 357. (One PhyloPalm locus could not be retrieved and is therefore not included in this analysis.)

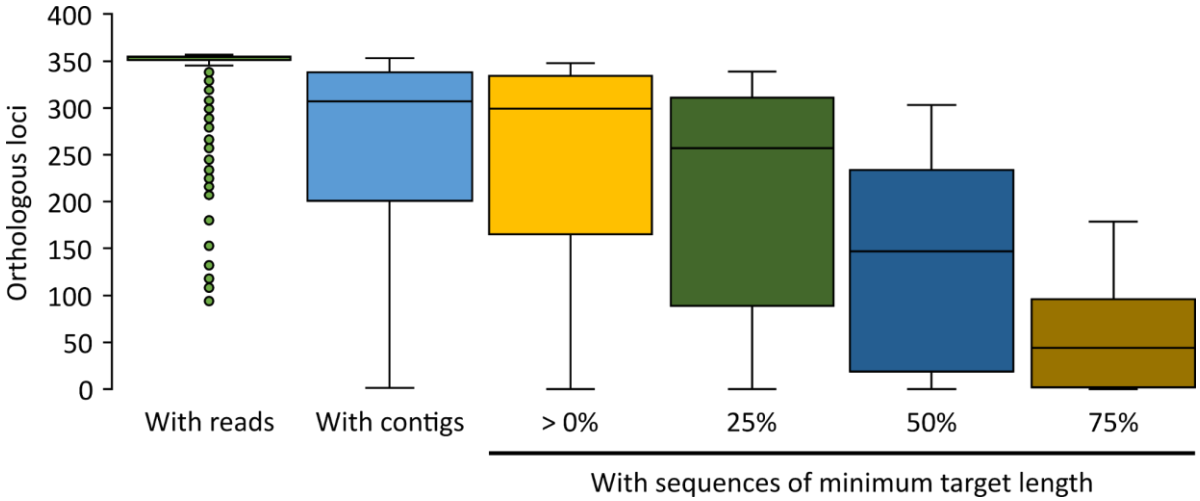



**Figure 3.3. Recovered orthologous loci.** Number of orthologs for which reads, contigs and genes (in relation to the target length of the PhyloPalm loci) were recovered for 653 libraries of calamoid palm specimens.

### 3.3.2 Effects of alignment filtering

Multiple sequence alignments of the remaining 354 putative orthologous PhyloPalm loci included 151 “Heyduk” loci (Heyduk et al., 2015) and 203 other PhyloPalm loci (Loiseau et al., 2019). The alignments had a total length of 7,906,765 base pairs, a median of 88.9% missing bases, a total of 1,365,154 parsimony informative sites (17.2%), and included 640 samples (Table 3.4). Filtering of sites with more than 70% gaps resulted in a reduction of the total length of the alignments to 952,442 base pairs, the median of missing bases to 35.3%, and the total number of parsimony informative sites to 791,566. Removal of hypervariable sequence segments and fragmentary sequences drastically reduced the number of parsimony informative sites by over 40%, and led to a reduction of the number of samples to 593. The exclusion of samples occurring in less than 100 alignments resulted in the removal of a further 104 samples. The resulting alignments had a total length of 905,250 base pairs (median 2421), a median of 31.5% missing bases, a total of 451,264 parsimony informative

sites (median 1054), and included a total of 489 samples (median 414) representing 390 species and subspecies (Table 3.4). Of note, filtered alignments of the Heyduk loci were shorter (median 1716) and had fewer parsimony informative sites (median 752) compared to the other PhyloPalm loci, which had 77% longer alignment lengths (median 3037) and 90% more parsimony informative sites (median 1427).

### 3.3.3 Phylogenetic placement of data-deficient samples

This phylogenomic study included 653 samples representing 502 species (90%) of the Calamoideae, the majority of which had never been sequenced before. Out of the 653 samples, 593 (91%) had sufficient sequence data to be included in at least one gene tree (Supplementary Fig. 3.1). Phylogenetic placements of samples that were most data-deficient (i.e. present in less than 100 gene trees) were assessed using a preliminary phylogeny including all samples (Supplementary Fig. 3.1). In some cases, these samples were resolved in plausible positions (e.g., *Calamus deerratus*, *Korthalsia brassii*, *Salacca griffithii*). More often, however, data-deficient samples had implausible phylogenetic placements (e.g., *Calamus anomalus*, *C. rotang*, *C. sumatranus*, *Mauritia carana*), were suspiciously frequently resolved as sister to larger clades (e.g., *Calamus adspersus*, *C. bangiiensis*, *C. burkillianus*, *C. kjellbergii*, *C. minahassae*, *Korthalsia rogersii*, *Plectocomia longistigma*), and sometimes clustered into groups apparently solely defined by data deficiency (e.g., *Calamus asperrimus*, *C. heteroideus*, *C. samian*) (Supplementary Fig. 3.1). This underlines the need for the conservative approach to data filtering taken here, i.e. the removal of samples present in less than 100 gene trees.

**Table 3.4. Alignment statistics.** Summary statistics for different steps of alignment filtering of 354 orthologous genes. Filtering steps: None = original unfiltered alignments, Sites = sites with less than 30% occupancy removed, Seqs = hypervariable sequence parts and sequences shorter than 20% of the alignment length removed, Samples = samples occurring in less than 100 alignments removed. Min. = minimum, Q1 = 25% quartile, Med. = median, Q3 = 75% quartile, Max. = maximum. Total = total number of taxa, alignment lengths, variable sites, or parsimony informative sites across all alignments. Missing = missing or unidentified bases, Pars. inf. sites = parsimony informative sites, GC = guanine and cytosine bases.

| Filtering | Statistic | Taxa | Length | Missing | Variable sites | | Pars. inf. sites | | GC |
|---|---|---|---|---|---|---|---|---|---|
| | | | | % | # | % | # | % | % |
| None | **Min.** | 25 | 1,509 | 47.0 | 286 | 7.6 | 57 | 3.6 | 30.7 |
| | **Q1** | 420 | 11,278 | 86.1 | 3,882 | 27.2 | 2,349 | 15.5 | 38.5 |
| | **Med.** | 475 | 18,328 | 88.9 | 6,057 | 33.0 | 3,576 | 19.8 | 40.6 |
| | **Q3** | 513 | 28,314 | 91.9 | 8,403 | 38.4 | 5,011 | 23.9 | 43.7 |
| | **Max.** | 584 | 154,021 | 97.3 | 27,374 | 51.6 | 14,107 | 33.0 | 60.2 |
| | **Total** | 640 | 7,906,765 | - | 2,304,845 | - | 1,365,154 | - | - |
| Sites | **Min.** | 25 | 370 | 16.6 | 265 | 21.8 | 56 | 4.6 | 31.0 |
| | **Q1** | 420 | 1,716 | 30.9 | 1,598 | 93.0 | 1,387 | 78.5 | 38.5 |
| | **Med.** | 475 | 2,570 | 35.3 | 2,418 | 95.2 | 2,106 | 83.6 | 40.8 |
| | **Q3** | 513 | 3,470 | 39.6 | 3,278 | 97.4 | 2,970 | 88.4 | 44.2 |
| | **Max.** | 584 | 7,419 | 57.1 | 7,187 | 99.8 | 6,411 | 98.2 | 60.5 |
| | **Total** | 640 | 952,442 | - | 902,480 | - | 791,566 | - | - |
| Sites, Seqs | **Min.** | 18 | 368 | 10.3 | 113 | 14.4 | 47 | 3.9 | 30.0 |
| | **Q1** | 364 | 1,560 | 28.0 | 1,025 | 64.8 | 643 | 40.3 | 38.3 |
| | **Med.** | 417 | 2,421 | 32.1 | 1,642 | 70.0 | 1,070 | 45.6 | 40.4 |
| | **Q3** | 460 | 3,423 | 36.0 | 2,504 | 75.6 | 1,679 | 51.8 | 43.8 |
| | **Max.** | 558 | 7,421 | 47.4 | 6,640 | 92.6 | 5,508 | 82.8 | 60.9 |
| | **Total** | 593 | 905,250 | - | 658,192 | - | 453,390 | - | - |
| Sites, Seqs, Samples | **Min.** | 17 | 368 | 8.2 | 113 | 14.4 | 47 | 3.9 | 30.0 |
| | **Q1** | 357 | 1,560 | 26.6 | 1,023 | 64.5 | 642 | 40.2 | 38.3 |
| | **Med.** | 414 | 2,421 | 31.5 | 1,605 | 69.5 | 1,054 | 45.5 | 40.4 |
| | **Q3** | 451 | 3,423 | 35.7 | 2,489 | 75.3 | 1,676 | 51.6 | 43.7 |
| | **Max.** | 485 | 7,421 | 47.1 | 6,638 | 92.3 | 5,502 | 82.4 | 60.4 |
| | **Total** | 489 | 905,250 | - | 655,503 | - | 451,264 | - | - |

### 3.3.4 Species-level phylogenetic relationships of the Calamoideae

The phylogeny excluding data-deficient samples contained 482 samples, corresponding to 381 species (68%) of the Calamoideae (Supplementary Fig. 3.2). Their relationships were overall strongly supported, with 55% of clades having a local posterior probability of 1, and an average local posterior probability of 0.87 across all nodes. All genera were resolved as monophyletic (Supplementary Fig. 3.2), and all relationships at generic, subtribal and tribal level were identical with a sparsely sampled backbone phylogeny covering the taxonomic diversity of all genera of the subfamily (Kuhnhäuser et al., 2021). Higher-level relationships at genus level and above were supported by a clear majority of gene trees (Supplementary Fig. 3.2). The only exception was subtribe Plectocomiinae, in which the generic relationships were poorly supported despite the strong support of the monophyly of all genera.

#### *Infrageneric relationships*

Within genera, many clades were strongly supported by a clear majority of gene trees (Supplementary Fig. 3.2). In *Raphia*, *R. regalis* was resolved as sister to the remaining species. In *Eugeissona*, two geographically separated sister clades were resolved. In *Korthalsia*, two large sister clades, and several strongly supported clades within these clades, were resolved. *Salacca secunda* was resolved as sister to the remaining species of the genus. *Salacca griffithii* was data-deficient but plausibly resolved as sister to *S. secunda* in the preliminary phylogeny (Supplementary Fig. 3.1). Species of section *Leiosalacca* formed a clade nested within the genus (Supplementary Fig. 3.2). The synonymised genus *Lophospatha*, now *Salacca lophospatha*, was resolved as nested within the genus and sister to *S. clemensiana*. In *Metroxylon*, all species of section *Coelococcus* formed a clade that was

sister to *M. sagu*, the only species in section *Metroxylon*. In *Plectocomiopsis*, *P. geminiflora* was resolved as sister to a clade of *P. songthanhensis* and *P. mira*. In *Plectocomia*, section *Plectocomia* subsection *Trigonae* was resolved as paraphyletic, with subsections *Pedicelliformes* and *Plectocomia* nested within. *Plectocomia pygmaea*, the only species in section *Penicillatae*, had insufficient data for phylogenetic placement.

In *Calamus*, a clade defined by *C. erinaceus* and *C. zollingeri* was resolved as sister to the remaining species. Within the remaining species, a clade defined by *C. spiralis* and *C. rhabdocladus* was resolved as sister to the rest. Within the remaining clade, a clade defined by *C. dongnaiensis* and *C. flagellum* was resolved as sister to all other species. *Calamus dumetosus* (syn. *Retispatha dumetosa*) was resolved as sister to the remaining genus (Supplementary Fig. 3.2). The synonymised genera *Ceratolobus*, *Daemonorops* and *Pogonotium* formed a clade nested within *Calamus*. The species of synonym *Daemonorops* were resolved in two monophyletic groups interspersed by clades formed by the species of synonyms *Ceratolobus* and *Pogonotium*, rendering *Daemonorops* polyphyletic. *Calamus deerratus*, the only African species of the genus, was resolved as nested within Calamus and sister to a clade consisting of the former genera *Ceratolobus*, *Daemoronops* and *Pogonotium*, as well as *C. diepenhorstii* (Supplementary Fig. 3.2). Species of the former genera *Calospatha*, *Cornera* and *Zalaccella* were resolved as closely related and nested within a flagellate clade in *Calamus* (Supplementary Fig. 3.2). Within the remaining species, several other strongly supported clades were resolved (Supplementary Fig. 3.2).

Whilst most relationships were strongly supported, some relationships were poorly supported and characterised by considerable gene tree conflict. In *Eremospatha*, *Oncocalamus* and *Laccosperma*, most resolved relationships were only supported by a small

majority of gene trees (Supplementary Fig. 3.2). Furthermore, the placement of an undescribed species of Plectocomiinae (Plectocomiinae sp. nov. RBL178, specimen Kuhnhäuser 35 [K]) with uncertain generic affinities was poorly supported and associated with high gene tree conflict.

***Species monophyly***

For 36 species, the phylogeny included more than one specimen (Supplementary Table 3.2). In 14 of these cases, the multiple accessions per species formed strictly monophyletic groups. In the remaining 22 cases, the different samples per species had a very close phylogenetic placement but were paraphyletic (e.g., *Calamus erioacanthus*) or polyphyletic (e.g., *C. tetradactylus* or *C. moseleyanus*). This included 20 cases of species complexes that had been synonymised in a recent taxonomic revision of the genus *Calamus* (Henderson, 2020). The other two cases were samples with limited sequence data, i.e. samples represented in less than half of the 354 gene trees.

### 3.3.5 A time-calibrated phylogeny of the Calamoideae

***Summary statistics***

For the dated phylogeny, only one sample per species or subspecies was retained, leading to the exclusion of 92 samples (Supplementary Table 3.2). The resulting gene trees contained a total of 390 species and subspecies (corresponding to 381 different species) and had a median of median ultrafast bootstrap support of 87 %, median tree length of 2.75 expected

substitutions per site across all branches, median root-to-tip variance of 0.0001, and median similarity of the individual gene trees to the coalescent species tree of 0.15 (Table 3.5).

**Table 3.5. Gene tree statistics.** Summary statistics of all 354 gene trees for 390 species and subspecies of calamoid palms, as well as for the 10 gene trees selected using SortaDate. Branch support = ultrafast bootstrap support, Tree length = expected substitutions per site across all branches, Root-to-tip variance = variance of root-to-tip lengths, Bipartition proportion = similarity to ASTRAL species tree (1 would indicate identical topology).

| | | Mean branch support | Median branch support | Tree length | Root-to-tip variance | Bipartition proportion |
|---|---|---|---|---|---|---|
| **All 354 orthologs** | **Minimum** | 36 | 24 | 0.10 | 0.00001 | 0.00 |
| | **25% quartile** | 71 | 80 | 2.12 | 0.00006 | 0.09 |
| | **Median** | 77 | 87 | 2.75 | 0.00010 | 0.15 |
| | **75 % quartile** | 81 | 92 | 3.65 | 0.00024 | 0.22 |
| | **Maximum** | 91 | 100 | 13.49 | 0.01448 | 0.47 |
| **SortaDate top 10 gene trees** | **EGU105037938** | 87 | 98 | 3.18 | 0.00017 | 0.47 |
| | **EGU105039403** | 85 | 97 | 2.57 | 0.00004 | 0.39 |
| | **EGU105039783** | 82 | 94 | 1.98 | 0.00003 | 0.38 |
| | **EGU105044481** | 82 | 94 | 1.92 | 0.00004 | 0.37 |
| | **EGU105047940** | 82 | 96 | 1.12 | 0.00002 | 0.41 |
| | **EGU105051795** | 87 | 97 | 3.88 | 0.00008 | 0.36 |
| | **EGU105051860** | 87 | 97 | 2.68 | 0.00004 | 0.44 |
| | **EGU105052804** | 82 | 94 | 2.75 | 0.00013 | 0.38 |
| | **EGU105057553** | 77 | 96 | 1.48 | 0.00002 | 0.39 |
| | **HEY1050** | 83 | 94 | 2.71 | 0.00006 | 0.39 |

The ten gene trees selected based on similarity to the species tree had a median ultrafast bootstrap support between 94 and 98 %, tree length of 1.12 to 3.88 expected substitutions per site across all branches, root-to-tip variance from 0.00002 to 0.00017, and similarity to the species tree of 0.36 to 0.47 (Table 3.5) and contained between 379 and 390 taxa. Over 90% of the taxa (363 samples) occurred in all trees, 20 taxa in nine trees, and the remaining seven taxa in eight trees. The lengths of the alignments on which the trees were

based ranged from 2,279 to 5,588 base pairs (median 2,913.5), and the number of parsimony informative sites from 869 to 2,339 (median 1,123). After concatenation of the alignments of the ten genes for molecular dating analyses, their total length comprised 34,002 base pairs, of which 12,580 (37%) were parsimony informative sites. The topology of the phylogeny estimated from the ten concatenated alignments was highly similar to the topology of the ASTRAL species tree based on all 354 genes (Supplementary Fig. 3.3).

***Differences between calibration strategies***

Divergence time estimates were highly similar for both maximum age constrained analyses (i.e. with a maximum age of 127.4 Ma for the crown node of Calamoideae), with either the more conservative assignment of *Dicolpopollis* to the stem node of Metroxylinae, Pigafettinae, Plectocomiinae and Calaminae, or the narrower assignment to the stem node of Plectocomiinae and Calaminae (Fig. 3.4). The highest posterior density intervals (HPDs) of node age estimates were generally strongly overlapping between both analyses, but always somewhat older in the analysis with the narrower assignment of *Dicolpopollis* (Fig. 3.4). The only exception was the backbone of tribe Calameae, where the HPDs were strongly disjunct. Analyses without a maximum age constraint resulted in older age estimates and broader HPDs, especially for higher-level backbone nodes. HPDs of the unconstrained analysis with a conservative assignment of *Dicolpopollis* generally had a broad overlap with HPDs of both maximum age constrained analyses, whereas the unconstrained analysis with narrow assignment of *Dicolpopollis* often had considerably older HPDs that were often not or only slightly overlapping with the constrained analyses (Fig. 3.4).

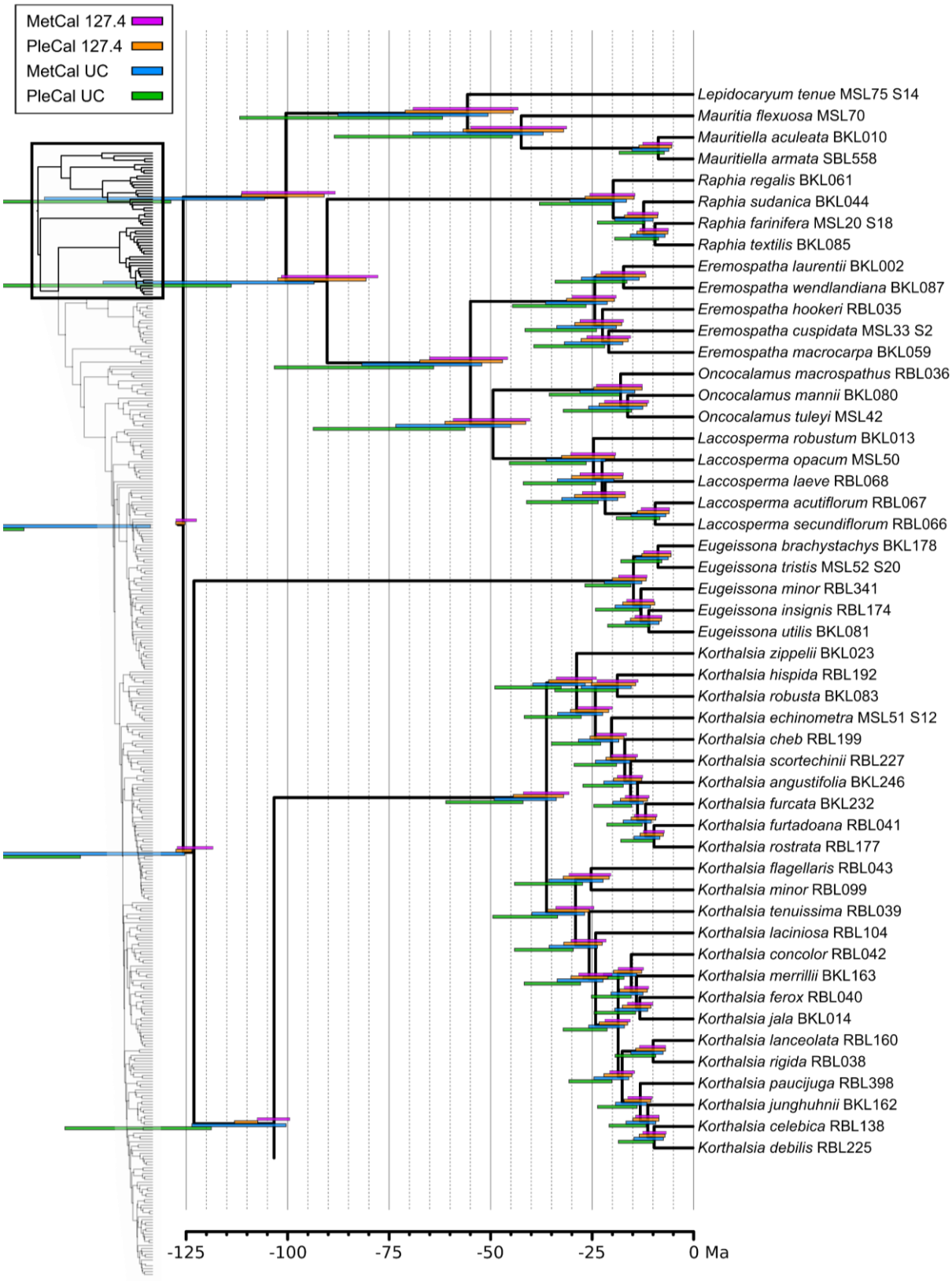



**Figure 3.4. Dated species-level phylogeny of 381 calamoid palm species.** Divergence time estimates based on 10 genes selected using SortaDate. Highest posterior density intervals (HPDs) indicated on each node for four different dating strategies (Table 3.2). Ma = million years ago.

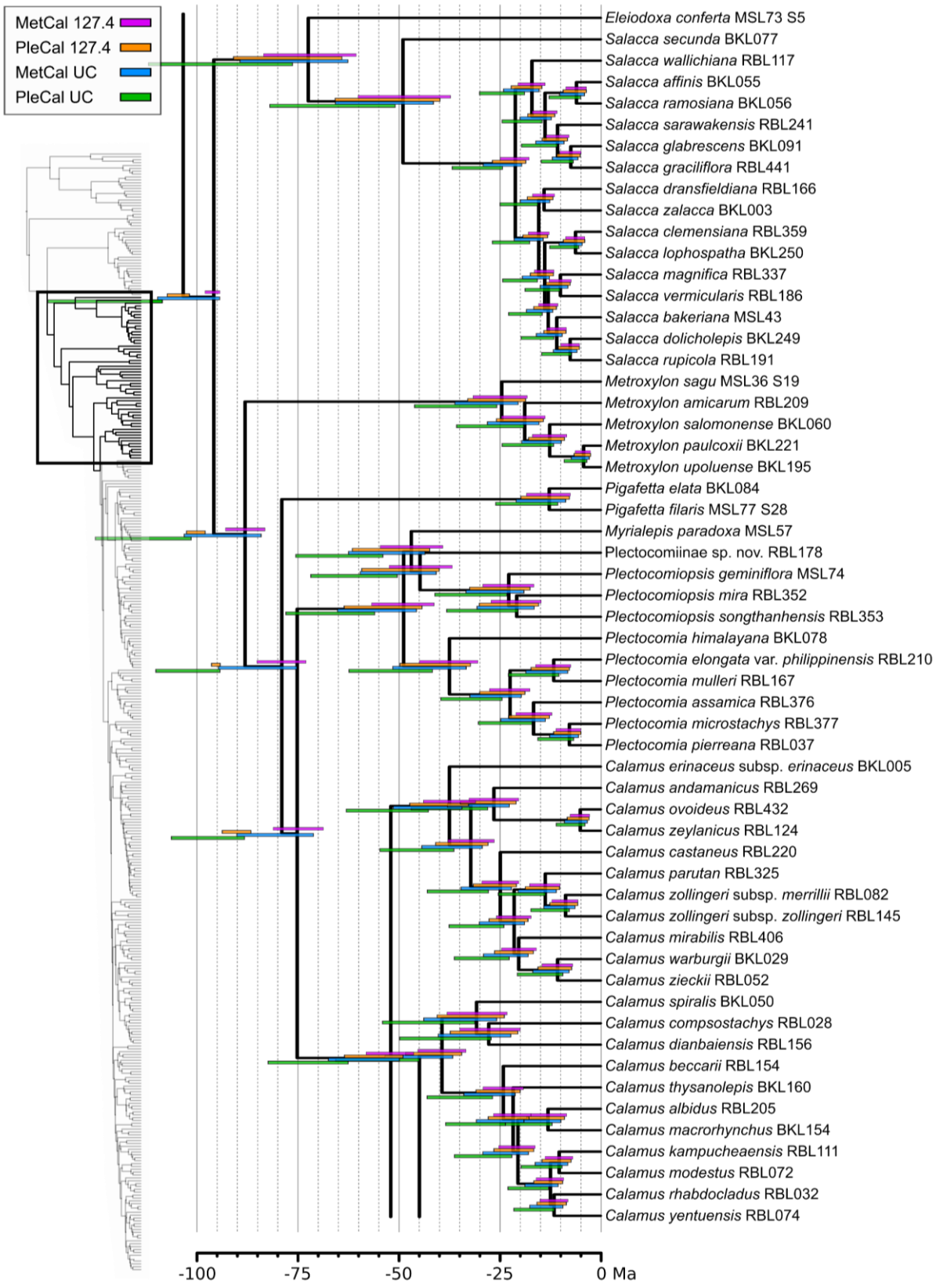



**Figure 3.4** (*continued*)

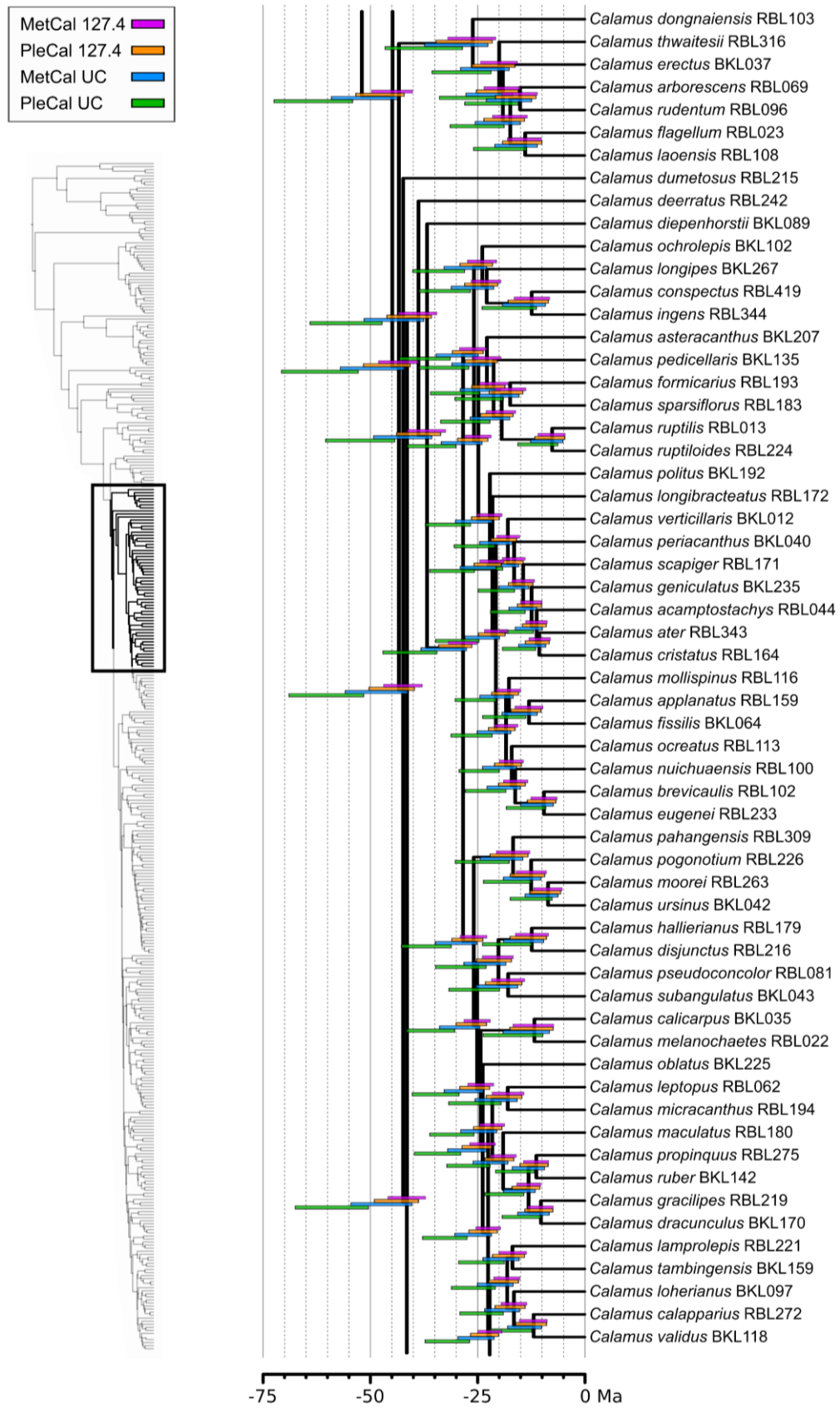



**Figure 3.4** (*continued*)

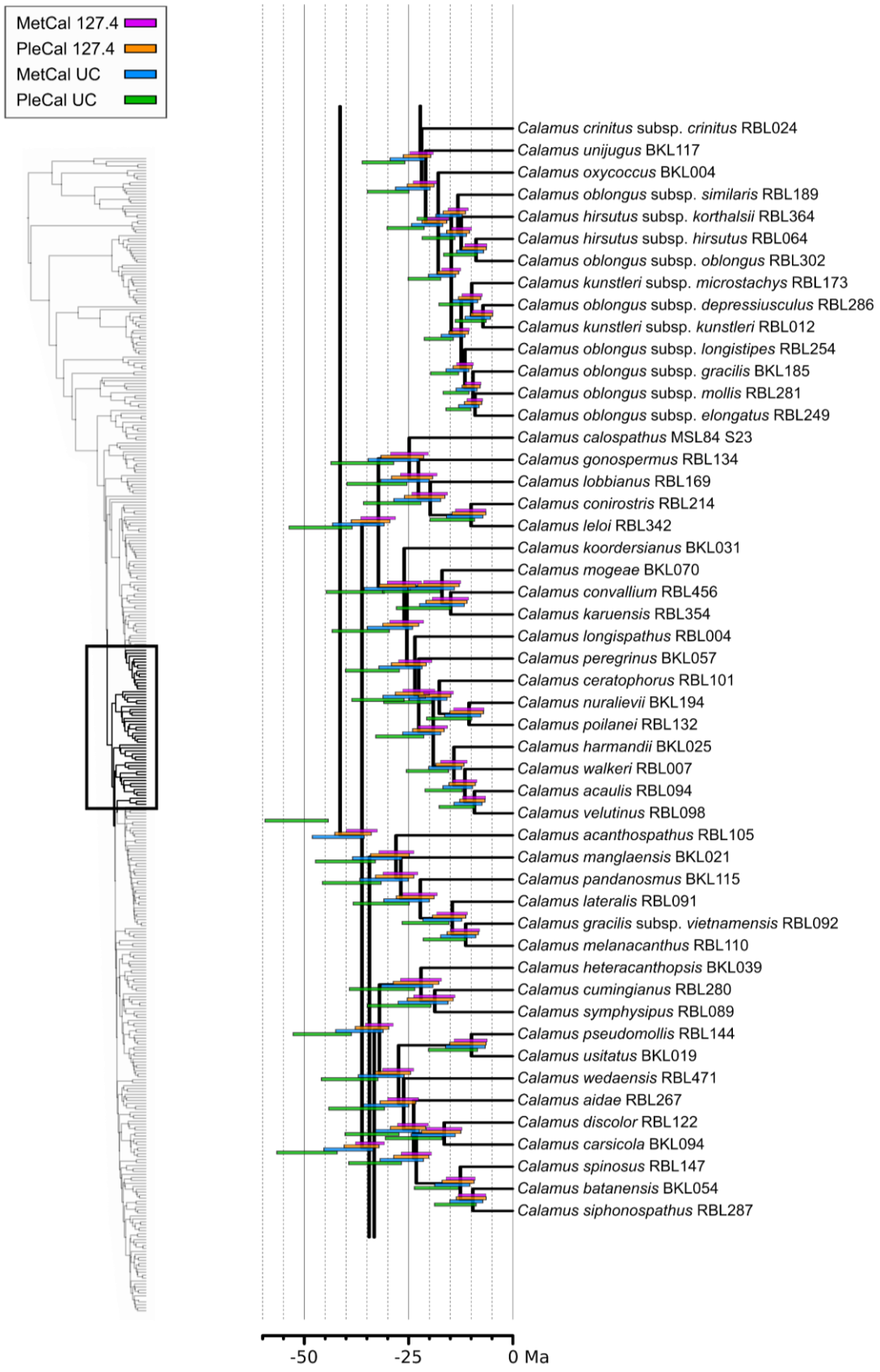



**Figure 3.4** (*continued*)

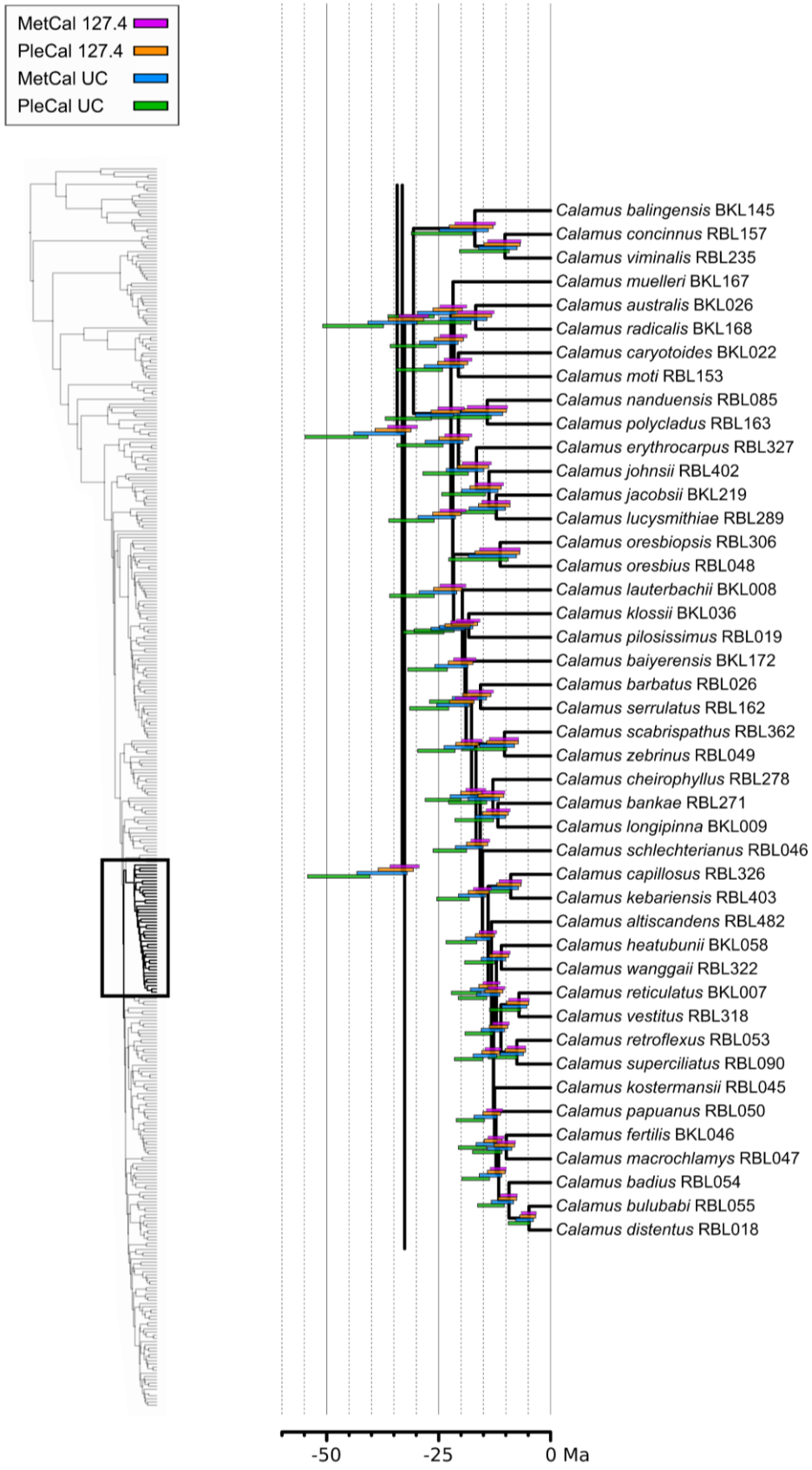



**Figure 3.4** (*continued*)

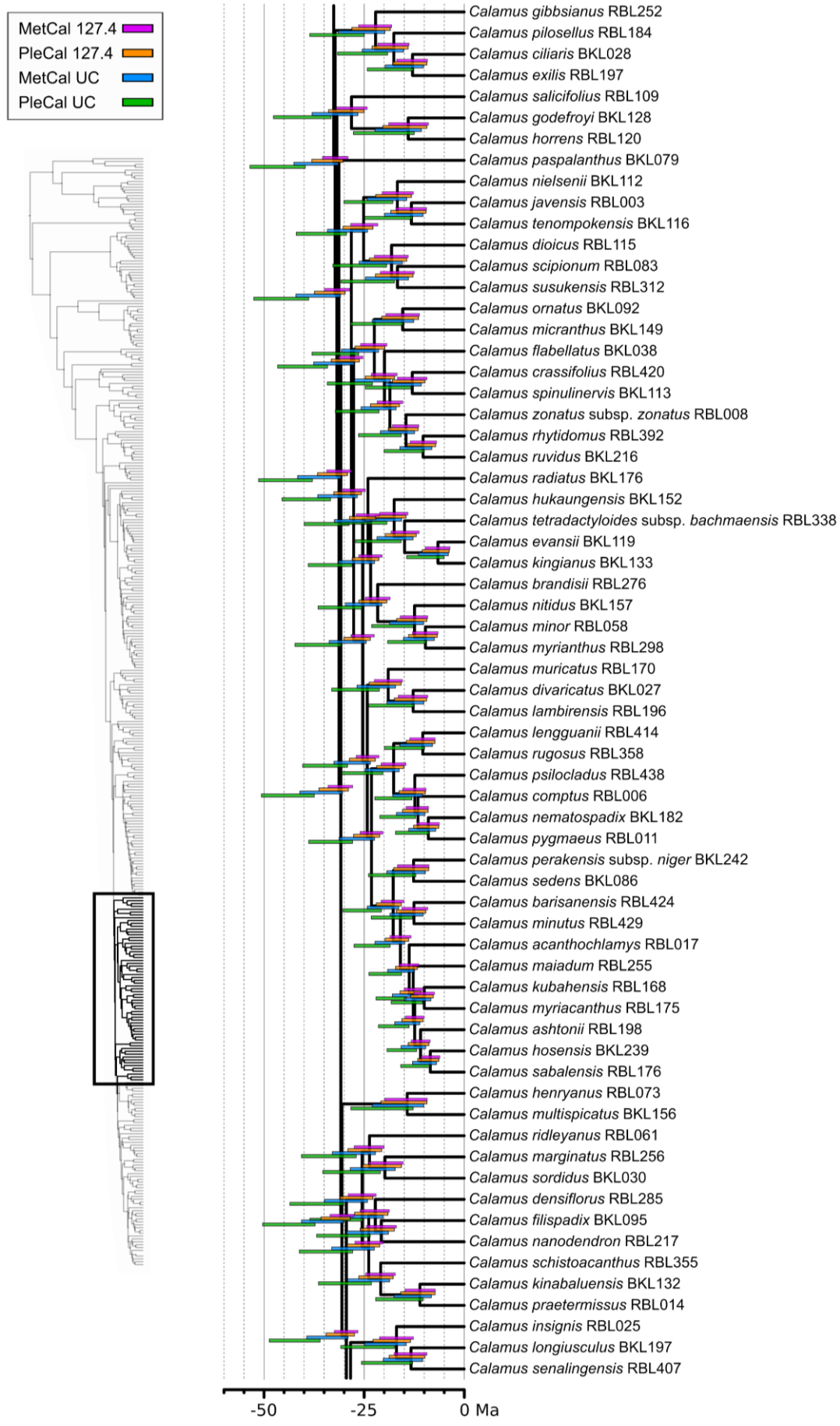



**Figure 3.4** (*continued*)

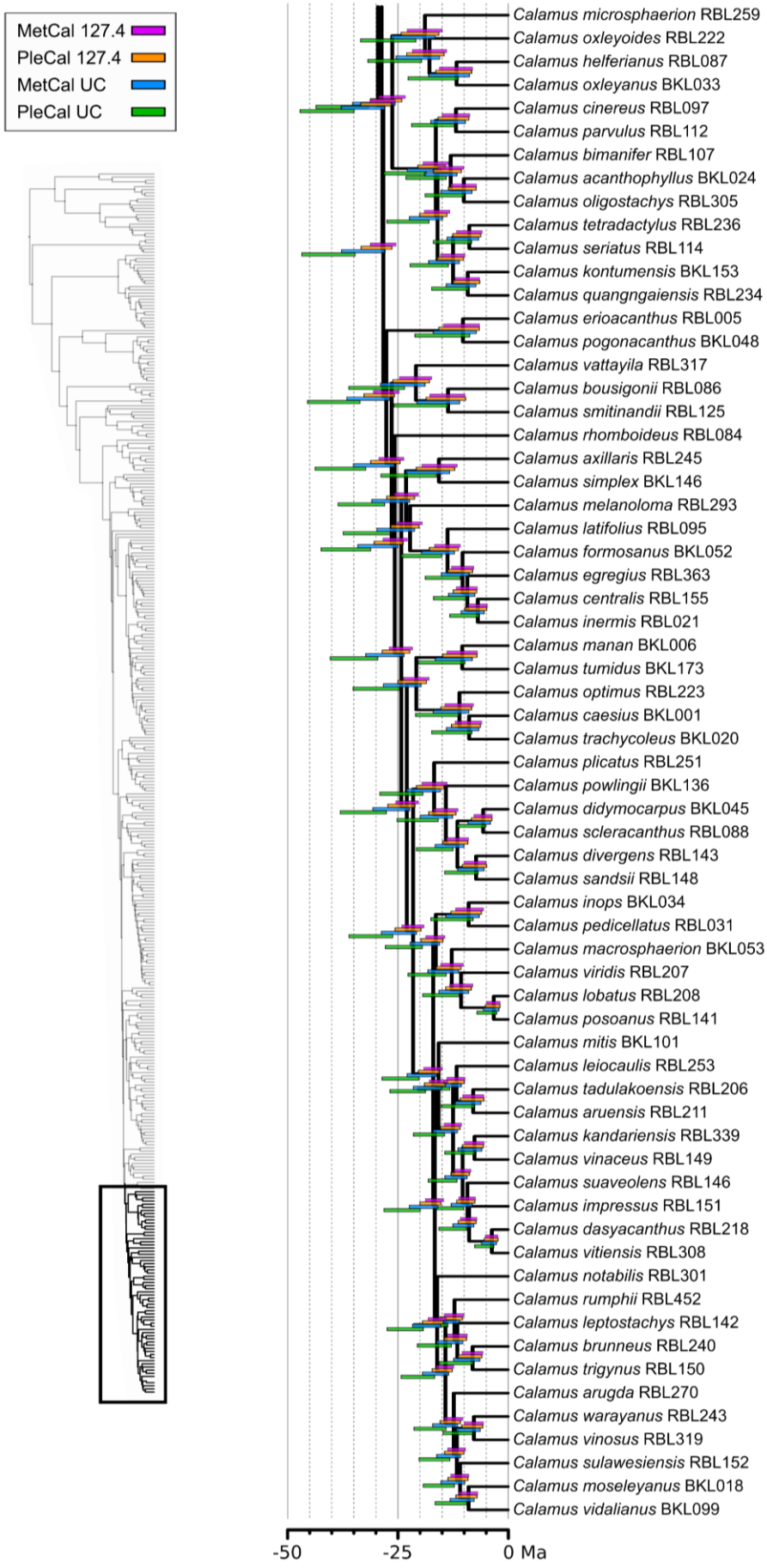



**Figure 3.4** (*continued*)

In both maximum age constrained analyses, the estimated Calamoideae crown node age distributions were pushing against this limit (Supplementary Fig. 3.4 a, b, Table 3.6), whereas age estimates of the internal fossil calibrated nodes were not constrained by the prior (Supplementary Fig. 3.4 c - f). In contrast, the analyses without a maximum age constraint for the crown node of Calamoideae were limited by the implemented minimum stem node age of Metroxylinae, Pigafettinae, Plectocomiinae and Calaminae (Supplementary Fig. 3.4 g) or Plectocomiinae and Calaminae (Supplementary Fig. 3.4 h), respectively, whereas age estimates for the stem node of Mauritiinae were not constrained by the prior (Supplementary Fig. 3.4 i, j).

***Divergence time estimates within the Calamoideae***

Robust divergence time estimates for the extant Calamoideae can be derived from a comparison of highest posterior density intervals across all dating analyses (Fig. 3.4, Table 3.6). The minimum estimated crown node age for the Calamoideae was 122.3 Ma, whilst the maximum age was up to 231.6 Ma.

In Lepidocaryeae, stem node age estimates of the tribe varied between 122.3 and 231.6 Ma, and crown node age estimates between 88.1 and 201.4 Ma. Age estimates for the crown node of Mauritiinae were 43.1 to 111.6 Ma, for the crown node of the sister clade of *Mauritia* and *Mauritiella* 31.1 to 88.4 Ma, and for the crown node of *Mauritiella* 5.1 to 18.2 Ma. The crown node age of the clade formed by Raphiinae and Ancistrophyllinae was estimated to be 77.6 to 182.9 Ma, and the crown node age inferred for Ancistrophyllinae

was 45.6 to 103.1 Ma. For the genera *Raphia*, *Eremospatha*, *Oncocalamus* and *Laccosperma*, similar crown node ages ranging from 12.5 to 45.2 Ma were estimated.

For tribes Eugeissoneae and Calameae, the estimated stem node age was 118.2 to 213.7 Ma. Crown node ages of *Eugeissona*, the only genus in Eugeissoneae, varied from 11.3 to 26.6 Ma. For Calameae, the inferred age of the crown node (corresponding to the stem node of Korthalsiinae) was 99.3 to 154.7 Ma. The estimated crown node age of Korthalsiinae was 30.5 to 60.9 Ma. The crown node age of the clade defined by Salaccinae, Metroxylinae, Pigafettinae, Plectocomiinae and Calaminae (corresponding to the stem node age of Salaccinae) was inferred to be between 94.1 and 136.7 Ma. Age estimates for the crown node of Salaccinae ranged from 60.4 to 111.8 Ma, for crown node of *Salacca* from 37.1 to 81.8 Ma, and for the crown node of *Salacca* excluding *S. secunda* from 17.6 to 36.6 Ma. The clade defined by Metroxylinae, Pigafettinae, Plectocomiinae and Calaminae had an estimated crown node age (corresponding to the stem node age of Metroxylinae) of 83.0 to 124.9 Ma, and a crown node age of 18.1 to 46.0 Ma was inferred for Metroxylinae. The age of the crown node of the clade defined by Pigafettinae, Plectocomiinae and Calaminae (corresponding to the stem node of Pigafettinae) was estimated as 72.8 to 110.0 Ma, and the crown node of Pigafettinae as 7.4 to 25.9 Ma. For the sister clades Plectocomiinae and Calaminae, the inferred crown node age was 68.5 to 106.1 Ma. The crown node age of Plectocomiinae was inferred as 41.1 to 77.8 Ma, and estimates for the crown node ages of *Myrialepis*, *Plectocomiopsis* and *Plectocomia* were 39.0 to 75.4, 16.4 to 41.0, and 30.3 to 62.2 Ma, respectively. The inferred crown node age of Calaminae was 46.2 to 82.4 Ma.

**Table 3.6. Estimated crown node ages of selected clades of the Calamoideae.** Age estimates (in millions of years) for four different dating strategies (Table 3.2). min = lower boundary of highest posterior density interval (HPD), max = upper boundary of HPD. core *Salacca* = *Salacca* excluding *S. secunda*. For monogeneric tribes and subtribes, genus names are given in parentheses.

| Crown node | MetCal 127.4 | | | PleCal 127.4 | | | MetCal UC | | | PleCal UC | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | min | mean | max | min | mean | max | min | mean | max | min | mean | max |
| Calamoideae | 122.3 | 125.6 | 127.4 | 125.1 | 126.6 | 127.4 | 133.6 | 157.8 | 183.2 | 164.8 | 196.3 | 231.6 |
| Lepidocaryeae | 88.1 | 100.3 | 111.2 | 90.9 | 101.8 | 111.2 | 105.4 | 131.7 | 159.7 | 128.6 | 164.2 | 201.4 |
| Mauritiinae | 43.1 | 55.6 | 69.0 | 44.3 | 57.4 | 70.9 | 50.4 | 68.3 | 87.4 | 61.7 | 85.3 | 111.6 |
| *Mauritia* + *Mauritiella* | 31.1 | 42.3 | 54.8 | 31.9 | 43.5 | 56.7 | 36.8 | 52.1 | 69.0 | 44.5 | 65.3 | 88.4 |
| *Mauritiella* | 5.1 | 8.6 | 12.4 | 5.3 | 9.0 | 13.4 | 5.9 | 10.0 | 14.9 | 7.1 | 12.5 | 18.2 |
| Raphiinae + Ancistrophyllinae | 77.6 | 90.1 | 101.4 | 80.6 | 91.6 | 102.3 | 93.3 | 118.6 | 145.3 | 113.7 | 147.8 | 182.9 |
| Raphiinae (*Raphia*) | 14.3 | 19.7 | 25.5 | 14.5 | 20.7 | 26.6 | 16.3 | 22.9 | 30.2 | 19.8 | 28.6 | 37.8 |
| Ancistrophyllinae | 45.6 | 54.8 | 64.9 | 47.0 | 56.6 | 67.3 | 52.0 | 65.8 | 81.5 | 63.9 | 82.5 | 103.1 |
| *Eremospatha* | 18.9 | 24.2 | 29.8 | 19.4 | 25.2 | 31.2 | 21.0 | 28.1 | 36.2 | 26.4 | 35.1 | 44.5 |
| *Oncocalamus* + *Laccosperma* | 40.1 | 49.3 | 59.1 | 41.2 | 50.8 | 61.1 | 44.8 | 59.2 | 73.1 | 56.2 | 74.2 | 93.6 |
| *Oncocalamus* | 12.5 | 17.8 | 23.8 | 12.6 | 18.5 | 24.5 | 14.2 | 20.8 | 27.8 | 18.0 | 26.1 | 35.4 |
| *Laccosperma* | 19.0 | 24.5 | 30.2 | 19.3 | 25.9 | 32.4 | 21.5 | 28.5 | 36.1 | 26.3 | 35.7 | 45.2 |
| Eugeissoneae + Calameae | 118.2 | 122.9 | 127.1 | 123.4 | 125.6 | 127.4 | 125.1 | 146.0 | 170.7 | 150.9 | 180.7 | 213.7 |
| Eugeissoneae (*Eugeissona*) | 11.3 | 14.7 | 18.3 | 11.5 | 15.4 | 19.9 | 12.5 | 16.7 | 21.8 | 15.3 | 21.1 | 26.6 |
| Calameae | 99.3 | 103.2 | 107.4 | 107.2 | 110.1 | 112.9 | 100.2 | 110.4 | 123.4 | 118.6 | 135.1 | 154.7 |
| Korthalsiinae (*Korthalsia*) | 30.5 | 36.1 | 41.8 | 31.9 | 38.0 | 44.3 | 33.6 | 41.1 | 48.9 | 41.9 | 51.4 | 60.9 |
| Salaccinae + Metroxylinae + Pigafettinae + Plectocomiinae + Calaminae | 94.1 | 95.6 | 97.8 | 101.6 | 104.3 | 107.1 | 94.1 | 99.4 | 109.5 | 108.4 | 120.9 | 136.7 |
| Salaccinae | 60.4 | 72.3 | 83.4 | 63.9 | 78.1 | 90.8 | 62.4 | 76.7 | 89.1 | 76.2 | 93.9 | 111.8 |
| *Eleiodoxa* | 60.4 | 72.3 | 83.4 | 63.9 | 78.1 | 90.8 | 62.4 | 76.7 | 89.1 | 76.2 | 93.9 | 111.8 |
| *Salacca* | 37.1 | 48.8 | 60.0 | 39.7 | 52.6 | 65.6 | 41.2 | 53.3 | 65.5 | 50.8 | 65.4 | 81.8 |
| core *Salacca* | 17.6 | 21.0 | 24.8 | 18.4 | 22.3 | 26.7 | 19.3 | 24.1 | 28.9 | 24.2 | 30.0 | 36.6 |
| Metroxylinae + Pigafettinae + Plectocomiinae + Calaminae | 83.0 | 87.9 | 92.8 | 97.8 | 100.0 | 102.4 | 83.9 | 92.3 | 102.9 | 101.2 | 111.3 | 124.9 |
| Metroxylinae (*Metroxylon*) | 18.1 | 24.4 | 31.5 | 18.5 | 25.7 | 32.8 | 20.2 | 27.7 | 35.9 | 25.6 | 34.9 | 46.0 |
| Pigafettinae + Plectocomiinae + Calaminae | 72.8 | 78.8 | 85.0 | 94.0 | 95.0 | 96.2 | 74.9 | 83.6 | 94.4 | 94.1 | 99.5 | 110.0 |
| Pigafettinae (*Pigafetta*) | 7.4 | 12.6 | 18.3 | 7.7 | 13.4 | 19.7 | 8.5 | 14.2 | 20.7 | 10.5 | 17.8 | 25.9 |
| Plectocomiinae + Calaminae | 68.5 | 75.0 | 81.0 | 86.4 | 90.2 | 93.5 | 70.9 | 79.7 | 89.9 | 88.0 | 95.1 | 106.1 |
| Plectocomiinae | 41.1 | 48.7 | 56.6 | 44.4 | 53.7 | 63.4 | 45.4 | 54.5 | 65.0 | 55.8 | 66.7 | 77.8 |
| *Myrialepis* | 39.0 | 46.8 | 54.5 | 42.2 | 51.6 | 61.4 | 43.3 | 52.4 | 62.2 | 53.8 | 64.1 | 75.4 |
| *Plectocomiopsis* | 16.4 | 22.7 | 29.1 | 17.4 | 24.4 | 32.3 | 18.8 | 25.4 | 33.2 | 23.1 | 31.8 | 41.0 |
| *Plectocomia* | 30.3 | 37.3 | 44.8 | 32.1 | 40.8 | 49.6 | 33.1 | 42.0 | 51.3 | 41.5 | 51.6 | 62.2 |
| Calaminae (*Calamus*) | 46.2 | 51.9 | 58.0 | 48.8 | 55.7 | 63.3 | 49.6 | 58.3 | 67.2 | 62.4 | 71.8 | 82.4 |

## 3.4 Discussion

### 3.4.1 Towards a complete species-level phylogeny

Dated and densely sampled phylogenies hold tremendous value as comparative frameworks for a large range of research disciplines from taxonomy to macroevolution. With rapid advances in sequencing technologies and the unlocking of natural history collections, the ambitious goal of complete species-level phylogenies of large clades has come within reach (Heyduk et al., 2015; Barrett et al., 2016a; Hart et al., 2016; Lewin et al., 2018; de La Harpe et al., 2019; Loiseau et al., 2019; Baker et al., 2021).

The present study exemplifies the largest such attempt to date for a tropical rainforest lineage. Here, near-complete taxon sampling of the highly diverse palm subfamily Calamoideae could be achieved by collating samples from herbarium specimens from 24 herbaria across Asia, Europe, and North America (Supplementary Table 3.1). The sampling included 502 (90%) of the 555 species of the subfamily and was particularly complete for the sister tribes Eugeissoneae, with 100% of its six species covered, and Calameae, with 471 (95%) of its 498 species included. For tribe Lepidocaryeae, 25 (49%) of its 51 species were sampled.

On this basis, sequence data was generated using targeted sequence capture of 971 nuclear loci (Loiseau et al., 2019). While it was possible to retrieve at least one targeted locus for 640 (98%) out of 655 sequenced samples, highly data-deficient samples often had improbable phylogenetic placements (Supplementary Fig. 3.1), necessitating a conservative data filtering strategy leading to the exclusion of 121 (24%) of the 502 sampled species.

Despite the originally near-complete sampling, the final phylogeny of the Calamoideae presented here therefore contains 381 (68%) of the species. Whilst increased sequencing efforts based on existing specimens may generate enough data to allow the inclusion of some additional species in future phylogenies of the Calamoideae, this may not be an option for species for which only highly degraded plant material is available. Despite the enormous potential that natural history collections hold as a starting point for phylogenomic studies, targeted field work to collect high-quality samples may therefore remain a vital component of future phylogenomic research striving to include all extant species of the Calamoideae.

### 3.4.2 Paralogy in the PhyloPalm loci

This study targeted 971 putatively orthologous nuclear genomic "PhyloPalm" loci developed for the palm family Arecaceae (Loiseau et al., 2019). Because visual assessment of multiple sequence alignments indicated the presence of paralogous sequences, additional tests were conducted, revealing that as many as 613 (63%) of the PhyloPalm loci might be paralogous (Figure 3.2). There was substantial overlap between the paralogous loci identified using the genomes of *Elaeis guineensis* (Chan et al., 2017) and *Calamus inermis* (Zhao et al., 2018), with 338 paralogs identified in both genomes. In contrast, only 28 paralogous loci were detected using HybPiper (Figure 3.2).

The presence of widespread paralogy in the PhyloPalm loci is somewhat surprising given that orthology tests had been conducted in all studies on which the loci are based (Heyduk et al., 2015; de La Harpe et al., 2019). However, 13% of the 176 loci of Heyduk et al. (2015) were found here to be paralogous, and 74% of the additional 795 loci by de La Harpe

et al. (2019) that were included by Loiseau et al. (2019). While this in particular calls for a re-assessment of orthology of the loci selected by de La Harpe et al. (2019), it must also be noted that these loci were phylogenetically much more informative with a median alignment length of 3037 bases and median of 1427 parsimony informative sites compared to the loci of Heyduk et al. (2015), which had a median alignment length of 1716 bases and median of 752 parsimony informative sites. Overall, these findings highlight the need for the combination of different methodological approaches to reliably detect paralogy in phylogenomic data sets.

Because the subfamily Calamoideae (represented here by *Calamus*) is sister to all other palm subfamilies (represented here by *Elaeis*), the finding of a large number of paralogous loci for both subfamilies supports previous evidence of a genome duplication event common to the ancestor of all palms (Barrett et al., 2019). The fact that over 100 paralogs were uniquely identified in *Calamus* and *Elaeis*, respectively, is probably not surprising given the divergence of both lineages at least 120 million years ago, i.e. before the crown age of the Calamoideae (Table 3.6). It could be either explained by independent gene or genome duplications in both lineages, or by the differential loss of paralogous genes (De Smet et al., 2013; Albalat and Cañestro, 2016).

A comparison of the phylogeny presented in this chapter (Fig. 3.4), which was based only on the 354 orthologous loci, and the phylogeny presented in chapter 2 (Fig. 2.3), which was based on all PhyloPalm loci and therefore also includes the loci inferred here as paralogous, allows an assessment of the impacts of paralogy on phylogeny estimation. Interestingly, the topologies of both phylogenies are highly similar (Supplementary Fig. 3.5). All inferred relationships at and above genus level, as well as species-level relationships in all

subtribes except Plectocomiinae and Calaminae, are identical. In Plectocomiinae and Calaminae, species-level relationships are almost identical and differ only in small details for few species (Supplementary Fig. 3.5). Overall, these findings support the view that methods robust to incomplete lineage sorting, which were used here for phylogeny reconstruction (Zhang et al., 2018), are also robust to the presence of paralogs (Smith and Hahn, 2021).

### 3.4.3 Systematic implications

This is the first phylogenomic study with a dense sampling of species across all genera in the Calamoideae, allowing a validation of the current classification of the subfamily. All genera were strongly supported as monophyletic, validating the existing classification at genus level and above (Baker and Dransfield, 2016). Further, relationships at genus level and above were identical compared to a phylogenomic backbone of the Calamoideae (Kuhnhäuser et al., 2021).

Within genera, many of the clades were morphologically well-defined (Supplementary Fig. 3.2). In *Raphia*, the phylogenetic reconstruction corresponds to differences in inflorescence structure (Helmstetter et al., 2020), with *R. regalis*, an acaulescent species with erect inflorescences that are subtended by leaves (section *Erectae*), resolved as sister to the other species, which have different inflorescences (Supplementary Fig. 3.2). In *Eugeissona*, two clades differing in habit and distribution were resolved, with the stemless *E. tristis* and *E. brachystachys* occurring in the Malay Peninsula, and the erect *E. minor*, *E. insignis* and *E. utilis* constrained to Borneo (Supplementary Fig. 3.2). The poorly known Bornean species *E. ambigua* was data-deficient but plausibly resolved in the preliminary phylogeny as sister to the other Bornean species (Supplementary Fig. 3.1). In

*Korthalsia*, differences in the structure of the ocrea, an extension of the leaf sheath beyond the insertion of the leaf stalk that is characteristic for the genus (Dransfield, 1981b), appear of systematic significance. All species with an "inflated" ocrea form a clade, to which both species with an ocrea diverging from the stem are the sister clade (Supplementary Fig. 3.2). *Salacca secunda*, a species that is morphologically intermediate between *Eleiodoxa* and the other species of *Salacca* (Dransfield et al., 2008), was resolved here as sister to the remaining species of the genus (Supplementary Fig. 3.2). *Salacca griffithii*, the second species that is morphologically intermediate between both genera, was data-deficient but plausibly resolved as sister to *S. secunda* in the preliminary phylogeny (Supplementary Fig. 3.1). Species with fruits with adpressed scales (*S. affinis*, *S. ramosiana*), formerly classified into section *Leiosalacca* (Beccari, 1918), form a clade nested within the genus. All other species of which fruits are known belong to the former section *Salacca* based on their divergent and spine-like fruit scale tips (Supplementary Fig. 3.2). The synonymised genus *Lophospatha*, now *Salacca lophospatha*, is nested within the genus and was resolved as sister to *S. clemensiana*, confirming this presumed relationship based on morphology (Dransfield and Mogea, 1981). In *Metroxylon*, all species of section *Coelococcus*, defined by having fruits covered in 20 or more vertical rows of scales (McClatchey, 1998), formed a clade that was sister to *M. sagu*, the only species in section *Metroxylon*, which has less than 20 rows of fruit scales (Supplementary Fig. 3.2). In *Plectocomiopsis*, *P. geminiflora*, the only species with a disintegrating ocrea, was resolved as sister to a clade of *P. songthanhensis* and *P. mira*, both with a persistent ocrea (Supplementary Fig. 3.2). In *Plectocomia*, section *Plectocomia* subsection *Trigonae* was resolved as paraphyletic, with subsections *Pedicelliformes* and *Plectocomia* nested within (Supplementary Fig. 3.2). *Plectocomia*

*pygmaea*, the only species in section *Penicillatae* (Madulid, 1981), had insufficient data for phylogenetic placement.

In *Calamus*, a basal grade with stalked partial inflorescences (rachillae) comprising the informal *C. andamanicus*, *C. albidus* and *C. arborescens* species groups (Henderson, 2020) as well as *C. dumetosus* (syn. *Retispatha dumetosa*) was resolved (Supplementary Fig. 3.2). The synonymised genera *Ceratolobus*, *Daemonorops* and *Pogonotium*, all characterised by their inflated primary inflorescence bracts, formed a clade nested within *Calamus*. The species of synonym *Daemonorops* were resolved in two monophyletic groups interspersed by clades formed by the species of synonyms *Ceratolobus* and *Pogonotium*, rendering *Daemonorops* polyphyletic. Species of the former genera *Calospatha*, *Cornera* and *Zalaccella* were resolved as closely related and nested within a flagellate clade in *Calamus* (Supplementary Fig. 3.2). The highly specialised climbing organs in *Calamus* appear of high systematic significance, although ca. 10% of the species of *Calamus* do not possess climbing organs (Couvreur et al., 2015). Large clades and grades are defined by the presence of either a cirrus (an extension of the leaf rachis, equipped with hooked spines) or a flagellum (a modified inflorescence axis, equipped with hooked spines) (Supplementary Fig. 3.2). Whilst the presence of cirrus or flagellum is generally mutually exclusive in *Calamus*, some species possess both a cirrus and a rudimentary (vestigial) flagellum, as in the clade formed by the species of *C. erioacanthus* and *C. pogonacanthus*. The only larger clade defined by the absence of climbing organs is the *Calamus acanthochlamys* group. Synapomorphies of other clades that could be identified include recurvature of the rachillae (*C. ciliaris* group), crowdedness of male flowers (clade defined by *C. balingensis* and *C. viminalis*), joint apical leaflets (clade defined by *C. javensis* and C. *scipionum*), grouped arrangement of leaflets (*C. bimanifer* and *C. elegans* groups), short membranous ocreae (clade comprising *C. adspersus*,

*C. arugula*, *C. brunneus*, *C. didymocarpus*, *C. inops* groups), and large conspicuous ocreae (flagellate *Calamus* from New Guinea and Australia) (Supplementary Fig. 3.2).

### 3.4.4 Considerably older divergence time estimates compared to previous studies

The divergence time estimates for the Calamoideae based on 354 orthologous genes, 390 taxa, and the incorporation of new fossil evidence shed new light on the early evolution of palms and the build-up of extant diversity through time. The minimum crown node age of the Calamoideae across the four analyses conducted here was estimated as 122.3 Ma (Table 3.6), pre-dating previously widely used maximum crown node age estimates for the subfamily Calamoideae and family Arecaceae (Couvreur et al., 2011a) by over 30 and 10 million years, respectively. Within the Calamoideae, mean divergence times estimated here for crown nodes of calamoid subtribes were between 7.4 and 37.3 Ma older than previous mean divergence time estimates (Baker and Couvreur, 2013a). The substantially older divergence time estimates inferred here compared to previous studies can potentially be explained by three cumulative factors.

First, a revision of the fossil record of *Mauritiidites* (Bogotá-Ángel et al., 2021) and the recent discovery of Cenomanian fossils of *Dicolpopollis* (Totterdell and Mitchell, 2009; Macphail and Jordan, 2015) resulted in older ages for both key calamoid fossil taxa. Whereas the previous oldest reliable age estimate for *Mauritiidites* was Maastrichtian (72.1-66.0 Ma) (Iles et al., 2015), a comprehensive review of Mauritiinae-like pollen found that *Mauritiidites* is at least of Campanian (83.6-72.1 Ma) age (Bogotá-Ángel et al., 2021). Even more importantly, several recently discovered fossils of *Dicolpopollis* from the Cenomanian (100.5-93.9 Ma) of Australia (Totterdell and Mitchell, 2009; Macphail and Jordan, 2015)

substantiate earlier but rarely considered fossils from this time and region (Burger, 1990), and pre-date the previously established Maastrichtian age of *Dicolpopollis* (Harley, 2006; Iles et al., 2015) by over 20 million years.

Second, newly resolved and strongly supported backbone relationships of the Calamoideae (Kuhnhäuser et al., 2021), which were confirmed and substantially expanded here (Supplementary Fig. 3.2), established a robust phylogenomic framework for the placement of fossils. This allowed the more precise assignment of the fossil pollen taxon *Dicolpopollis kalewensis* to the stem node of the clade constituted by Plectocomiinae and Calaminae when considering both pollen macro- and microstructure, or more conservatively to the stem node of the clade constituted by Metroxylinae, Pigafettinae, Plectocomiinae and Calaminae when only considering pollen macrostructure (Figure 3.1). Both possible placements are nested within tribe Calameae, whereas *Dicolpopollis* had previously been assigned to the stem node of Calameae (Couvreur et al., 2011a; Iles et al., 2015).

Third, the maximum age constraint of 127.4 Ma implemented here was more lenient than in other dated phylogenies that included calamoid palms (Couvreur et al., 2011a; Baker and Couvreur, 2013a), thus expanding the range of possible divergence time estimates. Despite this relaxation, crown node age estimates of the Calamoideae were strongly limited by the age constraint (Supplementary Fig. 3.4 a, b, Table 3.6), and unconstrained analyses resulted in considerably older age estimates (Fig. 3.4, Table 3.6), indicating that the used age constraint may still be highly conservative. Maximum age constraints are equivalent to the assumption that a lineage had not evolved yet before that time (Ho and Phillips, 2009), an assumption that is difficult to justify because fossil preservation rates are highly variable in space and time and fossil discovery is dependent on rock exposure and sampling efforts

(Holland, 2016), potentially leading to long temporal gaps before the first fossils of a lineage are recorded. The absence of a known fossil record should therefore not be mis-interpreted as evidence for the absence of fossils (Sauquet, 2013). The fact that the Cenomanian fossils of *Dicolpopollis*, which can be confidently assigned to a clade nested within the Calamoideae, pre-date almost all previous age estimates for the crown node of the Calamoideae (Table 3.1), raises the possibility that previous maximum age constraints for the palm family may have been too restrictive.

Overall, the findings presented here strongly suggest an origin of crown Calamoideae by 122.3 Ma, and possibly earlier. Given the position of the Calamoideae as sister to all other palm subfamilies (Baker et al., 2009; Faurby et al., 2016), the much older age of the subfamily established here is likely to considerably push back minimum age estimates of the palm family Arecaceae in future studies, and will potentially also affect divergence time estimates of monocotyledonous plants and beyond.

### 3.4.5 Long lag times between lineage originations and build-up of extant diversity

Many calamoid lineages have long lag times between origination (stem node) and the build-up of extant species diversity (crown node), with stem branch lengths of over 50 million years being observed in *Eugeissona*, *Korthalsia*, *Metroxylon*, *Pigafetta*, and *Raphia* (Fig. 3.4, Table 3.6), and a post-Cretaceous origin of almost all extant calamoid species despite the Early Cretaceous crown node age and early diversification of the Calamoideae (Fig. 3.5).

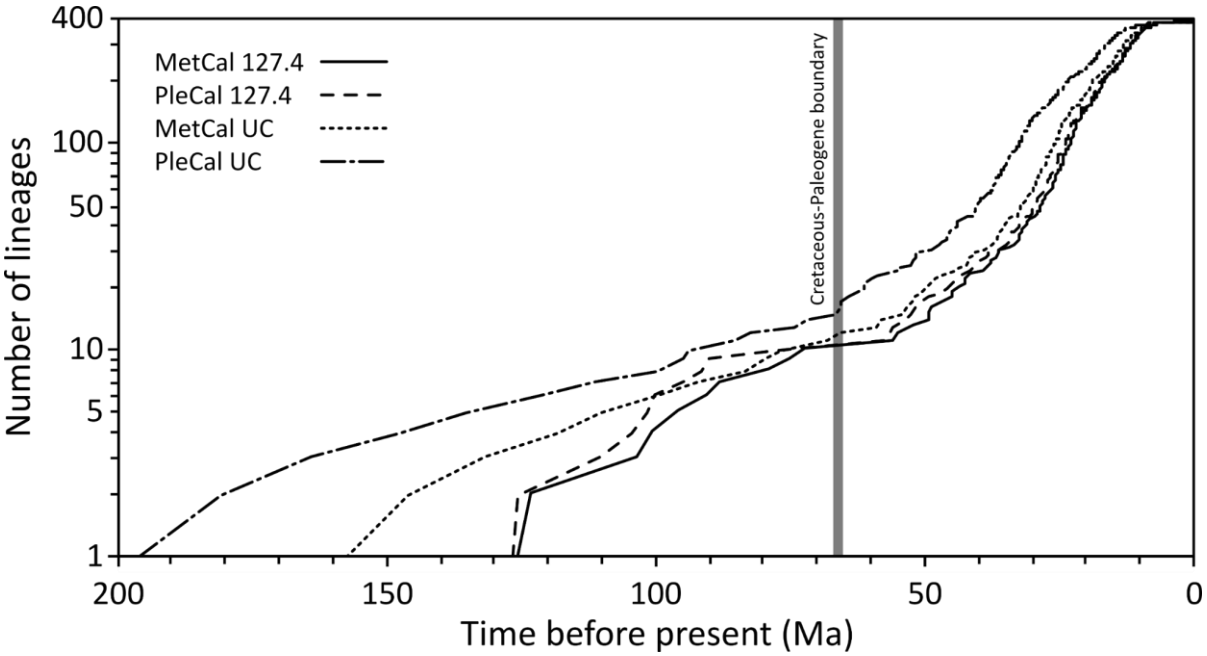



**Figure 3.5. Number of extant calamoid lineages through time.** Number of calamoid lineages was inferred from the dated phylogeny including 390 extant calamoid species and subspecies (Figure 3.4) and four dating strategies (Table 3.2). The Cretaceous-Paleogene boundary is marked by a grey vertical line.

The discrepancy between old ages of most lineages but recent origin of most extant species could be either due to low speciation rates, high extinction rates, or a combination of both. While the inference of speciation and extinction rates based on extant phylogenies is problematic (Rabosky, 2010; Louca and Pennell, 2020, 2021), the rich and distinctive fossil record of the Calamoideae (Morley, 2000; Harley, 2006; Iles et al., 2015; Bogotá-Ángel et al., 2021) can give some insights into the species diversity of the subfamily before the build-up of extant diversity in the recent past (Fig. 3.5). A comprehensive analysis of the diversity of calamoid fossil pollen taxa through time (Lim et al., 2021) shows that the tribes Lepidocaryeae, Eugeissoneae and Calameae had a much wider distribution during the Paleocene (66.0-56.0 Ma) and Eocene (56.0-33.9 Ma) than at present, with *Dicolpopollis* (assignable to subtribes Metroxylinae, Plectocomiinae and Calaminae in tribe Calameae)

present in Australia, Europe, Southeast Asia, and North America, *Quilonipollenites* (tribe Eugeissoneae) in India and Southeast Asia, and *Mauritiidites* (subtribe Mauritiinae in tribe Lepidocaryeae) in Africa, South America, North America, and India.

In contrast, no Calameae exist at present in Europe or North America, no Eugeissoneae in India, and no Mauritiinae in Africa, North America or India (Dransfield et al., 2008). The dramatic geographic range contractions of all calamoid tribes after the Eocene, potentially linked to global cooling and a global decline in tropical rainforest area from the Eocene onwards (Kissling et al., 2012b; Couvreur et al., 2021), were therefore possibly accompanied by high extinction rates and extinction of large local radiations in Africa, Europe, North America, and India, and may explain the long lag times between stem and crown node ages observed in the extant Calamoideae. Similar conclusions have been reached in other plant groups with long stem branches, such as cycads (Nagalingum et al., 2011; Condamine et al., 2015), gymnosperms (Crisp and Cook, 2011), or Chloranthaceae (Antonelli and Sanmartín, 2011).

### 3.4.6 Break-up of Gondwana linked to early phylogenetic history

The divergence times and phylogenetic relationships inferred here for the Calamoideae, combined with the locations of the oldest calamoid fossils, have substantial implications for the understanding of its pantropical distribution, which has been explained alternatively by a Gondwanan origin and spatial separation associated with the break-up of its landmasses (Baker and Dransfield, 2000), or by a Laurasian origin followed by a multiple long-distance dispersal events (Couvreur et al., 2011a; Baker and Couvreur, 2013a).

The African-American tribe Lepidocaryeae and its Asian sister clade, the tribes Eugeissoneae and Calameae, were inferred here to have diverged no later than 122.3 Ma (Table 3.6), followed by the divergence of the Eugeissoneae and Calameae soon after, i.e. no later than 118.2 Ma (Table 3.6). No later than 88.1 Ma, the Lepidocaryeae had diverged into the American subtribe Mauritiinae and the African clade composed of subtribes Raphiinae and Ancistrophyllinae (Table 3.6).

The locations of the oldest calamoid fossils allow some insights into the geographical context in which this diversification took place. The earliest reliable fossil records for tribe Calameae are *Dicolpopollis* pollen fossils from the Cenomanian (100.5-93.9 Ma) of Australia (Burger, 1990; Totterdell and Mitchell, 2009; Macphail and Jordan, 2015). For Eugeissoneae, the oldest unambiguous fossils are *Quilonipollenites* pollen from the Paleocene (66.0-56.0 Ma) of India (Morley, 1998), i.e. from a time when India had separated from Gondawana and Madagascar, and was drifting across the Indian ocean (Müller et al., 2016). The oldest reliable fossils of tribe Lepidocaryeae, the pollen fossil taxon *Mauritiidites* (assigned to subtribe Mauritiinae), are from the Campanian (83.6-72.1 Ma) of Africa (Bogotá-Ángel et al., 2021).

Sequence and timing of the early divergence of the Calamoideae and the geographical location of the oldest fossils fit remarkably well with the sequence and timing of the break-up of Gondwana into East Gondwana (Australia, Antarctica and India) and West Gondwana (Africa and South America) between ca. 160 and 120 Ma (although the southernmost tip of South America remained connected to Western Antarctica until ca. 30 Ma), the subsequent separation of India from Australia and Antarctica between ca. 130 and

110 Ma, and the disjunction of South America from Africa between ca. 120 and 100 Ma (Müller et al., 2016) (Fig. 3.6).

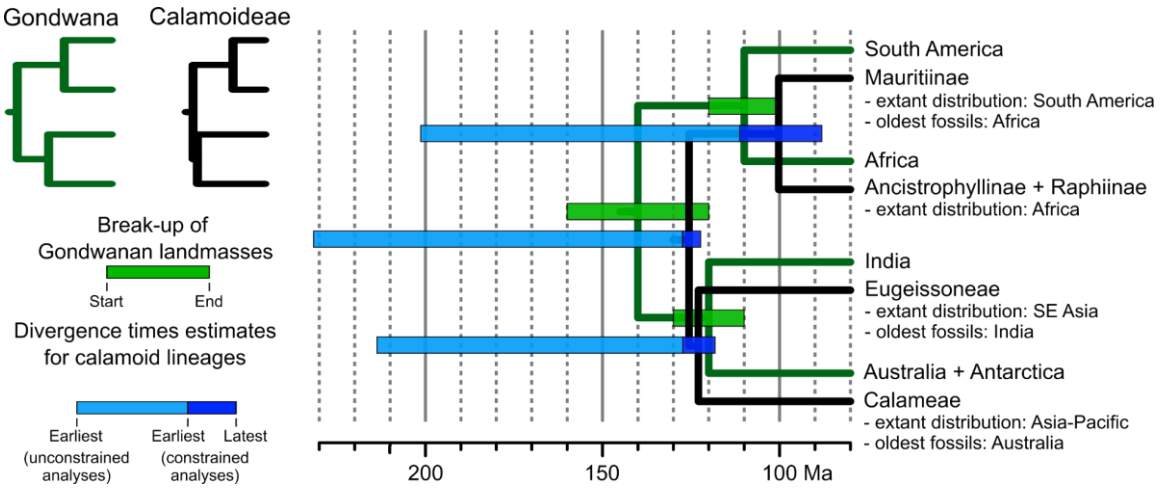



**Figure 3.6. Correlation of Gondwanan break-up and divergence of calamoid lineages.** Sequence and timing of break-up of Gondwanan landmasses and relationships and divergence time estimates for calamoid lineages. Green bars indicate the interval between initiation of separation of Gondwanan landmasses and completion of their separation (Müller et al., 2016). Note, however, that the southernmost tip of South America likely remained connected to Western Antarctica until ca. 30 Ma (Müller et al., 2016). Blue bars indicate highest posterior density intervals for divergence time estimates based on molecular dating analyses with a maximum age constraint for the crown node of Calamoideae of 127.4 Ma (analyses MetCal 127.4 and PleCal 127.4), whereas grey bars extend to the upper end of the highest posterior density intervals of analyses without a maximum age constraint (analyses MetCal UC and PleCal UC).

Overall, the combined evidence is consistent with the hypothesis of a Gondwanan origin of the Calamoideae and central role of the break-up of Gondwana in the early diversification and biogeographical disjunction of the subfamily (Baker and Dransfield, 2000). Given the sister relationship of the Calamoideae to all other palms, this finding potentially requires refining the hypothesis of a Laurasian (Uhl and Dransfield, 1987; Couvreur et al.,

2011a; Baker and Couvreur, 2013a) or South American (Moore, 1973b; Antonelli et al., 2017) origin of the palm family. It further raises the possibility that the geographical disjunction of the other pantropical subfamilies Arecoideae, Ceroxyloideae and Coryphoideae may also be due to former land connections before the break-up of Gondwana, as has been hypothesised earlier (Moore, 1973a, b; Uhl and Dransfield, 1987; Baker and Dransfield, 2000; Hahn, 2002b; Trénel et al., 2007), and questions the presumed overarching role of transoceanic dispersals in achieving their worldwide distribution (Baker and Couvreur, 2013a). New divergence time estimates for the palm family under incorporation of the growing Cenomanian fossil record of the Calamoideae (Burger, 1990; Totterdell and Mitchell, 2009; Macphail and Jordan, 2015) are needed to re-evaluate this hypothesis.

## 3.5 Conclusion

This study represents the largest densely sampled, dated species-level phylogeny of a tropical rainforest plant lineage to date. The phylogeny is well-resolved and strongly supported, allowing highly detailed insights into evolutionary relationships among calamoid species. Divergence time estimates incorporate recently discovered pollen fossils that represent the oldest palm fossils known to date and are substantially older compared to previous studies, establishing an Early Cretaceous age of the crown Calamoideae predating previously established maximum crown node age estimates for both the Calamoideae and the palm family Arecaceae by over 20 and 10 million years, respectively. Contrasting with the ancient age of the crown Calamoideae, ages inferred for the vast majority of extant calamoid lineages substantially post-date the Cretaceous. Overall, this study sheds new light

on the temporal origin and species-level relationships of palms and provides a highly detailed framework for future comparative studies.

## 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, an Emily Holmes Memorial Scholarship by the Amar-Franses & Foster-Jenkins Trust, and two Bentham-Moxon Trust grants (BMT1-2017 and BMT36- 2018) to the author. Labwork was facilitated by a grant from IKEA of Sweden AB to William J. Baker 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.

## Acknowledgements

This research would not have been possible without the generous provision of plant material by the Royal Botanic Gardens, Kew (UK), Andrew Henderson at the New York Botanical Garden (USA), Nicolien Sol at Naturalis (Netherlands), Jiro Adorador at the Forestry Herbarium at the University of the Philippines at Los Baños, Lorenzo Cecchi and Chiara Nepi at Herbarium Universitatis Florentinae (Italy), Himmah Rustiami at Herbarium Bogoriense (Indonesia), Anna Stalter and Kevin Nixon at the Bailey Hortorium Herbarium (USA), the Forest Research Institute Malaysia, James Solomon at Missouri Botanical Garden (USA), Bazilah Mohd Ibrahim at the Singapore Botanic Gardens, Brendan Lepschi and Jo Palmer at

the Australian National Herbarium, Brett Jestrow at Fairchild Tropical Botanic Garden (USA), Peter Homot at Papua New Guinea National Herbarium, Myriam Gaudeul at the Herbarium of the Muséum National d'Histoire Naturelle (France), Mohizah Mohamad at the Sarawak Herbarium (Malaysia), Eric Schuettpelz and Sue Lutz at the United States National Herbarium (USA), Michaela Schmull at the Herbarium of the Arnold Arboretum (USA), Finn Borchsenius at the Herbarium of Aarhus University (Denmark), Phattaravee Prommanut at Bangkok Herbarium (Thailand), Norbert Holstein at the British Museum, London (UK), the Conservatoire et Jardin botaniques de la Ville de Genève (Switzerland), Larry Noblick at the Montgomery Botanical Center (USA), Arne Anderberg at the Swedish Museum of Natural History, Thomas Couvreur (Université de Montpellier), and Ángela Cano (University of Cambridge). In addition, I thank Imalka Kahandawala, Edith Kapinos and Rosie Woods for providing samples from the DNA bank at the Royal Botanic Gardens, Kew. I am grateful for support and advice in conducting DNA extractions by László Csiba, support in library preparation by Rowan Schley, and general phylogenomic lab support by Robyn Cowan, Panagiota Malakasi and Dion Devey. Andrew Henderson and John Dransfield gave valuable advice on specimen identification and the plausibility of phylogenetic placements in a preliminary species-level phylogeny. Michael Macphail kindly provided details on Australian *Dicolpopollis* fossils and gave permission for use of unpublished images. Finally, I would like to thank Robert Morley for providing advice on verification, age, and affinities of pollen fossils. Computational analyses were conducted on the high performance computing facilities of the Royal Botanic Gardens, Kew, and the “UK’s Crop Diversity Bioinformatics HPC” (BBSRC grant BB/S019669/1) at The James Hutton Institute and NIAB.

## Supplementary material

**Supplementary Table 3.1. Voucher information for specimens sequenced in this thesis.** Names in parentheses indicate species recognized by Vorontsova et al. (2016a) that were synonymised in the taxonomic revision of *Calamus* by Henderson (2020). Herbarium codes in the voucher information follow Thiers (2020). Seq. ID = Sequence ID. Herbarium specimens for which the species identity was corrected during this study are marked with an asterisk.

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *acamptostachys* | Baker 703 (K) | RBL044 |
| *Calamus* | *acanthochlamys* (*temburongii*) | Said 15623 (K) | BKL230 |
| *Calamus* | *acanthochlamys* | Cultivated 1989-2701 (K) | RBL017 |
| *Calamus* | *acanthophyllus* | Khamphone 141 (K) | BKL024 |
| *Calamus* | *acanthospathus* | Henderson 3784 (NY) | RBL105 |
| *Calamus* | *acaulis* | Henderson 3608 (NY) | RBL094 |
| *Calamus* | *adspersus* | Hooker 93 (K) | RBL265 |
| *Calamus* | *aidae* | Baja-Lapis 123 (K) | RBL267 |
| *Calamus* | *albidus* | Henderson 4028 (NY) | RBL205 |
| *Calamus* | *altiscandens* | Brass 7327 (K) | RBL482 |
| *Calamus* | *andamanicus* | Renuka 4063 (K) | RBL269 |
| *Calamus* | *anomalus* | Brass 5298 (NY) | RBL440 |
| *Calamus* | *applanatus* | Henderson 3220 (NY) | RBL159 |
| *Calamus* | *arborescens* | Henderson 3182 (NY) | RBL069 |
| *Calamus* | *aruensis* | Baker 568 (K) | 14931-S33 |
| *Calamus* | *aruensis* | Dransfield 7571 (K) | 15786-S14 |
| *Calamus* | *aruensis* | Baker 568 (K) | RBL211 |
| *Calamus* | *aruensis* | Dransfield 7571 (K) | RBL231 |
| *Calamus* | *arugda* | Fernando 359 (K) | RBL270 |
| *Calamus* | *ashtonii* | Kuhnhäuser 81 (K) | RBL198 |
| *Calamus* | *asperrimus* (*benkulensis*) | Dransfield 3602 (L) | BKL175 |
| *Calamus* | *asperrimus* (*benkulensis*) | Dransfield 3602 (L) | RBL323 |
| *Calamus* | *asteracanthus* | Johns 6521 (K) | BKL207 |
| *Calamus* | *ater* | Eiserhardt 154 (K) | RBL343 |
| *Calamus* | *australis* | Cultivated 1988-186 (K) | BKL026 |
| *Calamus* | aff. *axillaris* | Heyne 24 (L)* | RBL331 |
| *Calamus* | *axillaris* | Wong 1620 (K) | RBL245 |
| *Calamus* | *badius* | Dransfield 7661 (K) | RBL054 |
| *Calamus* | *baiyerensis* | Zieck NGF 36252 (CANB) | BKL172 |
| *Calamus* | *balingensis* (*setulosus*) | Dransfield 7339 (K) | BKL145 |
| *Calamus* | *balingensis* | Henderson 4005 (K) | RBL126 |
| *Calamus* | *banggiensis* | Diwol & Sugau SAN 135133 (K) | BKL208 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *bankae* | Baker 1097 (K) | RBL271 |
| *Calamus* | *barbatus* | Gardiner 436 (K) | RBL026 |
| *Calamus* | *barfodii* | Barfod 464 (K) | BKL263 |
| *Calamus* | *barisanensis* (*perakensis*) | Dransfield 4063 (K) | RBL424 |
| *Calamus* | *batanensis* | Jeanson 93 (NY) | BKL054 |
| *Calamus* | *beccarii* | Jeanson 89 (NY) | RBL154 |
| *Calamus* | *bicolor* | Elmer 10541 (K) | RBL425 |
| *Calamus* | *billitonensis* | Heyne 12 (L) | RBL484 |
| *Calamus* | *bimanifer* | Henderson 3833 (NY) | RBL107 |
| *Calamus* | *boniensis* | Heyne 2596 (BO) | RBL464 |
| *Calamus* | *bousigonii* | Henderson 3628 (NY) | RBL086 |
| *Calamus* | *brandisii* | Nair 6144 (K) | RBL276 |
| *Calamus* | *brevicaulis* | Henderson 3861 (NY) | RBL102 |
| *Calamus* | *brunneus* | Henderson 4281 (NY) | RBL240 |
| *Calamus* | *bulubabi* | Dransfield 7688 (K) | RBL055 |
| *Calamus* | *burkillianus* | Dransfield 5195 (K) | RBL457 |
| *Calamus* | *caesius* (*tapa*) | Giesen s.n. (K) | RBL397 |
| *Calamus* | *caesius* | Cultivated 1989-1214 (K) | BKL001 |
| *Calamus* | *calapparius* (*beguinii*) | de Vogel 3107 (K) | RBL272 |
| *Calamus* | *calapparius* (*komsaryi*) | Baker 1386 (K) | BKL047 |
| *Calamus* | *calapparius* (*sarasinorum*) | de Vogel 3817 (K) | RBL395 |
| *Calamus* | *calapparius* | Mogea 5299 (K) | RBL458 |
| *Calamus* | *calciphilus* (*marginatus*) | de Vogel 2054 (K) | RBL426 |
| *Calamus* | *calicarpus* | Baker 506 (K) | BKL035 |
| *Calamus* | *calospathus* | Cultivated 1990-2783 (K) | BKL011 |
| *Calamus* | *calospathus* | Cultivated 1990-2783 (K) | MSL84-S23 |
| *Calamus* | *capillosus* | Wanda Ave 4048 (L) | RBL326 |
| *Calamus* | *carsicola* | Adorador 72 (LBC) | BKL094 |
| *Calamus* | *caryotoides* | Cultivated 1988-196 (K) | BKL022 |
| *Calamus* | *castaneus* (*inopinatus*) | Furtado 301919 (L) | BKL266 |
| *Calamus* | *castaneus* | Baker 507 (K) | 17250-S44 |
| *Calamus* | *castaneus* | Baker 507 (K) | RBL220 |
| *Calamus* | *caurinus* | Dransfield 3300 (K) | BKL264 |
| *Calamus* | *centralis* | Nguyen Quoc Dung 2027 (NY) | RBL155 |
| *Calamus* | *ceratophorus* | Henderson 3690 (NY) | RBL101 |
| *Calamus* | *cheirophyllus* | Zieck NGF 36516 (K) | RBL278 |
| *Calamus* | *ciliaris* | Cultivated 1992-2048 (K) | BKL028 |
| *Calamus* | *cinereus* (*tetradactylus*) | Henderson 3625 (NY) | RBL097 |
| *Calamus* | *clivorum* | Clemens 4485 (P) | RBL378 |
| *Calamus* | *cockburnii* | Cockburn FRI 10949 (BH) | RBL444 |
| *Calamus* | *compsostachys* | RITF s.n. (K) | RBL028 |
| *Calamus* | aff. *comptus* | Kuhnhäuser 63 (K)* | RBL190 |
| *Calamus* | *comptus* | Baker 540 (K) | RBL006 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *concinnus* | Watson 604 (FTG) | RBL157 |
| *Calamus* | *confusus* (*acehensis*) | Dransfield 2015 (BO) | RBL463 |
| *Calamus* | *confusus* | Nur 7308 (K) | RBL459 |
| *Calamus* | *conirostris* | Baker 532 (K) | 14971-S55 |
| *Calamus* | *conirostris* | Baker 532 (K) | RBL214 |
| *Calamus* | *conjugatus* | Dransfield 5864 (K) | RBL340 |
| *Calamus* | *conspectus* | Dransfield 5328 (K) | RBL419 |
| *Calamus* | aff. *convallium* | Sugau 145693 (K)* | BKL179 |
| *Calamus* | *convallium* | Dransfield 6136 (K) | RBL456 |
| *Calamus* | *crassifolius* | Dransfield 6035 (K) | RBL420 |
| *Calamus* | *crinitus* | Toroes 4291 (K) | RBL248 |
| *Calamus* | *crinitus* subsp. *crinitus* (*mirabilis*) | Winkler 2602 (FI) | BKL199 |
| *Calamus* | *crinitus* subsp. *crinitus* (*sabut*) | Rajasegar 8 (K) | RBL024 |
| *Calamus* | *crinitus* subsp. *forbesii* (*forbesii*) | Dransfield 3576 (K) | BKL125 |
| *Calamus* | *crinitus* subsp. *sabut* (*macrophyllus*) | Dransfield 4970 (K) | RBL290 |
| *Calamus* | *crinitus* subsp. *sabut* (*oligophyllus*) | Avé 7 (K) | RBL304 |
| *Calamus* | *cristatus* | Kuhnhäuser 3 (K) | RBL164 |
| *Calamus* | *croftii* | Zieck NGF 36542 (K) | RBL279 |
| *Calamus* | *cumingianus* | Fernando 339 (K) | RBL280 |
| *Calamus* | *cuthbertsonii* | Croft 65028 (K) | RBL282 |
| *Calamus* | *dasyacanthus* | Baker 827 (K) | 16260-S24 |
| *Calamus* | *dasyacanthus* | Baker 827 (K) | RBL218 |
| *Calamus* | *deerratus* | Tsiforkor s.n. (K) | 14968-S63 |
| *Calamus* | *deerratus* | Dümmer 2650 (K) | BKL174 |
| *Calamus* | *deerratus* | Tsiforkor s.n. (K) | RBL213 |
| *Calamus* | *deerratus* | Sunderland 1754 (K) | RBL242 |
| *Calamus* | *delicatulus* | Thwaites 3159 (K) | RBL284 |
| *Calamus* | *densifloropsis* | Mogea 1542 (AAU) | RBL443 |
| *Calamus* | *densiflorus* | Dransfield 4951 (K) | RBL285 |
| *Calamus* | *dianbaiensis* | Nguyen Quoc Dung 2031 (NY) | RBL156 |
| *Calamus* | *didymocarpus* | Henderson 4308 (NY) | BKL045 |
| *Calamus* | *diepenhorstii* | Baker 725 (K) | BKL089 |
| *Calamus* | *diepenhorstii* | Witono 174 (K) | BKL245 |
| *Calamus* | *digitatus* | de Zoysa 30 (K) | BKL063 |
| *Calamus* | *dilaceratus* | Man s.n. (FI) | BKL253 |
| *Calamus* | *dioicus* | Henderson 3891 (NY) | RBL115 |
| *Calamus* | *discolor* | Henderson 3996 (K) | RBL122 |
| *Calamus* | *disjunctus* (*concolor*) | Baker 559 (K) | 15002-S15 |
| *Calamus* | *disjunctus* (*concolor*) | Baker 559 (K) | RBL216 |
| *Calamus* | *disjunctus* (*concolor*) | Dransfield 4352 (K) | RBL246 |
| *Calamus* | *distentus* | Baker 637 (K) | RBL018 |
| *Calamus* | *divaricatus* | Baker 556 (K) | BKL027 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *divergens* | Henderson 4315 (NY) | RBL143 |
| *Calamus* | *dongnaiensis* | Henderson 3714 (NY) | RBL103 |
| *Calamus* | *draco* | Boeea 7980 (K) | BKL181 |
| *Calamus* | *dracunculus* (*siberutensis*) | Afriastini 2560 (BO) | BKL170 |
| *Calamus* | *dracunculus* | Boden-Kloss 10585 (K) | BKL233 |
| *Calamus* | *dumetosus* | Baker 530 (K) | 14987-S34 |
| *Calamus* | *dumetosus* | Baker 530 (K) | RBL215 |
| *Calamus* | *egregius* (*austroguangxiensis*) | Tsang 22299 (A) | BKL164 |
| *Calamus* | *egregius* | Gilbert 201 (MO) | RBL363 |
| *Calamus* | *elegans* (*viridispinus*) | Haniff 8124 (SING) | RBL475 |
| *Calamus* | *erectus* | Baker 814 (K) | BKL037 |
| *Calamus* | *erinaceus* subsp. *daemonoropoides* (*daemonoropoides*) | Fernando 386 (K) | RBL283 |
| *Calamus* | *erinaceus* subsp. *erinaceus* (*erinaceus*) | Baker 554 (K) | BKL005 |
| *Calamus* | *erioacanthus* (*endauensis*) | Dransfield 5084 (K) | BKL234 |
| *Calamus* | *erioacanthus* (*hepburnii*) | Dewol & Lee SAN 132453 (K) | RBL421 |
| *Calamus* | *erioacanthus* (*pseudoulur*) | Petoe 42 (K) | RBL200 |
| *Calamus* | *erioacanthus* (*semoi*) | Dransfield 7117 (K) | BKL229 |
| *Calamus* | *erioacanthus* (*ulur*) | Dransfield 5041 (K) | BKL217 |
| *Calamus* | *erioacanthus* | Baker 511 (K) | RBL005 |
| *Calamus* | *erythrocarpus* | Zieck NGF 36176 (L) | RBL327 |
| *Calamus* | *eugenei* (*brevicaulis*) | Henderson 3693 (NY) | RBL233 |
| *Calamus* | *evansii* | Khamphone 375 (K) | BKL119 |
| *Calamus* | *exiguus* (*hirsutus*) | Dransfield 5153 (K) | BKL148 |
| *Calamus* | *exilis* (*hispidulus*) | Kuhnhäuser 80 (K) | RBL197 |
| *Calamus* | *exilis* | Henderson 3998 (K) | RBL123 |
| *Calamus* | *fertilis* | Baker 648 (K) | BKL046 |
| *Calamus* | *filipendulus* | Dransfield 4520 (K) | BKL120 |
| *Calamus* | *filispadix* | Adorador 75 (LBC) | BKL095 |
| *Calamus* | *fissilis* | Henderson 3289 (K) | BKL064 |
| *Calamus* | *flabellatus* | Baker 518 (K) | BKL038 |
| *Calamus* | *flagellum* | Oulathong 200 (K) | RBL023 |
| *Calamus* | *floribundus* | Hooker & Thompson s.n. (K) | BKL121 |
| *Calamus* | *formicarius* | Kuhnhäuser 69 (K) | RBL193 |
| *Calamus* | *formosanus* | Jeanson 87b (NY) | BKL052 |
| *Calamus* | *gaharuensis* (*nematospadix*) | Dransfield 6100 (K) | RBL439 |
| *Calamus* | *gamblei* | Gamble 14405 (K) | BKL127 |
| *Calamus* | *geniculatus* (*collarifer*) | Kuhnhäuser 72 (K) | RBL195 |
| *Calamus* | *geniculatus* | Trichon 797 (K) | BKL235 |
| *Calamus* | *gibbsianus* | Salick, Alim & Benidict 9021 (K) | RBL252 |
| *Calamus* | *glaucescens* | Dransfield 1104 (K) | BKL187 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *godefroyi* | Evans 153 (K) | BKL128 |
| *Calamus* | *gonospermus* | Henderson 4057 (K) | RBL134 |
| *Calamus* | *goramensis* | Chai 36136 (BH) | RBL445 |
| *Calamus* | *gracilipes* (*didymophyllus*) | Baker 692 (K) | 17182-S25 |
| *Calamus* | *gracilipes* (*didymophyllus*) | Baker 692 (K) | RBL219 |
| *Calamus* | *gracilipes* | Dransfield 3578 (K) | BKL129 |
| *Calamus* | *gracilis* subsp. *vietnamensis* (*gracilis*) | Henderson 3546 (NY) | RBL092 |
| *Calamus* | *griseus* | Kerr 11760 (K) | BKL237 |
| *Calamus* | *guruba* | Parker 2483 (K) | BKL238 |
| *Calamus* | aff. *hallierianus* | Peters s.n. (K) | RBL366 |
| *Calamus* | *hallierianus* (*georgei*) | Dransfield 5198 (K) | BKL236 |
| *Calamus* | *hallierianus* | Kuhnhäuser 37 (K) | RBL179 |
| *Calamus* | *harmandii* | Khamphone 398 (K) | BKL025 |
| *Calamus* | *heatubunii* | Baker 1392 (K) | BKL058 |
| *Calamus* | *helferianus* (*oxleyanus*) | Henderson 3974 (NY) | RBL087 |
| *Calamus* | *helferianus* (*viridispinus*) | Dransfield 5393 (K) | RBL320 |
| *Calamus* | *henryanus* | Henderson 3599 (NY) | RBL073 |
| *Calamus* | *heteracanthopsis* (*heteracanthus*) | Dransfield 7575 (K) | BKL039 |
| *Calamus* | *heteroideus* | Dransfield 950 (L) | RBL435 |
| *Calamus* | *hirsutus* subsp. *hirsutus* (*hirsutus*) | Rajasegar 22 (K) | RBL064 |
| *Calamus* | *hirsutus* subsp. *korthalsii* (*korthalsii*) | van Valkenburg 1011 (K) | RBL364 |
| *Calamus* | *holttumii* | Dransfield 5180 (K) | BKL130 |
| *Calamus* | *hookerianus* | Menon 2503 (K) | RBL400 |
| *Calamus* | *horrens* (*tenuis*) | Henderson 3966 (NY) | RBL120 |
| *Calamus* | *hosensis* (*myriacanthus*) | Unyong S. 21213 (K) | BKL239 |
| *Calamus* | *hukaungensis* | Henderson 3174 (NY) | BKL152 |
| *Calamus* | *impressus* | Henderson 4368 (NY) | RBL151 |
| *Calamus* | *inermis* (*palustris*) | Khamphone 403 (K) | RBL060 |
| *Calamus* | *inermis* (*platyacanthoides*) | Henderson 3793 (NY) | RBL106 |
| *Calamus* | *inermis* (*simplicifolius*) | Chen 10883 (MO) | RBL375 |
| *Calamus* | *inermis* (*wailong*) | Oulathong 213 (K) | RBL021 |
| *Calamus* | *ingens* | Eiserhardt 155 (K) | RBL344 |
| *Calamus* | aff. *inops* | Henderson 4382 (NY) | BKL051 |
| *Calamus* | *inops* (*orthostachyus*) | Rustiami 446 (K) | BKL114 |
| *Calamus* | *inops* | Henderson 4274 (NY) | BKL034 |
| *Calamus* | *insignis* | Rajasegar 23 (K) | RBL025 |
| *Calamus* | *insularis* | Bidault 1095 (MO) | RBL374 |
| *Calamus* | *interruptus* | Maturbongs 281 (K) | RBL401 |
| *Calamus* | *jacobsii* | Jacobs 9561 (L) | BKL219 |
| *Calamus* | *javensis* (*acuminatus*) | Baker 551 (K) | RBL003 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *javensis* (*amplijugus*) | Lomudin Tadong 197 (K) | BKL206 |
| *Calamus* | *javensis* (*corrugatus*) | Kuhnhäuser 41 (K) | RBL181 |
| *Calamus* | *javensis* (*elopurensis*) | Sani Sambuling 87 (K) | RBL250 |
| *Calamus* | *javensis* (*hypertrichosus*) | Dransfield 4637 (K) | BKL209 |
| *Calamus* | *javensis* (*impar*) | Hallier 2033 (FI) | RBL472 |
| *Calamus* | *javensis* | Baker 542 (KEP) | RBL010 |
| *Calamus* | *johanis* | Mogea 4418 (K) | BKL240 |
| *Calamus* | *johndransfieldii* | Dransfield 2757 (BO) | RBL446 |
| *Calamus* | *johnsii* | Johns 11279 (K) | RBL402 |
| *Calamus* | *kampucheaensis* | Henderson 3842 (NY) | RBL111 |
| *Calamus* | *kandariensis* (*paucijugus*) | Boni 193 (FI) | RBL370 |
| *Calamus* | *kandariensis* (*paucijugus*) | ex Heynes 6 (BO) | RBL466 |
| *Calamus* | *kandariensis* | Henderson 4230 (NY) | RBL339 |
| *Calamus* | *karuensis* | Sangaji and Barrow 31 (K) | RBL354 |
| *Calamus* | *katikii* | Katik LAE 74954 (LAE) | RBL469 |
| *Calamus* | *kebariensis* | Maturbongs 75 (K) | RBL403 |
| *Calamus* | *kinabaluensis* | Tadong 520 (K) | BKL132 |
| *Calamus* | *kingianus* (*spicatus*) | Henderson 3138 (NY) | BKL158 |
| *Calamus* | *kingianus* | Khamphone 367 (K) | BKL133 |
| *Calamus* | *kjellbergii* | Kjellberg 2467 (S) | RBL474 |
| *Calamus* | aff. *klossii* | Dransfield 7698 (K) | RBL056 |
| *Calamus* | *klossii* | Baker 840 (K) | BKL036 |
| *Calamus* | *kontumensis* | Nuraliev 1544 (NY) | BKL153 |
| *Calamus* | *koordersianus* | Kramadibrata s.n. (K) | BKL031 |
| *Calamus* | *kostermansii* | Baker 848 (K) | RBL045 |
| *Calamus* | *kubahensis* | Kuhnhäuser 11 (K) | RBL168 |
| *Calamus* | *kunstleri* subsp. *kunstleri* (*kunstleri*) | Baker 496 (K) | RBL012 |
| *Calamus* | *kunstleri* subsp. *microstachys* (*microstachys*) | Kuhnhäuser 21 (K) | RBL173 |
| *Calamus* | *lambirensis* | Kuhnhäuser 79 (K) | RBL196 |
| *Calamus* | *lamprolepis* (*takanensis*) | Kramadibrata 28 (BO) | RBL468 |
| *Calamus* | *lamprolepis* | Henderson 4310 (NY) | RBL221 |
| *Calamus* | *lamprolepis* | Heyne 390 (FI) | RBL381 |
| *Calamus* | *laoensis* | Henderson 3834 (NY) | RBL108 |
| *Calamus* | *lateralis* | Henderson 3392 (NY) | RBL091 |
| *Calamus* | *latifolius* | Henderson 3613 (NY) | RBL095 |
| *Calamus* | *lauterbachii* | Baker 645 (K) | BKL008 |
| *Calamus* | *laxissimus* | Dransfield 5072 (K) | RBL404 |
| *Calamus* | *leiocaulis* | Anggana & Dali 25 (K) | RBL253 |
| *Calamus* | *leloi* | Petoe 65 (K) | RBL342 |
| *Calamus* | *lengguanii* | Saw FRI 48330 (KEP) | RBL414 |
| *Calamus* | *leptopus* | Rajasegar 10 (K) | RBL062 |
| *Calamus* | *leptospadix* | Griffith s.n. (K) | BKL189 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *leptostachys* | Henderson 4312 (NY) | RBL142 |
| *Calamus* | *lobatus* | Henderson 4279 (NY) | RBL208 |
| *Calamus* | *lobbianus* | Kuhnhäuser 12 (K) | RBL169 |
| *Calamus* | *loherianus* (*oligolepis*) | Elmer 11757 (K) | RBL431 |
| *Calamus* | *loherianus* (*pannosus*) | Elmer 11600 (K) | RBL310 |
| *Calamus* | *loherianus* (*urdanetanus*) | Adorador 77 (LBC) | BKL097 |
| *Calamus* | *loherianus* | Baja-Lapis 82 (K) | RBL288 |
| *Calamus* | *longibracteatus* | Kuhnhäuser 20 (K) | RBL172 |
| *Calamus* | *longipes* (*pumilus*) | van Valkenburg 1396 (L) | BKL267 |
| *Calamus* | *longipes* (*sabensis*) | Ridley 9038 (K) | RBL394 |
| *Calamus* | *longipinna* | Cultivated 1964-27301 (K) | BKL009 |
| *Calamus* | *longisetus* | Henderson 3963 (NY) | RBL119 |
| *Calamus* | *longispathus* | Baker 493 (K) | RBL004 |
| *Calamus* | *longiusculus* (*insignis var. longispinosus*) | Hodel 1615 (BH) | BKL197 |
| *Calamus* | *lucysmithiae* | Zieck NGF 36554 (K) | RBL289 |
| *Calamus* | *luridus* | Gianno E264 (K) | RBL428 |
| *Calamus* | *macrochlamys* | Baker 1124 (K) | RBL047 |
| *Calamus* | *macrorhynchus* | Merrill 11118 (NY) | BKL154 |
| *Calamus* | *macrosphaerion* | Henderson 4271 (NY) | BKL053 |
| *Calamus* | *maculatus* | Kuhnhäuser 39 (K) | RBL180 |
| *Calamus* | *maiadum* | Dransfield 7027 (K) | RBL255 |
| *Calamus* | *malawaliensis* | Philip & Dewol SAN 89981 (K) | RBL481 |
| *Calamus* | *manan* | Baker 561 (K) | BKL006 |
| *Calamus* | *manglaensis* | Henderson 3879 (NY) | BKL021 |
| *Calamus* | *manillensis* | Elmer 14011 (K) | RBL291 |
| *Calamus* | aff. *marginatus* | Anggana & Dali 52 (K)* | RBL396 |
| *Calamus* | *marginatus* | Duaneh 128 (K) | RBL256 |
| *Calamus* | *maturbongsii* | Maturbongs 32 (K) | BKL180 |
| *Calamus* | *megaphyllus* | Fernando 427 (K) | RBL292 |
| *Calamus* | *meghalayensis* | Koelz 29229 (BH) | BKL166 |
| *Calamus* | *melanacanthus* (*mellitus*) | Henderson 3840 (NY) | RBL110 |
| *Calamus* | *melanacanthus* | Hodel 1957 (NY) | BKL155 |
| *Calamus* | *melanochaetes* (*binnendijkii*) | Furtado 30832 (K) | RBL274 |
| *Calamus* | *melanochaetes* (*manii*) | King's collector s.n. (K) | BKL213 |
| *Calamus* | *melanochaetes* (*treubianus*) | Treub s.n. (FI) | RBL448 |
| *Calamus* | *melanochaetes* (*fissus*) | Baker 546 (K) | BKL082 |
| *Calamus* | *melanochaetes* (*pachyrostris*) | Hallier 1310 (BO) | RBL465 |
| *Calamus* | *melanochaetes* (*grandis*) | Rajasegar 29 (K) | RBL065 |
| *Calamus* | *melanochaetes* (*lewisianus*) | Vongkaluang 430 (K) | RBL427 |
| *Calamus* | *melanochaetes* (*monticola*) | Dransfield 5409 (K) | RBL297 |
| *Calamus* | *melanochaetes* (*melanochaetes*) | Henderson 3992 (NY) | RBL238 |
| *Calamus* | *melanochaetes* (*melanochaetes*) | Noblick 5199 (K) | RBL365 |
| *Calamus* | *melanochaetes* (*jenkinsianus*) | Oulathong 202 (K) | RBL022 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *melanochaetes* (*sepal*) | Dransfield 5395 (K) | BKL144 |
| *Calamus* | *melanochaetes* (*singalanus*) | Sangaji 16 (K) | BKL147 |
| *Calamus* | *melanochaetes* (*stenophyllus*) | Sangaji 14 (K) | BKL151 |
| *Calamus* | *melanochaetes* (*kurzianus*) | Henderson 3981 (NY) | RBL121 |
| *Calamus* | *melanoloma* | Watanabe 93 (K) | RBL293 |
| *Calamus* | *melanorhynchus* | Elmer 11708 (K) | RBL294 |
| *Calamus* | *metzianus* | Thwaites 3914 (K) | RBL295 |
| *Calamus* | *micracanthus* | Kuhnhäuser 71 (K) | RBL194 |
| *Calamus* | *micranthus* (*speciosissimus*) | Sutisna 87 (K) | BKL149 |
| *Calamus* | *microsphaerion* | Stone 302 (K) | RBL259 |
| *Calamus* | *minahassae* | Musser M11 (K) | RBL260 |
| *Calamus* | aff. *mindorensis* | Adorador 76 (LBC) | BKL096 |
| *Calamus* | *minor* | Khamphone 319 (K) | RBL058 |
| *Calamus* | *minor* | Khamphone 319 (K) | RBL296 |
| *Calamus* | *minutus* | Dransfield 5151 (K) | RBL429 |
| *Calamus* | *mirabilis* (*polystachys*) | Dransfield 4587 (K) | RBL406 |
| *Calamus* | *mitis* (*elmerianus*) | Adorador 86 (LBC) | BKL101 |
| *Calamus* | *mitis* | Barbon 1611 (K) | BKL214 |
| *Calamus* | *modestus* | Henderson 3589 (NY) | RBL072 |
| *Calamus* | *mogeae* | Dransfield 2025 (L) | BKL070 |
| *Calamus* | *mollispinus* | Henderson 3931 (NY) | RBL116 |
| *Calamus* | *moorei* | Dransfield 6102 (K) | RBL263 |
| *Calamus* | *moorhousei* | Moorhouse s.n. (SING) | RBL476 |
| *Calamus* | *moseleyanus* (*grandifolius*) | Madulid 5107 (K) | BKL188 |
| *Calamus* | *moseleyanus* (*jenningsianus*) | Coode 5758 (K) | BKL215 |
| *Calamus* | *moseleyanus* (*multinervis*) | Adorador 71 (LBC) | BKL093 |
| *Calamus* | *moseleyanus* (*reyesianus*) | Jacobs 7774 (K) | BKL141 |
| *Calamus* | *moseleyanus* (*spinifolius*) | Fernando 342 (K) | BKL150 |
| *Calamus* | *moseleyanus* (*subinermis*) | Baker 545 (K) | BKL018 |
| *Calamus* | *moseleyanus* (*viridissimus*) | Elmer 11938 (K) | RBL434 |
| *Calamus* | *moseleyanus* | Moseley s.n. (K) | RBL264 |
| *Calamus* | *moti* | Dowe 794 (MBC) | RBL153 |
| *Calamus* | *muelleri* | Moore 9230 (BH) | BKL167 |
| *Calamus* | *multispicatus* (*henryanus*) | Guo Lixiu 165 (NY) | BKL156 |
| *Calamus* | *muricatus* | Kuhnhäuser 14 (K) | RBL170 |
| *Calamus* | aff. *myriacanthus* | Kuhnhäuser 45 (K) | RBL182 |
| *Calamus* | *myriacanthus* | Kuhnhäuser 23 (K) | RBL175 |
| *Calamus* | *myrianthus* | Barfod 45254 (K) | RBL298 |
| *Calamus* | *nanduensis* (*pintaudii*) | Pintaud 671 (K) | RBL357 |
| *Calamus* | *nanduensis* | Baker 678 (K) | RBL085 |
| *Calamus* | *nannostachys* | Zieck NGF 36544 (K) | RBL299 |
| *Calamus* | *nanodendron* | Baker 720 (K) | 15014-S73 |
| *Calamus* | *nanodendron* | Baker 720 (K) | RBL217 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *nematospadix* | Bingshan S 73908 (K) | BKL182 |
| *Calamus* | *nicobaricus* | Rogers 49 (K) | RBL300 |
| *Calamus* | *nielsenii* | Nielsen 761 (K) | BKL112 |
| *Calamus* | *nitidus* (*guruba*) | Hodel 1948 (NY) | BKL157 |
| *Calamus* | *notabilis* | Milliken 1046X (K) | RBL301 |
| *Calamus* | *novae-georgii* | Waterhouse 22 (K) | BKL066 |
| *Calamus* | *nuichuaensis* | Henderson 3646 (NY) | RBL100 |
| *Calamus* | *nuralievii* | Nuraliev 1658 (NY) | BKL194 |
| *Calamus* | *obiensis* | de Vogel 4118 (MO) | RBL470 |
| *Calamus* | *oblatus* | Dransfield 5333 (K) | BKL225 |
| *Calamus* | *oblongus* subsp. *depressiusculus* (*depressiusculus*) | Dransfield 4135 (K) | RBL286 |
| *Calamus* | *oblongus* subsp. *elongatus* (*elongatus*) | Stockdale 339 (K) | RBL249 |
| *Calamus* | *oblongus* subsp. *gracilis* (*flexilis*) | Ridsdale SMHI 282 (K) | BKL185 |
| *Calamus* | *oblongus* subsp. *longistipes* (*longistipes*) | Duaneh 136 (K) | RBL254 |
| *Calamus* | *oblongus* subsp. *mollis* (*affinis*) | Elmer 13978 (K) | RBL266 |
| *Calamus* | *oblongus* subsp. *mollis* (*clemensianus*) | Clemens 1227 (FI) | RBL368 |
| *Calamus* | *oblongus* subsp. *mollis* (*curranii*) | Fernando 640 (K) | RBL281 |
| *Calamus* | *oblongus* subsp. *mollis* (*mollis*) | Adorador 70 (LBC) | RBL244 |
| *Calamus* | *oblongus* subsp. *mollis* (*riedelianus*) | Dransfield 3787 (K) | RBL393 |
| *Calamus* | *oblongus* subsp. *oblongus* (*oblongus*) | Watanabe 49 (K) | RBL302 |
| *Calamus* | *oblongus* subsp. *sedisspirituum* (*sedisspirituum*) | Dransfield 4172 (K) | BKL193 |
| *Calamus* | *oblongus* subsp. *similaris* (*korthalsii*) | Kuhnhäuser 62 (K) | RBL189 |
| *Calamus* | *occidentalis* | Dransfield 1483 (K) | RBL430 |
| *Calamus* | *ochrolepis* | Adorador 94 (LBC) | BKL102 |
| *Calamus* | *ocreatus* | Henderson 3862 (NY) | RBL113 |
| *Calamus* | *oligostachys* | Khamphone 324 (K) | RBL305 |
| *Calamus* | *optimus* | Kuhnhäuser 59 (K) | RBL223 |
| *Calamus* | *optimus* | Watanabe 40 (K) | RBL390 |
| *Calamus* | *oresbiopsis* | Zieck NGF 36520 (K) | RBL306 |
| *Calamus* | *oresbius* | Baker 1131 (K) | RBL048 |
| *Calamus* | *ornatus* | Dransfield 7628 (KEP) | BKL092 |
| *Calamus* | *ovoideus* | Thwaites s.n. (K) | RBL432 |
| *Calamus* | *oxleyanus* | Rajasegar 20 (K) | BKL033 |
| *Calamus* | *oxleyoides* | Kuhnhäuser 24 (K) | RBL222 |
| *Calamus* | *oxycarpus* | Henry 10475 (K) | RBL433 |
| *Calamus* | *oxycoccus* | Baker 523 (K) | BKL004 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *pachystemonus* | Thwaites 2334 (K) | RBL389 |
| *Calamus* | *padangensis* | Kiah bin Hj. 33376 (SING) | RBL450 |
| *Calamus* | *pahangensis* | Saw FRI 44659 (K) | RBL309 |
| *Calamus* | *pandanosmus* | Ambriansyah & Arifin 882 (K) | BKL115 |
| *Calamus* | *papuanus* | Baker 1382 (K) | RBL050 |
| *Calamus* | *parutan* | Dransfield 3534 (K) | BKL134 |
| *Calamus* | *parutan* | Dransfield 3534 (L) | RBL325 |
| *Calamus* | *parvulus* | Henderson 3860 (NY) | RBL112 |
| *Calamus* | *paspalanthus* | Baker 693 (K) | BKL079 |
| *Calamus* | *paulii* | Chai S 36834 (K) | BKL226 |
| *Calamus* | *pedicellaris* | Siebert S8 (K) | BKL135 |
| *Calamus* | *pedicellatus* | Henderson 4361 (NY) | RBL031 |
| *Calamus* | *penicillatus* | Burkill 750 (K) | BKL241 |
| *Calamus* | *perakensis* subsp. *niger* (*perakensis*) | Dransfield 5074 (K) | BKL242 |
| *Calamus* | *peregrinus* | Henderson 3959 (NY) | BKL057 |
| *Calamus* | *periacanthus* | Baker 527 (K) | BKL040 |
| *Calamus* | *pertenuis* | de Vogel 7595 (K) | RBL405 |
| *Calamus* | *pholidostachys* | Kairo 109 (K) | BKL191 |
| *Calamus* | *pilosellus* | Kuhnhäuser 52 (K) | RBL184 |
| *Calamus* | *pilosissimus* | Baker 653 (K) | RBL019 |
| *Calamus* | *platyspathus* | Henderson 3957 (NY) | RBL118 |
| *Calamus* | *plicatus* (*fimbriatus*) | Ambri & Arifin 409 (K) | RBL251 |
| *Calamus* | *plicatus* (*laevigatus*) | Baker 499 (K) | BKL017 |
| *Calamus* | *plicatus* (*mattanensis*) | Kuhnhäuser 5 (K) | RBL165 |
| *Calamus* | *poensis* | Awa & Paie S 47340 (K) | RBL437 |
| *Calamus* | *pogonacanthus* | Baker 522 (K) | BKL048 |
| *Calamus* | *pogonotium* | Kuhnhäuser 58 (K) | RBL226 |
| *Calamus* | *poilanei* | Henderson 4032 (NY) | RBL132 |
| *Calamus* | *politus* | Fernando 719 (K) | BKL192 |
| *Calamus* | *polycladus* | Baker 680 (K) | RBL163 |
| *Calamus* | *posoanus* | Henderson 4298 (NY) | RBL141 |
| *Calamus* | *powlingii* | Powling 30+31 (K) | BKL136 |
| *Calamus* | *praetermissus* | Baker 557 (K) | RBL014 |
| *Calamus* | *propinquus* (*brachypodus*) | Trichon 802 (K) | RBL275 |
| *Calamus* | *pseudoconcolor* | Cultivated 1975-3398 (K) | RBL081 |
| *Calamus* | *pseudomollis* | Henderson 4317 (NY) | RBL144 |
| *Calamus* | *pseudotenuis* subsp. *karnatakensis* (*karnatakensis*) | Ramamoorthy 1442 (US) | RBL379 |
| *Calamus* | *pseudotenuis* subsp. *pseudotenuis* (*pseudotenuis*) | Thwaites 2335 (K) | BKL137 |
| *Calamus* | *pseudozebrinus* | Brass 3923 (NY) | RBL367 |
| *Calamus* | *psilocladus* | Dransfield 6137 (K) | RBL438 |
| *Calamus* | *pycnocarpus* | Sinclair & Saleh 40844 (K) | BKL138 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *pygmaeus* | Baker 510 (K) | RBL011 |
| *Calamus* | *quangngaiensis* | Henderson 3811 (NY) | RBL234 |
| *Calamus* | *radiatus* | de Zoysa 3 (K) | BKL176 |
| *Calamus* | *radicalis* | Clemens s.n. (BH) | BKL168 |
| *Calamus* | *radulosus* | King's collector 8171 (K) | BKL139 |
| *Calamus* | *reinwardtii* | Blume 62 (L) | BKL220 |
| *Calamus* | *reticulatus* | Baker 644 (K) | BKL007 |
| *Calamus* | *retroflexus* | Baker 1396 (K) | RBL053 |
| *Calamus* | aff. *reyesianus* | Adorador 79 (LBC) | BKL098 |
| *Calamus* | aff. *rhabdocladus* | Wang & Green 2093 (K)* | RBL315 |
| *Calamus* | *rhabdocladus* | Henderson 3763 (NY) | RBL032 |
| *Calamus* | *rheedei* | Unknown 291 (K) | BKL065 |
| *Calamus* | *rhomboideus* (*blumei*) | Baker 565 (K) | RBL084 |
| *Calamus* | *rhomboideus* | Dransfield 4007 (K) | BKL227 |
| *Calamus* | *rhytidomus* | van Valkenburg 1186 (K) | RBL392 |
| *Calamus* | *ridleyanus* | Rajasegar 9 (K) | RBL061 |
| *Calamus* | *robinsonianus* | Robinson 1613 (K) | BKL177 |
| *Calamus* | *rotang* | Unknown s.n. (K) | BKL067 |
| *Calamus* | *ruber* | Noblick 5178 (K) | BKL142 |
| *Calamus* | *rudentum* | Henderson 3619 (NY) | RBL096 |
| *Calamus* | *rugosus* | Dransfield 5220 (K) | RBL358 |
| *Calamus* | *rumphii* | Robinson 24 (US) | RBL452 |
| *Calamus* | *ruptilis* | Baker 555 (K) | RBL013 |
| *Calamus* | *ruptiloides* | Kuhnhäuser 75 (K) | RBL224 |
| *Calamus* | *ruvidus* (*tanakadatei*) | Dransfield 5032 (K) | BKL216 |
| *Calamus* | *ruvidus* | Kuhnhäuser 57 (K) | RBL187 |
| *Calamus* | *sabalensis* | Kuhnhäuser 31 (K) | RBL176 |
| *Calamus* | *salicifolius* | Henderson 3835 (NY) | RBL109 |
| *Calamus* | *samian* | Elmer 11336 (NY) | RBL408 |
| *Calamus* | *sandsii* | Henderson 4354 (NY) | RBL148 |
| *Calamus* | *sashae* | Dransfield 7601 (K) | BKL143 |
| *Calamus* | *scabridulus* | Sutisna 69 (K) | BKL228 |
| *Calamus* | *scabrispathus* (*papyraceus*) | Pullen 1506 (CANB) | RBL362 |
| *Calamus* | aff. *scapiger* | Henderson 4063 (K)* | RBL135 |
| *Calamus* | *scapiger* | Kuhnhäuser 18 (K) | RBL171 |
| *Calamus* | *schistoacanthus* | Giesen s.n. (K) | RBL355 |
| *Calamus* | *schlechterianus* | Baker 1119 (K) | RBL046 |
| *Calamus* | *scipionum* | Baker 549 (K) | RBL083 |
| *Calamus* | *scleracanthus* (*toli-toliensis*) | Heyne 2606 (FI) | BKL255 |
| *Calamus* | *scleracanthus* | Henderson 4235 (NY) | RBL088 |
| *Calamus* | *sedens* | Baker 494 (K) | BKL086 |
| *Calamus* | *senalingensis* | Dransfield 4991 (K) | RBL407 |
| *Calamus* | *septimus* | Jacobs 4470 (SING) | RBL451 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *seriatus* | Henderson 3864 (NY) | RBL114 |
| *Calamus* | *seropakensis* | Banyeng 43662 (K) | BKL243 |
| *Calamus* | *serpentinus* | Dransfield 5790 (K) | RBL461 |
| *Calamus* | *serrulatus* | Baker 628 (K) | RBL162 |
| *Calamus* | *sessiliflorus* | Dransfield 2730 (BO) | RBL467 |
| *Calamus* | *simplex* | Dransfield 5238 (K) | BKL146 |
| *Calamus* | *siphonospathus* (*dimorphacanthus*) | Reynoso 3593 (K) | RBL287 |
| *Calamus* | *siphonospathus* | Henderson 4006 (K) | RBL127 |
| *Calamus* | *smitinandii* (*bousigonii* subsp. *smitinandii*) | Henderson 4003 (K) | RBL125 |
| *Calamus* | *sordidus* | Baker 541 (K) | BKL030 |
| *Calamus* | sp. | Baker 529 (K)* | BKL032 |
| *Calamus* | sp. | Baja-Lapis 123 (K [silica])* | BKL041 |
| *Calamus* | sp. | Baja-Lapis 75 (K)* | BKL140 |
| *Calamus* | sp. | Siebert 30 (K)* | RBL030 |
| *Calamus* | sp. | Rajasegar 15 (K)* | RBL063 |
| *Calamus* | sp. | Henderson 4020 (K)* | RBL130 |
| *Calamus* | sp. | Henderson 4011 (K)* | RBL239 |
| *Calamus* | *spanostachys* | v. Leeuwen 10473 (K) | BKL244 |
| *Calamus* | *sparsiflorus* | Kuhnhäuser 51 (K) | RBL183 |
| *Calamus* | *spiculifer* | Brass 13341 (L) | RBL436 |
| *Calamus* | *spinosus* | Henderson 4353 (NY) | BKL049 |
| *Calamus* | *spinosus* | Henderson 4353 (NY) | RBL147 |
| *Calamus* | *spinulinervis* (*nigricans*) | Ambri & Arifin 475 (K) | BKL113 |
| *Calamus* | *spiralis* | Henderson 3883 (NY) | BKL050 |
| *Calamus* | *suaveolens* | Henderson 4333 (NY) | RBL146 |
| *Calamus* | *subangulatus* | Baker 500 (K) | BKL043 |
| *Calamus* | *subsolanus* | Dransfield 3577 (K) | RBL460 |
| *Calamus* | *sulawesiensis* (*mindorensis*) | Henderson 4402 (NY) | RBL152 |
| *Calamus* | *sulawesiensis* (*mindorensis*) | Musser R2 (K) | RBL261 |
| *Calamus* | *sumatranus* | Dransfield 3998 (K) | BKL265 |
| *Calamus* | *superciliatus* | Baker 1385 (K) | RBL090 |
| *Calamus* | *susukensis* | Dransfield 4754 (K) | RBL312 |
| *Calamus* | *symphysipus* | Henderson 4234 (NY) | RBL089 |
| *Calamus* | *tadulakoensis* | Henderson 4256 (NY) | RBL206 |
| *Calamus* | *tambingensis* | Henderson 4276 (NY) | BKL159 |
| *Calamus* | *tapanensis* | Dransfield 4078 (K) | RBL313 |
| *Calamus* | *temii* | Smitinand s.n. (K) | RBL314 |
| *Calamus* | *tenompokensis* | Soibeh 353 (K) | BKL116 |
| *Calamus* | *tenuis* | Harsakh 22851 (K) | BKL183 |
| *Calamus* | *tetradactyloides* subsp. *bachmaensis* (*bachmaensis*) | Henderson 4043 (NY) | RBL338 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *tetradactylus* (*batoensis*) | Henderson 3881 (K) | RBL237 |
| *Calamus* | *tetradactylus* (*crispus*) | Henderson 3273 (K) | RBL232 |
| *Calamus* | *tetradactylus* (*flavinervis*) | Henderson 3869 (NY) | RBL236 |
| *Calamus* | *tetradactylus* (*solitarius*) | Khamphone 317 (K) | RBL057 |
| *Calamus* | *tetradactylus* | Henderson 3605 (NY) | RBL093 |
| *Calamus* | *thwaitesii* | Read 2161 (K) | RBL316 |
| *Calamus* | *thysanolepis* | Shui Ying Hu 12415 (K) | BKL160 |
| *Calamus* | *tomentosus* | van Valkenburg 1110 (K) | BKL231 |
| *Calamus* | *trachycoleus* | Baker 560 (K) | BKL020 |
| *Calamus* | *trigynus* | Henderson 4364 (NY) | RBL150 |
| *Calamus* | *tumidus* | Baker 1006 (K) | BKL173 |
| *Calamus* | *unifarius* | Lütjeharms 4359 (L) | BKL268 |
| *Calamus* | *unijugus* | Dransfield 5894 (K) | BKL117 |
| *Calamus* | *ursinus* | Baker 517 (K) | BKL042 |
| *Calamus* | aff. *usitatus* | Adorador 83 (LBC) | BKL100 |
| *Calamus* | *usitatus* | Baker 564 (K) | BKL019 |
| *Calamus* | *validus* (*mogeanus*) | Mogea 1356 (K) | RBL262 |
| *Calamus* | *validus* | Musser R5 (K) | BKL118 |
| *Calamus* | *vattayila* | Renuka 4001 (K) | RBL317 |
| *Calamus* | *velutinus* | Henderson 3627 (NY) | RBL098 |
| *Calamus* | *verticillaris* | Cultivated 1991-1094 (K) | BKL012 |
| *Calamus* | *vestitus* | Maturbongs 294 (K) | RBL318 |
| *Calamus* | *vidalianus* | Adorador 81 (LBC) | BKL099 |
| *Calamus* | *viminalis* (*phuocbinhensis*) | Henderson 3853 (NY) | RBL235 |
| *Calamus* | *viminalis* (*siamensis*) | Khamphone 401 (K) | RBL059 |
| *Calamus* | *viminalis* | Cultivated 1992-3475 (K) | RBL016 |
| *Calamus* | *vinaceus* | Henderson 4359 (NY) | RBL149 |
| *Calamus* | *vinosus* | Elmer 14158 (K) | RBL319 |
| *Calamus* | *viridis* | Henderson 4272 (NY) | RBL207 |
| *Calamus* | *vitiensis* (*pachypus*) | Dransfield 7600 (K) | RBL308 |
| *Calamus* | aff. *walkeri* | Baker & Utteridge 13 (K) | 14967-S5 |
| *Calamus* | aff. *walkeri* | Baker & Utteridge 13 (K) | RBL212 |
| *Calamus* | *walkeri* (*faberi*) | Eberhardt 1485 (FI) | BKL254 |
| *Calamus* | *walkeri* | Baker & Utteridge 2 (K) | RBL007 |
| *Calamus* | *wanggaii* | Barrow 129 (K) | RBL322 |
| *Calamus* | *warayanus* | Adorador 22 (LBC) | RBL243 |
| *Calamus* | *warburgii* | Baker 578 (K) | BKL029 |
| *Calamus* | *wedaensis* | Bangun 889 (MO) | RBL471 |
| *Calamus* | *wightii* | Gamble 14315 (K) | BKL222 |
| *Calamus* | *womersleyi* | Moore 9285 (LAE) | BKL198 |
| *Calamus* | *wuliangshanensis* | Henderson 3116 (NY) | BKL161 |
| *Calamus* | *yentuensis* | Henderson 3601 (NY) | RBL074 |
| *Calamus* | *zebrinus* | Baker 1375 (K) | RBL049 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Calamus* | *zeylanicus* | Henderson 4002 (K) | RBL124 |
| *Calamus* | *zieckii* | Baker 1393 (K) | RBL052 |
| *Calamus* | *zollingeri* subsp. *foxworthyi* (*foxworthyi*) | Dransfield 6190 (K) | BKL186 |
| *Calamus* | *zollingeri* subsp. *merrillii* (*merrillii*) | Baker 548 (K) | RBL082 |
| *Calamus* | *zollingeri* subsp. *zollingeri* (*zollingeri*) | Henderson 4318 (NY) | RBL145 |
| *Calamus* | *zonatus* subsp. *corneri* (*corneri*) | Corner 30058 (K) | RBL483 |
| *Calamus* | *zonatus* subsp. *zonatus* (*zonatus*) | Baker 538 (K) | RBL008 |
| *Eleiodoxa* | *conferta* | Dransfield 6514 (K) | MSL73-S5 |
| *Eremospatha* | *cuspidata* | Bruneau 1074 (K) | MSL33-S2 |
| *Eremospatha* | *dransfieldii* | Sunderland 2261 (K) | BKL071 |
| *Eremospatha* | *hookeri* | Sunderland 1801 (K) | RBL035 |
| *Eremospatha* | *laurentii* | Cultivated 1984-1058 (K) | BKL002 |
| *Eremospatha* | *macrocarpa* | Sunderland 1800 (K) | BKL059 |
| *Eremospatha* | *wendlandiana* | Dransfield 7004 (K) | BKL087 |
| *Eugeissona* | *ambigua* | Beccari 3343 (FI) | RBL473 |
| *Eugeissona* | *brachystachys* | Gianno 328 (K) | BKL178 |
| *Eugeissona* | *insignis* | Kuhnhäuser 22 (K) | RBL174 |
| *Eugeissona* | *minor* | Petoe 59 (K) | RBL341 |
| *Eugeissona* | *tristis* | Baker 501 (K) | MSL52-S20 |
| *Eugeissona* | *utilis* | Baker 712 (SAR) | BKL081 |
| *Korthalsia* | *angustifolia* | Giesen 89 (K) | BKL246 |
| *Korthalsia* | *bejaudii* | Bejaud s.n. (P) | RBL449 |
| *Korthalsia* | *brassii* | Brass 6864 (BM) | BKL252 |
| *Korthalsia* | *celebica* | Henderson 4270 (NY) | RBL138 |
| *Korthalsia* | *cheb* | Petoe 40 (K) | RBL199 |
| *Korthalsia* | *concolor* | Baker 562 (K) | RBL042 |
| *Korthalsia* | *debilis* | Kuhnhäuser 55 (K) | RBL225 |
| *Korthalsia* | *echinometra* | Baker 509 (K) | MSL51-S12 |
| *Korthalsia* | *ferox* | Baker 515 (K) | RBL040 |
| *Korthalsia* | *flagellaris* | Baker 702 (K) | RBL043 |
| *Korthalsia* | *furcata* | Dransfield 6075 (K) | BKL232 |
| *Korthalsia* | *furtadoana* | Baker 553 (K) | RBL041 |
| *Korthalsia* | *hispida* | Kuhnhäuser 67 (K) | RBL192 |
| *Korthalsia* | *jala* | Baker 558 (K) | BKL014 |
| *Korthalsia* | *junghuhnii* | Watanabe 48 (K) | BKL162 |
| *Korthalsia* | *laciniosa* | Henderson 3741 (NY) | RBL104 |
| *Korthalsia* | *lanceolata* | Baker 504 (KEP) | RBL160 |
| *Korthalsia* | *merrillii* | Fernando 638 (K) | BKL163 |
| *Korthalsia* | *minor* | Henderson 3632 (NY) | RBL099 |
| *Korthalsia* | *paucijuga* | Dransfield 4641 (K) | RBL398 |
| *Korthalsia* | *rigida* | Baker 498 (KEP) | RBL038 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Korthalsia* | *robusta* | Baker 552 (K) | BKL083 |
| *Korthalsia* | *rogersii* | Rogers 143 (K) | BKL247 |
| *Korthalsia* | *rostrata* | Kuhnhäuser 34 (K) | RBL177 |
| *Korthalsia* | *scaphigeroides* | Baja-Lapis 112 (K) | RBL409 |
| *Korthalsia* | *scortechinii* | Henderson 4014 (K) | RBL227 |
| *Korthalsia* | *tenuissima* | Baker 505 (KEP) | RBL039 |
| *Korthalsia* | *zippelii* | Baker 597 (K) | BKL023 |
| *Laccosperma* | *acutiflorum* | Sunderland 1764 (K) | RBL067 |
| *Laccosperma* | *laeve* | van Valkenburg 2629 (K) | RBL068 |
| *Laccosperma* | *opacum* | Sunderland 1750 (K) | MSL50 |
| *Laccosperma* | *robustum* | Dransfield 7006 (K) | BKL013 |
| *Laccosperma* | *secundiflorum* | Sunderland 1763 (K) | RBL066 |
| *Lepidocaryum* | *tenue* | Dransfield 7012 (K) | MSL75-S14 |
| *Mauritia* | *carana* | Traill 1103 (K) | BKL072 |
| *Mauritia* | *flexuosa* | Ely 17 (K) | MSL70 |
| *Mauritiella* | *aculeata* | Cultivated 1988-4331 (K) | BKL010 |
| *Mauritiella* | *armata* | Irwin 24627 (K) | BKL069 |
| *Mauritiella* | *armata* | Couvreur & Vargas 257 (K) | SBL558 |
| *Metroxylon* | *amicarum* | Henderson 4000 (K) | RBL209 |
| *Metroxylon* | *paulcoxii* | McClatchy 1882 (K) | BKL221 |
| *Metroxylon* | *sagu* | Baker 550 (K) | MSL36-S19 |
| *Metroxylon* | *salomonense* | Zona 651 (FTG) | BKL060 |
| *Metroxylon* | *upoluense* | Moore 9985 (K) | BKL195 |
| *Metroxylon* | *vitiense* | Seemann 658 (K) | BKL223 |
| *Metroxylon* | *warburgii* | Bennett s.n. (K) | BKL224 |
| *Myrialepis* | *paradoxa* | Baker 491 (K) | BKL015 |
| *Myrialepis* | *paradoxa* | Baker 491 (K) | MSL57 |
| *Oncocalamus* | *macrospathus* | van Valkenburg 2628 (K) | RBL036 |
| *Oncocalamus* | *mannii* | Sunderland 1768 (K) | BKL080 |
| *Oncocalamus* | *tuleyi* | Dransfield 7007 (K) | BKL016 |
| *Oncocalamus* | *tuleyi* | Dransfield 7007 (K) | MSL42 |
| *Pigafetta* | *elata* | Baker 508 (K) | BKL084 |
| *Pigafetta* | *filaris* | Dransfield 7610 (K) | MSL77-S28 |
| *Plectocomia* | *assamica* | Henderson 3133 (NY) | RBL376 |
| *Plectocomia* | *bractealis* | Falconer 36 (K) | BKL256 |
| *Plectocomia* | *dransfieldiana* | Dransfield 4518 (K) | RBL372 |
| *Plectocomia* | *elongata var. philippinensis* (*elongata*) | Dransfield 6200 (K) | 14926-S4 |
| *Plectocomia* | *elongata var. philippinensis* (*elongata*) | Dransfield 6200 (K) | RBL210 |
| *Plectocomia* | *himalayana* | Baker 815 (K) | BKL078 |
| *Plectocomia* | *khasyana* | Hooker & Thompson s.n. (K) | RBL373 |
| *Plectocomia* | *longistigma* | Dransfield 3540 (K) | BKL262 |
| *Plectocomia* | *macrostachya* | Unknown s.n. (K) | RBL413 |

**Supplementary Table 3.1** (*continued*)

| Genus | Species (Synonym) | Herbarium voucher | Seq. ID |
|---|---|---|---|
| *Plectocomia* | *microstachys* | Guo Lixiu 144 (NY) | RBL377 |
| *Plectocomia* | *mulleri* | Baker 563 (K) | MSL34 |
| *Plectocomia* | *mulleri* | Kuhnhäuser 9 (K) | RBL167 |
| *Plectocomia* | *pierreana* | Khamphone 377 (K) | RBL037 |
| *Plectocomia* | *pygmaea* | Mondi 75 (K) | BKL076 |
| Plectocomiinae | sp. nov. | Kuhnhäuser 35 (K) | RBL178 |
| *Plectocomiopsis* | *corneri* | Dransfield 5075 (K) | BKL205 |
| *Plectocomiopsis* | *geminiflora* | Baker 492 (K) | MSL74 |
| *Plectocomiopsis* | aff. *mira* | Kuhnhäuser 16 (K) | BKL204 |
| *Plectocomiopsis* | *mira* | Kuhnhäuser 44 (K) | RBL352 |
| *Plectocomiopsis* | *songthanhensis* | Henderson 3594 (NY) | RBL353 |
| *Plectocomiopsis* | *wrayi* | Dransfield 4940 (K) | BKL073 |
| *Raphia* | *farinifera* | Dransfield 7516 (K) | MSL20-S18 |
| *Raphia* | *hookeri* | Couvreur 984 (WAG) | BKL088 |
| *Raphia* | *monbuttorum* | Mogue 4 (WAG) | BKL090 |
| *Raphia* | *regalis* | Couvreur 753 (K) | BKL061 |
| *Raphia* | *sudanica* | Michon 56 (G) | BKL044 |
| *Raphia* | *textilis* | Couvreur 1075 (K) | BKL085 |
| *Salacca* | *acehensis* | Mogea 568 (K) | BKL248 |
| *Salacca* | *affinis* | Baker 708 (SAR) | BKL055 |
| *Salacca* | *bakeriana* | Baker 711 (K) | MSL43 |
| *Salacca* | *clemensiana* | Soibeh 613 (K) | RBL359 |
| *Salacca* | *dolicholepis* | Duaneh 124 (K) | BKL249 |
| *Salacca* | *dransfieldiana* | Kuhnhäuser 6 (K) | RBL166 |
| *Salacca* | *flabellata* | Corner 30525 (K) | RBL462 |
| *Salacca* | *glabrescens* | Cultivated 1984-3791 (K) | BKL091 |
| *Salacca* | *graciliflora* | Dransfield 5086 (K) | RBL441 |
| *Salacca* | *griffithii* | Smith 20 (K) | BKL074 |
| *Salacca* | *lophospatha* | Clemens 26380 (K) | BKL250 |
| *Salacca* | *magnifica* | Henderson 4015 (K) | RBL337 |
| *Salacca* | *minuta* | Dransfield 5079 (K) | BKL251 |
| *Salacca* | *multiflora* | Dransfield 5137 (K) | RBL410 |
| *Salacca* | *ramosiana* | Cultivated 1979-4409 (K) | BKL056 |
| *Salacca* | *rupicola* | Kuhnhäuser 64 (K) | RBL191 |
| *Salacca* | *sarawakensis* | Dransfield 5916 (K) | BKL075 |
| *Salacca* | *sarawakensis* | Kuhnhäuser 7 (K) | RBL241 |
| *Salacca* | *secunda* | Henderson 3176 (K) | BKL077 |
| *Salacca* | *stolonifera* | Hodel 1629 (BK) | RBL447 |
| *Salacca* | *sumatrana* | Toroes 5242 (K) | RBL442 |
| *Salacca* | *vermicularis* | Kuhnhäuser 56 (K) | RBL186 |
| *Salacca* | *wallichiana* | Henderson 3951A (NY) | RBL117 |
| *Salacca* | *zalacca* | Cultivated 1984-3376 (K) | BKL003 |

**Supplementary Table 3.2. Monophyly assessment of species represented by more than one specimen.** The species name includes information on infraspecific variation where applicable. If a species was synonymised during a recent taxonomic revision of *Calamus* (Henderson, 2020), the species name assigned to the specimen before the taxonomic revision is indicated in parentheses. Genes = number of gene trees in which a sample was represented, Monophyletic? = indication whether all samples belonging to a species formed a clade, Exclude? = indication which samples were removed in downstream analyses to reduce the sampling to one sample per species. Specimens with taxonomic uncertainty in the assignment to a species ("aff.") were not included in this analysis.

| Species | Seq. ID | Genes | Monophyletic? | Exclude? |
|---|---|---|---|---|
| *Calamus aruensis* | RBL231 | 335 | Yes | x |
| *Calamus aruensis* | RBL211 | 338 | Yes | |
| *Calamus balingensis* | RBL126 | 319 | Yes | x |
| *Calamus balingensis* (*setulosus*) | BKL145 | 225 | Yes | |
| *Calamus caesius* | BKL001 | 322 | No | |
| *Calamus caesius* (*tapa*) | RBL397 | 254 | No | x |
| *Calamus calapparius* (*beguinii*) | RBL272 | 167 | Yes | |
| *Calamus calapparius* (*komsaryi*) | BKL047 | 126 | Yes | x |
| *Calamus crinitus* subsp. *sabut* (*macrophyllus*) | RBL290 | 315 | Yes | x |
| *Calamus crinitus* subsp. *sabut* (*oligophyllus*) | RBL304 | 293 | Yes | x |
| *Calamus crinitus* subsp. *crinitus* (*sabut*) | RBL024 | 341 | Yes | |
| *Calamus deerratus* | RBL213 | 338 | Yes | x |
| *Calamus deerratus* | RBL242 | 339 | Yes | |
| *Calamus diepenhorstii* | BKL089 | 341 | Yes | |
| *Calamus diepenhorstii* | BKL245 | 197 | Yes | x |
| *Calamus egregius* | RBL363 | 305 | No | |
| *Calamus egregius* (*austroguangxiensis*) | BKL164 | 181 | No | x |
| *Calamus erioacanthus* | RBL005 | 343 | No | |
| *Calamus erioacanthus* (*endauensis*) | BKL234 | 328 | No | x |
| *Calamus erioacanthus* (*hepburnii*) | RBL421 | 298 | No | x |
| *Calamus erioacanthus* (*pseudoulur*) | RBL200 | 324 | No | x |
| *Calamus erioacanthus* (*semoi*) | BKL229 | 303 | No | x |
| *Calamus erioacanthus* (*ulur*) | BKL217 | 272 | No | x |
| *Calamus exilis* | RBL123 | 313 | Yes | x |
| *Calamus exilis* (*hispidulus*) | RBL197 | 167 | Yes | |
| *Calamus geniculatus* | BKL235 | 254 | No | |
| *Calamus geniculatus* (*collarifer*) | RBL195 | 247 | No | x |
| *Calamus gracilipes* | BKL129 | 189 | No | x |
| *Calamus gracilipes* (*didymophyllus*) | RBL219 | 333 | No | |
| *Calamus hallierianus* | RBL179 | 263 | No | |
| *Calamus hallierianus* (*georgei*) | BKL236 | 176 | No | x |
| *Calamus helferianus* (*oxleyanus*) | RBL087 | 335 | No | |

**Supplementary Table 3.2** (*continued*)

| Species | Seq. ID | Genes | Monophyletic? | Exclude? |
|---|---|---|---|---|
| *Calamus helferianus* (*viridispinus*) | RBL320 | 237 | No | x |
| *Calamus hirsutus* subsp. *hirsutus* (*hirsutus*) | RBL064 | 317 | No | |
| *Calamus hirsutus* subsp. *korthalsii* (*korthalsii*) | RBL364 | 214 | No | |
| *Calamus inermis* (*palustris*) | RBL060 | 317 | No | x |
| *Calamus inermis* (*platyacanthoides*) | RBL106 | 251 | No | x |
| *Calamus inermis* (*wailong*) | RBL021 | 340 | No | |
| *Calamus inops* | BKL034 | 233 | Yes | |
| *Calamus inops* (*orthostachyus*) | BKL114 | 225 | Yes | x |
| *Calamus javensis* | RBL010 | 341 | No | x |
| *Calamus javensis* (*acuminatus*) | RBL003 | 342 | No | |
| *Calamus javensis* (*corrugatus*) | RBL181 | 256 | No | x |
| *Calamus javensis* (*elopurensis*) | RBL250 | 338 | No | x |
| *Calamus javensis* (*hypertrichosus*) | BKL209 | 313 | No | x |
| *Calamus kingianus* | BKL133 | 295 | No | |
| *Calamus kingianus* (*spicatus*) | BKL158 | 277 | No | x |
| *Calamus kunstleri* subsp. *kunstleri* (*kunstleri*) | RBL012 | 338 | No | |
| *Calamus kunstleri* subsp. *microstachys* (*microstachys*) | RBL173 | 152 | No | |
| *Calamus loherianus* | RBL288 | 286 | No | x |
| *Calamus loherianus* (*urdanetanus*) | BKL097 | 310 | No | |
| *Calamus melanacanthus* | BKL155 | 173 | Yes | x |
| *Calamus melanacanthus* (*mellitus*) | RBL110 | 277 | Yes | |
| *Calamus melanochaetes* (*binnendijkii*) | RBL274 | 182 | No | x |
| *Calamus melanochaetes* (*fissus*) | BKL082 | 327 | No | x |
| *Calamus melanochaetes* (*grandis*) | RBL065 | 324 | No | x |
| *Calamus melanochaetes* (*lewisianus*) | RBL427 | 210 | No | x |
| *Calamus melanochaetes* (*monticola*) | RBL297 | 310 | No | x |
| *Calamus melanochaetes* (*melanochaetes*) | RBL238 | 335 | No | x |
| *Calamus melanochaetes* (*melanochaetes*) | RBL365 | 318 | No | x |
| *Calamus melanochaetes* (*jenkinsianus*) | RBL022 | 339 | No | |
| *Calamus melanochaetes* (*sepal*) | BKL144 | 119 | No | x |
| *Calamus melanochaetes* (*singalanus*) | BKL147 | 330 | No | x |
| *Calamus melanochaetes* (*stenophyllus*) | BKL151 | 328 | No | x |
| *Calamus melanochaetes* (*kurzianus*) | RBL121 | 256 | No | x |
| *Calamus mitis* | BKL214 | 332 | Yes | x |
| *Calamus mitis* (*elmerianus*) | BKL101 | 337 | Yes | |
| *Calamus moseleyanus* | RBL264 | 148 | No | x |
| *Calamus moseleyanus* (*jenningsianus*) | BKL215 | 109 | No | x |
| *Calamus moseleyanus* (*multinervis*) | BKL093 | 337 | No | x |
| *Calamus moseleyanus* (*spinifolius*) | BKL150 | 291 | No | x |
| *Calamus moseleyanus* (*subinermis*) | BKL018 | 337 | No | |
| *Calamus nanduensis* | RBL085 | 340 | No | |
| *Calamus nanduensis* (*pintaudii*) | RBL357 | 201 | No | x |

**Supplementary Table 3.2** (*continued*)

| Species | Seq. ID | Genes | Monophyletic? | Exclude? |
|---|---|---|---|---|
| *Calamus oblongus* subsp. *depressiusculus* (*depressiusculus*) | RBL286 | 205 | No | |
| *Calamus oblongus* subsp. *elongatus* (*elongatus*) | RBL249 | 277 | No | |
| *Calamus oblongus* subsp. *gracilis* (*flexilis*) | BKL185 | 109 | No | |
| *Calamus oblongus* subsp. *longistipes* (*longistipes*) | RBL254 | 301 | No | |
| *Calamus oblongus* subsp. *mollis* (*curranii*) | RBL281 | 329 | No | |
| *Calamus oblongus* subsp. *mollis* (*mollis*) | RBL244 | 291 | No | x |
| *Calamus oblongus* subsp. *oblongus* (*oblongus*) | RBL302 | 322 | No | |
| *Calamus oblongus* subsp. *similaris* (*korthalsii*) | RBL189 | 171 | No | |
| *Calamus optimus* | RBL223 | 341 | No | |
| *Calamus optimus* | RBL390 | 176 | No | x |
| *Calamus plicatus* (*fimbriatus*) | RBL251 | 332 | Yes | |
| *Calamus plicatus* (*laevigatus*) | BKL017 | 322 | Yes | x |
| *Calamus plicatus* (*mattanensis*) | RBL165 | 224 | Yes | x |
| *Calamus ruvidus* | RBL187 | 201 | No | x |
| *Calamus ruvidus* (*tanakadatei*) | BKL216 | 255 | No | |
| *Calamus siphonospathus* | RBL127 | 220 | Yes | x |
| *Calamus siphonospathus* (*dimorphacanthus*) | RBL287 | 306 | Yes | |
| *Calamus sulawesiensis* (*mindorensis*) | RBL152 | 308 | Yes | |
| *Calamus sulawesiensis* (*mindorensis*) | RBL261 | 278 | Yes | x |
| *Calamus tetradactylus* | RBL093 | 301 | No | x |
| *Calamus tetradactylus* (*batoensis*) | RBL237 | 340 | No | x |
| *Calamus tetradactylus* (*crispus*) | RBL232 | 334 | No | x |
| *Calamus tetradactylus* (*flavinervis*) | RBL236 | 343 | No | |
| *Calamus tetradactylus* (*solitarius*) | RBL057 | 317 | No | x |
| *Calamus viminalis* | RBL016 | 338 | No | x |
| *Calamus viminalis* (*phuocbinhensis*) | RBL235 | 338 | No | |
| *Calamus viminalis* (*siamensis*) | RBL059 | 332 | No | x |
| *Calamus zollingeri* subsp. *merrillii* (*merrillii*) | RBL082 | 340 | Yes | |
| *Calamus zollingeri* subsp. *zollingeri* (*zollingeri*) | RBL145 | 335 | Yes | |
| *Plectocomia mulleri* | MSL34 | 122 | No | x |
| *Plectocomia mulleri* | RBL167 | 174 | No | |

**Supplementary List 3.1. Paralogous PhyloPalm loci.** Paralogy was inferred using HybPiper, or by mapping the PhyloPalm target loci using BLAST against the genomes of *Elaeis guineensis* and *Calamus inermis*.

**Paralogs HybPiper**

EGU105033626, EGU105037800, EGU105039449, EGU105042168, EGU105043827, EGU105044758, EGU105044846, EGU105045946, EGU105046168, EGU105048015, EGU105048357, EGU105049690, EGU105052947, EGU105053263, EGU105054898, EGU105057015, EGU105057019, EGU105058798, EGU105059108, EGU105059479, EGU105059594, EGU105059636, EGU105061025, HEY1013, HEY125, HEY2550, HEY728, HEY825

**Paralogs BLAST PhyloPalm vs. *Elaeis guineensis***

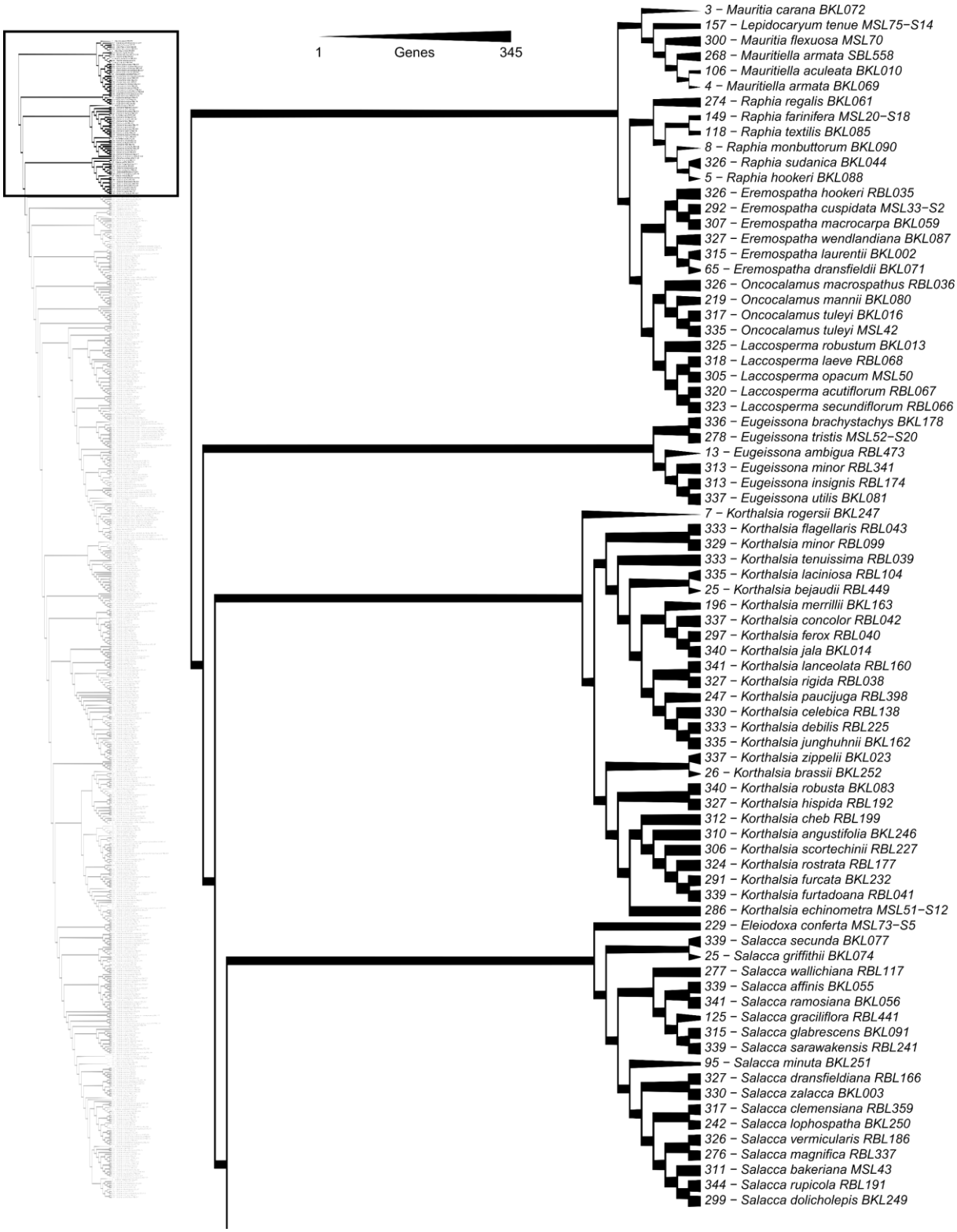
EGU105032175, EGU105032229, EGU105032337, EGU105033626, EGU105034616, EGU105034993, EGU105035196, EGU105035989, EGU105036031, EGU105036385, EGU105037749, EGU105037800, EGU105037890, EGU105037930, EGU105038008, EGU105038036, EGU105038098, EGU105038123, EGU105038179, EGU105038234, EGU105038245, EGU105038310, EGU105038382, EGU105038400, EGU105038419, EGU105038431, EGU105038499, EGU105038603, EGU105038680, EGU105038794, EGU105038882, EGU105038986, EGU105038988, EGU105039062, EGU105039067, EGU105039101, EGU105039221, EGU105039236, EGU105039282, EGU105039298, EGU105039313, EGU105039431, EGU105039542, EGU105039587, EGU105039595, EGU105039609, EGU105039685, EGU105039809, EGU105039947, EGU105039962, EGU105040073, EGU105040114, EGU105040125, EGU105040139, EGU105040189, EGU105040206, EGU105040281, EGU105040302, EGU105040308, EGU105040359, EGU105040368, EGU105040452, EGU105040530, EGU105040583, EGU105040684, EGU105040690, EGU105040700, EGU105040756, EGU105040758, EGU105040813, EGU105040863, EGU105040887, EGU105040914, EGU105040922, EGU105040957, EGU105041055, EGU105041117, EGU105041125, EGU105041133, EGU105041182, EGU105041246, EGU105041283, EGU105041337, EGU105041353, EGU105041650, EGU105041680, EGU105041687, EGU105041816, EGU105041902, EGU105042113, EGU105042128, EGU105042147, EGU105042205, EGU105042290, EGU105042307, EGU105042323, EGU105042329, EGU105042368, EGU105042422, EGU105042560, EGU105042651, EGU105042664, EGU105042722, EGU105042808, EGU105042820, EGU105042965, EGU105043011, EGU105043042, EGU105043061, EGU105043069, EGU105043119, EGU105043155, EGU105043164, EGU105043193, EGU105043320, EGU105043338, EGU105043419, EGU105043430, EGU105043601, EGU105043666, EGU105043730, EGU105043816, EGU105043827, EGU105043926, EGU105043975, EGU105044063, EGU105044133, EGU105044182, EGU105044309, EGU105044378, EGU105044400, EGU105044407, EGU105044446, EGU105044588, EGU105044613, EGU105044614, EGU105044676, EGU105044758, EGU105044844, EGU105044854, EGU105044885, EGU105044893, EGU105044896, EGU105044983, EGU105044984, EGU105045005, EGU105045043, EGU105045070, EGU105045137, EGU105045232, EGU105045248, EGU105045282, EGU105045310, EGU105045358, EGU105045367, EGU105045424, EGU105045464, EGU105045467, EGU105045489, EGU105045509, EGU105045514, EGU105045529, EGU105045658, EGU105045703, EGU105045782, EGU105045827, EGU105045828, EGU105045835, EGU105045932, EGU105045946, EGU105046050, EGU105046099, EGU105046360, EGU105046387, EGU105046401, EGU105046449, EGU105046454, EGU105046503, EGU105046518, EGU105046549, EGU105046562, EGU105046574, EGU105046630, EGU105046632, EGU105046766, EGU105046786, EGU105046827, EGU105046875, EGU105047024, EGU105047029, EGU105047253, EGU105047288, EGU105047293, EGU105047357, EGU105047362, EGU105047379, EGU105047395, EGU105047433, EGU105047519, EGU105047533, EGU105047546, EGU105047553, EGU105047578, EGU105047585, EGU105047597, EGU105047621, EGU105047644, EGU105047689, EGU105047751, EGU105047777, EGU105047916, EGU105047922, EGU105047945, EGU105048009, EGU105048015, EGU105048028, EGU105048054, EGU105048056, EGU105048129, EGU105048159, EGU105048182, EGU105048199, EGU105048357, EGU105048410, EGU105048474, EGU105048479, EGU105048484, EGU105048486, EGU105048493, EGU105048527, EGU105048581, EGU105048751, EGU105048839, EGU105048867, EGU105048898, EGU105048915, EGU105048926, EGU105049020, EGU105049025, EGU105049097, EGU105049274, EGU105049318, EGU105049360, EGU105049426, EGU105049543, EGU105049583, EGU105049729, EGU105049737, EGU105049761, EGU105049827, EGU105049902, EGU105049947, EGU105050012, EGU105050023, EGU105050036, EGU105050058, EGU105050114, EGU105050202, EGU105050328, EGU105050366, EGU105050404, EGU105050450, EGU105050644, EGU105050680, EGU105050682, EGU105050831, EGU105050841, EGU105051403, EGU105051499, EGU105051503, EGU105051560, EGU105051564, EGU105051704, EGU105051748, EGU105051802, EGU105051832, EGU105051857, EGU105051891, EGU105051924, EGU105051953, EGU105052070, EGU105052170, EGU105052178, EGU105052304, EGU105052346, EGU105052351, EGU105052386, EGU105052389, EGU105052394, EGU105052428, EGU105052446, EGU105052483, EGU105052495, EGU105052527, EGU105052538, EGU105052552, EGU105052573, EGU105052623, EGU105052650, EGU105052694, EGU105052704, EGU105052739, EGU105052743, EGU105052947, EGU105053055, EGU105053059, EGU105053105, EGU105053136, EGU105053204, EGU105053227, EGU105053263, EGU105053403, EGU105053457, EGU105053465, EGU105053654, EGU105053735, EGU105053747, EGU105053770, EGU105053835, EGU105053866, EGU105053961, EGU105053980,

EGU105054124, EGU105054130, EGU105054204, EGU105054293, EGU105054435, EGU105054440, EGU105054455, EGU105054457, EGU105054469, EGU105054486, EGU105054498, EGU105054529, EGU105054534, EGU105054668, EGU105054723, EGU105054765, EGU105054827, EGU105054845, EGU105054864, EGU105054896, EGU105054898, EGU105054924, EGU105054972, EGU105055008, EGU105055015, EGU105055023, EGU105055024, EGU105055030, EGU105055047, EGU105055052, EGU105055077, EGU105055090, EGU105055130, EGU105055144, EGU105055283, EGU105055433, EGU105055438, EGU105055490, EGU105055507, EGU105055550, EGU105055569, EGU105055709, EGU105055755, EGU105055761, EGU105055883, EGU105055889, EGU105055908, EGU105055912, EGU105055913, EGU105056032, EGU105056151, EGU105056269, EGU105056287, EGU105056313, EGU105056323, EGU105056365, EGU105056393, EGU105056460, EGU105056530, EGU105056539, EGU105056684, EGU105056688, EGU105056714, EGU105056817, EGU105056943, EGU105056960, EGU105056998, EGU105057013, EGU105057019, EGU105057090, EGU105057110, EGU105057235, EGU105057256, EGU105057335, EGU105057357, EGU105057669, EGU105057742, EGU105057941, EGU105058078, EGU105058083, EGU105058094, EGU105058107, EGU105058175, EGU105058202, EGU105058241, EGU105058245, EGU105058326, EGU105058377, EGU105058418, EGU105058499, EGU105058547, EGU105058556, EGU105058592, EGU105058598, EGU105058614, EGU105058633, EGU105058682, EGU105058683, EGU105058723, EGU105058731, EGU105058781, EGU105058863, EGU105058894, EGU105058904, EGU105058990, EGU105059008, EGU105059035, EGU105059112, EGU105059113, EGU105059131, EGU105059193, EGU105059276, EGU105059367, EGU105059381, EGU105059453, EGU105059458, EGU105059479, EGU105059490, EGU105059570, EGU105059573, EGU105059575, EGU105059587, EGU105059594, EGU105059612, EGU105059636, EGU105059671, EGU105059900, EGU105059996, EGU105061427, HEY1119, HEY136, HEY369, HEY736

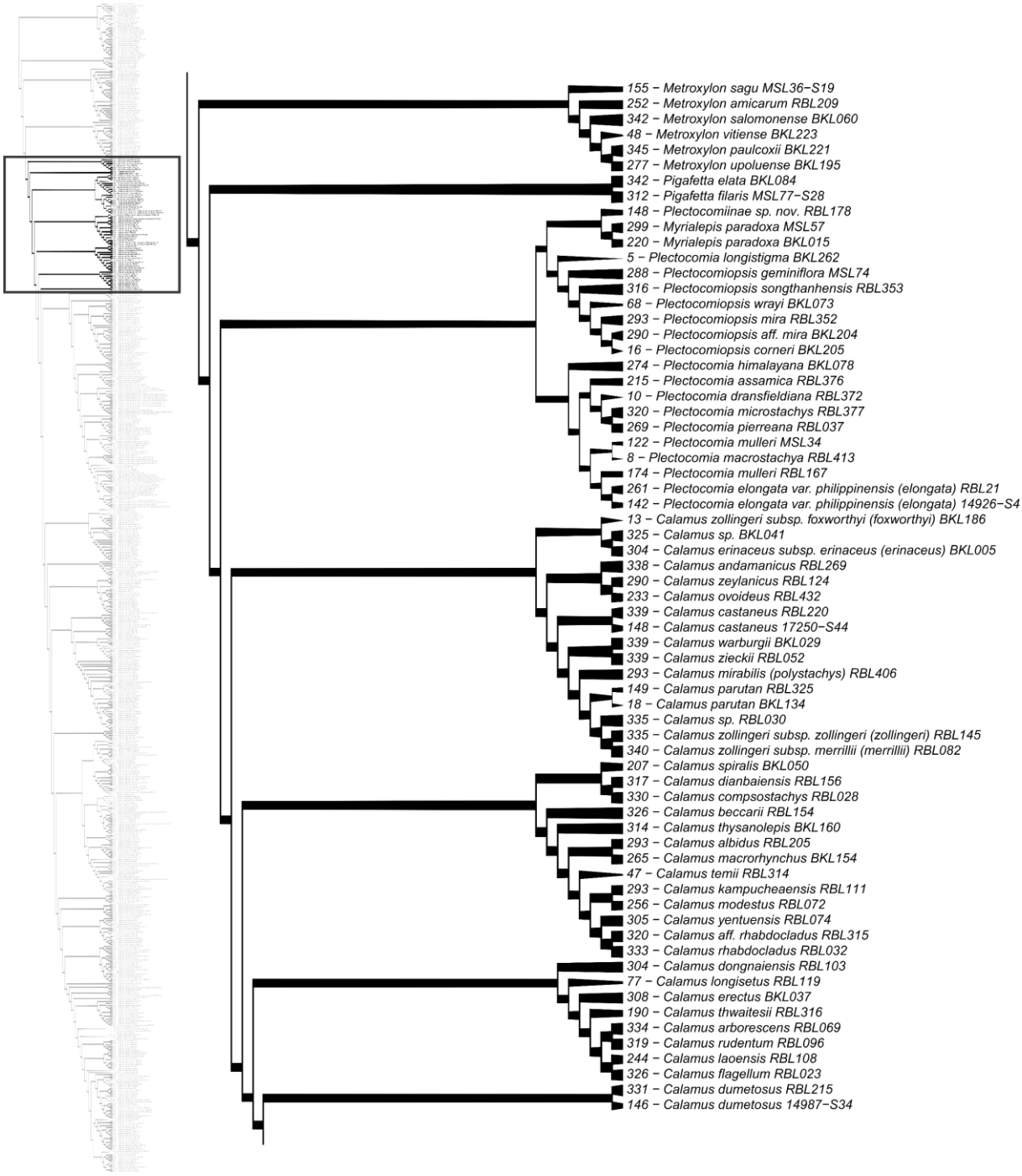
**Paralogs BLAST PhyloPalm vs. *Calamus inermis***

EGU105032229, EGU105032337, EGU105032379, EGU105033063, EGU105033626, EGU105034121, EGU105034893, EGU105035196, EGU105035462, EGU105036385, EGU105037749, EGU105037890, EGU105037902, EGU105037930, EGU105038036, EGU105038098, EGU105038099, EGU105038100, EGU105038118, EGU105038123, EGU105038179, EGU105038228, EGU105038245, EGU105038310, EGU105038382, EGU105038400, EGU105038419, EGU105038431, EGU105038499, EGU105038571, EGU105038603, EGU105038631, EGU105038680, EGU105038693, EGU105038794, EGU105038882, EGU105038986, EGU105039013, EGU105039067, EGU105039101, EGU105039107, EGU105039178, EGU105039221, EGU105039282, EGU105039313, EGU105039460, EGU105039542, EGU105039587, EGU105039660, EGU105039685, EGU105039763, EGU105039809, EGU105039947, EGU105039962, EGU105040073, EGU105040088, EGU105040099, EGU105040114, EGU105040125, EGU105040139, EGU105040206, EGU105040207, EGU105040281, EGU105040302, EGU105040308, EGU105040368, EGU105040426, EGU105040452, EGU105040530, EGU105040583, EGU105040675, EGU105040684, EGU105040690, EGU105040700, EGU105040758, EGU105040813, EGU105040863, EGU105040887, EGU105040957, EGU105041055, EGU105041117, EGU105041133, EGU105041182, EGU105041246, EGU105041283, EGU105041353, EGU105041650, EGU105041665, EGU105041687, EGU105041807, EGU105041816, EGU105041902, EGU105041982, EGU105042113, EGU105042128, EGU105042147, EGU105042205, EGU105042290, EGU105042307, EGU105042323, EGU105042329, EGU105042525, EGU105042560, EGU105042633, EGU105042722, EGU105042808, EGU105042873, EGU105042965, EGU105043011, EGU105043042, EGU105043061, EGU105043119, EGU105043155, EGU105043164, EGU105043193, EGU105043320, EGU105043338, EGU105043374, EGU105043419, EGU105043430, EGU105043499, EGU105043601, EGU105043666, EGU105043685, EGU105043730, EGU105043816, EGU105043827, EGU105044120, EGU105044174, EGU105044182, EGU105044183, EGU105044281, EGU105044309, EGU105044350, EGU105044378, EGU105044407, EGU105044445, EGU105044446, EGU105044588, EGU105044613, EGU105044614, EGU105044676, EGU105044758, EGU105044844, EGU105044846, EGU105044885, EGU105044893, EGU105044896, EGU105044978, EGU105044982, EGU105045005, EGU105045043, EGU105045070, EGU105045137, EGU105045248, EGU105045358, EGU105045424, EGU105045464, EGU105045467, EGU105045489, EGU105045507, EGU105045520, EGU105045529, EGU105045544, EGU105045640, EGU105045703, EGU105045760, EGU105045788, EGU105045827, EGU105045828, EGU105045835, EGU105045932, EGU105045946, EGU105046030, EGU105046056, EGU105046099, EGU105046147, EGU105046245, EGU105046297, EGU105046360, EGU105046387, EGU105046393, EGU105046401, EGU105046449, EGU105046454, EGU105046503, EGU105046530, EGU105046549, EGU105046559, EGU105046574, EGU105046632, EGU105046696, EGU105046735, EGU105046827, EGU105046875, EGU105046918, EGU105047024, EGU105047029, EGU105047253, EGU105047293, EGU105047342, EGU105047357, EGU105047362, EGU105047379, EGU105047385, EGU105047433, EGU105047446, EGU105047519, EGU105047533, EGU105047553, EGU105047578, EGU105047585, EGU105047597, EGU105047621, EGU105047644, EGU105047662, EGU105047751, EGU105047777, EGU105047790, EGU105047907, EGU105047916, EGU105047922, EGU105047945, EGU105047970, EGU105048015, EGU105048028, EGU105048054, EGU105048056, EGU105048129, EGU105048130, EGU105048182, EGU105048199, EGU105048357, EGU105048410, EGU105048474, EGU105048484, EGU105048486, EGU105048493, EGU105048527, EGU105048541, EGU105048581, EGU105048612, EGU105048725, EGU105048751, EGU105048839, EGU105048886, EGU105048915, EGU105048968, EGU105049020, EGU105049025, EGU105049052, EGU105049097, EGU105049274, EGU105049312, EGU105049360, EGU105049426, EGU105049543, EGU105049690, EGU105049729, EGU105049737, EGU105049761, EGU105049827, EGU105049902, EGU105049903, EGU105049947, EGU105050012, EGU105050036, EGU105050058, EGU105050202, EGU105050207, EGU105050328, EGU105050362, EGU105050404, EGU105050644, EGU105050680, EGU105050681, EGU105050841, EGU105050853, EGU105050970, EGU105050972,

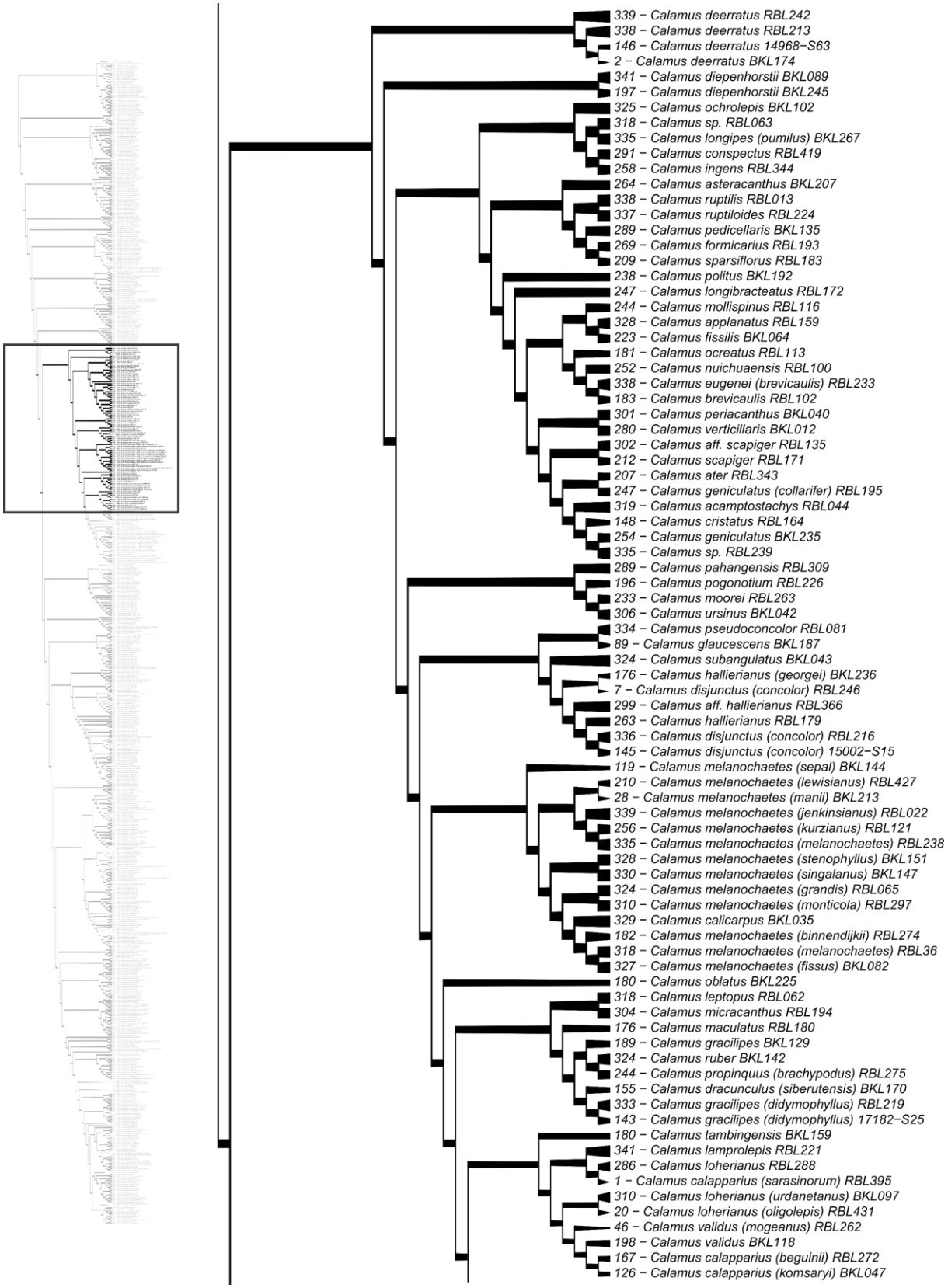
EGU105051345, EGU105051366, EGU105051373, EGU105051391, EGU105051395, EGU105051499, EGU105051503, EGU105051614, EGU105051677, EGU105051704, EGU105051740, EGU105051748, EGU105051802, EGU105051821, EGU105051823, EGU105051832, EGU105051857, EGU105051953, EGU105052035, EGU105052070, EGU105052178, EGU105052304, EGU105052351, EGU105052386, EGU105052389, EGU105052428, EGU105052446, EGU105052476, EGU105052483, EGU105052495, EGU105052529, EGU105052538, EGU105052552, EGU105052573, EGU105052623, EGU105052704, EGU105052739, EGU105052743, EGU105052750, EGU105052771, EGU105052849, EGU105052947, EGU105053055, EGU105053059, EGU105053136, EGU105053204, EGU105053263, EGU105053403, EGU105053426, EGU105053457, EGU105053642, EGU105053654, EGU105053735, EGU105053932, EGU105053969, EGU105053974, EGU105054124, EGU105054130, EGU105054204, EGU105054293, EGU105054435, EGU105054455, EGU105054457, EGU105054469, EGU105054486, EGU105054529, EGU105054534, EGU105054649, EGU105054653, EGU105054668, EGU105054827, EGU105054845, EGU105054898, EGU105054924, EGU105055008, EGU105055015, EGU105055024, EGU105055030, EGU105055052, EGU105055077, EGU105055090, EGU105055093, EGU105055098, EGU105055144, EGU105055283, EGU105055433, EGU105055550, EGU105055569, EGU105055621, EGU105055634, EGU105055709, EGU105055761, EGU105055800, EGU105055862, EGU105055883, EGU105055889, EGU105055908, EGU105055912, EGU105055913, EGU105056032, EGU105056151, EGU105056269, EGU105056313, EGU105056323, EGU105056393, EGU105056460, EGU105056468, EGU105056534, EGU105056539, EGU105056662, EGU105056684, EGU105056688, EGU105056714, EGU105056817, EGU105056881, EGU105056943, EGU105057013, EGU105057015, EGU105057019, EGU105057090, EGU105057110, EGU105057130, EGU105057256, EGU105057335, EGU105057357, EGU105057669, EGU105057742, EGU105057912, EGU105057941, EGU105058078, EGU105058107, EGU105058131, EGU105058175, EGU105058202, EGU105058241, EGU105058245, EGU105058326, EGU105058366, EGU105058377, EGU105058418, EGU105058499, EGU105058547, EGU105058567, EGU105058582, EGU105058592, EGU105058598, EGU105058614, EGU105058633, EGU105058682, EGU105058683, EGU105058687, EGU105058723, EGU105058731, EGU105058781, EGU105058802, EGU105058863, EGU105058889, EGU105058890, EGU105058894, EGU105058904, EGU105058990, EGU105059003, EGU105059035, EGU105059042, EGU105059108, EGU105059112, EGU105059131, EGU105059193, EGU105059276, EGU105059342, EGU105059367, EGU105059453, EGU105059479, EGU105059490, EGU105059570, EGU105059573, EGU105059575, EGU105059587, EGU105059594, EGU105059612, EGU105059636, EGU105059639, EGU105059671, EGU105059900, EGU105059996, EGU105060095, EGU105060589, EGU105061025, EGU105061385, EGU105061427, HEY1119, HEY1171, HEY139, HEY218, HEY225, HEY310, HEY369, HEY391, HEY429, HEY673, HEY727, HEY736, HEY757, HEY758, HEY762, HEY785, HEY808



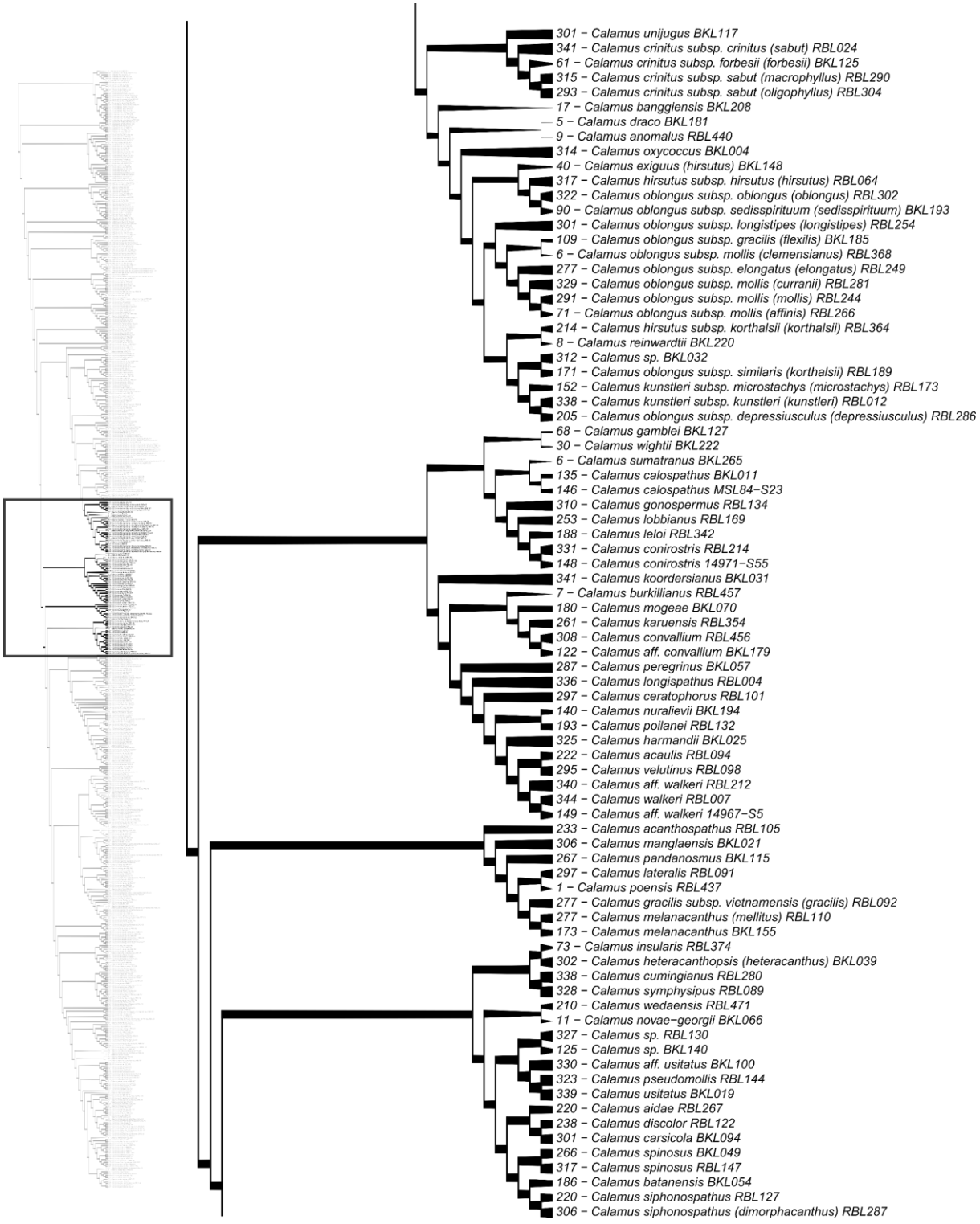

**Supplementary Figure 3.1. Species-level phylogeny of the calamoid palms including data-deficient samples.** ASTRAL phylogeny of 593 samples of 502 calamoid palm species based on 354 gene trees. Thickness of terminal branch tips corresponds to the number of gene trees in which a sample was represented, thin branches therefore indicate data deficiency. The number preceding each sample name indicates the precise number of gene trees in which a sample was represented.



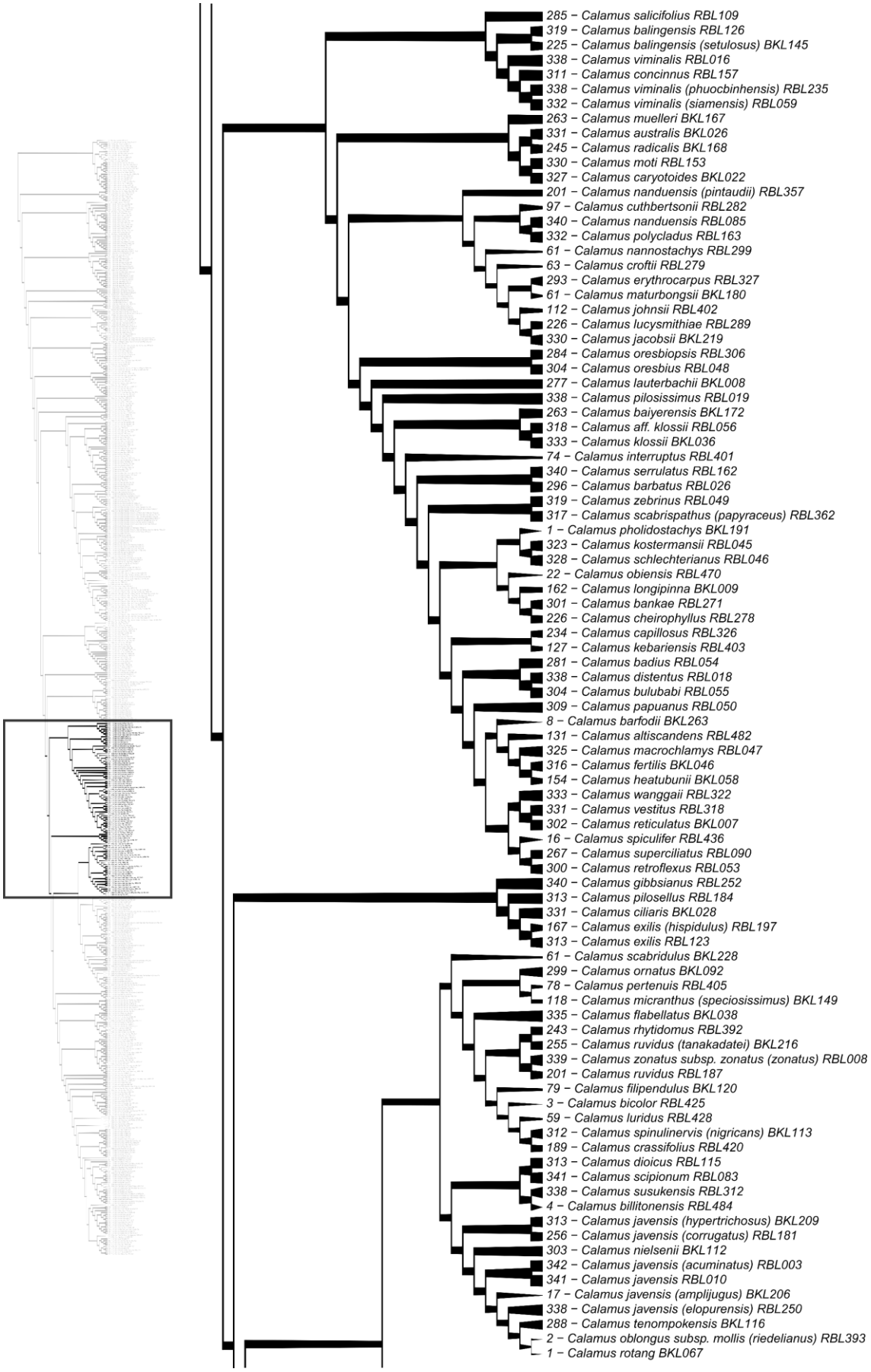

**Supplementary Figure 3.1** (*continued*). Names in parentheses indicate species recognized by Vorontsova et al. (2016a) that were synonymised in the taxonomic revision of *Calamus* by Henderson (2020).



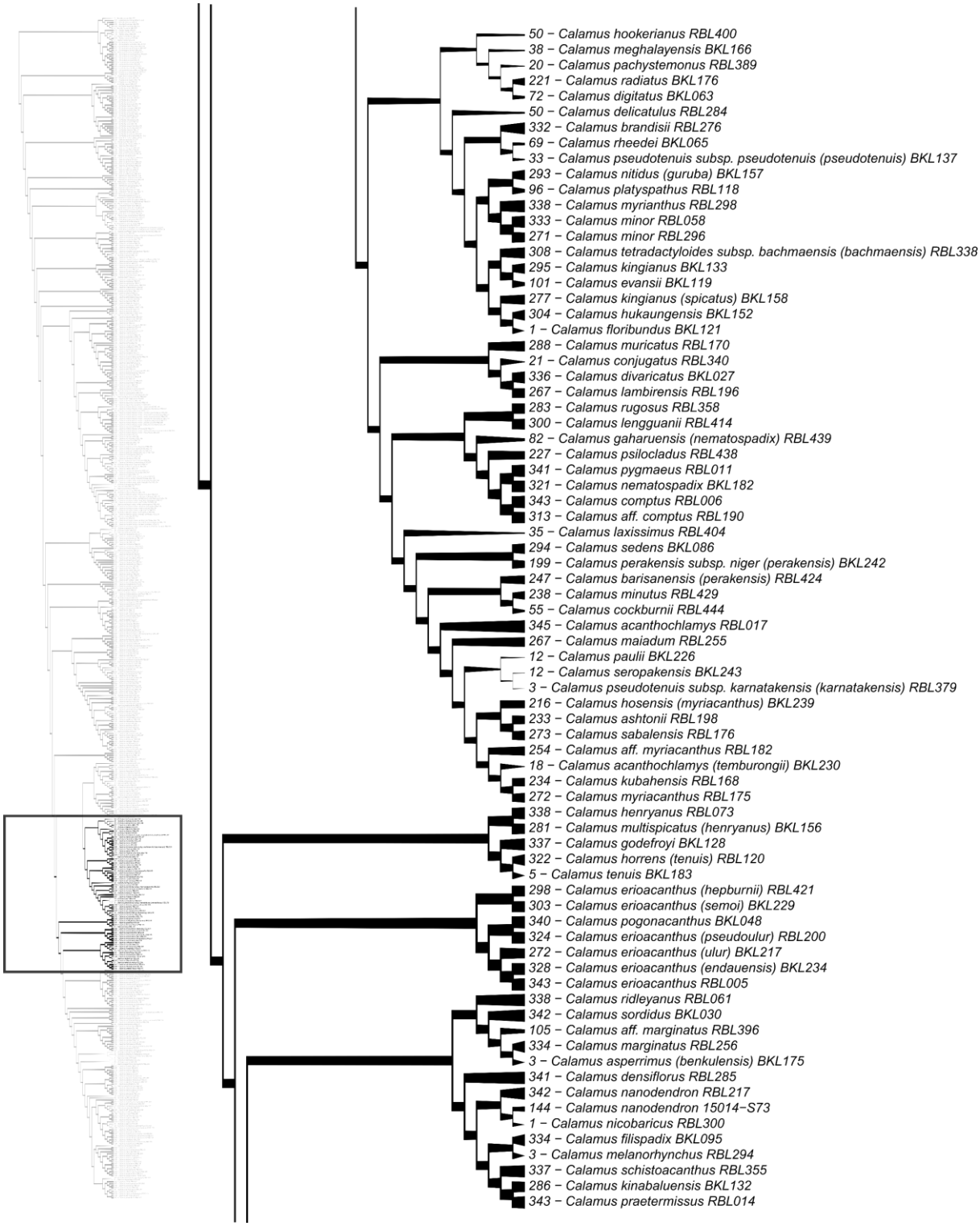

**Supplementary Figure 3.1** (*continued*)




**Supplementary Figure 3.1** (*continued*)



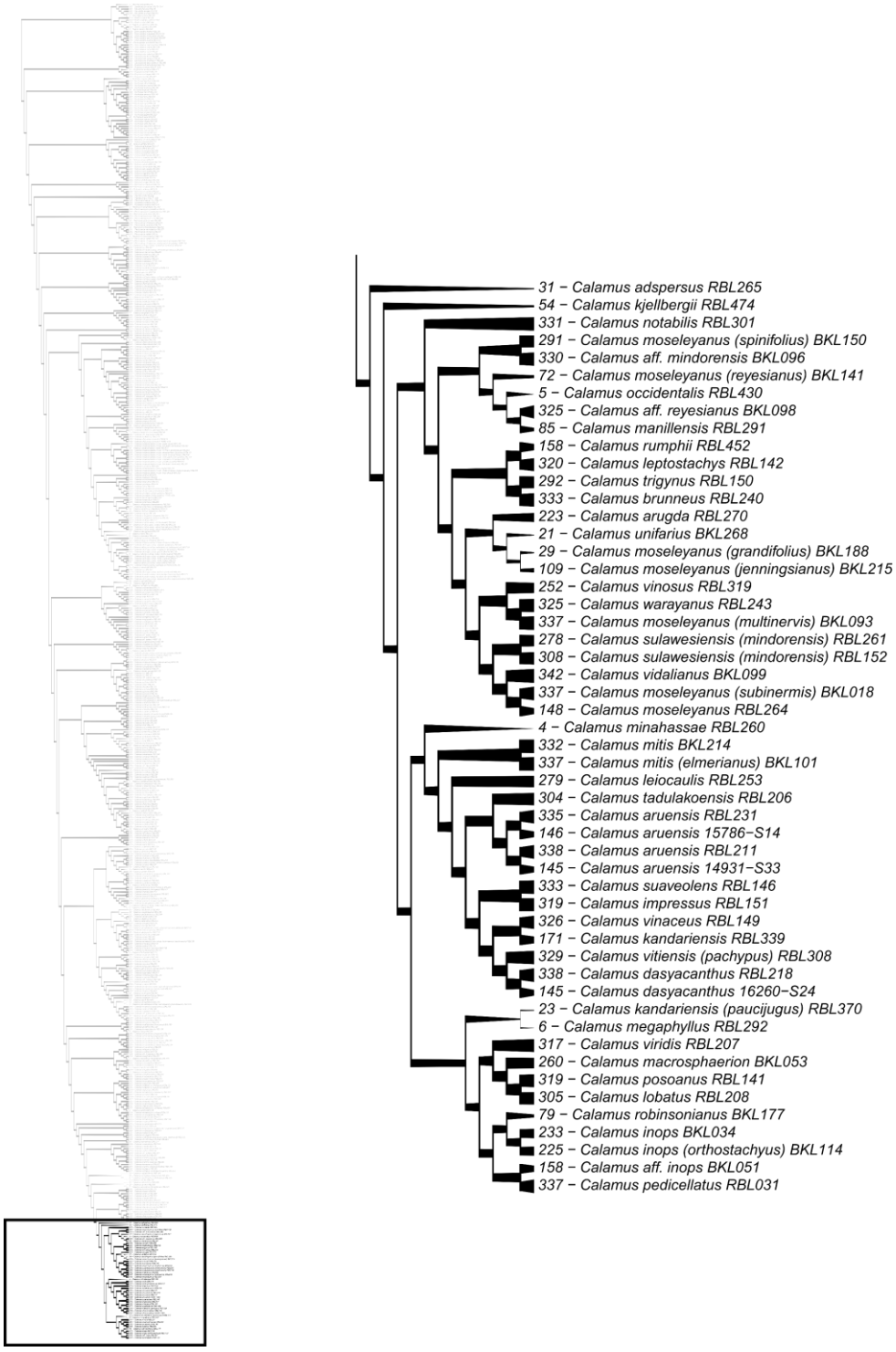

**Supplementary Figure 3.1** (*continued*)



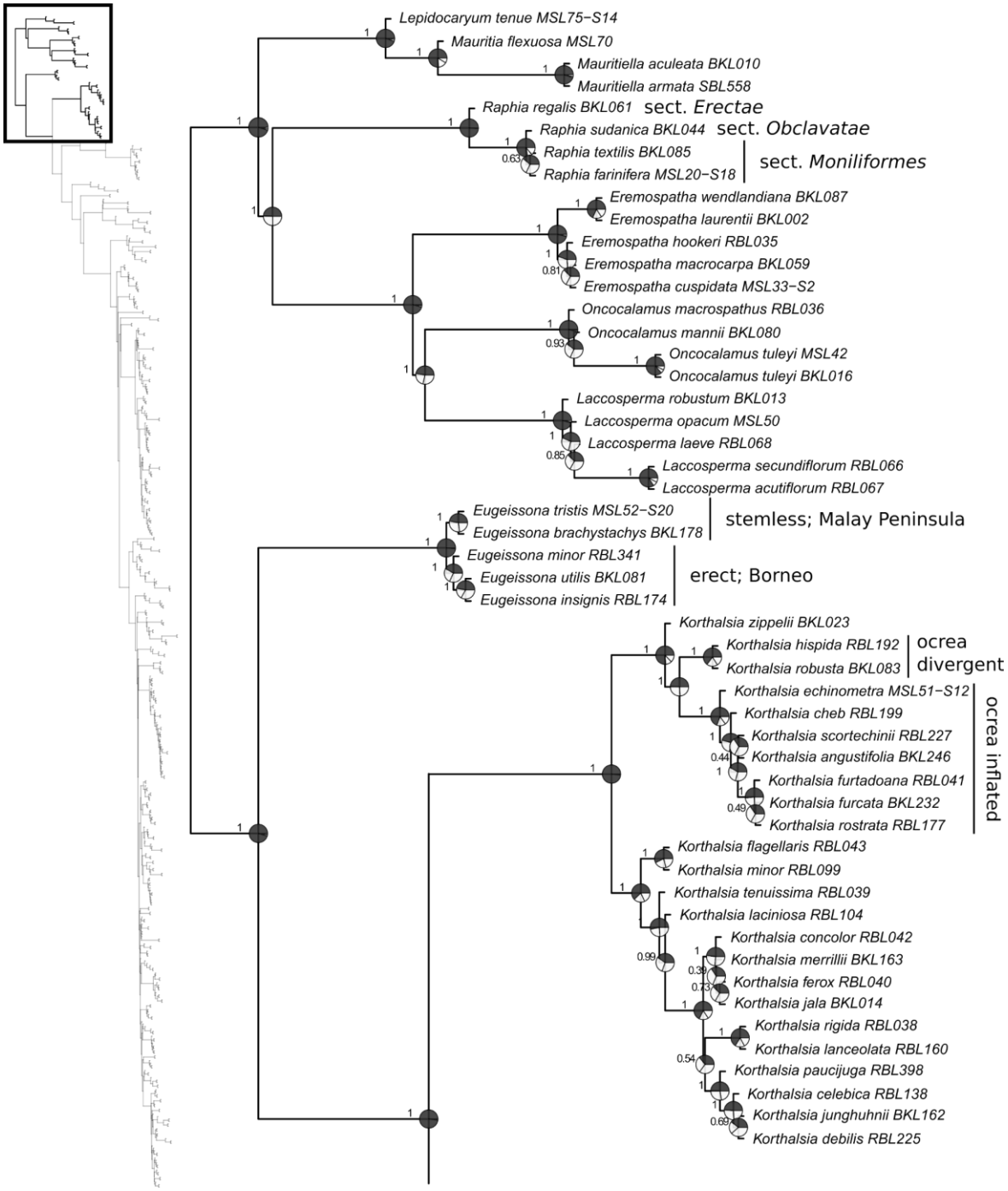

**Supplementary Figure 3.1** (*continued*)



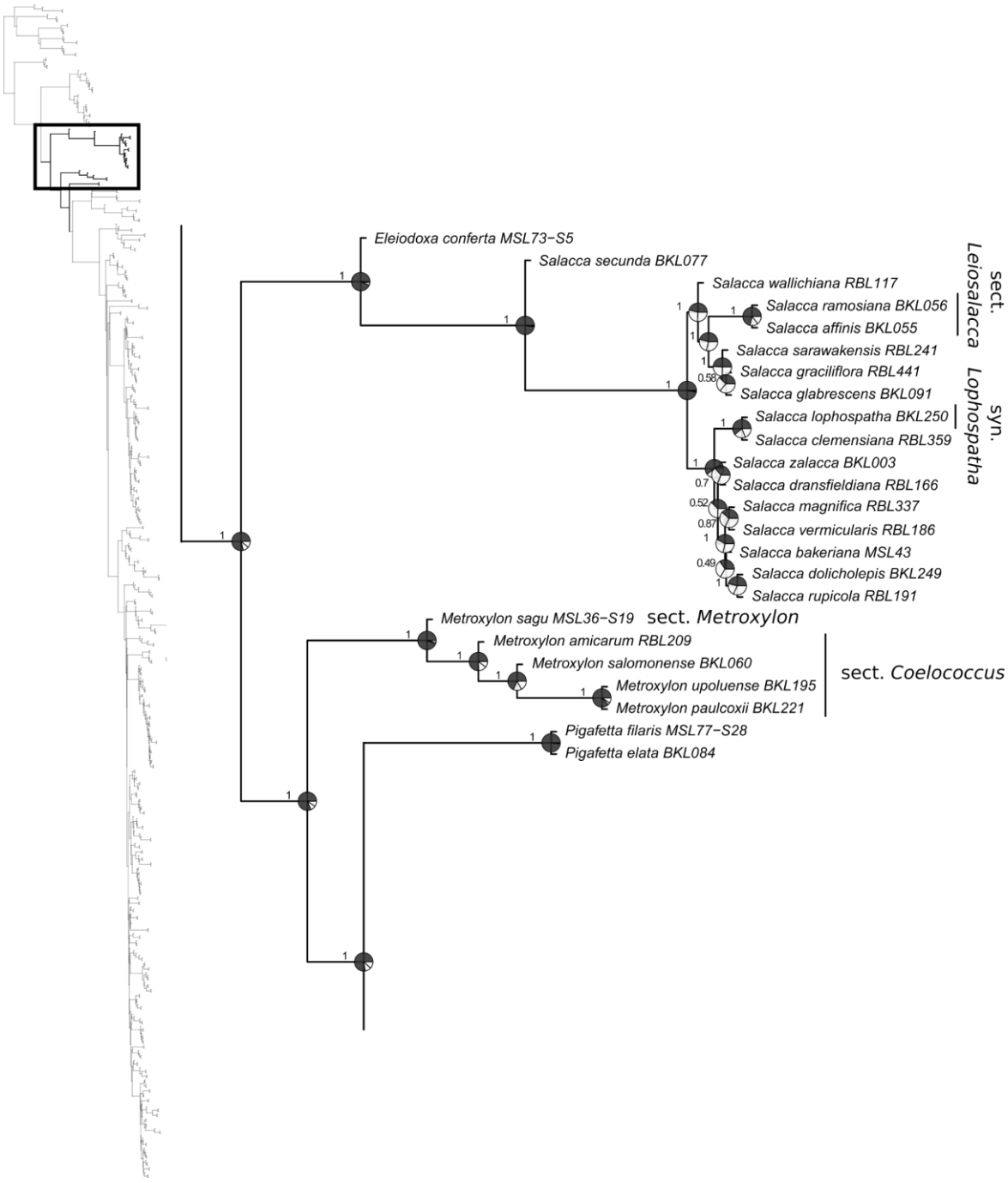

**Supplementary Figure 3.1** (*continued*)



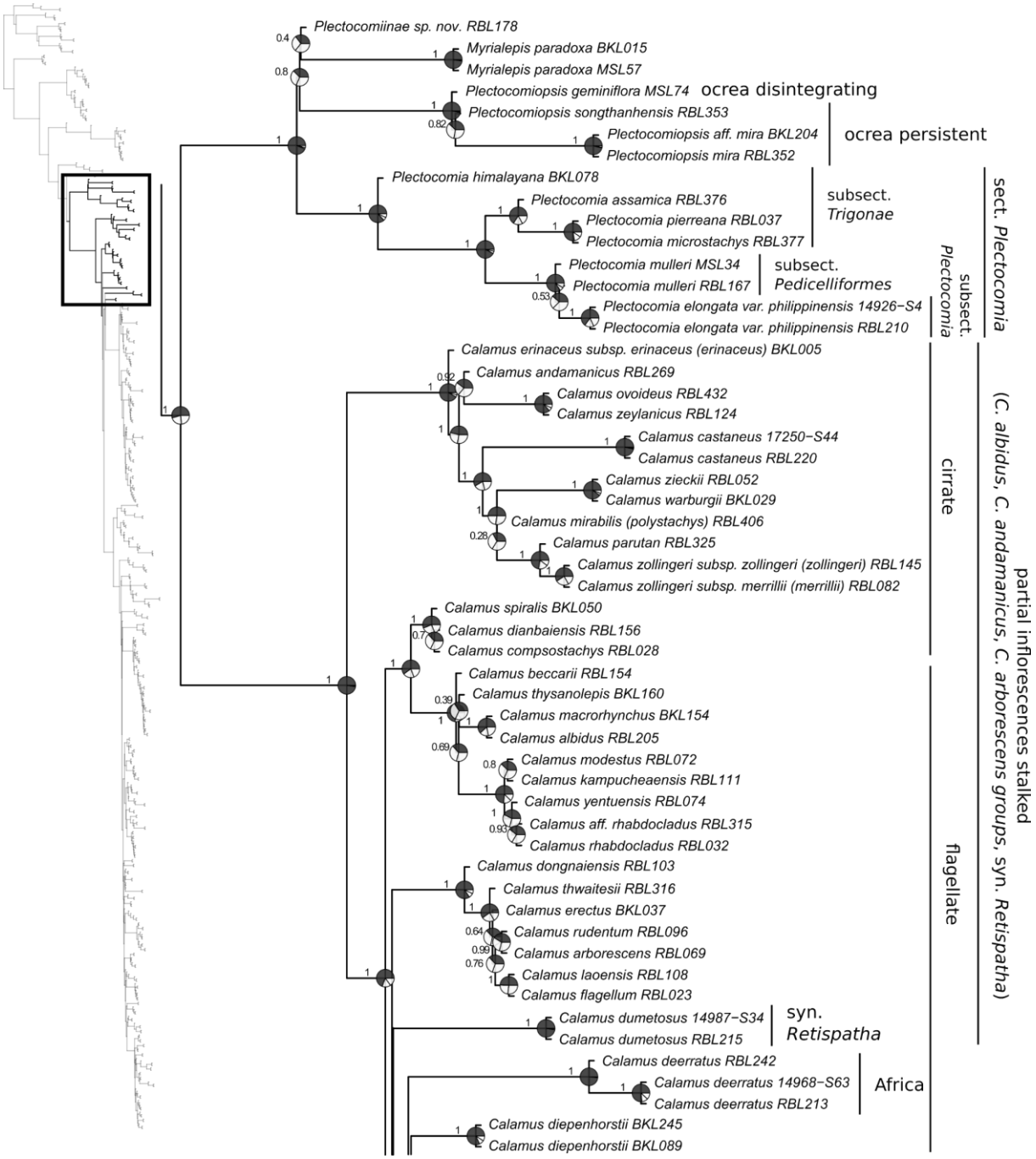

**Supplementary Figure 3.1** (*continued*)



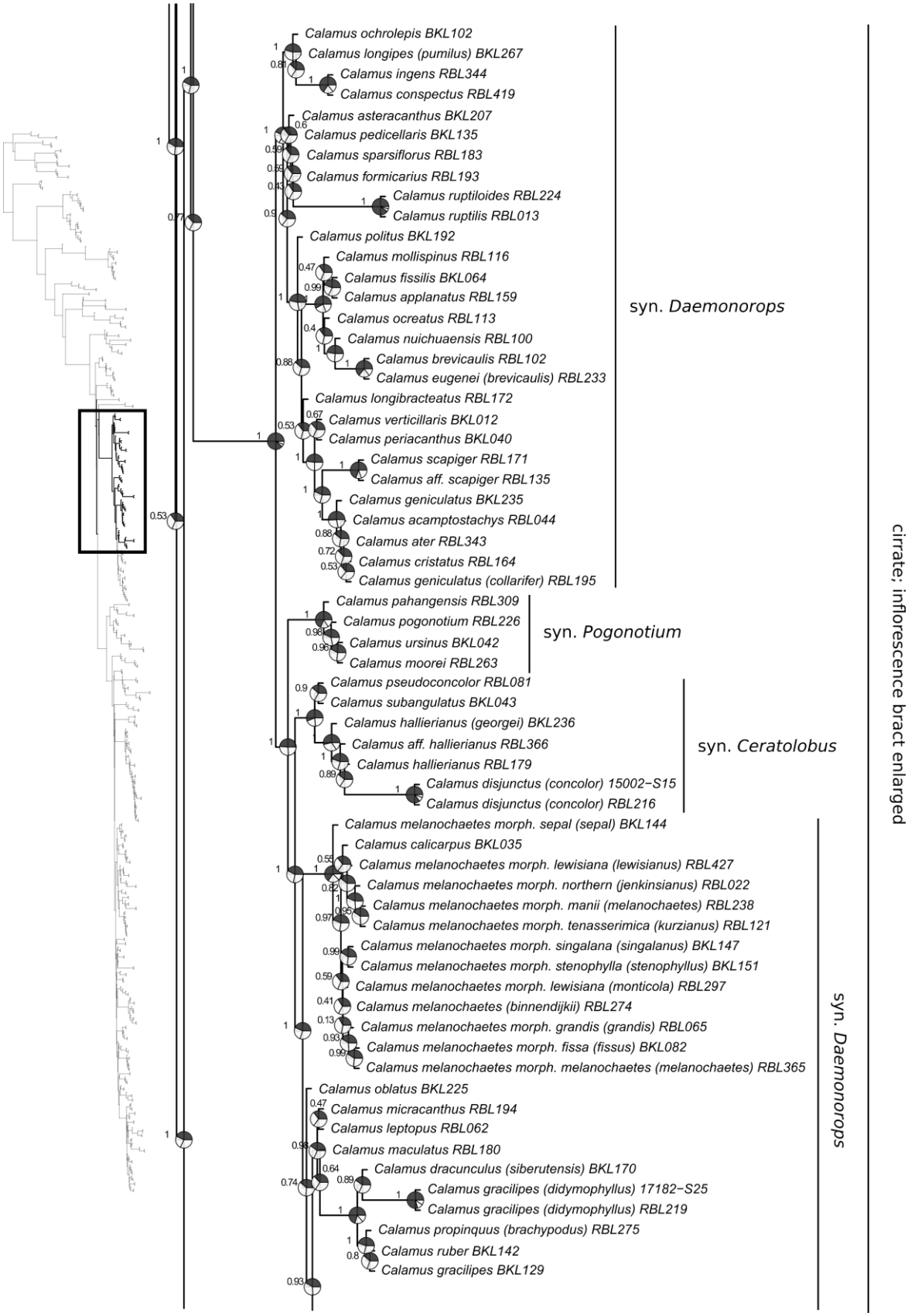

**Supplementary Figure 3.2. Species-level phylogeny of the calamoid palms.** Coalescent-based ASTRAL phylogeny of 482 samples representing 381 calamoid palm species based on 354 gene trees. Node support values given next to nodes are local posterior probabilities. Pie charts represent the proportion of gene trees supporting the main topology (black section) and the two possible alternative topologies (white sections). Terminal branch lengths are not estimated by ASTRAL and were set to a constant, arbitrary length. Names in parentheses indicate species recognized by Vorontsova et al. (2016a) that were synonymised in the taxonomic revision of *Calamus* by Henderson (2020).



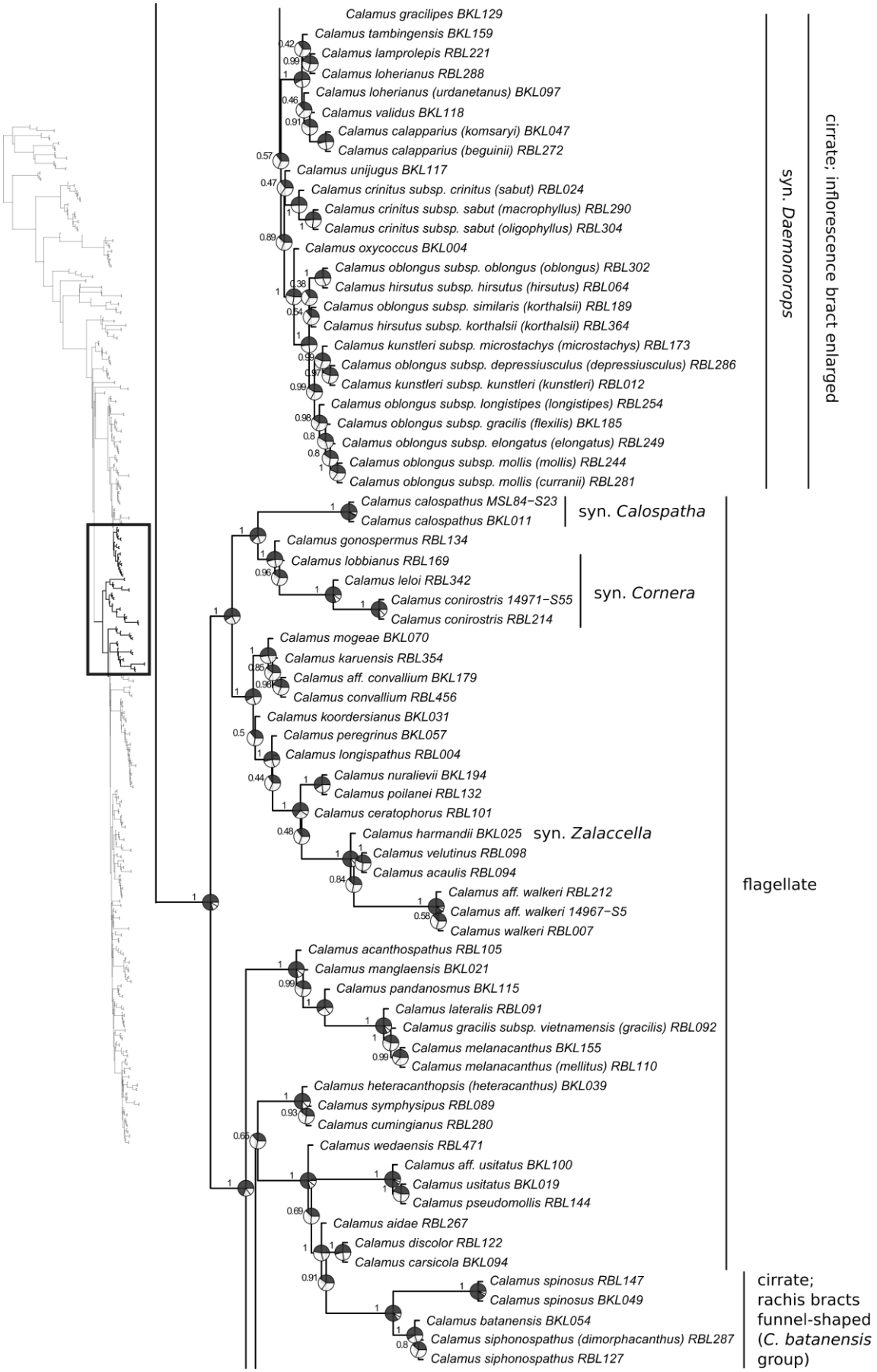

**Supplementary Figure 3.2** (*continued*)



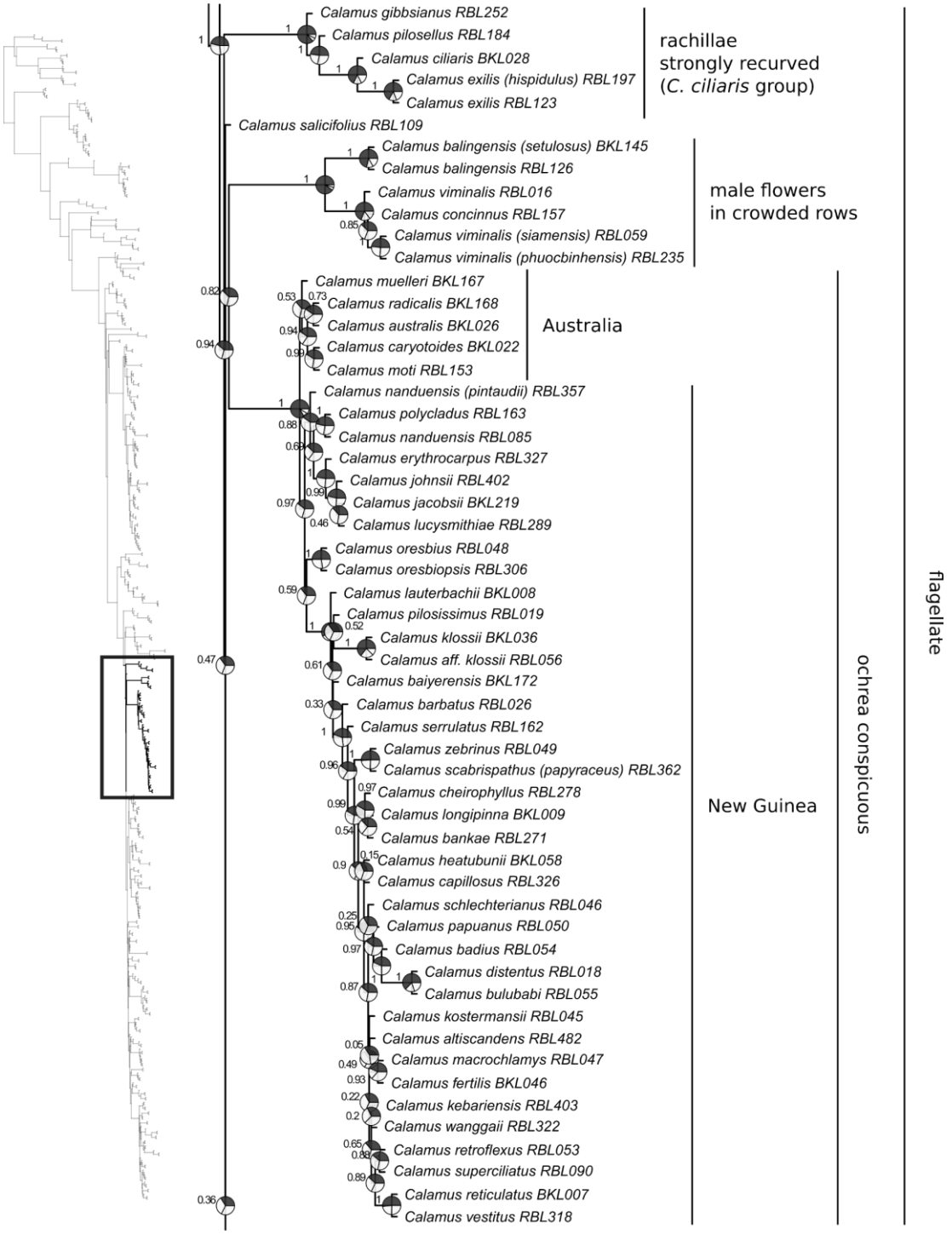

**Supplementary Figure 3.2** (*continued*)



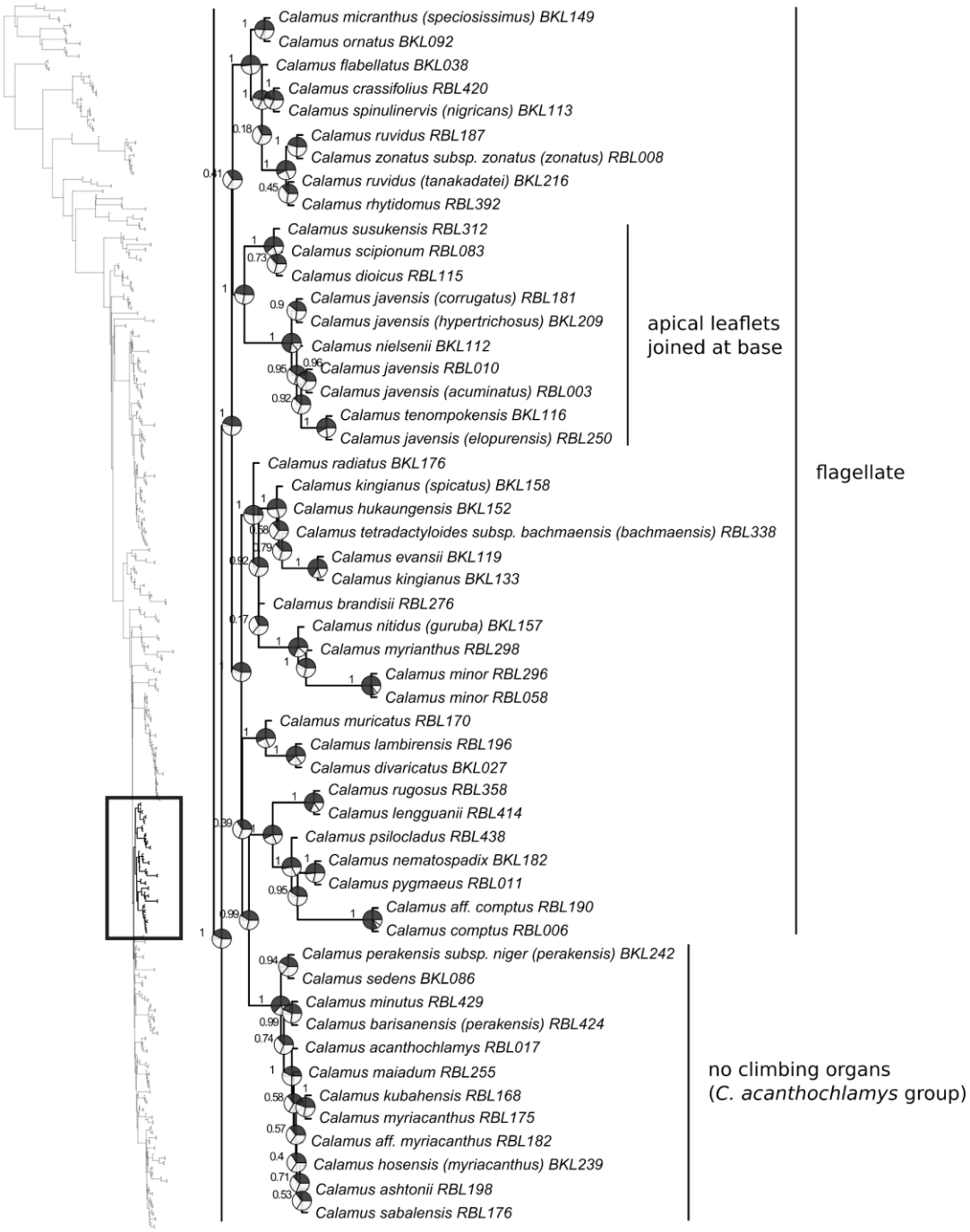

**Supplementary Figure 3.2** (*continued*)



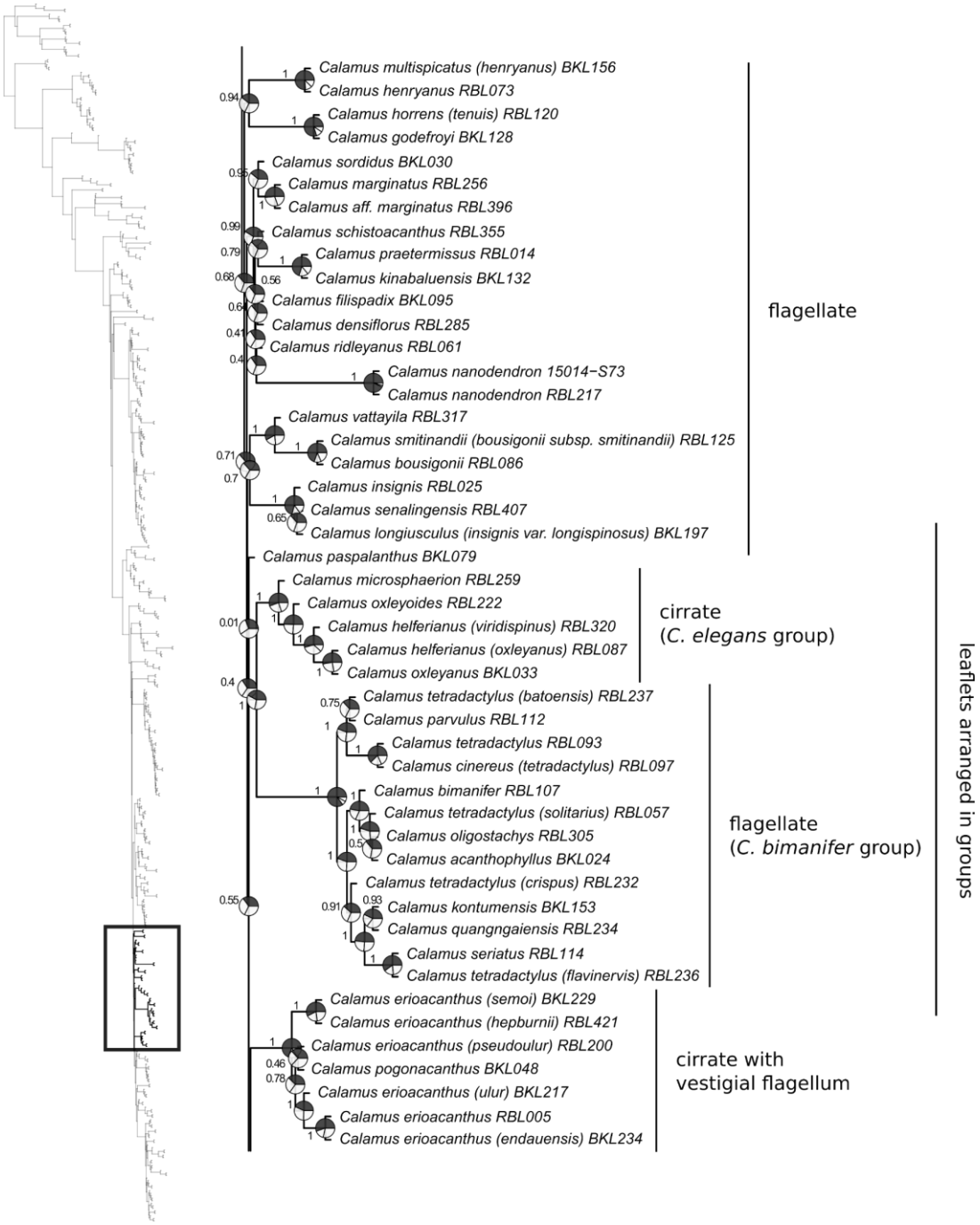

**Supplementary Figure 3.2** (*continued*)



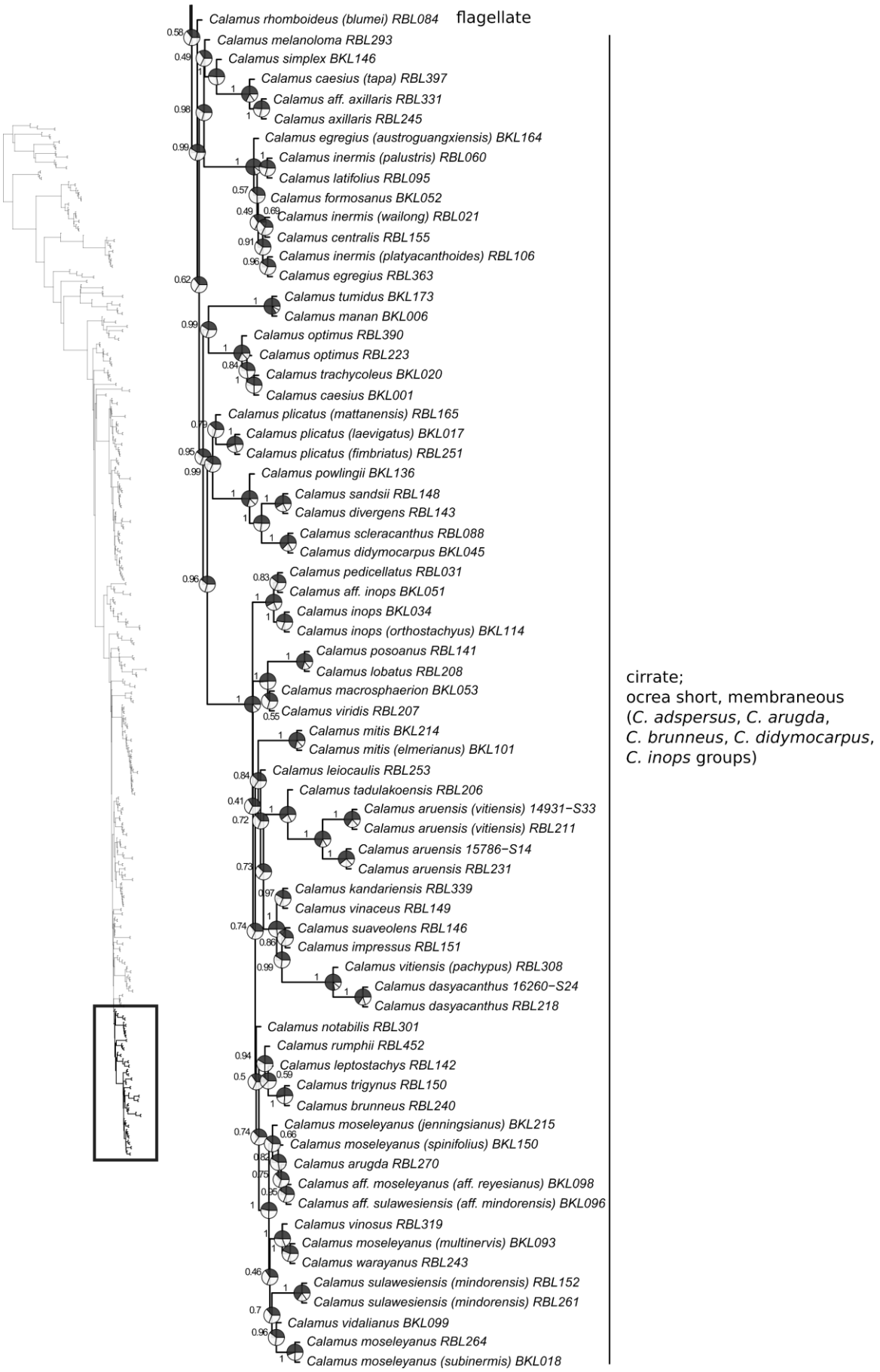

**Supplementary Figure 3.2** (*continued*)



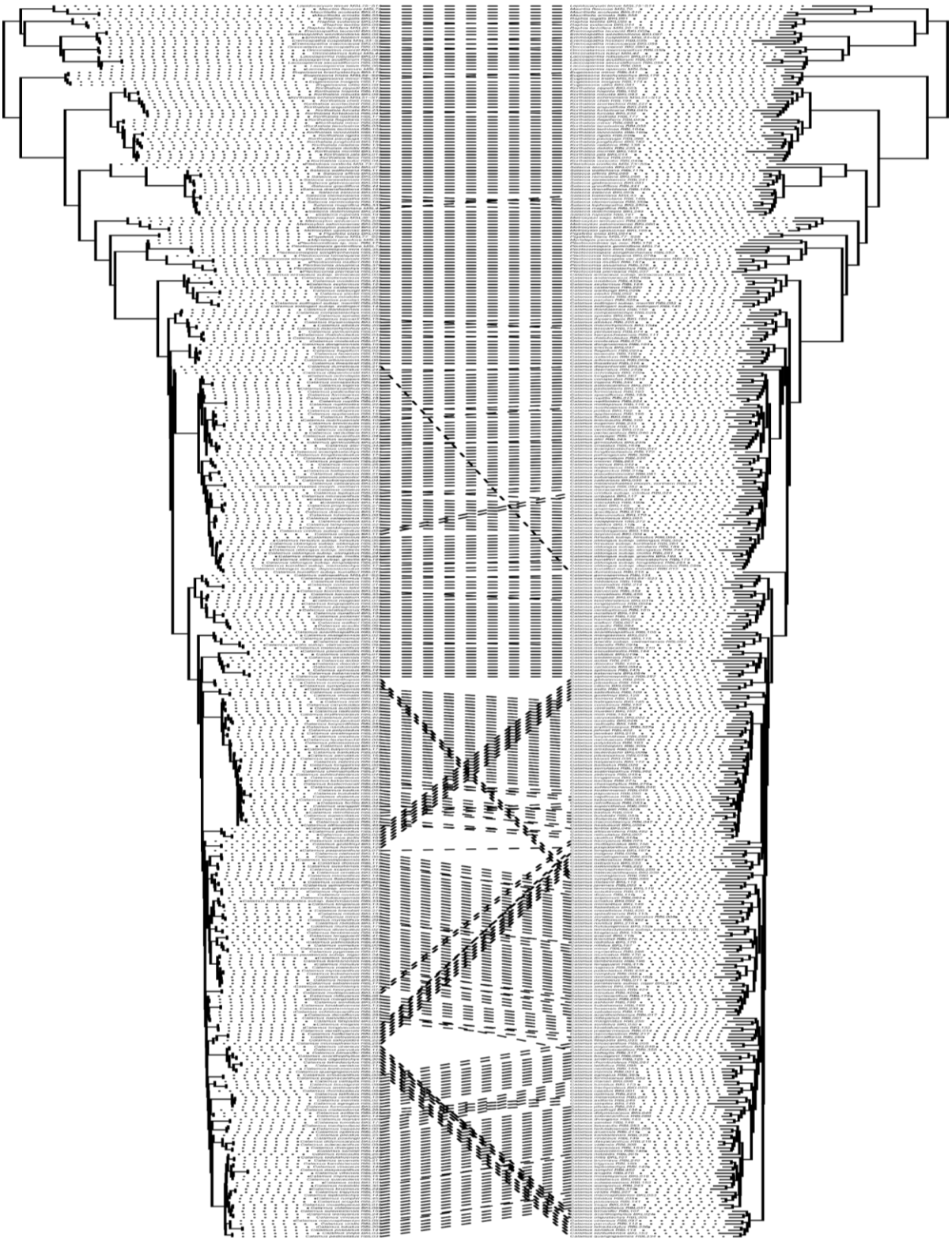

**Supplementary Figure 3.2** (*continued*)



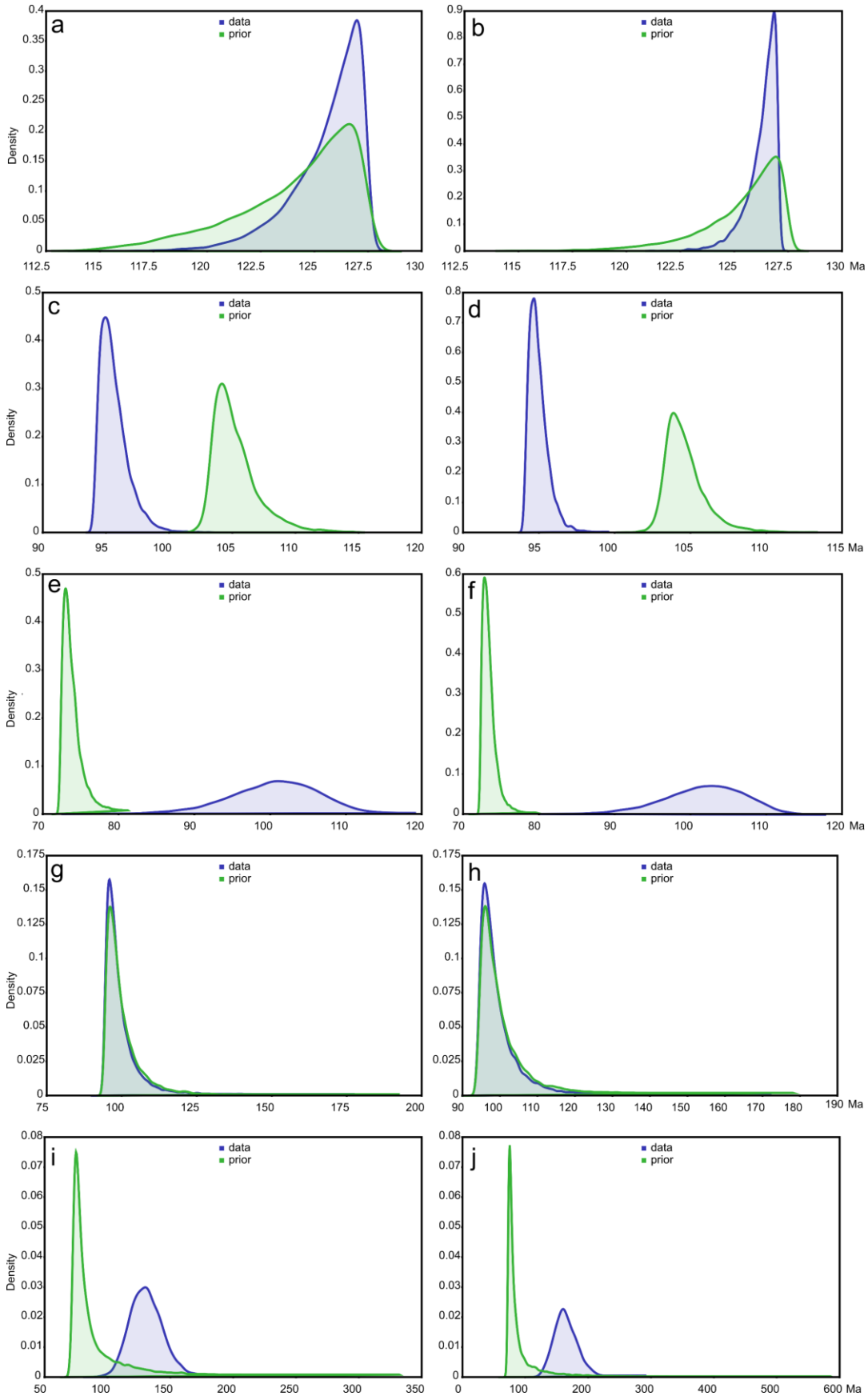

**Supplementary Figure 3.2** (*continued*)




**Supplementary Figure 3.2** (*continued*)

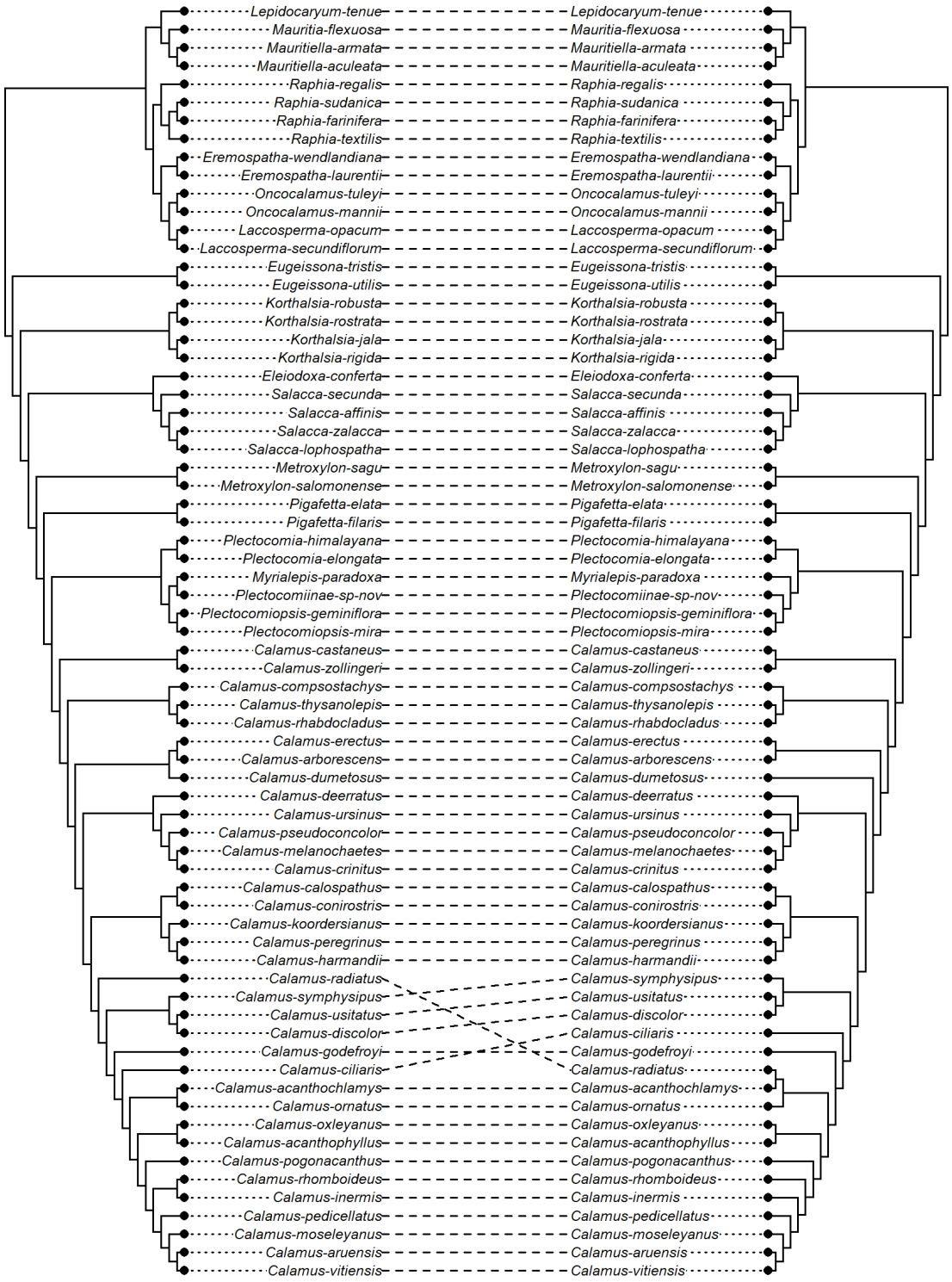


**Supplementary Figure 3.3. Co-phylogenetic plot of ASTRAL and IQ-TREE species trees for 390 species and subspecies of the Calamoideae.** ASTRAL analysis (left) is based on all 354 genes (Fig. 3.4), IQ-TREE analysis (right) is based on concatenated alignments of the 10 genes selected for dating analyses (Table 3.5).




**Supplementary Figure 3.4. Comparison of estimated age distributions of calibrated nodes based on data (blue) or sampled from the prior (green).** (*Figure caption on next page.*)

**Supplementary Figure 3.4.** (*Figure on previous page.*) **a**, crown node age of Calamoideae under strategy MetCal 127.4. **b**, crown node age of Calamoideae under strategy PleCal 127.4. **c**, stem node age of the clade defined by Metroxylinae, Pigafettinae, Plectocomiinae and Calaminae under strategy MetCal 127.4. **d**, stem node age of the clade defined by Plectocomiinae and Calaminae under strategy PleCal 127.4. **e**, stem node age of Mauritiinae under strategy MetCal 127.4. **f**, stem node age of Mauritiinae under strategy PleCal 127.4. **g**, stem node age of the clade defined by Metroxylinae, Pigafettinae, Plectocomiinae and Calaminae under strategy MetCal UC. **h**, stem node age of the clade defined by Plectocomiinae and Calaminae under strategy PleCal UC. **i**, stem node age of Mauritiinae under strategy MetCal UC. **j**, stem node age of Mauritiinae under strategy PleCal UC. For abbreviations of dating strategies, see Table 3.2.




**Supplementary Figure 3.5. Co-phylogenetic plot of phylogenies based on all PhyloPalm loci or orthologous PhyloPalm loci only**. The phylogeny on the left is based on all 971 PhyloPalm loci (Fig. 2.3), the phylogeny on the right is based only on 354 orthologous loci (Fig. 3.4). Phylogenies were pruned to only contain species represented in both trees.

# 4 Historical biogeography of the Calamoideae in the Asia-Pacific

## Abstract

The Asia-Pacific region hosts highly diverse tropical rainforests in a complex geological setting that is centred around the largest tropical archipelago on earth. Previous attempts to unravel the biogeographic history of tropical rainforests of the Asia-Pacific region have largely been based on distribution patterns of extant and fossil taxa, and on dated molecular phylogenies of small tropical rainforest lineages. While this has allowed the identification of some general biogeographic patterns, densely sampled dated phylogenies of large tropical rainforest lineages occurring across the Asia-Pacific are necessary to provide more detailed insights into the biogeographic history of the region.

Here, I utilised a dated phylogeny of 369 species and subspecies of the Indo-Pacific Calamoideae, a palm clade comprising tribes Eugeissoneae and Calameae and occurring across the entire range of the Asia-Pacific. Based on the comparison of 198 biogeographic models, I identified the model that best explains the biogeographic history of the Calamoideae, namely a time-uniform Dispersal-Extinction-Cladogenesis model allowing for "jump speciation" events and accounting for region distance and connectivity.

Using biogeographic stochastic mapping based on ancestral range estimation, migration events between regions were quantified, revealing that the biogeographic history of the extant Calamoideae in the Asia-Pacific largely unfolded during Oligocene and Miocene and may be closely associated with the convergence of Australia and Asia during this time. The widespread distribution of the Calamoideae could largely be explained by stepping-stone dispersal out of Sundaland (Southeast Asia, Sumatra, Java and Borneo) into the

periphery of the Asia-Pacific region (56.4 inferred dispersal events), with 4.6 times fewer dispersals (12.3 inferred dispersal events) in the opposite direction. Dispersal events within Sundaland (185.1 inferred dispersal events) were 5.9 times more frequent than across the deep ocean barriers in the east of the Asia-Pacific (31.5 inferred dispersal events), although only 3.9 dispersals across the shallow seas separating New Guinea from Australia were inferred.

Proportions of *in situ* diversification and emigration were inferred from the number of dispersal events and species diversity associated with regions and allowed insights into the relative contributions of different regions to the distribution and diversity of the Calamoideae. Borneo was identified as a "radiator" with high *in situ* diversification and high emigration rates; Indochina, Sulawesi, and Papuasia as "incubators" with high *in situ* diversification in relative isolation; the Thai-Malay Peninsula, Sumatra, and the Moluccas as "corridors" that connect other regions; and India, Java, Palawan, the Philippines, and Australia as "containers" mainly shaped by immigration.

The species-level biogeography of the calamoid palms presented here is the largest densely sampled biogeographic study of any plant or animal lineage in the Asia-Pacific region to date. It reveals the biogeographic history of a characteristic tropical rainforest clade distributed across the entirety of the Asia-Pacific region in unprecedented detail and will serve as a basis for future biogeographic research in the Asia-Pacific region and beyond.

## 4.1 Introduction

The Asia-Pacific is a highly diverse and geologically complex region (Slik et al., 2015; Hall, 2017; Cámara-Leret et al., 2020) that has attracted interest in its biogeographical history since its exploration by Alfred Russell Wallace (Wallace, 1860). In its centre lies the largest tropical archipelago on earth (Fig. 4.1), functioning as a contact zone between the tropical rainforests of mainland Asia in the west and those of northeast Australia and the Pacific in the east (Crayn et al., 2015; Kooyman et al., 2019; Joyce et al., 2020a).

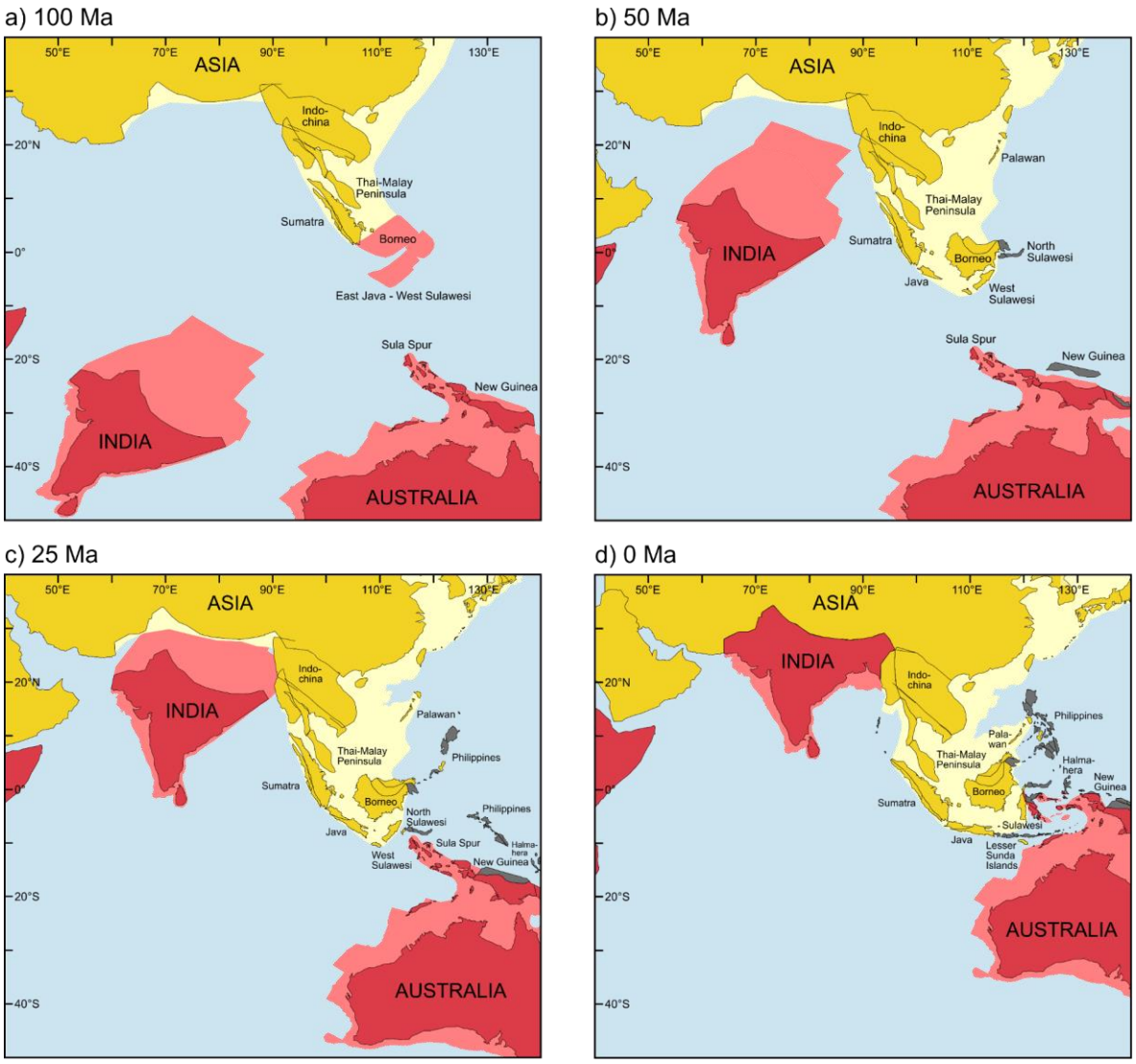



**Figure 4.1. Geological history of the Asia-Pacific region.** Red = Gondwanan plate fragments, yellow = Asian plate, grey = landmasses of other origins. Light shades indicate where plates are submerged. Modified from Hall (2017).

Plant distribution patterns, the fossil record, and biogeographic analyses based on dated phylogenies have provided first valuable insights into the biogeographic history of the region. However, a detailed understanding of how major tropical rainforests lineages attained their broad distribution and high diversity across the Asia-Pacific region remains elusive. In particular, dispersal pathways and the roles of different biogeographic regions in shaping the tropical rainforest assemblages are understood only in vague outlines (Kooyman et al., 2019).

The biogeography of the Asia-Pacific cannot be understood without consideration of the complex geological history of the region (Fig. 4.1). Around 100 million years ago (Ma), fragments that had separated from Australia during the Late Triassic collided with southeastern Asia, forming Borneo, Java and West Sulawesi (Hall, 2017). At ca. 45 Ma, India started to collide with Asia, and from around 25 Ma, the northwest of the northward moving Australian plate collided with the southeast of the Asian plate. After the initiation of the collision of the Australian and Asian plates, several additional changes in the geographical configuration of the Asia-Pacific occurred (Hall, 2017). The continued northward movement of the Australian plate brought it even closer to the Asian plate and was associated with an anti-clockwise rotation of approx. 45 degrees of the area comprising Sumatra, Java, Borneo and Sulawesi until 10 Ma, resulting in the southward moving of Sumatra and Java, and northward moving of Borneo and Sulawesi. Contrasting the movement of Borneo, Palawan moved south from its original position near southeast China to attain its present-day proximity to northern Borneo at around 10 Ma. The Lesser Sunda Islands only formed in the last 10 million years (myr). Within the last 5 myr, West Sulawesi, North Sulawesi and the Sula Spur united to form Sulawesi. Both the Philippines and the Moluccas attained their present composition only in the most recent past, with an island arc comprising the eastern

Philippines and Halmahera moving from 25 Ma rapidly northwest from the north of New Guinea to their current positions.

Overall, these geological events brought together biogeographic realms that had been isolated for tens of millions of years. The collision of India with the Asian plate directly introduced Gondwanan elements that had rafted across the Tethys sea (Morley, 1998). In contrast, the collision of the Australian and Asian plates had a more indirect effect on the interchange of biota as a direct land connection between both regions never existed (Hall, 2017). However, dispersal opportunities substantially increased, both by some islands potentially acting as rafts, and by the increasing density of islands providing additional and closer stepping stones between the Australian and Asian realms (Kooyman et al., 2019).

Finally, differences in ocean depth between islands are of crucial importance for understanding potential dispersal pathways during low sea levels (Weigelt et al., 2016). Southeast Asia, Sumatra, Java and Borneo form the Sunda shelf (also called Sundaland), an area separated by shallow seas less than 100 m deep (Voris, 2000) and connected by continuous land during the periods of lower sea levels that frequently occurred as a result of cooler global climates in the Quaternary (de Bruyn et al., 2014), potentially facilitating the dispersal of plants and animals. However, it has also been suggested that sandy sea-bed soils between western Sundaland (Thai-Malay Peninsula and Sumatra) and eastern Sundaland (Borneo) have acted as a substantial dispersal barrier (Slik et al., 2011). Similarly to Sundaland, New Guinea and Australia were also connected by land at low sea levels, making up the Sahul shelf. In contrast, the islands between the Sunda and Sahul shelves, i.e. the Philippines and Wallacea (consisting of Sulawesi, the Moluccas, and the Lesser Sunda

Islands), are surrounded by deep oceans and remained isolated even during the lowest sea levels, substantially impeding dispersal to and from these islands regions.

The current knowledge on the tropical rainforest biogeography of the Asia-Pacific is largely based on distribution data for extant lineages (Dransfield, 1981a; Whitmore, 1982; Baker et al., 1998; Baker and Couvreur, 2012) and fossil taxa (Morley, 1998; Morley, 2000, 2018), as well as biogeographical analyses based on dated phylogenies of densely sampled small to medium-sized tropical rainforest lineages (Nauheimer et al., 2012; Grudinski et al., 2014; Strijk et al., 2020; Joyce et al., 2021) or sparsely sampled large lineages (Thomas et al., 2012; Baker and Couvreur, 2013a; Koenen et al., 2015). Several biogeographic patterns and processes have been discerned. First, the immigration of tropical rainforest plants from India after its collision with Asia may have substantially shaped the composition and diversity of Asian rainforests (Morley, 1998). Second, it is widely assumed that dispersal between the Sunda and Sahul shelves mainly occurred after the collision of the Australian and Asian plates (Crayn et al., 2015), with the Philippines and Wallacea acting as stepping stones between both regions (Kooyman et al., 2019). Third, analyses indicate that the exchange of tropical rainforest plants between Asia and Australia may have been dominated by eastwards dispersals from Asia into Australia (Sniderman and Jordan, 2011; Crayn et al., 2015; Koenen et al., 2015; Strijk et al., 2020), with mainland Southeast Asia (Indochina) and Borneo identified as important source regions of migration events (de Bruyn et al., 2014). Finally, plant distributions in the Asia-Pacific region are commonly discussed in the light of biogeographic boundaries such as Wallace's line (Wallace, 1860; Schuster, 1972; Dransfield, 1981a; Whitmore, 1982; Dransfield, 1987; Baker et al., 1998; Baker and Dransfield, 2000; van Welzen et al., 2011; Baker and Couvreur, 2012; Kissling et al., 2012b; Baker and Couvreur, 2013a; Kooyman et al., 2019; Joyce et al., 2020a; Joyce et al., 2020b).

Formal ancestral range inferences incorporating densely sampled and reliably dated molecular phylogenies of large tropical rainforest clades distributed across the Asia-Pacific have potential to provide more detailed insights into the historical biogeography of the tropical rainforests of the Asia-Pacific (Kooyman et al., 2019). One promising candidate that fulfils these criteria is the palm subfamily Calamoideae, specifically the lineage formed by the tribes Eugeissoneae and Calameae. All except one of the 504 extant species of this clade occur in the Asia-Pacific, ranging from India to Australia and Fiji. The Calamoideae are highly constrained to wet tropical and subtropical climates (Reichgelt et al., 2018) and almost exclusively occur in forest habitats due to their predominantly climbing habit (Couvreur et al., 2015), making them good indicators of wet tropical rainforests.

Whereas the extant distribution of the Calamoideae in the Asia-Pacific region is well understood (Dransfield, 1981a, 1987; Baker et al., 1998; Baker and Dransfield, 2000; Dransfield et al., 2008; Baker and Couvreur, 2012; Baker and Couvreur, 2013a), the past distribution of the subfamily is incompletely known despite its rich fossil record (Morley, 2000; Dransfield et al., 2008). The oldest unambiguous fossils of *Eugeissona* are *Quilonipollenites* pollen from the Paleocene of India (Morley, 1998), i.e. before the collision with Asia. From the Eocene onwards, *Eugeissona* has also been recorded from Sundaland (Morley, 1998), whereas it has disappeared from India since the Pliocene (Lim et al., 2021). In Calameae, the oldest fossils are the pollen taxon *Dicolpopollis* from the Late Cretaceous of Australia, where a continuous record up to the present exists (Truswell and Owen, 1988; Dransfield et al., 2008; Macphail and Jordan, 2015; Macphail and Hill, 2018). In the Paleocene to Miocene, it was also widely distributed across Eurasia and North America, and has been recorded in Africa from the Oligocene onwards. Since the Pliocene, *Dicolpopollis* has only been recorded from Africa and the Asia-Pacific region (Lim et al., 2021).

At present, most genera and species occur in Sundaland, but a secondary centre of diversity exists in Papuasia (Baker and Couvreur, 2012). *Eugeissona*, *Eleiodoxa*, *Salacca*, *Plectocomiopsis* and *Myrialepis* only occur in Sundaland. *Plectocomia* is also essentially of Sundaland distribution, but one species also occurs on Mindanao Island in the southern Philippines. In contrast to these western distributions, *Metroxylon* and *Pigafetta* occur in Wallacea and Papuasia. Only two genera have broader ranges spanning most of the Asia-Pacific: most species of *Korthalsia* occur in Sundaland, but some species also occur in Wallacea and Papuasia. *Calamus*, the most diverse genus of palms with over 400 species, occurs across the entirety of the wet tropical Asia-Pacific region, with most species concentrated in Sundaland and Papuasia. A single species of *Calamus* occurs in Africa.

The historical biogeography of the Calamoideae has been studied for over four decades. The first biogeographic studies were based on distribution and diversity patterns of genera and their presumed relationships based on morphological systematics (Moore, 1973a; Baker et al., 1999b), leading to 1) the proposition of dispersal of the Calamoideae from both Eurasia and Australia into Malesia to explain the disjunct distribution of *Metroxylon* and *Pigafetta* (Dransfield, 1981a), and 2) the explanation of the broad distribution of *Calamus* and *Korthalsia* by eastwards dispersal from Sundaland across Malesia after the collision of the Australian and Asian plates (Dransfield, 1981a, 1987; Baker et al., 1998). When the first molecular phylogeny of the Calamoideae became available (Baker et al., 2000a; Baker et al., 2000c), a new biogeographical narrative incorporating new evidence from the fossil record (Morley, 1998) and plate tectonic reconstructions (Smith et al., 1994) was attempted, proposing a Gondwanan origin of the Calamoideae, rafting of Eugeissoneae via India into Asia, and the colonisation of the Asia-Pacific region from Eurasia (Baker and Dransfield, 2000).

Following the first dated genus-level phylogeny of the Arecaceae (Baker et al., 2009; Couvreur et al., 2011a), these narratives were revisited (Baker and Couvreur, 2012), ruling out a Gondwanan origin of the Eugeissoneae and Calameae based on incompatible divergence time estimates but reinforcing the hypothesis of multiple dispersals of the Calameae from Sundaland eastwards across Malesia. Finally, formal biogeographical analyses of the Arecaceae on a global scale were conducted that included three study regions in the Asia-Pacific (India, Eurasia, and the region east of Wallace's line), inferring an Eurasian origin of the Eugeissoneae and Calameae and a complex pattern of multiple dispersals back and forth across Wallace's line from as early as 50 Ma (Baker and Couvreur, 2013a). It must be noted, however, that the latest biogeographic inferences (Baker and Couvreur, 2012; Baker and Couvreur, 2013a) could not satisfactorily account for the presence of Eugeissoneae in India and of Calameae in Australia long before the collision of these plates with Asia, as indicated by the fossil record (Macphail and Jordan, 2015; Macphail and Hill, 2018).

Here, I used a densely sampled, dated species-level phylogeny of the Calamoideae to address the following questions in the biogeographical history of the Calamoideae in the Asia-Pacific region: 1) what are the ancestral ranges of extant calamoid lineages?, 2) what were the main pathways of dispersal for the Calamoideae across the Asia-Pacific region?, 3) what roles did different biogeographic regions play in shaping the current distribution and diversity of the Calamoideae? This study represents the most detailed biogeographic analysis of any major tropical rainforest lineage in the Asia-Pacific to date, providing unprecedented insights into dispersal patterns and roles of biogeographic regions in one of the most complex geological settings on earth.

## 4.2 Material and Methods

### 4.2.1 Dated phylogeny

The biogeographical analyses are based on a dated species-level phylogeny of the Calamoideae (Fig. 3.4, analysis MetCal 127.4). Analysis MetCal 127.4 (Table 3.2) was selected because of its conservative placement of fossils and implementation of a maximum age constraint for the crown node age of the Calamoideae, as unconstrained analyses resulted in implausibly old divergence time estimates for the oldest nodes in the phylogeny (Table 3.6). Because of the regional focus of the biogeographical analyses, the African-American tribe Lepidocaryeae, which is sister to the Asian-Pacific tribes Eugeissoneae and Calameae, was pruned from the dated phylogeny using phyx (Brown et al., 2017). The resulting phylogeny contained 369 species and subspecies, corresponding to 360 species.

### 4.2.2 Biogeographic analyses

#### *Distribution data*

Species-level distribution data for all taxa were retrieved from the World Checklist of Arecaceae (Govaerts et al., 2021) at level 3 of the World Geographical Scheme for Recording Plant Distributions, i.e. “botanical countries” (Brummitt et al., 2001), and non-native occurrences were excluded. For *Calamus*, the distribution data were updated using a recent revision of the genus (Henderson, 2020) that is not yet fully incorporated in the World Checklist of Arecaceae.

Following de Bruyn et al. (2014), the distribution data were grouped into 14 biogeographic regions based on the areas of endemism proposed by Turner et al. (2001),

which also closely reflect the complex geological history of the region (Fig. 4.1): 1) India including Sri Lanka, 2) Indochina in the sense of “Indo-Burma” of Myers et al. (2000) (Himalayas, Southeast Asia, Andaman and Nicobar Islands), but with eastwards inclusion of Southeast China, Hainan and Taiwan, 3) Thai-Malay Peninsula up to the Isthmus of Kra, 4) Sumatra, 5) Java including Bali, 6) Lesser Sunda Islands, 7) Borneo, 8) Palawan, 9) Philippines including oceanic islands in the North up to Lanyu Island, 10) Sulawesi, 11) Moluccas, 12) Papuasia, i.e. New Guinea with inclusion of eastwards Pacific islands up to Fiji, 13) Australia, 14) Africa. Species recorded from botanical countries (Brummitt et al., 2001) that cross the boundaries between these biogeographical regions were assigned to one or both regions based on more detailed species distributions published in the taxonomic literature (Dransfield, 1981b; Henderson, 2020) as follows: Thailand was divided at the Isthmus of Kra into Indochina and Thai-Malay peninsula, Bali was assigned from the Lesser Sunda Islands to Java, Palawan was split off from the Philippines as a separate region, and the oceanic island Lanyu was assigned from Taiwan to the Philippines. Biogeographic regions occupied by only a single taxon were excluded, leading to the removal of the regions Lesser Sunda Islands and Africa, and the pruning of the only African species of tribe Calameae, *Calamus deerratus*, from the phylogeny.

Three main analyses were conducted in which the inference of ancestral ranges was constrained to a maximum of two, three or four regions. The distributions of species occurring in more regions than the set maximum were constrained to the set maximum based on the distribution of their closest relatives. This approach likely leads to a slight underestimation of the number of dispersal events, and is thus considered here as conservative. In some species of *Calamus*, the clade of closest relatives occurred in more than the maximum number of regions, and in this case the distribution was reduced to the

regions with the highest densities of collection records (Henderson, 2020) based on the assumption that these are most likely to contain the ancestral region of a species.

***Ancestral range estimation***

Probabilistic inference of ancestral geographic ranges was conducted for the six models DEC, DEC+J, DIVALIKE, DIVALIKE+J, BAYAREALIKE, and BAYAREALIKE+J using BioGeoBEARS v. 1.1.2 (Matzke, 2013; Matzke, 2014) first as unconstrained analyses and then with various modifications of the distance and connectivity parameters. Of note, a previous critique of the DEC+J model (Ree and Sanmartín, 2018) has been thoroughly refuted (Matzke, 2021).

In BioGeoBEARS models implementing distance, dispersal probability d is calculated as d_actual = d_base * distance ^ x, where x is a free parameter (http://phylo.wikidot.com/biogeobears#distance_matrix2; last accessed on 14 October 2021). If x is estimated to be 0, distance has no effect on dispersal probability. Positive values would signify that dispersals are more likely with increasing distance, whereas negative values would signify that dispersals are less likely with increasing distance. For example, a value of x = -2 can be interpreted as dispersals between regions 10 times as distant having a 100 times (10 ^ -2 = 0.01) lower dispersal probability. Following Landis et al. (2013), and because of substantially fluctuating coastlines in the recent and distant geological past (Voris, 2000; Hall, 2012b; Morley, 2018), distances between biogeographic regions were calculated based on region centroids rather than shortest distances between their land masses. Geographical coordinates of region centroids were retrieved in Google Earth Pro v. 7.3.3.7786 (www.google.com/earth) and rounded to one degree latitude and

longitude. For Australia, the centre bottom of the Gulf of Carpentaria, located in the Northeast of the continent, was used instead of the region centroid because of the large size of the continent, and because extant calamoid palms are constrained to the Northeast of Australia. Pairwise great circle distances between regions were calculated in the R package geosphere v. 1.5.10 (Hijmans et al., 2019) using the function distVincentyEllipsoid (Supplementary Table 4.1). Following Klaus and Matzke (2019), all distances were divided by the smallest observed distance to obtain relative distance matrices for use in the analyses.

Differences in connectivity between islands were implemented as dispersal multipliers. In BioGeoBEARS, dispersal multipliers modify the dispersal probability d as a multiplier with an optional exponential parameter w, with d_actual = d_base * dispersal_multiplier ^ w (http://phylo.wikidot.com/biogeobears#dispersal_multipliers; last accessed on 14 October 2021). Depending on the tested model, w was either fixed to 1 (the default) or set to be free. Regions between which a continuous land connection existed at sea levels of 100m below present (Voris, 2000) were assigned a multiplier of 1 and regions that were not connected by land assigned a multiplier of 0.1 (Supplementary Table 4.2). If w is fixed to 1, this corresponds to the assumption that dispersal between regions that remained isolated during low sea levels was ten times less likely than between regions that were connected. If w was set free, this probability was estimated by the model. For example, an estimated parameter w = 0.5 can be interpreted as dispersals between isolated regions as three times less likely (0.1 ^ 0.5 = 0.32) compared to dispersals between connected regions.

To take the dynamic geological history of the Asia-Pacific into account, additional analyses that implemented changing distances and connectivity between the regions through the past 125 million years were conducted (Supplementary Tables 4.1 - 4.2). This

time interval was divided into eight bins, with five bins of 10 million years from the present to 50 million years ago, and older ages divided into three 25-million-year bins to take into account the increasing uncertainty of reconstructions of the location and connectivity of regions in the deep geological past. Connectivity between biogeographic regions in the past was inferred from paleogeographic reconstructions (Hall, 2009, 2012b, 2017; Morley, 2018). Distances for the first time bin (0 - 10 Ma) were estimated as described above, i.e. rounded to 1 degree, whereas distances for older time bins were based on centroid coordinates retrieved from geological reconstructions of the Asia-Pacific (Hall, 2017), rounded to 5 degrees. To allow direct comparison with time-uniform models (i.e. models based on present distances only), distances were always estimated for the more recent end of the time bin. Borneo, the Philippines, Sulawesi and the Moluccas are complex amalgamates of smaller regions with different geological origins that were assembled through time (Hall, 2017). The positionally most stable component of these regions, i.e. Southwest Borneo, the Northwest Philippines, West Sulawesi and the South Moluccas, was chosen for times prior to the coalescence of the different main constituents to the current setting, which occurred ca. 5 Ma in the Philippines and the Moluccas, ca. 10 Ma in Sulawesi, and ca. 85 Ma in Borneo.

Overall, 66 biogeographical reconstructions were conducted for each of the three main analyses (with maximum range occupancy of two, three, or four regions), resulting in a total of 198 biogeographical analyses. For each main analysis, the model best explaining the data was inferred by comparison of their likelihoods using the Akaike information criterion (AIC). Note that models based on the different data of the three main analyses are not directly comparable in this way. For the best model explaining the dataset closest to the real distribution of extant calamoid species, i.e. the main analysis based on a maximum area occupancy of four regions, the number of dispersal events between biogeographic regions

was estimated in BioGeoBEARS using biogeographic stochastic mapping with 50 replicates for probabilistic inference of ancestral ranges (Dupin et al., 2017). For each biogeographical region, the number of immigration events was calculated as the sum of all dispersal events into the region, and the number of emigration events was calculated as the sum of all dispersal events out of the region. *In situ* diversification, i.e. the number of species that originated within a region, was inferred as the difference between the total number of species in a region and the number of immigration events.

## 4.3 Results

### 4.3.1 DEC+J best model under all scenarios incorporating distance or connectivity

Across the three main analyses with maximum range occupancy of two, three or four regions, the model DEC+J with implementation of the parameters distance and connectivity (with a free parameter w), and assuming time uniformity (i.e. no time stratification), explained the data best (Table 4.1, Supplementary Tables 4.3 - 4.4). DEC+J was identified as the best of the six compared model classes (DEC, DEC+J, DIVALIKE, DIVALIKE+J, BAYAREALIKE, BAYAREALIKE+J) for 31 of the 33 different combinations of the parameters maximum range occupancy, distance, connectivity, free connectivity parameter, and time stratification. Only in the most simple scenarios (models not taking distance or connectivity into account) with a maximum range occupancy of three or four regions, BAYAREALIKE+J had the highest statistical support (Table 4.1, Supplementary Table 4.4).

Models implementing connectivity had slightly higher likelihoods if the parameter w was set to free compared to when it was fixed, but models that implemented distance had considerably higher likelihoods than any models only implementing connectivity. Models

implementing both distance and connectivity with a free parameter w had the highest likelihoods. When comparing models with identical parameters except the factor time (time-stratified vs. time-uniform), time-uniform analyses had higher likelihoods for all DEC, DEC+J, DIVALIKE, DIVALIKE+J models, and most BAYAREALIKE and BAYAREALIKE+J models.

Because models of the three main analyses are not comparable as they are based on different datasets constrained to a maximum range of two, three or four regions, the best model most closely approaching real species distributions, i.e. with a maximum range of four regions, was selected for subsequent analyses.

**Table 4.1. Comparison of 66 biogeographic models for the distribution of the Calamoideae based on maximum area occupancy of four regions.** Dist. = distance dependence, Con. = connectivity dependence, w free = implementation of a free connectivity parameter w, lnL = natural logarithm of likelihood, np = number of parameters, d = probability of dispersal (range expansion), e = probability of extirpation (range reduction), j = weight of jump dispersal, x = distance parameter, w = connectivity parameter, AIC = Akaike information criterion. AIC weight calculated across all models. Model best explaining the data in bold.

| Dist. | Con. | w free | Time | Base model | lnL | np | d | e | j | x | w | AIC | AIC weight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| no | no | no | no | DEC | -1414.0 | 2 | 0.01 | 0.01 | 0 | 0 | 0 | 2832 | 0% |
| | | | | DEC+J | -1090.0 | 3 | 0.002 | 1E-12 | 0.013 | 0 | 0 | 2187 | 0% |
| | | | | DIVALIKE | -1412.0 | 2 | 0.01 | 0.01 | 0 | 0 | 0 | 2827 | 0% |
| | | | | DIVALIKE+J | -1412.0 | 3 | 0.01 | 0.01 | 0.0001 | 0 | 0 | 2829 | 0% |
| | | | | BAYAREALIKE | -1571.0 | 2 | 0.01 | 0.01 | 0 | 0 | 0 | 3146 | 0% |
| | | | | BAYAREALIKE+J | -1067.0 | 3 | 0.0008 | 0.014 | 0.011 | 0 | 0 | 2139 | 0% |
| yes | no | no | no | DEC | -1008.0 | 3 | 0.029 | 0.0025 | 0 | -2.35 | 0 | 2022 | 0% |
| | | | | DEC+J | -947.8 | 4 | 0.021 | 1E-12 | 0.15 | -2.3 | 0 | 1904 | 0% |
| | | | | DIVALIKE | -1025.0 | 3 | 0.03 | 0.0016 | 0 | -2.17 | 0 | 2056 | 0% |
| | | | | DIVALIKE+J | -966.1 | 4 | 0.023 | 1E-12 | 0.14 | -2.25 | 0 | 1940 | 0% |
| | | | | BAYAREALIKE | -1067.0 | 3 | 0.025 | 0.028 | 0 | -2.61 | 0 | 2140 | 0% |
| | | | | BAYAREALIKE+J | -976.0 | 4 | 0.017 | 1E-07 | 0.2 | -2.23 | 0 | 1960 | 0% |
| yes | no | no | yes | DEC | -1011.0 | 3 | 0.023 | 0.0018 | 0 | -2.16 | 0 | 2028 | 0% |
| | | | | DEC+J | -952.3 | 4 | 0.017 | 1E-12 | 0.12 | -2.11 | 0 | 1913 | 0% |
| | | | | DIVALIKE | -1029.0 | 3 | 0.025 | 0.0011 | 0 | -2.03 | 0 | 2063 | 0% |
| | | | | DIVALIKE+J | -969.9 | 4 | 0.018 | 1E-12 | 0.11 | -2.04 | 0 | 1948 | 0% |
| | | | | BAYAREALIKE | -1070.0 | 3 | 0.019 | 0.029 | 0 | -2.4 | 0 | 2145 | 0% |
| | | | | BAYAREALIKE+J | -983.0 | 4 | 0.013 | 1E-07 | 0.15 | -1.98 | 0 | 1974 | 0% |
| no | yes | no | no | DEC | -1078.0 | 2 | 0.0068 | 0.0033 | 0 | 0 | 0 | 2160 | 0% |
| | | | | DEC+J | -1016.0 | 3 | 0.005 | 1E-12 | 0.036 | 0 | 0 | 2037 | 0% |
| | | | | DIVALIKE | -1091.0 | 2 | 0.0078 | 0.0023 | 0 | 0 | 0 | 2187 | 0% |
| | | | | DIVALIKE+J | -1030.0 | 3 | 0.0055 | 1E-12 | 0.034 | 0 | 0 | 2065 | 0% |
| | | | | BAYAREALIKE | -1238.0 | 2 | 0.01 | 0.01 | 0 | 0 | 0 | 2481 | 0% |
| | | | | BAYAREALIKE+J | -1042.0 | 3 | 0.0041 | 1E-07 | 0.049 | 0 | 0 | 2089 | 0% |
| no | yes | no | yes | DEC | -1079.0 | 2 | 0.0073 | 0.0034 | 0 | 0 | 0 | 2161 | 0% |
| | | | | DEC+J | -1018.0 | 3 | 0.0054 | 1E-12 | 0.038 | 0 | 0 | 2042 | 0% |
| | | | | DIVALIKE | -1091.0 | 2 | 0.0084 | 0.0024 | 0 | 0 | 0 | 2186 | 0% |
| | | | | DIVALIKE+J | -1030.0 | 3 | 0.0059 | 1E-12 | 0.036 | 0 | 0 | 2067 | 0% |
| | | | | BAYAREALIKE | -1153.0 | 2 | 0.0056 | 0.03 | 0 | 0 | 0 | 2310 | 0% |
| | | | | BAYAREALIKE+J | -1036.0 | 3 | 0.0034 | 0.0017 | 0.063 | 0 | 0 | 2078 | 0% |
| no | yes | yes | no | DEC | -1072.0 | 3 | 0.0059 | 0.0033 | 0 | 0 | 0.73 | 2150 | 0% |
| | | | | DEC+J | -1010.0 | 4 | 0.0045 | 1E-12 | 0.036 | 0 | 0.8 | 2029 | 0% |
| | | | | DIVALIKE | -1085.0 | 3 | 0.0068 | 0.0022 | 0 | 0 | 0.74 | 2175 | 0% |
| | | | | DIVALIKE+J | -1023.0 | 4 | 0.0049 | 1E-12 | 0.029 | 0 | 0.76 | 2054 | 0% |
| | | | | BAYAREALIKE | -1180.0 | 3 | 0.01 | 0.031 | 0 | 0 | 0.97 | 2366 | 0% |
| | | | | BAYAREALIKE+J | -1033.0 | 4 | 0.0036 | 1E-07 | 0.042 | 0 | 0.74 | 2075 | 0% |

**Table 4.1** (*continued*)

| Dist. | Con. | w free | Time | Base model | lnL | np | d | e | j | x | w | AIC | AIC weight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| no | yes | yes | yes | DEC | -1074.0 | 3 | 0.0064 | 0.0033 | 0 | 0 | 0.76 | 2154 | 0% |
| | | | | DEC+J | -1013.0 | 4 | 0.0048 | 1E-12 | 0.034 | 0 | 0.79 | 2034 | 0% |
| | | | | DIVALIKE | -1086.0 | 3 | 0.0075 | 0.0024 | 0 | 0 | 0.79 | 2177 | 0% |
| | | | | DIVALIKE+J | -1025.0 | 4 | 0.0053 | 1E-12 | 0.031 | 0 | 0.78 | 2058 | 0% |
| | | | | BAYAREALIKE | -1143.0 | 3 | 0.0039 | 0.032 | 0 | 0 | 0.75 | 2292 | 0% |
| | | | | BAYAREALIKE+J | -1029.0 | 4 | 0.0029 | 0.0016 | 0.052 | 0 | 0.73 | 2065 | 0% |
| yes | yes | no | no | DEC | -975.9 | 3 | 0.066 | 0.0031 | 0 | -2.39 | 0 | 1958 | 0% |
| | | | | DEC+J | -908.2 | 4 | 0.045 | 1E-12 | 0.35 | -2.3 | 0 | 1824 | 0% |
| | | | | DIVALIKE | -984.5 | 3 | 0.073 | 0.0024 | 0 | -2.32 | 0 | 1975 | 0% |
| | | | | DIVALIKE+J | -920.1 | 4 | 0.049 | 1E-12 | 0.28 | -2.29 | 0 | 1848 | 0% |
| | | | | BAYAREALIKE | -1073.0 | 3 | 0.097 | 0.028 | 0 | -2.98 | 0 | 2153 | 0% |
| | | | | BAYAREALIKE+J | -943.4 | 4 | 0.037 | 1E-07 | 0.35 | -2.27 | 0 | 1895 | 0% |
| yes | yes | no | yes | DEC | -992.6 | 3 | 0.046 | 0.003 | 0 | -1.94 | 0 | 1991 | 0% |
| | | | | DEC+J | -930.2 | 4 | 0.033 | 1E-12 | 0.26 | -1.91 | 0 | 1868 | 0% |
| | | | | DIVALIKE | -1003.0 | 3 | 0.048 | 0.0024 | 0 | -1.81 | 0 | 2012 | 0% |
| | | | | DIVALIKE+J | -941.3 | 4 | 0.035 | 1E-12 | 0.22 | -1.85 | 0 | 1891 | 0% |
| | | | | BAYAREALIKE | -1088.0 | 3 | 0.044 | 0.03 | 0 | -2.11 | 0 | 2181 | 0% |
| | | | | BAYAREALIKE+J | -963.0 | 4 | 0.024 | 1E-07 | 0.28 | -1.74 | 0 | 1934 | 0% |
| **yes** | **yes** | **yes** | **no** | DEC | -960.9 | 4 | 0.055 | 0.003 | 0 | -2.39 | 0.6 | 1930 | 0% |
| | | | | **DEC+J** | **-895.4** | **5** | **0.044** | **1E-12** | **0.33** | **-2.48** | **0.65** | **1801** | **100%** |
| | | | | DIVALIKE | -968.6 | 4 | 0.058 | 0.002 | 0 | -2.26 | 0.61 | 1945 | 0% |
| | | | | DIVALIKE+J | -906.2 | 5 | 0.041 | 1E-12 | 0.24 | -2.25 | 0.64 | 1822 | 0% |
| | | | | BAYAREALIKE | -1056.0 | 4 | 0.063 | 0.028 | 0 | -3.01 | 0.46 | 2119 | 0% |
| | | | | BAYAREALIKE+J | -927.6 | 5 | 0.031 | 1E-07 | 0.31 | -2.27 | 0.62 | 1865 | 0% |
| yes | yes | yes | yes | DEC | -975.8 | 4 | 0.039 | 0.0026 | 0 | -2.02 | 0.56 | 1960 | 0% |
| | | | | DEC+J | -913.3 | 5 | 0.028 | 1E-12 | 0.21 | -1.97 | 0.58 | 1837 | 0% |
| | | | | DIVALIKE | -984.7 | 4 | 0.044 | 0.0019 | 0 | -1.94 | 0.61 | 1977 | 0% |
| | | | | DIVALIKE+J | -923.9 | 5 | 0.03 | 1E-12 | 0.18 | -1.89 | 0.6 | 1858 | 0% |
| | | | | BAYAREALIKE | -1063.0 | 4 | 0.026 | 0.03 | 0 | -2.12 | 0.33 | 2134 | 0% |
| | | | | BAYAREALIKE+J | -943.7 | 5 | 0.021 | 1E-07 | 0.24 | -1.83 | 0.59 | 1897 | 0% |

### 4.3.2 Ancestral ranges and regional radiations

For most clades above genus level, ancestral ranges could not be resolved (Fig. 4.2). The only exceptions were subtribe Plectocomiinae and the sister clade of Plectocomiinae and Calaminae, for both of which an origin in Indochina was inferred.

In contrast, ancestral ranges at genus level and below were generally highly resolved (Fig. 4.2). *Eugeissona* was estimated to have originated in the combined area of the Thai-Malay peninsula and Borneo (both in Sundaland), and Indochina was estimated as the most likely region of origin for subtribes Salaccinae, Plectocomiinae and Calaminae. For *Salacca*,

an ancestral range spanning all of Sundaland (Indochina, Thai-Malay peninsula, Sumatra, Borneo) was inferred for the backbone nodes. The most likely ancestral area of occurrence of *Metroxylon* was Papuasia, and for *Pigafetta* the range comprising Sulawesi, the Moluccas and Papuasia. In *Calamus*, the ancestral area of occurrence of the three basalmost backbone nodes was Indochina, whereas Borneo was estimated as the most likely ancestral range for the majority of other backbone nodes.

Ancestral range estimates indicate a multitude of endemic *in situ* diversification events, especially in Indochina, Borneo, Sulawesi and Papuasia, over the last ca. 50 million years (Fig. 4.2). The first inferred radiation in the Calamoideae occurred in Indochina in subtribe Plectocomiinae around 50 Ma, leading to the formation of the genera *Plectocomia*, *Plectocomiopsis* and *Myrialepis*, and was followed by nine further radiations in *Calamus*, mostly between 30 and 10 million years ago (Ma). In Borneo, 14 radiations occurred across the phylogenetically divergent genera *Eugeissona*, *Korthalsia*, *Salacca* and *Calamus* in the interval from ca. 30 to 10 Ma. Sulawesi was the stage of five radiations in *Calamus* at 30 to 5 Ma. In Papuasia, two radiations in *Metroxylon* and *Calamus* formed between ca. 25 and 5 Ma, with the *Calamus* radiation being notable for being the largest single regional radiation across the Asian Calamoideae. Sister to the large Papuasian radiation was a smaller radiation in Australia comprising five species, which formed around 20 Ma. In the Thai-Malay peninsula, ancestral range estimations suggest two radiations in *Eugeissona* and *Salacca* between 15 and 5 Ma. In the Philippines, two radiations occurred from ca. 25 to 10 Ma in *Calamus*. In India, one radiation comprising two species occurred at ca. 5 Ma in *Calamus*. No radiations were observed in Sumatra, Java, Palawan, and the Moluccas.

These regional radiations often were the point of origin of dispersals into adjacent regions. From Indochina, dispersals mainly occurred into India and the Thai-Malay peninsula. Borneo was the origin of a multitude of dispersals into all neighbouring regions, i.e. the Thai-Malay peninsula, Sumatra, Java, Palawan, the Philippines, and Sulawesi. From Sulawesi, several dispersals into the Philippines and Moluccas occurred. In contrast, the Papuasian and Australian radiations of *Calamus* were revealed to be fully endemic clades without dispersals to other regions.

Despite the generally observed strong regional clustering in the Asian Calamoideae, some exceptions from this pattern were observed. The small clade defined by *Calamus erinaceus* and *C. andamanicus*, sister to all other species of the genus and represented here with 10 species, spans the entire range of the genus in the Asia-Pacific region, with almost every species occurring in a different region. Although *Calamus deerratus*, the only African species of the genus, was not included in the formal ancestral range estimates (see Methods), its phylogenetic position as sister to the clade defined by *C. diepenhorstii* and *C. acamptostachys* (Fig. 3.4) suggests an origin in Borneo because this region was identified with a high probability as the ancestral range of both the crown and stem node of this clade (Fig. 4.2).

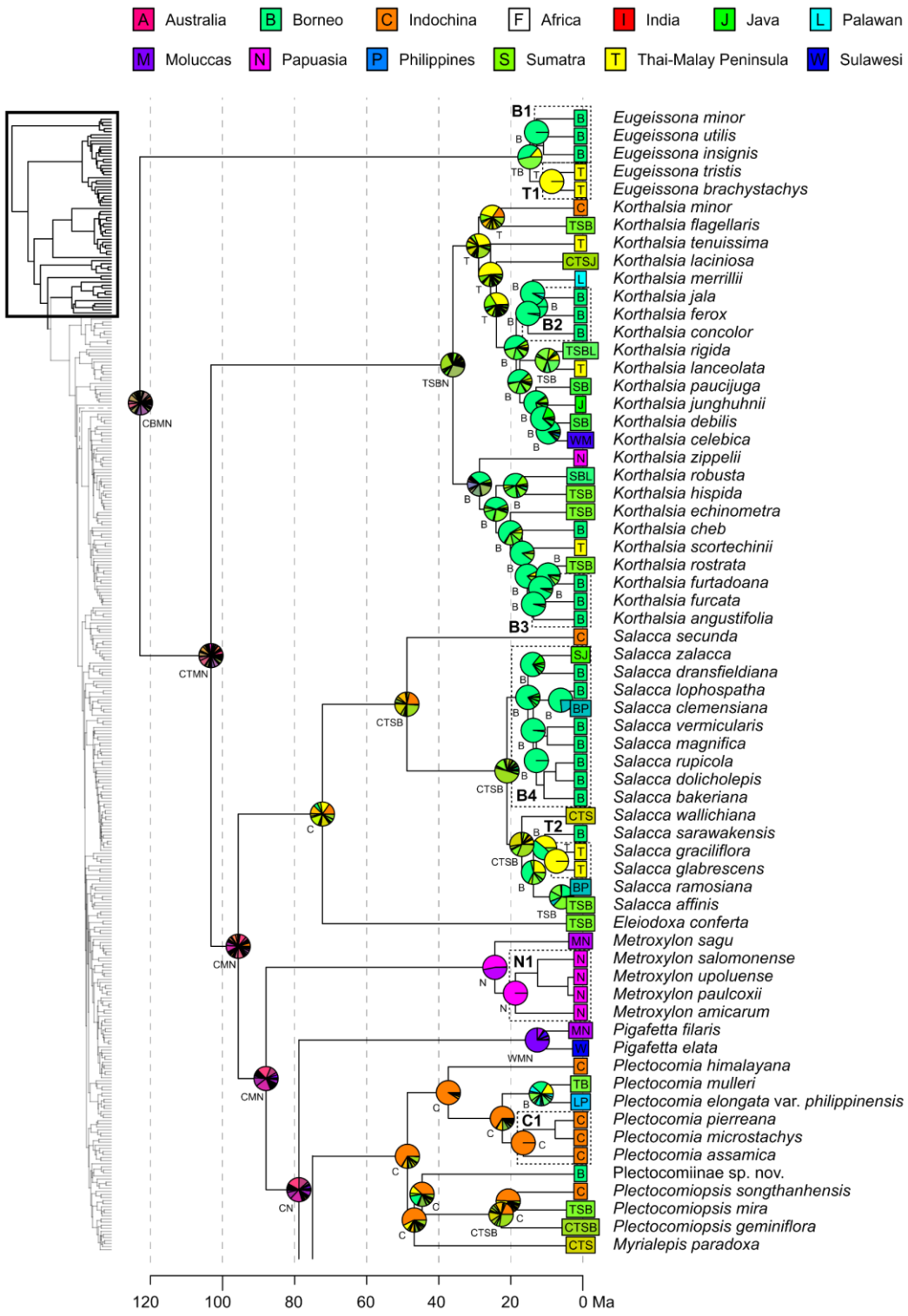



**Figure 4.2. Ancestral range estimates for 390 species and subspecies of the Calamoideae in the Asia-Pacific.** Ancestral range estimates based on model DEC+J with maximum range occupancy of four regions, implementation of the parameters distance and connectivity (with a free parameter w), time-uniform. Boxes at the tips of the phylogeny indicate extant species distributions and pies indicate inferred ancestral ranges, with pie section sizes corresponding to probabilities. The most probable ancestral range for each node is indicated next to each node.

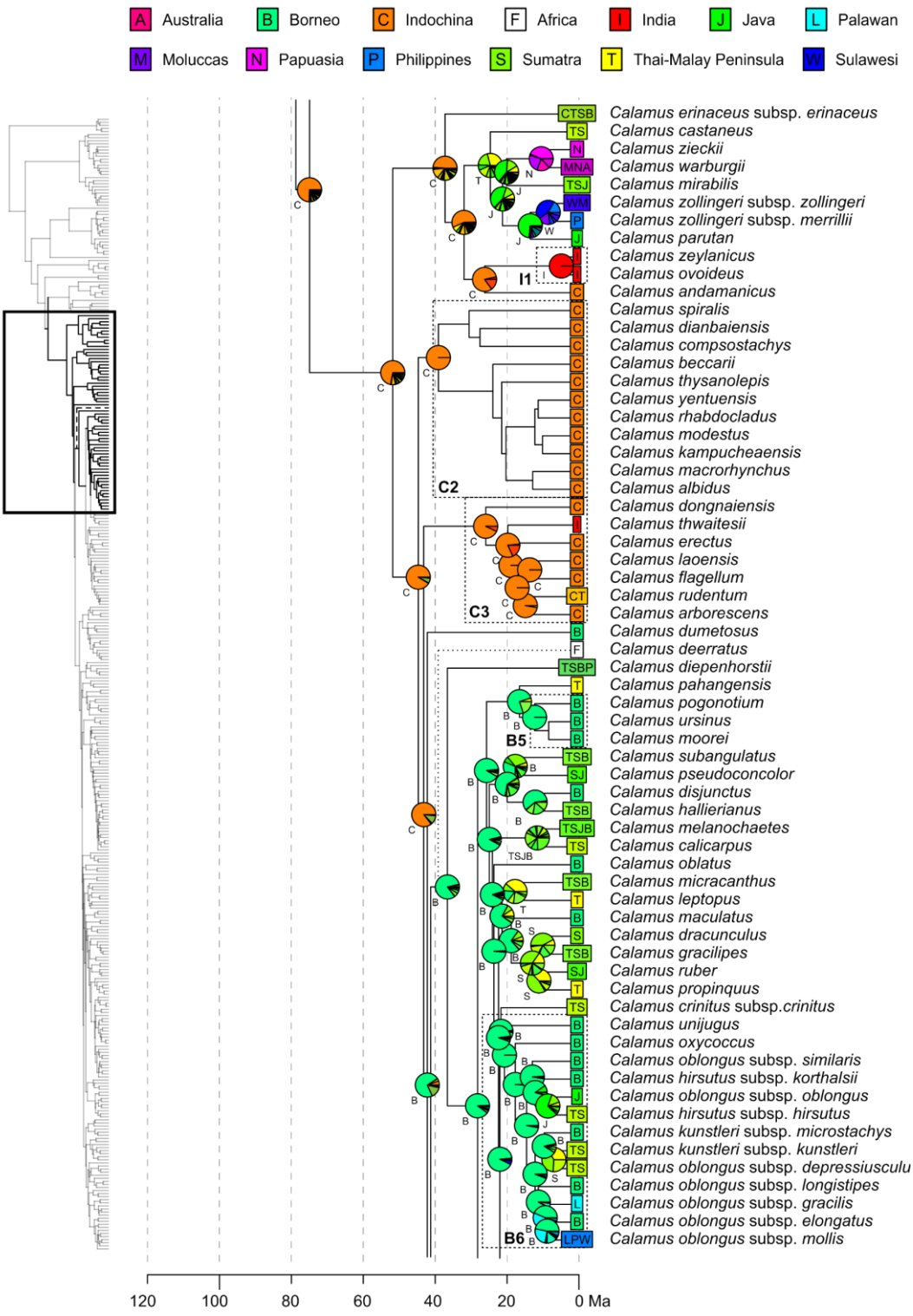



**Figure 4.2** (*continued*). The phylogenetic position of *Calamus deerratus*, the only African species of the genus, is indicated by a dotted line as it was not included in ancestral range estimations. Local radiations are indicated by dashed outlines and numerated, e.g. B1 – B14 indicate radiations on Borneo. Ma = millions of years ago.

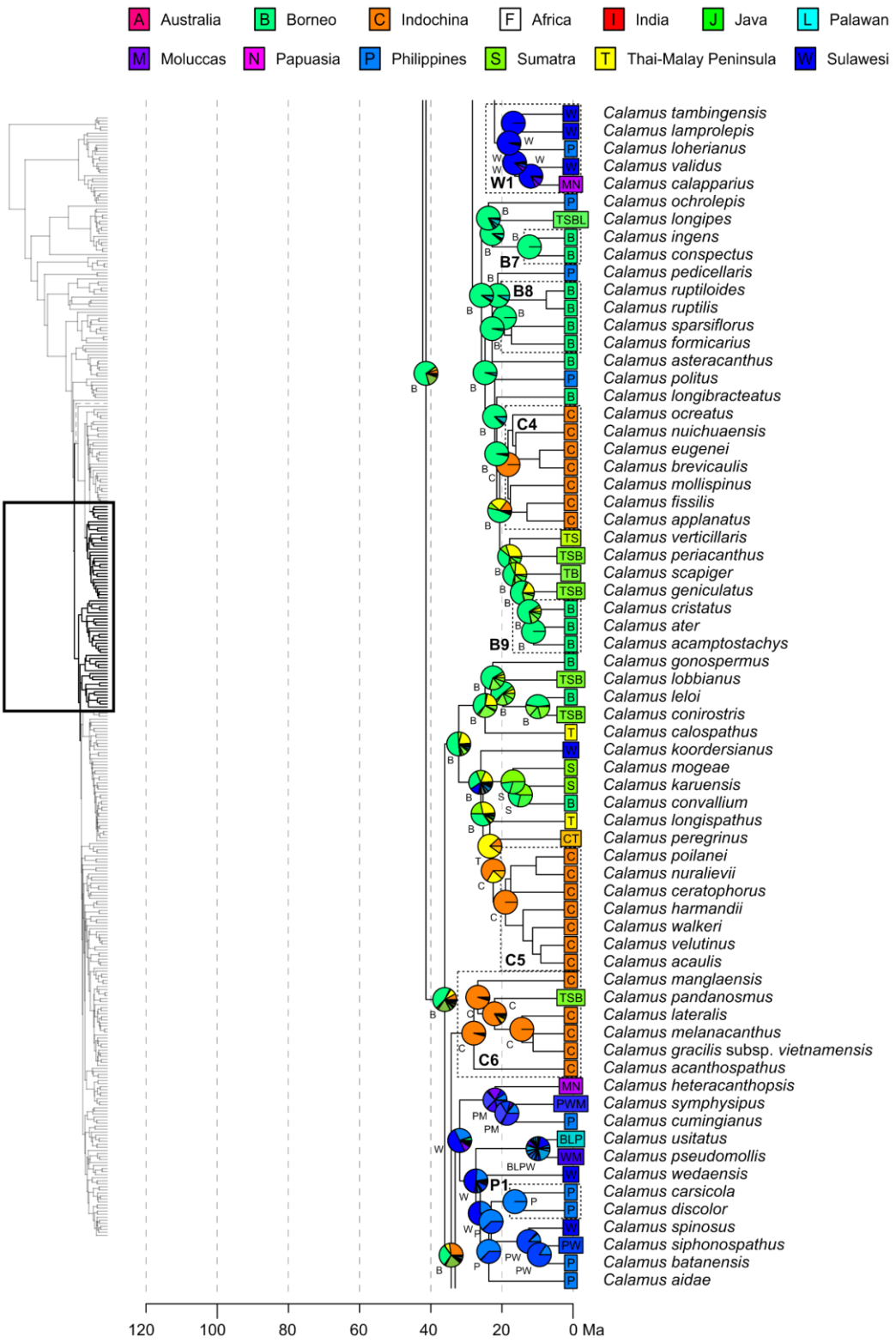



**Figure 4.2** (*continued*)




**Figure 4.2** (*continued*)

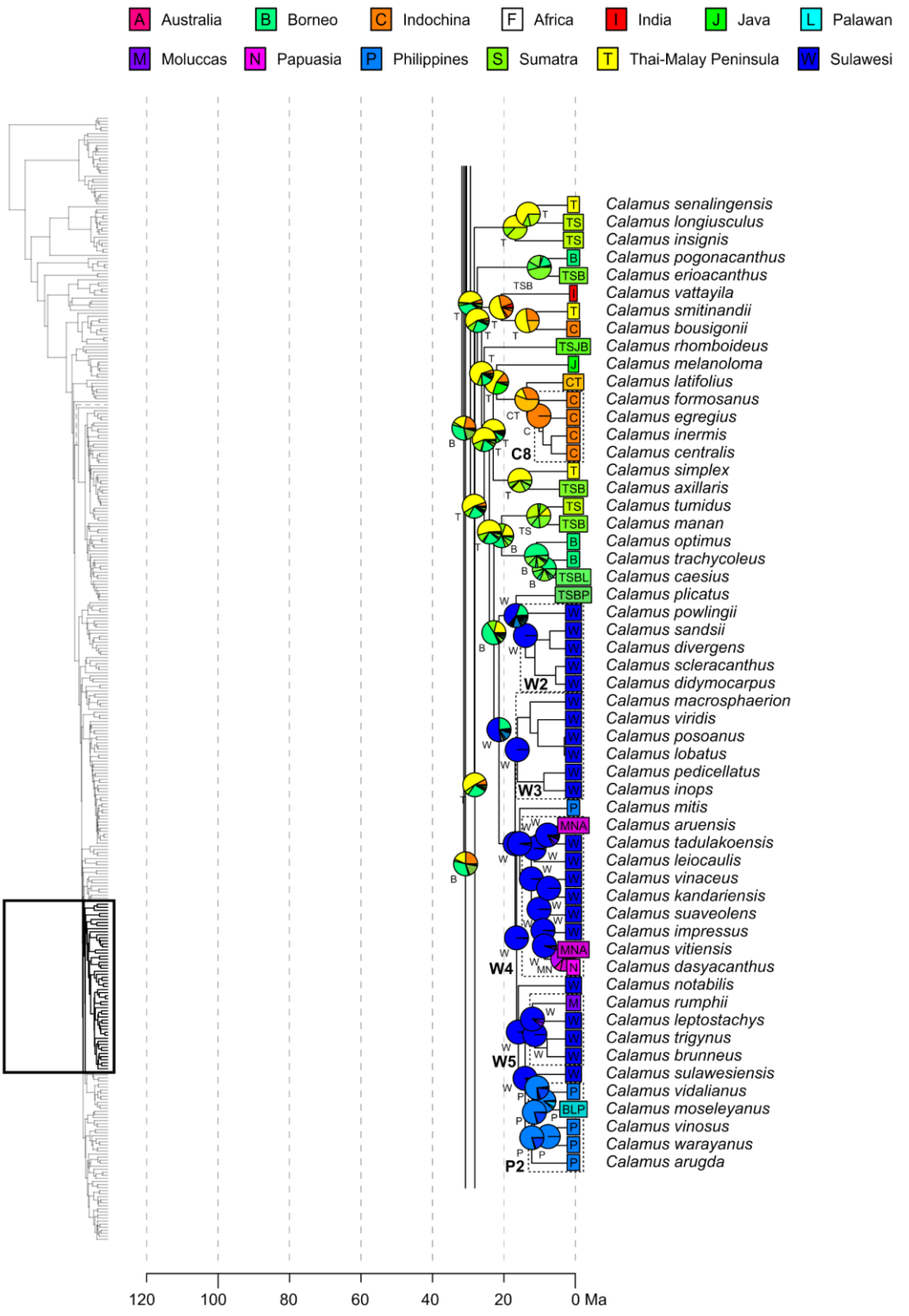



**Figure 4.2** (*continued*)

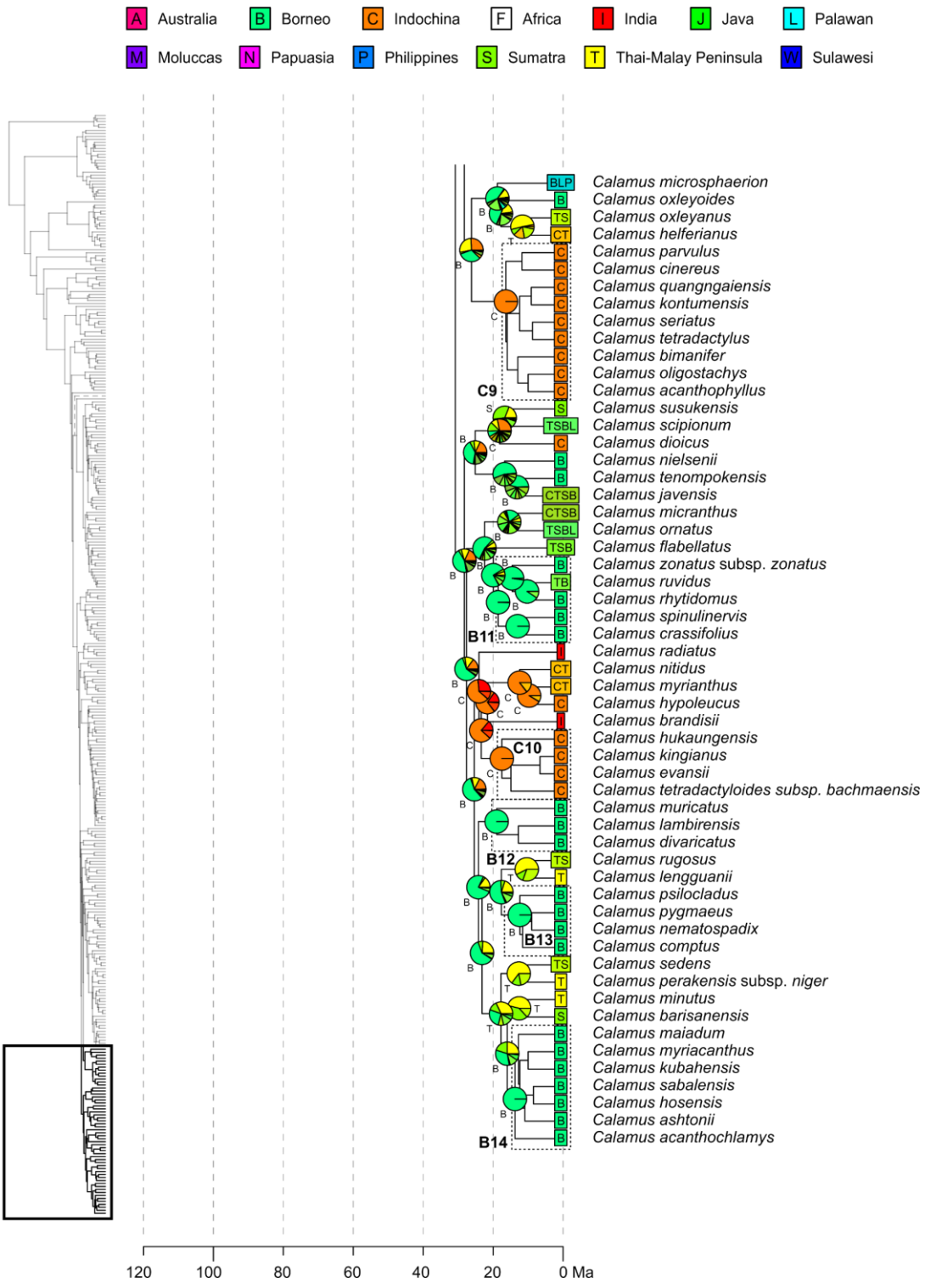



**Figure 4.2** (*continued*)

### 4.3.3 Dispersals patterns across the Asia-Pacific

To reveal overarching biogeographical patterns of dispersal, the number and direction of dispersal events between regions across all Asian Calamoideae were inferred from ancestral range estimations using biogeographical stochastic mapping (Table 4.2).

**Table 4.2. Dispersal of the Calamoideae between biogeographic regions in the Asia-Pacific.** Dispersal counts were averaged across 50 biogeographic stochastic mappings based on ancestral range estimates of 369 species and subspecies of the Calamoideae (Fig. 4.2). I = India, C = Indochina, T = Thai-Malay Peninsula, S = Sumatra, J = Java, B = Borneo, L = Palawan, P = Philippines, W = Sulawesi, M = Moluccas, N = Papuasia, A = Australia. Values of at least 1 are highlighted in bold.

| | | **Dispersed to** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **I** | **C** | **T** | **S** | **J** | **B** | **L** | **P** | **W** | **M** | **N** | **A** |
| **Dispersed from** | **I** | 0 | 0.50 | 0.02 | 0.02 | 0.04 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | **C** | **4.42** | 0 | **12.26** | **2.48** | 0.98 | **4.14** | 0.22 | 0.34 | 0.10 | 0.04 | 0.32 | 0.18 |
| | **T** | **1.00** | **10.00** | 0 | **32.36** | **2.82** | **14.20** | 0.92 | 0.46 | 0.54 | 0.12 | 0.32 | 0.16 |
| | **S** | 0.02 | **2.24** | **28.44** | 0 | **5.06** | **12.46** | 0.62 | 0.24 | 0.36 | 0.10 | 0.24 | 0.06 |
| | **J** | 0.06 | 0.24 | 0.92 | **2.94** | 0 | **1.12** | 0 | 0.24 | 0.60 | 0.20 | 0.50 | 0.04 |
| | **B** | 0.24 | **4.32** | **22.66** | **21.06** | **4.44** | 0 | **9.68** | **7.14** | **4.24** | 0.22 | 0.64 | 0.02 |
| | **L** | 0 | 0 | 0.08 | 0.04 | 0 | **1.38** | 0 | **3.52** | 0.74 | 0 | 0 | 0 |
| | **P** | 0 | 0 | 0.04 | 0.02 | 0 | 0.96 | **2.46** | 0 | **2.34** | 0.80 | 0.04 | 0 |
| | **W** | 0 | 0 | 0.08 | 0.16 | 0.08 | 0.82 | 0.46 | **4.88** | 0 | **6.80** | 0.58 | 0.08 |
| | **M** | 0 | 0.08 | 0 | 0 | 0.06 | 0.06 | 0.02 | 0.72 | **1.74** | 0 | **4.16** | 0.44 |
| | **N** | 0.02 | 0.08 | 0.02 | 0.06 | 0 | 0.04 | 0 | 0 | 0.20 | **2.06** | 0 | **3.02** |
| | **A** | 0 | 0.02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.26 | 0.92 | 0 |

The majority of dispersal events occurred between the regions West of Wallace's line (India, Indochina, Thai-Malay peninsula, Sumatra, Java, Borneo, Palawan), which were connected by intermittent land connections in the recent geological past (Fig. 4.3). In particular, this is true for the Thai-Malay peninsula, Sumatra and Borneo, with many

dispersals inferred between all three regions, and the Thai-Malay peninsula and Sumatra showing the closest connection (i.e. highest numbers of dispersal events) across the Asia-Pacific. In contrast, dispersals between regions that were never connected by land were linked by much fewer dispersal events, even if they were geographically very close (e.g., Borneo and Sulawesi). Similarly, very few dispersal events were inferred between Papuasia (New Guinea and surrounding islands) and Australia despite being separated by only shallow seas.

In addition to connectivity and distance, a predominance of dispersals away from the geographical centre of the Asia-Pacific, i.e. Borneo, and into peripheral regions, can be noted (Fig. 4.3). The majority of dispersals between Borneo, Sulawesi, the Moluccas and Papuasia were consistently from west to east, whereas there were more westbound dispersals from Indochina to India than in the other direction. In the North, the connection of Palawan and the Philippines with other regions was dominated by northbound dispersals, and in the South, there were more southward dispersals into Java and Australia than in the other direction. The only cases with similar numbers of dispersals in both directions between regions were Indochina - Thai-Malay peninsula, Indochina - Borneo, and Thai-Malay peninsula - Sumatra.

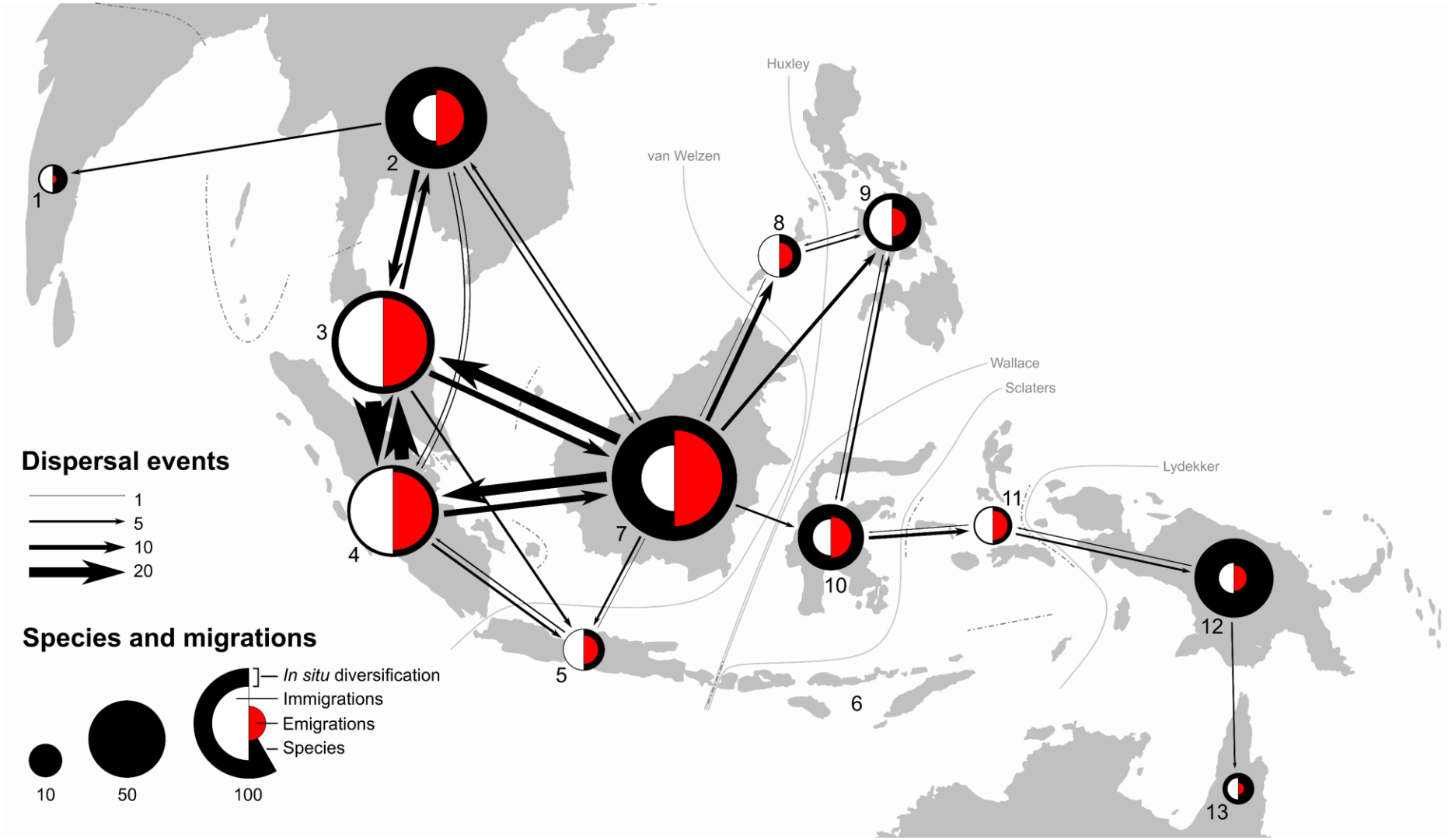



**Figure 4.3. Migration events of the Calamoideae in the Asia-Pacific.** (*Figure caption on next page.*)

**Figure 4.3. Migration events of the Calamoideae in the Asia-Pacific.** (*Figure on previous page*.) Dispersal events between biogeographic regions were inferred from ancestral range estimates for 369 species and subspecies of the Calamoideae (Fig. 4.2) using biogeographic stochastic mapping (Table 4.2). Arrow width indicates the number of inferred dispersals between two regions, with dispersal frequencies < 1 not shown (note that dispersal frequencies > 0 but < 1 are commonly inferred using stochastic mapping). Selected biogeographic boundaries that have been proposed for the Asia-Pacific redrawn from Simpson (1977) and (van Welzen et al., 2011). Numbers indicate regions (see Chapter 4.2.2). Where ambiguous, region boundaries are denoted by interrupted lines.

Finally, more dispersals generally occurred from species-rich regions (e.g., Borneo, Indochina, Papuasia) to species-poor regions (e.g., Sulawesi, India, Australia), with species richness often loosely correlated with region size (Fig. 4.3). In particular, there were considerably more dispersals out of Borneo than into Borneo. However, this trend was reversed in some cases. For example, more dispersals were inferred from Palawan to the Philippines than in the other direction, and more dispersals from the Moluccas to Papuasia than opposite.

### 4.3.4 Differing roles of biogeographic regions in shaping the distribution of the Calamoideae in the Asia-Pacific

To identify the contributions of the different regions to the rich diversity of the Calamoideae across the Asia-Pacific, the proportions of *in situ* diversification and emigration (i.e. dispersal events from a region into other regions) relative to the number of species included in this study were therefore put in relation for each region (Fig. 4.4). On this basis, four categories were identified:

First, regions with high *in situ* diversification and a high proportion of emigrations ("radiators") signify local radiations with a key role in influencing species assemblages of other areas. Borneo is the only region that falls into this category.

Second, regions with high *in situ* diversification but a low proportion of emigrations ("incubators") characterise local radiations that happen largely in isolation and therefore have limited influence on other regions. The most extreme example of an incubator is Papuasia, largely consisting of a single radiation of *Calamus* solely constrained to this region. Other incubators identified here are Indochina and Sulawesi.

Third, regions with little *in situ* diversification but a high proportion of emigrations ("corridors") have an important role in connecting different regions despite limited local radiation. The Thai-Malay peninsula, Sumatra, and the Moluccas are therefore corridors.

Fourth, regions with little *in situ* diversification and a low proportion of emigration events ("containers") are mainly shaped by immigration events and exert little influence on other regions. India, Palawan and the Philippines fall into this category. Australia and Java are both in intermediate positions, with Australia taking the role of a "container incubator" (with intermediate *in situ* diversification and low emigration), and Java the role of a "container corridor" (with low *in situ* diversification and intermediate emigration).

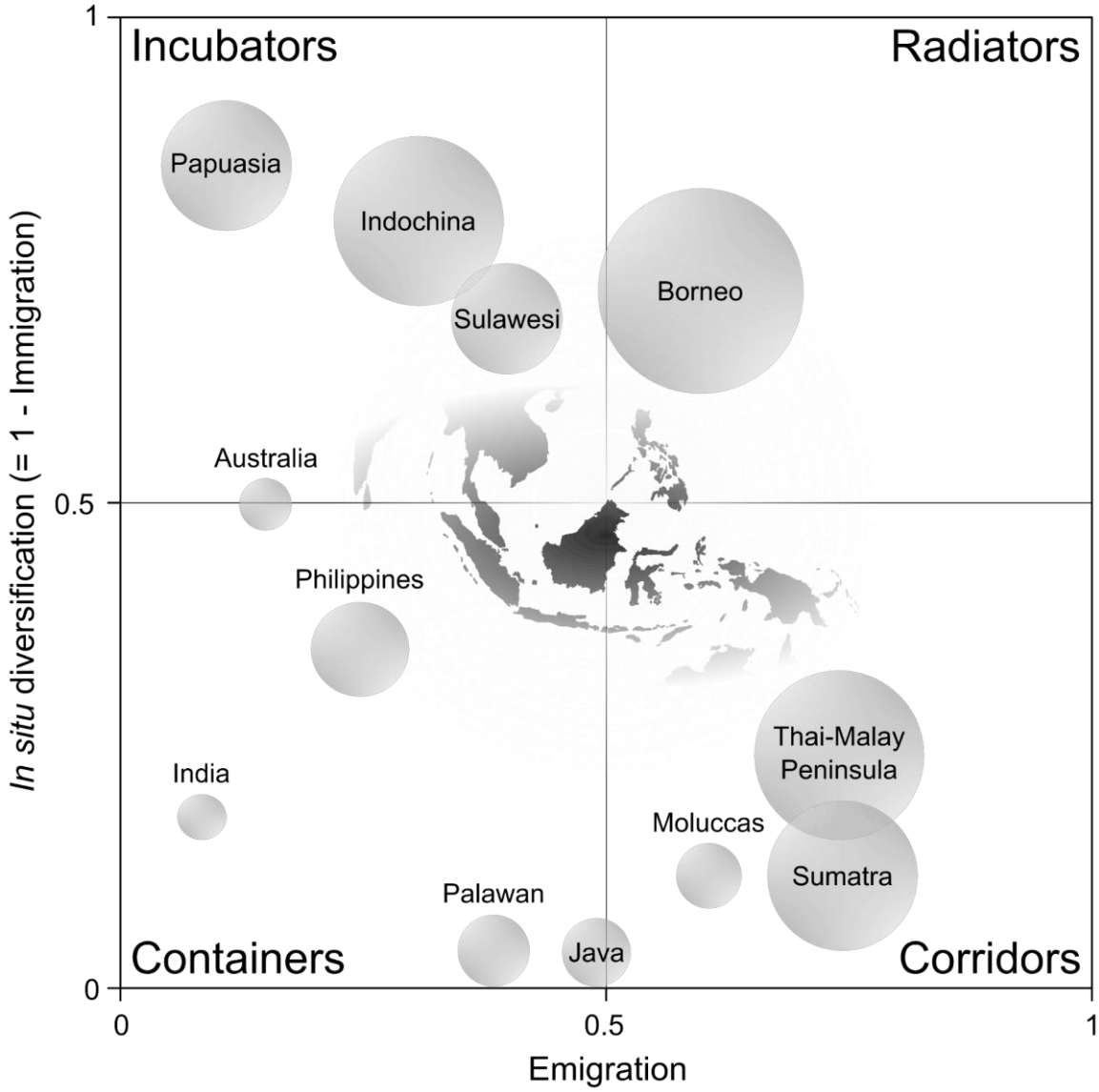



**Figure 4.4. Role of Asian-Pacific regions in the *in situ* diversification and emigration of the Calamoideae.** Proportions of *in situ* diversification (= 1 - immigration) and emigration relative to the number of species in a region, as inferred from ancestral range estimates for 369 species and subspecies of the Calamoideae (Fig. 4.2) using biogeographic stochastic mapping (Table 4.2).

## 4.4 Discussion

### 4.4.1 Biogeographic history of the Asian-Pacific Calamoideae

This study offers for the first time detailed insights into the biogeographic history of the Calamoideae in the Asia-Pacific and allows the validation of previous hypotheses. Mostly based on distribution and diversity patterns but also taking into account fossil and phylogenetic information, previous studies hypothesised the origin of most genera in Sundaland (with the exception of *Metroxylon* and *Pigafetta*, which only occur in Wallacea and Papuasia), followed by eastward dispersal across Malesia in the more widespread genera *Calamus* and *Korthalsia* (Dransfield, 1981a, 1987; Baker et al., 1998; Baker and Dransfield, 2000; Baker and Couvreur, 2012; Baker and Couvreur, 2013a). However, this hypothesis is contrasted by the continuous fossil record of *Dicolpopollis* in Australia from the Cenomanian onwards (Macphail and Jordan, 2015; Macphail and Hill, 2018). While the ancestral ranges of calamoid clades above genus level and of *Korthalsia* remain unresolved, the results presented here support the hypothesis of an origin of *Calamus* in Indochina and provide highly refined estimates of ancestral ranges within all calamoid different genera (Fig. 4.2).

Most previous studies assumed that the eastward dispersal of *Calamus* and *Korthalsia* out of Sundaland and across Malesia had taken place after the collision of the Australian and Asian plates around 25 Ma (Dransfield, 1981a, 1987; Baker et al., 1998; Baker and Dransfield, 2000; Baker and Couvreur, 2012). The divergence times and ancestral ranges estimated in this study are largely consistent with this theory because of the origin of Calamus in Indochina around 50 Ma and because almost all dispersal events out of Sundaland are younger than 25 million years. The only exception is one dispersal event into

Sulawesi that may have occurred before 30 Ma, giving rise to the radiation including *Calamus symphysipus* and *C. aidae* (Fig. 4.2), but may be explained by the close proximity of West Sulawesi to Sundaland already before the collision of Sunda and Sahul (Hall, 2017).

The relatively young crown node ages of the east Malesian genera *Metroxylon* and *Pigafetta* (Chapter 3, Table 3.6) preclude the inference of their historic biogeography before the collision of Australia and Asia. Previous hypotheses that either proposed a potential connection of their current distribution with the Eocene opening of the Makassar strait between Borneo and Sulawesi (Baker and Couvreur, 2012) or regarded dispersal of their ancestors out of Sundaland in the Late Tertiary as improbable (Dransfield, 1987) therefore can neither be confirmed nor disputed.

Biogeographic analyses at the genus level indicated that dispersals out of Sundaland and into east Malesia may have happened in the Calamoideae as early as 50 Ma and therefore long before the collision of Australia and Asia (Baker and Couvreur, 2013a). These older inferences based on limited sampling are refuted here, whereas the ancestral ranges of the oldest backbone nodes are unresolved (Fig. 4.2). The extent of dispersals between the Australian and Asian plates before their collision therefore remains unclear, although the presence of *Dicolpopollis* fossils in Australia from the Cenomanian onwards (Macphail and Jordan, 2015; Macphail and Hill, 2018) suggests some dispersal between both regions long before their collision. A proposed scenario of dispersal out of Australia, followed by subsequent extinction in Australia and later re-colonisation from Asia (Baker and Dransfield, 2000), would be compatible with the available evidence but remains speculative.

### 4.4.2 Strong biogeographic significance of Indian and Australian collisions with Asia

It has been suggested that the collisions of India and Australia, respectively, with Asia may have facilitated the exchange of biota between these previously isolated regions (Kooyman et al., 2019). For *Eugeissona*, this hypothesis is supported by the fossil record, with the oldest fossils occurring in India and only appearing in Asia around the time of the collision of both land masses (Morley, 1998), although extant species of Eugeissona only occur on the Thai-Malay peninsula and Borneo. In the Calameae, almost all dispersal events between Sundaland, the Philippines, Wallacea, and Papuasia occurred after 25 Ma, supporting the collision of Asia and Australia as a geological event of outstanding biogeographic significance.

### 4.4.3 Differing roles of land connectivity within Sundaland and Sahul

The high number of bi-directional dispersal events between the Thai-Malay Peninsula, Sumatra and Borneo (Fig. 4.3), as well as the broad distributions of many calamoid species spanning all these regions (Fig. 4.2), support the hypothesis that extensive land exposed at lower sea levels facilitated dispersal across Sundaland (Kooyman et al., 2019) and that continuous wet tropical rainforests may have connected these regions (Cannon et al., 2009; Morley, 2018). These findings further challenge the hypothesis that the sandy sea-bed soils between western Sundaland (Thai-Malay Peninsula and Sumatra) and Borneo may have acted as a strong dispersal barrier for tropical rainforest plants (Slik et al., 2011). Further, the inference of a multitude of dispersals between Indochina and the Thai-Malay peninsula but much fewer dispersals between Indochina and the islands of Sundaland (Sumatra, Java, Borneo) (Fig. 4.3) indicates that dispersals between mainland Asia and the islands of

Sundaland were largely constrained to the corridor of the Thai-Malay peninsula, supporting paleogeographic reconstructions of the region (Morley, 2018). For Java and Palawan, the smaller number of migration events and strong prevalence of immigration over emigration may be explained by their submergence during parts of the Miocene and Pliocene (Hall, 2009, 2012b; de Bruyn et al., 2014; Morley, 2018) but could also indicate much lower connectivity to other parts of Sundaland. Overall, the largely balanced number of migration events between Indochina, the Thai-Malay Peninsula, Sumatra, and Borneo shows that dispersals within most of Sundaland were easily possible and suggests that the widespread ranges of many species may be explained largely by their formerly continuous distribution across Sundaland during low sea levels and range contractions during times of high sea levels.

In Sahul, very few dispersal events between Papuasia (with most calamoid species occurring in New Guinea) and Australia were inferred despite the shallow sea levels comparable to Sundaland, and despite the large pool of species in New Guinea. In fact, fewer dispersal events occurred between Papuasia and Australia than between Papuasia and the Moluccas (Fig. 4.3). Moreover, dispersal between Papuasia and Australia was inferred as strongly dominated by migration into Australia. This hints at a stronger barrier to dispersal of tropical rainforest plants between both regions than previously assumed (Kooyman et al., 2019). However, the apparent lack of dispersals may also be explained by the severe contraction of tropical rainforest area in the recent past of Australia (Morley, 2000) that may have led to severe extinction in the Calamoideae. This interpretation is supported by the paradox that all Australian calamoid lineages are of recent age (Fig. 4.2) despite the continuous record of calamoid fossils in Australia since the Late Cretaceous (Truswell and Owen, 1988; Dransfield et al., 2008; Macphail and Jordan, 2015; Macphail and Hill, 2018).

### 4.4.4 van Welzen's line demarcates contrasting dispersal frequencies in the Asia-Pacific

The number of dispersal events between regions allows some assessment of the strength of the multitude of biogeographic boundaries that have been previously proposed (Wallace, 1860; Huxley, 1868; Lydekker, 1896; Sclater and Sclater, 1899; Simpson, 1977; van Welzen et al., 2011). Huxley's line (Huxley, 1868), a modification of Wallace's line (Wallace, 1860) that excludes the Philippines and essentially characterises the eastern boundary of Sundaland, and Lydekker's line (Lydekker, 1896), delineating the western edge of the Sahul shelf, have been regarded as most meaningful because of their geological significance (Simpson, 1977).

For the Calamoideae, van Welzen's line (van Welzen et al., 2011), running between Sumatra and Java, Borneo and Sulawesi, and Borneo and Palawan, indicates a strong demarcation of dispersal frequencies (Fig. 4.3). West of van Welzen's line, a total of 173 dispersals between regions were inferred, strongly contrasting with only 36 dispersals between regions east of the line (Table 4.2). East of van Welzen's line, dispersal frequencies across proposed biogeographic boundaries strongly declined further. Whereas 98 dispersals were inferred across the line separating Indochina, the Thai-Malay peninsula and Sumatra from Java, Borneo, Palawan and the Philippines, only 50 dispersals were inferred across van Welzen's line, 27 across Huxley's line, 21 across Wallace's line, 15 across Sclaters' line, and 11 across Lydekker's line (Table 4.2).

The striking differences in dispersal frequencies in the west and east of the Asia-Pacific region can at least partially be explained by the deep oceans separating most regions east of van Welzen's line. However, this does not account for the low dispersals of Java and Palawan with other regions on the Sunda shelf. It is also notable that Lydekker's line appears of little relevance to the Calamoideae as dispersal events between Papuasia and Australia (4

inferred dispersals) were rarer than between the Moluccas and Papuasia (7 inferred dispersals) (Fig. 4.3). Other factors, such as the submergence of Java during parts of the Miocene (Hall, 2009, 2012b; de Bruyn et al., 2014; Morley, 2018), or potential extinction in Australia in the recent past (Morley, 2000), may also play a role in shaping dispersal patterns that can be inferred from extant taxa.

### 4.4.5 Prevalence of stepping-stone dispersal

Almost all dispersal events of the Calamoideae were inferred to have happened between directly adjacent regions (Fig. 4.3), validating the assumed importance of stepping-stone dispersal in the biogeography of the Asia-Pacific region (Kooyman et al., 2019), and challenging the idea that few extreme events of long-distance dispersal may be more probable than multiple dispersal events across shorter distances (Crisp et al., 2011). In particular, ancestral range reconstructions for the Calamoideae show that Sulawesi and the Moluccas played central roles in connecting Sundaland and Sahul (Fig. 4.3). The only dispersals bypassing intermediate regions occurred in Sundaland between Indochina and Sumatra, and between the Thai-Malay Peninsula and Java (Fig. 4.3), and may have happened at times of lower sea levels when these regions were directly connected by tropical rainforests (Morley, 2018).

### 4.4.6 Predominance of dispersals from Borneo to the periphery of the Asia-Pacific

Across the Asia-Pacific region, migration events between regions were dominated by dispersals from Borneo to the periphery, with much fewer dispersals occurring in the other

direction (Fig. 4.3). The well-documented patter of predominantly eastward dispersal from Sundaland across Wallacea and into Sahul (Crayn et al., 2015; Strijk et al., 2020; Joyce et al., 2021), may therefore only be one part of a broader pattern spanning the Asia-Pacific region. The pattern is potentially closely linked to the unique role of Borneo as a radiator, i.e. as a region with high proportions of both *in situ* diversification and emigration (Fig. 4.4). The inferred predominance of dispersals "away" from Borneo may partially be explained by a diffusion process of dispersal that is shaped by oceanic barriers and species richness at the place of origin (Allen et al., 1991; Lomolino et al., 2016), potentially combined with the occupancy of ecological niches in source regions impeding establishment after dispersal back into these regions (Wiens and Donoghue, 2004). However, such an explanation can only partially account for this pattern as the genus *Calamus*, which accounts for ca. 80% of the Calamoideae in the Asia-Pacific region and thus dominates the dispersal patterns inferred here, originated in Indochina (Fig. 4.2). The at least partial submergence of Sumatra, Java and Palawan during parts of the Miocene (Hall, 2009, 2012b; de Bruyn et al., 2014; Morley, 2018) likely contributed to the predominance of inferred dispersals from Borneo into these regions. Other abiotic factors, such as the direction of wind and ocean currents (Cook and Crisp, 2005), may also be of importance.

### 4.4.7 Distribution better explained by present than past geology

Time-uniform models based on the present-day geological setting almost always had higher AIC values than time-stratified models with otherwise identical parameters (Table 4.1, Supplementary Tables 4.3 - 4.4), indicating that the present distances and connectivity between landmasses explain the data better than geological settings inferred for the past.

Given the considerable changes in the geology of the Asia-Pacific over the last 125 million years (Hall, 2009, 2012b; de Bruyn et al., 2014; Morley, 2018), this finding is somewhat surprising. However, a combination of factors may potentially explain this disparity. First, most species of the Calamoideae originated in the Oligocene and Miocene (Chapter 3, Fig. 3.5). The current geology of the Asia-Pacific may thus be a relatively good approximation of the geological setting in which dispersal events took place. Second, considerable uncertainty exists in geological reconstructions for the geological history of the Asia-Pacific, especially towards older ages (Hall, 2017). Third, the complex geology of several biogeographical regions that were assembled from different landmasses and reached their current composition only in the most recent past (e.g. Sulawesi, the Philippines, and the Moluccas) impedes an accurate representation of their geological past in biogeographical models.

### 4.4.8 Different roles of regions in establishing extant diversity

The analysis of *in situ* diversification and emigration revealed the respective roles of different regions in attaining the high diversity and broad distribution of the Calamoideae across the Asia-Pacific (Fig. 4.4).

#### *Central role of Borneo for in situ diversification and migration*

The identification of Borneo as the only radiator, having high proportions of both *in situ* diversification and emigration, underlines its central role in shaping the biogeography of the Calamoideae in the Asia-Pacific. Borneo hosts the largest number of calamoid species of any region in the Asia-Pacific. It was the stage of more radiations than in any other region, both

in terms of the absolute number of radiations and the number of radiations in different genera, giving rise to one radiation in *Eugeissona*, two in *Korthalsia*, one in *Salacca*, and at least ten in *Calamus* (Fig. 4.2). Borneo was the source of a multitude of migrations into all adjacent regions, with much more dispersals inferred out of Borneo than into Borneo. The only exception was the relationship of Borneo to Indochina, where almost equal numbers of dispersals in both directions were inferred (Fig. 4.3). The central role of Borneo identified here for the biogeography of the Calamoideae has also been found for some other tropical rainforest plant groups, such as the aroid genus *Alocasia* (Nauheimer et al., 2012), and in a cross-taxonomic study incorporating six plant and 55 animal groups from the Asia-Pacific (de Bruyn et al., 2014). The high number of species and large proportions of *in situ* diversification and emigration are potentially linked to the large size and old age of the island, and its position near the geographical centre of the Asia-Pacific. At least parts of Borneo have been continuously above sea level throughout the Cenozoic and have likely hosted tropical rainforests since at least 50 Ma (Morley, 2018), providing a stable environment and potential refugium for the Calamoideae at times of high sea levels or when other islands such as Sumatra or Java were submerged. The presence of a diverse relief including lowlands and mountainous areas may have provided a rich setting for *in situ* diversification, as potentially did recent geological mountain building (Merckx et al., 2015).

### *In situ diversification in isolation in Indochina, Sulawesi and Papuasia*

Based on their high proportions of *in situ* diversification but low proportions of emigration, Indochina, Sulawesi and Papuasia were identified as incubators (Fig. 4.4), i.e. regions generating considerable local diversity but with limited influence on adjacent regions. After

Borneo, Indochina is the region with most radiations, with one radiation in *Plectocomia* and nine radiations in *Calamus* (Fig. 4.2). The inference of Indochina as the ancestral area of both *Calamus* and subtribe Plectocomiinae highlights its importance for the origin of both clades. The high proportion of *in situ* diversification may be linked to old age and large area of the region, and potentially the uplift of the Himalayas resulting from the collision of India with Asia. The relatively low proportion of emigrations contrasts with previous findings largely based on animal lineages (de Bruyn et al., 2014) and may be connected to its peripheral position in the Northeast of the Asia-Pacific region, although a strong link to the Thai-Malay Peninsula exists (Fig. 4.3).

In Sulawesi, the relatively high proportion of *in situ* diversification, resulting in three local radiations in *Calamus*, may be linked to the fusion of West Sulawesi with North Sulawesi and the Sula Spur after the collision of the Asian and Australian plates (Hall, 2017), as well as to the high rates of uplift since 10 Ma that resulted in the build-up of high mountains and substantial increases of land above sea level (Nugraha and Hall, 2018). While Sulawesi was inferred here as an important source region for the Moluccas and also the Philippines and therefore played a crucial role as a stepping stone connecting the West and East of the Asia-Pacific region, the absence of migrations to Borneo is notable (Fig. 4.3). The relatively small proportion of emigrations from Sulawesi in relation to its species richness may be explained by the surrounding deep seas acting as considerable barriers to dispersal.

Papuasia is remarkable for having the highest proportion of *in situ* diversification of all regions, but the second-lowest proportion of emigration. Only two radiations occurred in Papuasia, a small radiation in *Metroxylon*, and by far the largest radiation across the Calamoideae, in *Calamus* (Fig. 4.2). The high proportion of *in situ* diversification may be

indirectly linked to the rapid uplift of New Guinea's central mountain range since 10 Ma, as previously found in *Exocelina* diving beetles (Toussaint et al., 2014). In contrast, the very low proportion of emigrations is striking, with no dispersal originating from within the *Metroxylon* and *Calamus* radiations, and very few dispersal events into the Moluccas and Australia inferred. The Moluccas were inferred as the only source region for Papuasia, with no immigrations from Australia (Fig. 4.3). Overall, these dispersal patterns provide evidence for a closer biogeographic link of Papuasia with the Moluccas than with Australia, and indicate that migration of tropical rainforest plants between New Guinea and Australia during low sea levels might not have been as easily possible as within Sundaland. As discussed above, there is also a possibility that the inferred lack of dispersals between Australia and New Guinea may be due to contraction of tropical rainforests in Australia (Morley, 2000) resulting in severe extinction in the Calamoideae that may make it impossible to infer earlier dispersal events from the extant calamoid species.

***Thai-Malay Peninsula, Sumatra and Moluccas as migration corridors***

For the Thai-Malay Peninsula, Sumatra, and the Moluccas, low proportions of *in situ* diversification (i.e. high proportions of immigration) but high proportions of emigration were inferred (Fig. 4.4), marking them as important corridors for migration between other regions. The Thai-Malay Peninsula and Sumatra provide important dispersal routes between Indochina, Borneo and Java, whereas the Moluccas connect Sulawesi and Papuasia (Fig. 4.3). Whilst small radiations likely occurred in the Thai-Malay peninsula in *Eugeissona*, *Korthalsia*, *Salacca*, and *Calamus*, none were inferred for Sumatra and the Moluccas (Fig. 4.2). The notable lack of *in situ* diversification in Sumatra despite its large size is potentially linked to

the almost complete submergence of the island during the early Miocene (de Bruyn et al., 2014; Morley, 2018). Although the absence of *in situ* diversification on the Moluccas may partially be explained by the small size of its islands, the palm flora of this region is only poorly known. It is therefore possible that future field work will uncover local radiations and lead to a different understanding of the role of the Moluccas in the Asia-Pacific. The reasons for the low inferred proportion of *in situ* diversification on the Thai-Malay Peninsula are somewhat unclear as the region has been covered by wet tropical rainforests at least since the Miocene (Morley, 2018) but may be connected to its high connectivity to surrounding regions that could mitigate against endemism constrained to the Thai-Malay Peninsula (Fig. 4.2).

***Prevalence of immigration in India, Java, Palawan, Philippines and Australia***

The low proportions of both *in situ* diversification and emigration in India, Java, Palawan, the Philippines, and Australia revealed that these regions can be regarded as containers that were primarily shaped by immigration and had very limited influence on other regions (Fig. 4.4). Of note, all containers identified here are located in the periphery of the Asia-Pacific, with India in the Northwest, Palawan and the Philippines in the North, Java in the South, and Australia in the Southwest, making it likely that their low emigration rates can at least in part be explained by their peripheral positions in the Asia-Pacific region (but note that the fossil record indicates the origin of *Eugeissona* in India, see discussion above). In India and Palawan, the low proportion of *in situ* diversification is potentially connected to the small area of available tropical rainforest biomes (Eiserhardt et al., 2017). Australia also has very little tropical rainforest area at present but was the location of a small radiation in *Calamus*

in the early Miocene (Fig. 4.2), i.e. during a time when tropical rainforests were much more widespread in Australia (Morley, 2000). It can therefore be regarded as a container incubator. The almost complete absence of *in situ* diversification in Java, which might be seen as an container corridor due to its intermediate proportion of emigration, can likely be explained by its complete submergence during most of the Miocene until around 10 Ma (Hall, 2009).

## 4.5 Conclusion

It has long been recognised that the consecutive collisions of India and Australia with Asia catalysed a vibrant exchange of tropical rainforest plant lineages between the regions involved. However, details of these interchanges have long remained obscure due to the complex geology of the Asia-Pacific region and the limited information that fossils and geographic ranges of taxa can provide on dispersal pathways.

Based on the most densely sampled dated molecular phylogeny of a tropical rainforest lineage to date, the present study is a milestone in improving our understanding of the biogeography of the Asia-Pacific, revealing the biogeographic history of a characteristic tropical rainforest clade distributed across the entirety of the study region in unprecedented detail. The ancestral ranges estimated here revealed that the biogeographic history of the Calamoideae in the Asia-Pacific was strongly influenced by the geographical configuration of the region and closely associated with the convergence of Australia and Asia from around 30 Ma. It must be noted, however, that the conclusions presented here are only valid for the extant Calamoideae, with most species having originated in the Oligocene and Miocene. The earlier biogeographic history of the Calamoideae, including the

biogeographic significance of the presence of calamoid fossils in Australia long before its collision with Asia, remains elusive.

Dispersals generally occurred out of Sundaland into peripheral regions, with intermediate regions acting as stepping stones. In contrast, many fewer dispersals were inferred in the opposite direction. This asymmetrical dispersal pattern puts the well-known prevalence of eastwards dispersals across Wallacea into context by identifying it as part of a larger pattern spanning the entirety of the Asia-Pacific region.

Migration events across the deep ocean barriers in the East of the Asia-Pacific region were much rarer than between the regions of Sundaland, which are separated by shallow waters and were connected by land during periods of lower sea levels. In contrast, the scarcity of dispersals between Papuasia and Australia hints at the presence of a previously underappreciated dispersal barrier within Sahul despite the shallow sea levels. Taken together, the dispersal patterns show that deep oceans have presented substantial dispersal barriers for the Calamoideae despite occasional migrations across these boundaries.

The proportions of *in situ* diversification and emigration of different regions in the Asia-Pacific shed light on their respective roles in achieving the broad distribution and diversity of the Calamoideae, highlighting the key role of Borneo as a radiator for *in situ* diversification and emigration; of Indochina, Sulawesi, and Papuasia as incubators with high *in situ* diversification in relative isolation; of the Thai-Malay Peninsula, Sumatra, and the Moluccas as corridors that connect other regions; and of India, Java, Palawan, the Philippines, and Australia as containers mainly shaped by immigration. These different regional contributions were shown to be closely associated with their location in the Asia-Pacific, their connectedness during low sea levels, the size of regions and the expanse of past

and present tropical rainforests, and, for intermittently submerged regions, age since the last emergence.

This study has demonstrated the potential of biogeographic analyses based on dated phylogenies of large and densely sampled clades to provide detailed insights even in highly complex geological contexts. To achieve a more complete picture of the biogeographic history of the Asia-Pacific and to distinguish idiosyncrasies only applicable to the Calamoideae from generalisable patterns across tropical rainforest lineages, more studies of other species-rich clades distributed across the Asia-Pacific are needed, such as the highly diverse palm lineages Areceae and Coryphoideae.

## 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, an Emily Holmes Memorial Scholarship by the Amar-Franses & Foster-Jenkins Trust, and two Bentham-Moxon Trust grants (BMT1-2017 and BMT36- 2018) to the author.

## Acknowledgements

The author would like to thank John Dransfield for valuable advice on species distributions and Jun Ying Lim for helpful discussions on the scope of this chapter. Computational analyses were conducted on the high performance computing facilities of the “UK’s Crop Diversity Bioinformatics HPC” (BBSRC grant BB/S019669/1) at The James Hutton Institute and NIAB.

## Supplementary Material

**Supplementary Table 4.1. Pairwise distances (km) between biogeographic regions in the Asia-Pacific region through time.** I = India, C = Indochina, T = Thai-Malay Peninsula, S = Sumatra, J = Java, B = Borneo, L = Palawan, P = Philippines, W = Sulawesi, M = Moluccas, N = Papuasia, A = Australia. To allow direct comparison with time-uniform models (i.e. models based on present distances only), distances were always estimated for the more recent ends of the time bins.

| 0-10 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 0 | 2300 | 3000 | 3300 | 4600 | 4500 | 4500 | 4900 | 5200 | 5900 | 7400 | 7700 |
| **C** | 2300 | 0 | 1400 | 2000 | 3000 | 2500 | 2200 | 2600 | 3100 | 3700 | 5100 | 5700 |
| **T** | 3000 | 1400 | 0 | 600 | 1700 | 1500 | 1900 | 2500 | 2300 | 3100 | 4500 | 4700 |
| **S** | 3300 | 2000 | 600 | 0 | 1300 | 1400 | 2100 | 2800 | 2100 | 3000 | 4400 | 4500 |
| **J** | 4600 | 3000 | 1700 | 1300 | 0 | 900 | 2000 | 2500 | 1200 | 2100 | 3300 | 3200 |
| **B** | 4500 | 2500 | 1500 | 1400 | 900 | 0 | 1100 | 1700 | 700 | 1600 | 3000 | 3200 |
| **L** | 4500 | 2200 | 1900 | 2100 | 2000 | 1100 | 0 | 600 | 1200 | 1600 | 3000 | 3600 |
| **P** | 4900 | 2600 | 2500 | 2800 | 2500 | 1700 | 600 | 0 | 1600 | 1500 | 2700 | 3600 |
| **W** | 5200 | 3100 | 2300 | 2100 | 1200 | 700 | 1200 | 1600 | 0 | 900 | 2300 | 2600 |
| **M** | 5900 | 3700 | 3100 | 3000 | 2100 | 1600 | 1600 | 1500 | 900 | 0 | 1400 | 2100 |
| **N** | 7400 | 5100 | 4500 | 4400 | 3300 | 3000 | 3000 | 2700 | 2300 | 1400 | 0 | 1200 |
| **A** | 7700 | 5700 | 4700 | 4500 | 3200 | 3200 | 3600 | 3600 | 2600 | 2100 | 1200 | 0 |

| 10-20 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 0 | 2600 | 3200 | 3500 | 4700 | 4900 | 4400 | 5500 | 5600 | 6100 | 7900 | 8200 |
| **C** | 2600 | 0 | 1700 | 2200 | 3000 | 2800 | 2000 | 2900 | 3500 | 3900 | 5500 | 6300 |
| **T** | 3200 | 1700 | 0 | 600 | 1600 | 1800 | 1700 | 2800 | 2500 | 3000 | 4700 | 5000 |
| **S** | 3500 | 2200 | 600 | 0 | 1200 | 1700 | 2000 | 3000 | 2300 | 2800 | 4600 | 4700 |
| **J** | 4700 | 3000 | 1600 | 1200 | 0 | 800 | 1700 | 2300 | 1100 | 1700 | 3400 | 3500 |
| **B** | 4900 | 2800 | 1800 | 1700 | 800 | 0 | 1100 | 1600 | 800 | 1200 | 3000 | 3500 |
| **L** | 4400 | 2000 | 1700 | 2000 | 1700 | 1100 | 0 | 1100 | 1700 | 2000 | 3500 | 4400 |
| **P** | 5500 | 2900 | 2800 | 3000 | 2300 | 1600 | 1100 | 0 | 1700 | 1700 | 2800 | 4000 |
| **W** | 5600 | 3500 | 2500 | 2300 | 1100 | 800 | 1700 | 1700 | 0 | 600 | 2300 | 2700 |
| **M** | 6100 | 3900 | 3000 | 2800 | 1700 | 1200 | 2000 | 1700 | 600 | 0 | 1700 | 2500 |
| **N** | 7900 | 5500 | 4700 | 4600 | 3400 | 3000 | 3500 | 2800 | 2300 | 1700 | 0 | 1700 |
| **A** | 8200 | 6300 | 5000 | 4700 | 3500 | 3500 | 4400 | 4000 | 2700 | 2500 | 1700 | 0 |

**Supplementary Table 4.1** (*continued*)

| 20-30 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 0 | 2700 | 3000 | 3200 | 4000 | 4200 | 4300 | 4900 | 4900 | 5700 | 7400 | 7900 |
| **C** | 2700 | 0 | 1700 | 2200 | 2800 | 2500 | 1700 | 2400 | 3200 | 4000 | 5400 | 6300 |
| **T** | 3000 | 1700 | 0 | 600 | 1200 | 1200 | 2000 | 2300 | 2000 | 2800 | 4400 | 5000 |
| **S** | 3200 | 2200 | 600 | 0 | 800 | 1100 | 2300 | 2500 | 1800 | 2500 | 4200 | 4700 |
| **J** | 4000 | 2800 | 1200 | 800 | 0 | 800 | 2500 | 2300 | 1100 | 1700 | 3500 | 3900 |
| **B** | 4200 | 2500 | 1200 | 1100 | 800 | 0 | 1700 | 1600 | 800 | 1600 | 3200 | 3900 |
| **L** | 4300 | 1700 | 2000 | 2300 | 2500 | 1700 | 0 | 800 | 2200 | 2800 | 4000 | 4900 |
| **P** | 4900 | 2400 | 2300 | 2500 | 2300 | 1600 | 800 | 0 | 1700 | 2200 | 3200 | 4200 |
| **W** | 4900 | 3200 | 2000 | 1800 | 1100 | 800 | 2200 | 1700 | 0 | 800 | 2500 | 3100 |
| **M** | 5700 | 4000 | 2800 | 2500 | 1700 | 1600 | 2800 | 2200 | 800 | 0 | 1700 | 2300 |
| **N** | 7400 | 5400 | 4400 | 4200 | 3500 | 3200 | 4000 | 3200 | 2500 | 1700 | 0 | 1100 |
| **A** | 7900 | 6300 | 5000 | 4700 | 3900 | 3900 | 4900 | 4200 | 3100 | 2300 | 1100 | 0 |

| 30-40 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 0 | 2200 | 3000 | 2500 | 4000 | 4200 | 4300 | 4900 | 5200 | 6000 | 7600 | 7700 |
| **C** | 2200 | 0 | 1700 | 1700 | 3000 | 2800 | 2200 | 2900 | 4000 | 4700 | 6200 | 6700 |
| **T** | 3000 | 1700 | 0 | 600 | 1200 | 1200 | 2000 | 2300 | 2300 | 3100 | 4700 | 5000 |
| **S** | 2500 | 1700 | 600 | 0 | 1600 | 1800 | 2500 | 2800 | 2800 | 3500 | 5200 | 5400 |
| **J** | 4000 | 3000 | 1200 | 1600 | 0 | 800 | 2500 | 2300 | 1200 | 2000 | 3600 | 3800 |
| **B** | 4200 | 2800 | 1200 | 1800 | 800 | 0 | 1700 | 1600 | 1200 | 2000 | 3500 | 3900 |
| **L** | 4300 | 2200 | 2000 | 2500 | 2500 | 1700 | 0 | 800 | 2800 | 3400 | 4400 | 5200 |
| **P** | 4900 | 2900 | 2300 | 2800 | 2300 | 1600 | 800 | 0 | 2300 | 2800 | 3700 | 4600 |
| **W** | 5200 | 4000 | 2300 | 2800 | 1200 | 1200 | 2800 | 2300 | 0 | 800 | 2400 | 2700 |
| **M** | 6000 | 4700 | 3100 | 3500 | 2000 | 2000 | 3400 | 2800 | 800 | 0 | 1700 | 2000 |
| **N** | 7600 | 6200 | 4700 | 5200 | 3600 | 3500 | 4400 | 3700 | 2400 | 1700 | 0 | 1200 |
| **A** | 7700 | 6700 | 5000 | 5400 | 3800 | 3900 | 5200 | 4600 | 2700 | 2000 | 1200 | 0 |

| 40-50 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 0 | 2900 | 3400 | 2800 | 4200 | 4600 | 4900 | 6000 | 5500 | 5900 | 7600 | 8000 |
| **C** | 2900 | 0 | 1700 | 1700 | 3000 | 2800 | 2100 | 3400 | 4000 | 4900 | 6300 | 7100 |
| **T** | 3400 | 1700 | 0 | 600 | 1200 | 1200 | 2300 | 2800 | 2300 | 3200 | 4600 | 5400 |
| **S** | 2800 | 1700 | 600 | 0 | 1600 | 1800 | 2700 | 3400 | 2800 | 3500 | 5000 | 5700 |
| **J** | 4200 | 3000 | 1200 | 1600 | 0 | 800 | 3000 | 2800 | 1200 | 2000 | 3500 | 4200 |
| **B** | 4600 | 2800 | 1200 | 1800 | 800 | 0 | 2300 | 2000 | 1200 | 2300 | 3500 | 4400 |
| **L** | 4900 | 2100 | 2300 | 2700 | 3000 | 2300 | 0 | 1500 | 3300 | 4400 | 5200 | 6300 |
| **P** | 6000 | 3400 | 2800 | 3400 | 2800 | 2000 | 1500 | 0 | 2500 | 3500 | 3900 | 5000 |
| **W** | 5500 | 4000 | 2300 | 2800 | 1200 | 1200 | 3300 | 2500 | 0 | 1100 | 2300 | 3200 |
| **M** | 5900 | 4900 | 3200 | 3500 | 2000 | 2300 | 4400 | 3500 | 1100 | 0 | 1600 | 2200 |
| **N** | 7600 | 6300 | 4600 | 5000 | 3500 | 3500 | 5200 | 3900 | 2300 | 1600 | 0 | 1100 |
| **A** | 8000 | 7100 | 5400 | 5700 | 4200 | 4400 | 6300 | 5000 | 3200 | 2200 | 1100 | 0 |

**Supplementary Table 4.1** (*continued*)

| 50-75 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 0 | 3200 | 3400 | 2800 | 4000 | 4500 | 5100 | 7800 | 5100 | 5900 | 7300 | 7900 |
| **C** | 3200 | 0 | 1200 | 1700 | 3000 | 2800 | 2100 | 5400 | 3500 | 5400 | 6300 | 7600 |
| **T** | 3400 | 1200 | 0 | 800 | 1700 | 1600 | 2000 | 4600 | 2300 | 4200 | 5100 | 6300 |
| **S** | 2800 | 1700 | 800 | 0 | 1600 | 1800 | 2700 | 5000 | 2500 | 4000 | 5000 | 6100 |
| **J** | 4000 | 3000 | 1700 | 1600 | 0 | 800 | 3000 | 3900 | 1100 | 2500 | 3500 | 4600 |
| **B** | 4500 | 2800 | 1600 | 1800 | 800 | 0 | 2300 | 3300 | 800 | 2800 | 3500 | 4900 |
| **L** | 5100 | 2100 | 2000 | 2700 | 3000 | 2300 | 0 | 3500 | 2800 | 5000 | 5200 | 6800 |
| **P** | 7800 | 5400 | 4600 | 5000 | 3900 | 3300 | 3500 | 0 | 2800 | 3900 | 3000 | 4500 |
| **W** | 5100 | 3500 | 2300 | 2500 | 1100 | 800 | 2800 | 2800 | 0 | 2200 | 2700 | 4200 |
| **M** | 5900 | 5400 | 4200 | 4000 | 2500 | 2800 | 5000 | 3900 | 2200 | 0 | 1500 | 2200 |
| **N** | 7300 | 6300 | 5100 | 5000 | 3500 | 3500 | 5200 | 3000 | 2700 | 1500 | 0 | 1700 |
| **A** | 7900 | 7600 | 6300 | 6100 | 4600 | 4900 | 6800 | 4500 | 4200 | 2200 | 1700 | 0 |

| 75-100 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 0 | 4700 | 3900 | 3500 | 4100 | 4600 | 6600 | 7900 | 5100 | 5100 | 6600 | 6500 |
| **C** | 4700 | 0 | 1100 | 1700 | 2500 | 2800 | 2200 | 5200 | 3100 | 5700 | 6600 | 7800 |
| **T** | 3900 | 1100 | 0 | 600 | 1600 | 2000 | 2700 | 5000 | 2500 | 4700 | 5800 | 6800 |
| **S** | 3500 | 1700 | 600 | 0 | 1200 | 1800 | 3100 | 5000 | 2300 | 4200 | 5400 | 6300 |
| **J** | 4100 | 2500 | 1600 | 1200 | 0 | 600 | 3000 | 3900 | 1100 | 3200 | 4200 | 5300 |
| **B** | 4600 | 2800 | 2000 | 1800 | 600 | 0 | 2800 | 3400 | 600 | 3000 | 3800 | 5100 |
| **L** | 6600 | 2200 | 2700 | 3100 | 3000 | 2800 | 0 | 3500 | 2800 | 5600 | 5900 | 7500 |
| **P** | 7900 | 5200 | 5000 | 5000 | 3900 | 3400 | 3500 | 0 | 2800 | 3900 | 3400 | 5000 |
| **W** | 5100 | 3100 | 2500 | 2300 | 1100 | 600 | 2800 | 2800 | 0 | 2800 | 3500 | 4800 |
| **M** | 5100 | 5700 | 4700 | 4200 | 3200 | 3000 | 5600 | 3900 | 2800 | 0 | 1400 | 2100 |
| **N** | 6600 | 6600 | 5800 | 5400 | 4200 | 3800 | 5900 | 3400 | 3500 | 1400 | 0 | 1700 |
| **A** | 6500 | 7800 | 6800 | 6300 | 5300 | 5100 | 7500 | 5000 | 4800 | 2100 | 1700 | 0 |

| 100-125 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 0 | 7100 | 6600 | 5800 | 5800 | 6600 | 7700 | 8600 | 5900 | 5400 | 6600 | 5800 |
| **C** | 7100 | 0 | 1200 | 2300 | 3000 | 2700 | 1600 | 5100 | 3700 | 5900 | 7100 | 8300 |
| **T** | 6600 | 1200 | 0 | 1100 | 1700 | 1600 | 1200 | 4200 | 2500 | 4700 | 5900 | 7100 |
| **S** | 5800 | 2300 | 1100 | 0 | 800 | 1100 | 2000 | 3900 | 1600 | 3700 | 5000 | 6100 |
| **J** | 5800 | 3000 | 1700 | 800 | 0 | 800 | 2300 | 3300 | 800 | 3000 | 4200 | 5400 |
| **B** | 6600 | 2700 | 1600 | 1100 | 800 | 0 | 1700 | 2800 | 1100 | 3400 | 4400 | 5700 |
| **L** | 7700 | 1600 | 1200 | 2000 | 2300 | 1700 | 0 | 3500 | 2800 | 5000 | 5900 | 7400 |
| **P** | 8600 | 5100 | 4200 | 3900 | 3300 | 2800 | 3500 | 0 | 2800 | 3500 | 3400 | 5100 |
| **W** | 5900 | 3700 | 2500 | 1600 | 800 | 1100 | 2800 | 2800 | 0 | 2300 | 3500 | 4700 |
| **M** | 5400 | 5900 | 4700 | 3700 | 3000 | 3400 | 5000 | 3500 | 2300 | 0 | 1600 | 2400 |
| **N** | 6600 | 7100 | 5900 | 5000 | 4200 | 4400 | 5900 | 3400 | 3500 | 1600 | 0 | 1700 |
| **A** | 5800 | 8300 | 7100 | 6100 | 5400 | 5700 | 7400 | 5100 | 4700 | 2400 | 1700 | 0 |

**Supplementary Table 4.2. Land connectivity between biogeographic regions in the Asia-Pacific region through time.** I = India, C = Indochina, T = Thai-Malay Peninsula, S = Sumatra, J = Java, B = Borneo, L = Palawan, P = Philippines, W = Sulawesi, M = Moluccas, N = Papuasia, A = Australia. 1 = regions connected by land, 0.1 = regions not connected by land. To allow direct comparison with time-uniform models (i.e. models based on present connectivity only), land connectivity was always inferred for the more recent ends of the time bins.

| 0-10 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **C** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **T** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **S** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **J** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **B** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **L** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **P** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **W** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 |
| **M** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 |
| **N** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |
| **A** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |

| 10-20 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 1 | 1 | 1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **C** | 1 | 1 | 1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **T** | 1 | 1 | 1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **S** | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **J** | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **B** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **L** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **P** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **W** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 |
| **M** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 |
| **N** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |
| **A** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |

**Supplementary Table 4.2** (*continued*)

| 20-30 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **C** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **T** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **S** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **J** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **B** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **L** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **P** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **W** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 |
| **M** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 |
| **N** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |
| **A** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |

| 30-40 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **C** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **T** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **S** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **J** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **B** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **L** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **P** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **W** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 |
| **M** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 |
| **N** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |
| **A** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |

| 40-50 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **C** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **T** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **S** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **J** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **B** | 1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **L** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **P** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **W** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 |
| **M** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 |
| **N** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |
| **A** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |

**Supplementary Table 4.2** (*continued*)

| 50-75 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **C** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **T** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **S** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **J** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **B** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **L** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **P** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **W** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 |
| **M** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 |
| **N** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |
| **A** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |

| 75-100 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **C** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **T** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **S** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **J** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **B** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **L** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **P** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **W** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 |
| **M** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 |
| **N** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |
| **A** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |

| 100-125 Ma | I | C | T | S | J | B | L | P | W | M | N | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **C** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **T** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **S** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **J** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **B** | 0.1 | 1 | 1 | 1 | 1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **L** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **P** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 | 0.1 |
| **W** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 | 0.1 |
| **M** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 0.1 | 0.1 |
| **N** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |
| **A** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1 | 1 |

**Supplementary Table 4.3. Comparison of 66 biogeographic models based on maximum area occupancy of two regions.** Dist. = distance dependence, Con. = connectivity dependence, w free = connectivity parameter w free (yes) or fixed (no), lnL = natural logarithm of likelihood, np = number of parameters, d = probability of range expansion, e = probability of range reduction, j = weight of jump dispersal, x = distance parameter, w = connectivity parameter, AIC = Akaike information criterion. AIC weight calculated across all models. Model best explaining the data in bold.

| Dist. | Con. | w free | Time | Base model | lnL | np | d | e | j | x | w | AIC | AIC weight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| no | no | no | no | DEC | -1119.0 | 2 | 0.01 | 0.01 | 0 | 0 | 0 | 2242 | 0% |
| | | | | DEC+J | -903.4 | 3 | 0.0016 | 1E-12 | 0.015 | 0 | 0 | 1813 | 0% |
| | | | | DIVALIKE | -1135.0 | 2 | 0.01 | 0.01 | 0 | 0 | 0 | 2274 | 0% |
| | | | | DIVALIKE+J | -1135.0 | 3 | 0.01 | 0.01 | 0.0001 | 0 | 0 | 2276 | 0% |
| | | | | BAYAREALIKE | -1211.0 | 2 | 0.01 | 0.01 | 0 | 0 | 0 | 2427 | 0% |
| | | | | BAYAREALIKE+J | -1105.0 | 3 | 0.01 | 0.029 | 0.0015 | 0 | 0 | 2216 | 0% |
| yes | no | no | no | DEC | -899.4 | 3 | 0.026 | 0.0045 | 0 | -2.02 | 0 | 1805 | 0% |
| | | | | DEC+J | -813.4 | 4 | 0.014 | 1E-12 | 0.14 | -2.01 | 0 | 1635 | 0% |
| | | | | DIVALIKE | -903.3 | 3 | 0.025 | 0.0031 | 0 | -1.89 | 0 | 1813 | 0% |
| | | | | DIVALIKE+J | -824.1 | 4 | 0.014 | 1E-12 | 0.11 | -1.93 | 0 | 1656 | 0% |
| | | | | BAYAREALIKE | -977.4 | 3 | 0.039 | 0.025 | 0 | -2.2 | 0 | 1961 | 0% |
| | | | | BAYAREALIKE+J | -827.2 | 4 | 0.0098 | 0.0002 | 0.14 | -1.91 | 0 | 1662 | 0% |
| yes | no | no | yes | DEC | -905.5 | 3 | 0.021 | 0.0043 | 0 | -1.79 | 0 | 1817 | 0% |
| | | | | DEC+J | -818.4 | 4 | 0.011 | 1E-12 | 0.11 | -1.77 | 0 | 1645 | 0% |
| | | | | DIVALIKE | -909.6 | 3 | 0.021 | 0.003 | 0 | -1.7 | 0 | 1825 | 0% |
| | | | | DIVALIKE+J | -828.9 | 4 | 0.011 | 1E-12 | 0.091 | -1.72 | 0 | 1666 | 0% |
| | | | | BAYAREALIKE | -984.4 | 3 | 0.03 | 0.025 | 0 | -1.95 | 0 | 1975 | 0% |
| | | | | BAYAREALIKE+J | -831.9 | 4 | 0.0078 | 1E-07 | 0.11 | -1.69 | 0 | 1672 | 0% |
| no | yes | no | no | DEC | -949.4 | 2 | 0.0079 | 0.0052 | 0 | 0 | 0 | 1903 | 0% |
| | | | | DEC+J | -857.9 | 3 | 0.004 | 1E-12 | 0.042 | 0 | 0 | 1722 | 0% |
| | | | | DIVALIKE | -950.8 | 2 | 0.0081 | 0.0033 | 0 | 0 | 0 | 1906 | 0% |
| | | | | DIVALIKE+J | -866.9 | 3 | 0.0043 | 1E-12 | 0.038 | 0 | 0 | 1740 | 0% |
| | | | | BAYAREALIKE | -1072.0 | 2 | 0.01 | 0.01 | 0 | 0 | 0 | 2148 | 0% |
| | | | | BAYAREALIKE+J | -869.5 | 3 | 0.0031 | 1E-07 | 0.048 | 0 | 0 | 1745 | 0% |
| no | yes | no | yes | DEC | -952.4 | 2 | 0.0084 | 0.0055 | 0 | 0 | 0 | 1909 | 0% |
| | | | | DEC+J | -860.4 | 3 | 0.0042 | 1E-12 | 0.046 | 0 | 0 | 1727 | 0% |
| | | | | DIVALIKE | -951.5 | 2 | 0.0086 | 0.0036 | 0 | 0 | 0 | 1907 | 0% |
| | | | | DIVALIKE+J | -867.4 | 3 | 0.0045 | 1E-12 | 0.042 | 0 | 0 | 1741 | 0% |
| | | | | BAYAREALIKE | -1037.0 | 2 | 0.011 | 0.024 | 0 | 0 | 0 | 2078 | 0% |
| | | | | BAYAREALIKE+J | -871.6 | 3 | 0.0033 | 1E-07 | 0.053 | 0 | 0 | 1749 | 0% |

**Supplementary Table 4.3** (*continued*)

| Dist. | Con. | w free | Time | Base model | lnL | np | d | e | j | x | w | AIC | AIC weight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| no | yes | yes | no | DEC | -940.8 | 3 | 0.0067 | 0.0051 | 0 | 0 | 0.69 | 1888 | 0% |
| | | | | DEC+J | -850.2 | 4 | 0.0034 | 1E-12 | 0.035 | 0 | 0.7 | 1708 | 0% |
| | | | | DIVALIKE | -941.4 | 3 | 0.0068 | 0.0032 | 0 | 0 | 0.68 | 1889 | 0% |
| | | | | DIVALIKE+J | -858.5 | 4 | 0.0037 | 1E-12 | 0.032 | 0 | 0.7 | 1725 | 0% |
| | | | | BAYAREALIKE | -1025.0 | 3 | 0.0076 | 0.024 | 0 | 0 | 0.54 | 2056 | 0% |
| | | | | BAYAREALIKE+J | -860.1 | 4 | 0.0026 | 1E-07 | 0.039 | 0 | 0.68 | 1728 | 0% |
| no | yes | yes | yes | DEC | -944.5 | 3 | 0.007 | 0.0054 | 0 | 0 | 0.7 | 1895 | 0% |
| | | | | DEC+J | -853.2 | 4 | 0.0036 | 1E-12 | 0.037 | 0 | 0.7 | 1714 | 0% |
| | | | | DIVALIKE | -943.3 | 3 | 0.0073 | 0.0035 | 0 | 0 | 0.7 | 1893 | 0% |
| | | | | DIVALIKE+J | -859.8 | 4 | 0.0039 | 1E-12 | 0.034 | 0 | 0.72 | 1728 | 0% |
| | | | | BAYAREALIKE | -1031.0 | 3 | 0.0083 | 0.023 | 0 | 0 | 0.78 | 2067 | 0% |
| | | | | BAYAREALIKE+J | -862.9 | 4 | 0.0028 | 1E-07 | 0.042 | 0 | 0.68 | 1734 | 0% |
| yes | yes | no | no | DEC | -884.2 | 3 | 0.058 | 0.005 | 0 | -2.02 | 0 | 1774 | 0% |
| | | | | DEC+J | -790.5 | 4 | 0.029 | 1E-12 | 0.3 | -2 | 0 | 1589 | 0% |
| | | | | DIVALIKE | -884.2 | 3 | 0.053 | 0.0034 | 0 | -1.85 | 0 | 1774 | 0% |
| | | | | DIVALIKE+J | -799.6 | 4 | 0.03 | 1E-12 | 0.26 | -1.96 | 0 | 1607 | 0% |
| | | | | BAYAREALIKE | -976.7 | 3 | 0.19 | 0.025 | 0 | -3.15 | 0 | 1959 | 0% |
| | | | | BAYAREALIKE+J | -808.6 | 4 | 0.021 | 1E-07 | 0.27 | -1.89 | 0 | 1625 | 0% |
| yes | yes | no | yes | DEC | -897.7 | 3 | 0.043 | 0.0052 | 0 | -1.65 | 0 | 1801 | 0% |
| | | | | DEC+J | -804.0 | 4 | 0.021 | 1E-12 | 0.23 | -1.6 | 0 | 1616 | 0% |
| | | | | DIVALIKE | -896.5 | 3 | 0.04 | 0.0035 | 0 | -1.52 | 0 | 1799 | 0% |
| | | | | DIVALIKE+J | -811.3 | 4 | 0.022 | 1E-12 | 0.2 | -1.53 | 0 | 1631 | 0% |
| | | | | BAYAREALIKE | -986.0 | 3 | 0.063 | 0.025 | 0 | -1.78 | 0 | 1978 | 0% |
| | | | | BAYAREALIKE+J | -821.7 | 4 | 0.015 | 1E-07 | 0.22 | -1.48 | 0 | 1651 | 0% |
| **yes** | **yes** | **yes** | **no** | DEC | -869.1 | 4 | 0.046 | 0.0047 | 0 | -1.97 | 0.55 | 1746 | 0% |
| | | | | **DEC+J** | **-777.1** | **5** | **0.024** | **1E-12** | **0.24** | **-1.98** | **0.6** | **1564** | **100%** |
| | | | | DIVALIKE | -867.2 | 4 | 0.047 | 0.0033 | 0 | -1.96 | 0.56 | 1742 | 0% |
| | | | | DIVALIKE+J | -785.0 | 5 | 0.025 | 1E-12 | 0.21 | -1.95 | 0.59 | 1580 | 0% |
| | | | | BAYAREALIKE | -956.7 | 4 | 0.063 | 0.025 | 0 | -2.13 | 0.47 | 1921 | 0% |
| | | | | BAYAREALIKE+J | -815.7 | 5 | 0.039 | 1E-07 | 0.43 | -3.08 | 0.61 | 1641 | 0% |
| yes | yes | yes | yes | DEC | -881.2 | 4 | 0.036 | 0.0053 | 0 | -1.71 | 0.53 | 1770 | 0% |
| | | | | DEC+J | -788.5 | 5 | 0.018 | 1E-12 | 0.19 | -1.67 | 0.56 | 1587 | 0% |
| | | | | DIVALIKE | -878.8 | 4 | 0.036 | 0.0033 | 0 | -1.66 | 0.54 | 1766 | 0% |
| | | | | DIVALIKE+J | -795.2 | 5 | 0.019 | 1E-12 | 0.16 | -1.61 | 0.57 | 1600 | 0% |
| | | | | BAYAREALIKE | -968.8 | 4 | 0.047 | 0.025 | 0 | -1.79 | 0.49 | 1946 | 0% |
| | | | | BAYAREALIKE+J | -803.9 | 5 | 0.012 | 1E-07 | 0.18 | -1.57 | 0.52 | 1618 | 0% |

**Supplementary Table 4.4. Comparison of 66 biogeographic models based on maximum area occupancy of three regions.** Dist. = distance dependence, Con. = connectivity dependence, w free = connectivity parameter w free (yes) or fixed (no), lnL = natural logarithm of likelihood, np = number of parameters, d = probability of range expansion, e = probability of range reduction, j = weight of jump dispersal, x = distance parameter, w = connectivity parameter, AIC = Akaike information criterion. AIC weight calculated across all models. Model best explaining the data in bold.

| Dist. | Con. | w free | Time | Base model | lnL | np | d | e | j | x | w | AIC | AIC weight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| no | no | no | no | DEC | -1317.0 | 2 | 0.01 | 0.01 | 0 | 0 | 0 | 2637 | 0% |
| | | | | DEC+J | -1043.0 | 3 | 0.002 | 1E-12 | 0.013 | 0 | 0 | 2091 | 0% |
| | | | | DIVALIKE | -1326.0 | 2 | 0.01 | 0.01 | 0 | 0 | 0 | 2655 | 0% |
| | | | | DIVALIKE+J | -1326.0 | 3 | 0.01 | 0.01 | 0.0001 | 0 | 0 | 2657 | 0% |
| | | | | BAYAREALIKE | -1444.0 | 2 | 0.01 | 0.01 | 0 | 0 | 0 | 2893 | 0% |
| | | | | BAYAREALIKE+J | -1019.0 | 3 | 0.0011 | 0.012 | 0.012 | 0 | 0 | 2044 | 0% |
| yes | no | no | no | DEC | -966.5 | 3 | 0.03 | 0.0021 | 0 | -2.32 | 0 | 1939 | 0% |
| | | | | DEC+J | -907.4 | 4 | 0.024 | 1E-12 | 0.17 | -2.4 | 0 | 1823 | 0% |
| | | | | DIVALIKE | -984.4 | 3 | 0.032 | 0.0014 | 0 | -2.21 | 0 | 1975 | 0% |
| | | | | DIVALIKE+J | -925.8 | 4 | 0.023 | 1E-12 | 0.14 | -2.23 | 0 | 1860 | 0% |
| | | | | BAYAREALIKE | -1031.0 | 3 | 0.032 | 0.028 | 0 | -2.64 | 0 | 2069 | 0% |
| | | | | BAYAREALIKE+J | -935.2 | 4 | 0.017 | 1E-07 | 0.2 | -2.24 | 0 | 1878 | 0% |
| yes | no | no | yes | DEC | -971.0 | 3 | 0.025 | 0.0017 | 0 | -2.15 | 0 | 1948 | 0% |
| | | | | DEC+J | -912.6 | 4 | 0.017 | 1E-12 | 0.12 | -2.09 | 0 | 1833 | 0% |
| | | | | DIVALIKE | -990.4 | 3 | 0.026 | 0.0012 | 0 | -2.01 | 0 | 1987 | 0% |
| | | | | DIVALIKE+J | -931.7 | 4 | 0.018 | 1E-12 | 0.11 | -2.01 | 0 | 1871 | 0% |
| | | | | BAYAREALIKE | -1035.0 | 3 | 0.023 | 0.028 | 0 | -2.35 | 0 | 2076 | 0% |
| | | | | BAYAREALIKE+J | -942.7 | 4 | 0.013 | 1E-07 | 0.15 | -1.97 | 0 | 1893 | 0% |
| no | yes | no | no | DEC | -1037.0 | 2 | 0.0072 | 0.0028 | 0 | 0 | 0 | 2077 | 0% |
| | | | | DEC+J | -974.0 | 3 | 0.0051 | 1E-12 | 0.036 | 0 | 0 | 1954 | 0% |
| | | | | DIVALIKE | -1051.0 | 2 | 0.0081 | 0.0018 | 0 | 0 | 0 | 2105 | 0% |
| | | | | DIVALIKE+J | -988.4 | 3 | 0.0055 | 1E-12 | 0.034 | 0 | 0 | 1983 | 0% |
| | | | | BAYAREALIKE | -1184.0 | 2 | 0.01 | 0.01 | 0 | 0 | 0 | 2371 | 0% |
| | | | | BAYAREALIKE+J | -999.1 | 3 | 0.004 | 1E-07 | 0.051 | 0 | 0 | 2004 | 0% |
| no | yes | no | yes | DEC | -1038.0 | 2 | 0.0077 | 0.003 | 0 | 0 | 0 | 2079 | 0% |
| | | | | DEC+J | -976.4 | 3 | 0.0055 | 1E-12 | 0.039 | 0 | 0 | 1959 | 0% |
| | | | | DIVALIKE | -1051.0 | 2 | 0.0087 | 0.0021 | 0 | 0 | 0 | 2105 | 0% |
| | | | | DIVALIKE+J | -989.0 | 3 | 0.0059 | 1E-12 | 0.037 | 0 | 0 | 1984 | 0% |
| | | | | BAYAREALIKE | -1114.0 | 2 | 0.007 | 0.028 | 0 | 0 | 0 | 2231 | 0% |
| | | | | BAYAREALIKE+J | -994.6 | 3 | 0.0035 | 0.0019 | 0.062 | 0 | 0 | 1995 | 0% |

**Supplementary Table 4.4** (*continued*)

| Dist. | Con. | w free | Time | Base model | lnL | np | d | e | j | x | w | AIC | AIC weight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| no | yes | yes | no | DEC | -1030.0 | 3 | 0.0063 | 0.0028 | 0 | 0 | 0.74 | 2066 | 0% |
| | | | | DEC+J | -968.1 | 4 | 0.0045 | 1E-12 | 0.032 | 0 | 0.76 | 1944 | 0% |
| | | | | DIVALIKE | -1044.0 | 3 | 0.0071 | 0.0018 | 0 | 0 | 0.75 | 2093 | 0% |
| | | | | DIVALIKE+J | -982.0 | 4 | 0.0049 | 1E-12 | 0.029 | 0 | 0.75 | 1972 | 0% |
| | | | | BAYAREALIKE | -1099.0 | 3 | 0.005 | 0.027 | 0 | 0 | 0.62 | 2204 | 0% |
| | | | | BAYAREALIKE+J | -991.5 | 4 | 0.0035 | 1E-07 | 0.043 | 0 | 0.73 | 1991 | 0% |
| no | yes | yes | yes | DEC | -1032.0 | 3 | 0.0067 | 0.003 | 0 | 0 | 0.75 | 2070 | 0% |
| | | | | DEC+J | -971.2 | 4 | 0.0049 | 1E-12 | 0.034 | 0 | 0.77 | 1950 | 0% |
| | | | | DIVALIKE | -1045.0 | 3 | 0.0077 | 0.002 | 0 | 0 | 0.79 | 2096 | 0% |
| | | | | DIVALIKE+J | -983.6 | 4 | 0.0053 | 1E-12 | 0.032 | 0 | 0.78 | 1975 | 0% |
| | | | | BAYAREALIKE | -1100.0 | 3 | 0.005 | 0.028 | 0 | 0 | 0.57 | 2206 | 0% |
| | | | | BAYAREALIKE+J | -987.5 | 4 | 0.003 | 0.0021 | 0.051 | 0 | 0.73 | 1983 | 0% |
| yes | yes | no | no | DEC | -940.4 | 3 | 0.066 | 0.0028 | 0 | -2.33 | 0 | 1887 | 0% |
| | | | | DEC+J | -873.0 | 4 | 0.044 | 1E-12 | 0.36 | -2.31 | 0 | 1754 | 0% |
| | | | | DIVALIKE | -950.4 | 3 | 0.071 | 0.002 | 0 | -2.23 | 0 | 1907 | 0% |
| | | | | DIVALIKE+J | -888.4 | 4 | 0.057 | 1E-12 | 0.44 | -2.65 | 0 | 1785 | 0% |
| | | | | BAYAREALIKE | -1038.0 | 3 | 0.085 | 0.027 | 0 | -2.63 | 0 | 2082 | 0% |
| | | | | BAYAREALIKE+J | -905.8 | 4 | 0.035 | 1E-07 | 0.35 | -2.21 | 0 | 1820 | 0% |
| yes | yes | no | yes | DEC | -956.8 | 3 | 0.048 | 0.0026 | 0 | -1.95 | 0 | 1920 | 0% |
| | | | | DEC+J | -893.7 | 4 | 0.032 | 1E-12 | 0.26 | -1.86 | 0 | 1795 | 0% |
| | | | | DIVALIKE | -968.9 | 3 | 0.049 | 0.002 | 0 | -1.77 | 0 | 1944 | 0% |
| | | | | DIVALIKE+J | -905.6 | 4 | 0.033 | 1E-12 | 0.23 | -1.8 | 0 | 1819 | 0% |
| | | | | BAYAREALIKE | -1053.0 | 3 | 0.053 | 0.029 | 0 | -2.15 | 0 | 2112 | 0% |
| | | | | BAYAREALIKE+J | -924.2 | 4 | 0.024 | 1E-07 | 0.28 | -1.73 | 0 | 1856 | 0% |
| **yes** | **yes** | **yes** | **no** | DEC | -925.2 | 4 | 0.052 | 0.0026 | 0 | -2.26 | 0.58 | 1858 | 0% |
| | | | | **DEC+J** | **-860.9** | **5** | **0.046** | **1E-12** | **0.35** | **-2.54** | **0.64** | **1732** | **100%** |
| | | | | DIVALIKE | -934.2 | 4 | 0.052 | 0.0017 | 0 | -2.1 | 0.58 | 1876 | 0% |
| | | | | DIVALIKE+J | -888.2 | 5 | 0.082 | 4E-05 | 0.47 | -3.17 | 0.75 | 1786 | 0% |
| | | | | BAYAREALIKE | -1019.0 | 4 | 0.052 | 0.028 | 0 | -2.61 | 0.42 | 2046 | 0% |
| | | | | BAYAREALIKE+J | -890.4 | 5 | 0.029 | 1E-07 | 0.29 | -2.2 | 0.61 | 1791 | 0% |
| yes | yes | yes | yes | DEC | -939.9 | 4 | 0.041 | 0.0023 | 0 | -2.02 | 0.55 | 1888 | 0% |
| | | | | DEC+J | -879.9 | 5 | 0.033 | 1E-12 | 0.18 | -2.09 | 0.63 | 1770 | 0% |
| | | | | DIVALIKE | -950.5 | 4 | 0.042 | 0.0015 | 0 | -1.87 | 0.57 | 1909 | 0% |
| | | | | DIVALIKE+J | -899.0 | 5 | 0.04 | 1E-12 | 0.24 | -2.45 | 0.47 | 1808 | 0% |
| | | | | BAYAREALIKE | -1029.0 | 4 | 0.033 | 0.03 | 0 | -2.09 | 0.37 | 2066 | 0% |
| | | | | BAYAREALIKE+J | -906.0 | 5 | 0.021 | 1E-07 | 0.24 | -1.84 | 0.56 | 1822 | 0% |

# 5 General Discussion

In this thesis, I set out to address the knowledge gap on the temporal and spatial origins of tropical rainforests by 1) establishing higher-level and species-level relationships of the pantropical and highly diverse palm subfamily Calamoideae, 2) generating age estimates for the Calamoideae, and 3) reconstructing the biogeographic history of the subfamily in the Asia-Pacific region.

I achieved these aims in four steps. First, I used targeted sequencing of 971 genes for 75 calamoid species representing all tribes, subtribes, and genera, as well as most infrageneric groupings, for a robust inference of their phylogenetic relationships. Second, I expanded the sampling to 653 specimens representing 502 (90%) of the 555 species of the Calamoideae, resulting after data filtering in a phylogeny comprising 381 species (68%) of the subfamily. Third, I inferred divergence times for the subfamily based on incorporation of four dating strategies and newly discovered fossils that represent the oldest palm fossils that can unambiguously be assigned to a lineage within the palm family. Fourth, I estimated ancestral ranges for 360 species of the Calamoideae across 12 biogeographic regions in the geologically complex Asia-Pacific region based on the comparison of 198 different biogeographic models, allowing the inference of dispersal pathways and roles of different biogeographic regions in shaping the broad distribution and high diversity of the Calamoideae.

This work presents the most densely sampled and dated species-level phylogeny of any major tropical rainforest lineage to date. It contributes significantly to our understanding of the age of tropical rainforest lineages and is of particular importance for elucidating the biogeographic history of tropical rainforests in the Asia-Pacific region. In this final chapter,

key findings of the preceding chapters are summarised, and prospects of pursuing research enabled by this work are outlined. Finally, broader implications for the origins of tropical rainforests are discussed.

## 5.1 The first robust phylogenomic framework of the Calamoideae

In Chapter 2, I present the first robust phylogenomic framework of the Calamoideae. This work is based on the targeted sequencing of 75 calamoid species and four outgroup taxa using the PhyloPalm probe kit, which targets 971 nuclear genes (Loiseau et al., 2019). The results indicate the existence of a single topology for the backbone relationships of the subfamily (Fig. 2.3). The number of genes analysed was two orders of magnitude greater compared to previous studies (Baker et al., 2000a; Baker et al., 2000c; Asmussen et al., 2006; Baker et al., 2009; Faye et al., 2014; Faurby et al., 2016; Faye et al., 2016), enabling phylogenetic inferences on a dramatically increased data foundation and allowing the first insights to date into gene tree conflict in the subfamily, revealing that most higher-level relationships are supported by a clear majority of gene trees (Fig. 2.4). The results also showed that some higher-level relationships, namely among genera in subtribes Ancistrophyllinae and Plectocomiinae, were characterised by high levels of gene tree conflict and could not be resolved unambiguously (Fig. 2.4 d, g). Similarly, the generic affinity of an undescribed species of Plectocomiinae (Plectocomiinae sp. nov., Kuhnhäuser 35 [K]) remains unresolved (Fig. 2.4 g). Further study of the substantial gene tree conflict in these cases is needed to disentangle different potential causes such as incomplete lineage sorting, hybridisation, gene duplications, and horizontal gene transfer (Degnan and Rosenberg, 2009; Naciri and Linder, 2015; Degnan, 2018).

This study represents the first application of the PhyloPalm probe kit outside the subfamily Arecoideae, for which it was originally developed (Loiseau et al., 2019). Although gene recovery rates for the Calamoideae were lower than for other palm subfamilies (Table 2.3), the retrieved sequence data were highly informative and enabled phylogenetic analyses based on hundreds of thousands of parsimony informative sites (Table 2.4). This demonstrates the great potential of the PhyloPalm probe kit for phylogeny reconstruction across the palm family and indicates that it might also be useful in other closely related families, such as the family Dasypogonaceae (Barrett et al., 2016b), for which currently no bespoke probe kits for targeted sequence capture exist. Although further analyses in Chapter 3 indicate that a proportion of the genes targeted by the PhyloPalm bait kit may be paralogous (Fig. 3.2), the paralogous genes seem to have minimal impact on phylogeny estimation (Supplementary Fig. 3.5). While the PhyloPalm probe kit incorporates all genes targeted by the first palm probe kit (Heyduk et al., 2015) and therefore facilitates backwards compatibility with earlier studies, future studies should also incorporate probes that are designed to be phylogenetically informative across broader taxonomic scales, such as the now widely used Angiosperm 353 probe kit (Johnson et al., 2019; Hendriks et al., 2021), to pave the way for attempts to establish the phylogenetic relationships of all life on earth (Hinchliff et al., 2015; Baker et al., 2021).

The strong improvement of the understanding of higher-level phylogenetic relationships of the Calamoideae achieved here can largely be attributed to the drastic increase in gene sampling compared to previous studies. However, estimates for the number of genes in palm genomes range from ca. 25,000 to over 50,000 (Singh et al., 2013; Chan et al., 2017; Zhao et al., 2018), signifying that the phylogenomic inferences made here were based on only between two and four percent of all available genes. With the continuing

improvement of DNA sequencing technologies (Heather and Chain, 2016), rapid progress towards the ambitious goals to sequence all life on earth (Lewin et al., 2018; Leebens-Mack et al., 2019; Baker et al., 2021), and the increasing availability of genomes for the Calamoideae (Zhao et al., 2018; Ghosh Dasgupta et al., 2021), it is likely that future phylogenies of the Calamoideae will be based on substantially more sequence data than the present study, and may allow even more refined insights into their higher-level relationships.

## 5.2 The most detailed dated phylogeny of a tropical rainforest lineage to date

In Chapter 3, I expand the phylogenomic framework established in Chapter 2 to the first dated and densely sampled species-level phylogeny of the Calamoideae to date. Based on the sequencing of 653 specimens from 24 herbaria worldwide, the sampling included 502 (90%) of the 555 species of the Calamoideae, comprising 471 species (95%) of tribe Calameae, six species (100%) of tribe Eugeissoneae, and 25 species (49%) of tribe Lepidocaryeae (Supplementary Table 3.1). After the application of several data filtering steps to exclude data-deficient samples, the final phylogeny comprised 381 species (68%) of the Calamoideae, most of which had never been sequenced before (Supplementary Fig. 3.2).

All calamoid genera were resolved as monophyletic. The resolved species-level relationships were overall highly supported, revealing a multitude of morphologically well-defined clades, and unveiling the phylogenetic placement of many species with previously uncertain affinities, such as *Calamus deerratus*, the only African species of the genus (Supplementary Fig. 3.2). Recent taxonomic decisions to synonymise several genera (Baker, 2015; Henderson and Floda, 2015) were validated based on their nested positions in *Calamus*.

Divergence times were estimated using four dating strategies incorporating newly discovered fossils of *Dicolpopollis* (Totterdell and Mitchell, 2009; Macphail and Jordan, 2015), which can unambiguously be assigned to the clade containing the subtribes Metroxylinae, Pigafettinae, Plectocomiinae and Calaminae (Harley and Morley, 1995; Harley, 2006) and represent the oldest known palm fossils that can be attributed to a lineage nested within the palm family. The age estimates were substantially older compared to previous studies, with the minimum age estimate of 122.3 Ma for the crown node of the Calamoideae (Table 3.6) predating previously established maximum age estimates for the crown nodes of the Calamoideae and the entire palm family by over 20 and 10 million years, respectively (Couvreur et al., 2011a; Baker and Couvreur, 2013a). Contrasting with the Early Cretaceous crown node age of the Calamoideae, the vast majority of extant lineages were estimated to have originated during the Oligocene and Miocene (Fig. 3.5), refuting the previously inferred constant origination of palm lineages through time based on a genus level phylogeny (Couvreur et al., 2011a).

This work demonstrates that near-complete species coverage even of highly diverse tropical rainforest lineages can be attained by collating specimens from the vast collections held by herbariums and museums across the world, but also shows the limitations of such attempts. Out of 502 species that were sequenced, 381 (76%) passed the filtering steps implemented here and were included in the final phylogeny (Fig. 3.4, Supplementary Fig. 3.2). These results indicate that utilisation of existing natural history collections can substantially advance efforts to establish detailed phylogenetic relationships of tropical rainforest lineages such as the palm family (Bellot et al., 2020). To achieve complete species coverage, however, field work to obtain high-quality plant material will likely remain an invaluable component of phylogenomic research.

There are many potential avenues for future research enabled by this dated species-level phylogeny, three of which are outlined here. First, a new infrageneric classification of the Calamoideae informed by the phylogenetic relationships resolved here would provide a valuable framework for future comparative research by allowing the explicit reference to morphologically well-defined clades within the subfamily. The only genera for which a satisfying classification currently exists are *Raphia* and *Metroxylon* (Otedoh, 1982; McClatchey, 1998; Helmstetter et al., 2020), whereas most previously described sections in *Calamus*, *Plectocomia* and *Salacca* (Beccari, 1908, 1911, 1918; Furtado, 1949, 1953, 1956; Madulid, 1981; Kramadibrata, 1992) were resolved in this thesis as paraphyletic or polyphyletic (Chapter 2, Supplementary Fig. 3.2). Further, no formal classifications are available for *Eremospatha*, *Eugeissona*, *Korthalsia*, *Laccosperma*, *Mauritiella*, *Oncocalamus*, and *Plectocomiopsis*. (The remaining genera *Eleiodoxa*, *Lepidocaryum*, *Mauritia*, *Myrialepis*, and *Pigafetta* only contain one or two species.)

Second, a detailed analysis of the evolution of the highly diverse morphology of the Calamoideae is possible for the first time. For example, the evolution of climbing in the palm family has currently only been investigated at the genus level (Couvreur et al., 2015). A new study of the Calamoideae, which has a high diversity of morphological adaptations to climbing (Isnard, 2006; Isnard and Rowe, 2008) and comprises almost all climbing palm species, could utilise both existing comprehensive morphological datasets for palms (Kissling et al., 2019; Henderson, 2020) and the species-level phylogeny generated here. This might provide not only a better understanding of when and how often climbing evolved, but also of the pace of morphological evolution in climbing clades (Onstein, 2019), and reveal whether reveal whether specialised climbing organs such as acanthophylls, grapnel spines, cirri and flagella (Fig. 1.7) represent “synnovations” of several morphological pre-

adaptations (Donoghue and Sanderson, 2015). Similarly, a study based on the relationships resolved here might improve the currently poor understanding of when and how the large variety of sexual systems in the Calamoideae, ranging from hermaphroditism to monoecy and dioecy (Baker et al., 1999b), evolved.

Third, the phylogenomic dataset generated here presents an excellent resource to study gene evolution over a long period of time in a morphologically diverse clade. Given that the PhyloPalm probes contain both phylogenetically informative but also functionally important genes (de La Harpe et al., 2019; Loiseau et al., 2019), the correlation of genetic and morphological evolution could be investigated to identify genes linked to important morphological developments, such as in the above-mentioned evolution of climbing or of different sexual systems.

## 5.3 Unparalleled insights into the biogeography of the Asia-Pacific region

In Chapter 4, I infer ancestral ranges for 360 species of the Calamoideae in the Asia-Pacific region based on 198 biogeographic models incorporating 12 biogeographic regions, identifying a time-uniform Dispersal-Extinction-Cladogenesis model allowing for "jump speciation" events and accounting for region distance and connectivity as the model best explaining the extant distribution of the Calamoideae (Table 4.1). Based on this model, I used biogeographic stochastic mapping to estimate dispersal events between regions and reveal the different roles that individual biogeographic regions played in shaping the high diversity and wide distribution of the subfamily across the Asia-Pacific region.

Whereas areas of origin of most higher-level clades remained ambiguous, ancestral ranges were generally highly resolved at and below genus level, showing that the diverse clade comprising subtribes Plectocomiinae and Calaminae originated in Indochina (Fig. 4.2). The widespread distribution of the extant Calamoideae across the Asia-Pacific region was initiated around the time of the convergence of the Australian and Asian plates and could largely be explained by stepping-stone dispersal out of Sundaland (Southeast Asia, Sumatra, Java, Borneo) into the periphery of the Asia-Pacific region, with dispersals in the opposite direction being almost five times less frequent (Table 4.2) and no dispersal events inferred that bypassed intermediate regions except within Sundaland (Fig. 4.3). Dispersals across the shallow seas of Sundaland were nearly six times more frequent than across the deep seas in the east of the Asia-Pacific region, highlighting the importance of oceanic barriers for the biogeographic history of the Calamoideae (Table 4.2).

Borneo was identified as the only "radiator", i.e. a region with exceptional influence on other regions based on the high proportions of both *in situ* diversification and emigration. Papuasia, Indochina and Sulawesi were "incubators" with high *in situ* diversification but low emigration rates, thus generating high diversity within the region but with limited influence on other regions. The Thai-Malay Peninsula, Sumatra, and the Moluccas were shown to be "corridors" with low in situ diversification rates (and thus high immigration rates) but high rates of emigration, and thus play an important role in facilitating migration between other regions. Finally, Australia, India, Java, Palawan, and the Philippines were identified as "containers" with low proportions of both *in situ* diversification and emigration, revealing that these regions were mainly shaped by immigration and have limited influence on other regions (Fig. 4.4).

Although the high potential of the Calamoideae to contribute to a better understanding of the biogeography of the Asia-Pacific region has long been recognised (Baker and Dransfield, 2000), the biogeographic history of the subfamily in the region has only been known in vague outlines until now (Dransfield, 1981a; Morley, 1998; Baker and Dransfield, 2000; Morley, 2000; Baker and Couvreur, 2013a). The analyses conducted here finally realise this potential, presenting the most detailed biogeographic analysis of any major tropical rainforest lineage in the Asia-Pacific region to date.

Given that the biogeographic history of the Calamoideae in the Asia-Pacific region is closely linked to geological events, such as the convergence of the Australian and Asian plates during the Oligocene and Miocene (Hall, 2017), fluctuating sea level changes in the Quaternary that established land connections between regions separated by shallow waters (Voris, 2000; de Bruyn et al., 2014), or the at least partial submergence of Sumatra and Java during parts of the Miocene (Hall, 2009, 2012b; Morley, 2018), it can be expected that the key patterns identified here will also apply to many other tropical rainforest lineages. For example, the pattern of the predominantly eastward dispersal between Sundaland and Sahul (Fig. 4.3) is also well-known in other tropical rainforest plant lineages (Crayn et al., 2015; Strijk et al., 2020; Joyce et al., 2021). Because the predominantly fleshy and relatively large fruits of the Calamoideae are mainly dispersed by mammals and birds (Zona and Henderson, 1989; Dransfield et al., 2008), it is likely that the dispersal patterns revealed here (Fig. 4.3) are also closely associated with the dispersal history of these animal vectors. Overall, the dispersal pathways and roles of biogeographic regions identified here constitute the most detailed biogeographic framework in the Asia-Pacific region to date and may serve as a point of reference for future studies.

Of the many potential prospects for future biogeographic analyses in the Calamoideae enabled by this work, two are discussed here. First, analyses at the island level may allow to test whether classical theories of island biogeography (MacArthur and Wilson, 1963; Wilson and MacArthur, 1967; Warren et al., 2015), such as relationships between island area or isolation and species richness, can also be applied to single clades, and contribute to a better understanding of the relative contributions of immigration and *in situ* diversification in shaping island assemblies. Given the complex geology of the Asia-Pacific region (Hall, 2017; Morley, 2018), differences in island age and land connectivity could also be taken into account.

Second, the Calamoideae could be used as a study system to test whether the claim that "evolutionary winners are ecological losers" among island plants (Fernández-Palacios et al., 2021), which is based on an analysis of the Canary islands, also applies to the tropical islands in the Asia-Pacific region. Measures of evolutionary success, such as rates of species origination and morphological adaptation, could be inferred based on the dated phylogeny generated here (Fig. 3.4) and the rich datasets on palm morphology (Kissling et al., 2019; Henderson, 2020), whereas measures of ecological success, such as geographic range, could be estimated from the well-known distribution of calamoid palms (Vorontsova et al., 2016b; Henderson, 2020; Govaerts et al., 2021).

## 5.4 New views on tropical rainforest evolution

The palm family Arecaceae is an established model group for tropical rainforest evolution (Couvreur and Baker, 2013), and the Calamoideae plays a key role for the origins of the family due to its relationship as sister to all other palms (Baker et al., 2009; Barrett et al.,

2016b; Faurby et al., 2016). Further, calamoid palms are highly constrained to wet tropical climates (Reichgelt et al., 2018) and require forest habitats due to their predominantly climbing habit (Couvreur et al., 2015), making them good indicators of the presence of tropical rainforests. This thesis has therefore substantial implications for the understanding of the temporal and spatial context of the origin of tropical rainforests.

First, the minimum crown node age estimate for the Calamoideae of 122.3 Ma (Table 3.6) shows that the palms had already diversified into several lineages by the Early Cretaceous. Given that the oldest fossils of all calamoid tribes are found on fragments of Gondwana (Morley, 1998; Macphail and Jordan, 2015; Bogotá-Ángel et al., 2021), the widespread occurrence of the Calamoideae on Gondwana before the break-up of the supercontinent seems likely. Whilst this should not be over-interpreted as proof of the existence of "modern" tropical rainforests, which potentially only originated at the end of the Cretaceous (Carvalho et al., 2021), it indicates that key lineages characteristic of extant tropical rainforests already existed and had diversified on Gondwana by the Early Cretaceous.

Second, the sequence and timing of the early divergence of the Calamoideae and the location of the oldest calamoid fossils fit remarkably well with the sequence and timing of the break-up of Gondwana (Fig. 3.6). This combined evidence is compatible with the hypothesis that the geographical disjunction of the subfamily was caused by the drifting apart of the landmasses that constituted Gondwana (Baker and Dransfield, 2000), and challenges the currently prevalent view that pantropical distributions of extant plant and animal lineages are largely due to transoceanic dispersal (Givnish and Renner, 2004; Baker and Couvreur, 2013a; de Queiroz, 2014). In part, biogeographic inferences ruling out a

connection of pantropical distributions with the break-up of Gondwana were based on divergence time estimates post-dating the separation of these landmasses (Baker and Couvreur, 2013a). However, such age estimates might be biased towards younger ages by the implementation of too strict maximum age constraints (Chapter 3.4.4, Supplementary Fig. 3.4), which are often hard to justify (Ho and Phillips, 2009; Sauquet, 2013; Holland, 2016; Sauquet et al., 2021). It is therefore possible that more lenient maximum age constraints in other pantropical groups, such as the palm subfamilies Arecoideae, Coryphoideae, and Ceroxyloideae, may indirectly lead to substantially refined biogeographic narratives.

Third, the divergence time estimates and biogeographic inferences made here (Figs. 3.4-3.5, 4.2), combined with the rich and distinctive fossil record of the Calamoideae (Harley and Morley, 1995; Morley, 1998; Morley, 2000; Harley, 2006; Iles et al., 2015; Macphail and Jordan, 2015; Macphail and Hill, 2018; Bogotá-Ángel et al., 2021), paint a highly dynamic picture of tropical rainforest evolution through time. The oldest fossils of the Mauritiinae are from Africa (Bogotá-Ángel et al., 2021), but at present the subtribe is constrained to South America. Similarly, the earliest fossils of *Eugeissona* are from India (Morley, 1998), but the genus now only exists in the Thai-Malay Peninsula and Borneo. Finally, a continuous record of the fossil taxon *Dicolpopollis*, characteristic for a clade in tribe Calameae, exists in Australia since the Late Cretaceous (Truswell and Owen, 1988; Dransfield et al., 2008; Macphail and Jordan, 2015; Macphail and Hill, 2018), but the biogeographic inferences made here strongly suggest that the extant Calamoideae dispersed to Australia only around 20 million years ago (Figure 4.2). This combined evidence suggests dramatic shifts of the spatial distribution of tropical rainforests through time associated with the complete extinction of many calamoid lineages in the regions where they originated. In particular, the complete absence of extant calamoid lineages older than ca. 25 million years from India and Australia

(Fig. 4.2) indicates that past reductions in the extent of tropical rainforests in these regions (Morley, 2000) may have been so extreme that they resulted in the complete extinction of at least some characteristic tropical rainforest lineages. The prevalent pattern of eastward distribution of tropical rainforest lineages from Sundaland into Sahul (Crayn et al., 2015; Strijk et al., 2020; Joyce et al., 2021) may therefore be simply explained by the almost complete extirpation of tropical rainforests in Australia in the past, which would have disabled Australia's ability to act as a source of dispersals of tropical rainforests into other regions.

Fourth, the young ages of the vast majority of calamoid lineages compared to the ancient crown node age of the Calamoideae (Fig. 3.5) challenge previous inferences of a constant diversification of palm lineages through time (Couvreur et al., 2011a). As this pattern can be repeatedly observed across phylogenetically divergent calamoid lineages in all regions of the world (Fig. 3.4), it can be assumed that recent origin of most calamoid lineages is not a peculiarity of the subfamily but may constitute a common characteristic of tropical rainforest lineages.

Of the many potential avenues of future research on the origins of tropical rainforest diversity, one prospect building on the phylogenomic and biogeographic frameworks established here (Fig. 4.2) is outlined in the following. While it is well-established that tropical rainforests contain almost half of all plant species on just 7% of earth's land surface (Eiserhardt et al., 2017), it remains poorly understood how so many species can co-exist. The calamoid palms would present an appropriate study system to address this knowledge gap. For example, over 100 species of the Calamoideae occur on the island of Borneo alone. Although several different calamoid species often co-occur in close proximity in the same

habitats (Petoe et al., 2020), most species are morphologically highly distinctive, allowing the clear differentiation even in their vegetative state (Dransfield, 1992; Henderson, 2020). However, it remains unclear how so many calamoid species with apparently similar ecological requirements can co-exist and how species boundaries can be maintained, particularly as most calamoid species are dioecious with obligatory outcrossing (Dransfield et al., 2008). The biogeographic analyses conducted here allow to select several study areas with contrasting species compositions, such as Borneo, Sumatra, New Guinea, and Palawan (Fig. 4.4), which may make it possible to distinguish general patterns from regional idiosyncrasies. The dated species-level phylogeny (Fig. 3.4) could then be used to reveal how closely related co-existing calamoid species are (Webb et al., 2002) and over what timeframe these communities formed. Finally, findings from these evolutionary analyses could be linked to measures of morphological similarity (Petchey and Gaston, 2002) by utilising extensive morphological datasets (Kissling et al., 2019; Henderson, 2020), enabling a new understanding of species co-existence by uniting evolution and ecology. For a better understanding of the maintenance of species boundaries, population genomic methods (Durand et al., 2011; Solís-Lemus et al., 2017) could then be employed to infer the degree of genetic exchange between species. In particular, lineages for which high degrees of gene tree conflict were inferred here, such as the Plectocomiinae (Fig. 2.4) or species complexes in *Calamus* (Supplementary Fig. 3.2), could provide a good starting point for such analyses.

## 5.5 Conclusion

The high diversity of tropical rainforests has enthralled researchers for over a century, but many details of their origins have remained nebulous. This thesis helps to fill this knowledge gap based on inferences on phylogenetic relationships, ages, and dispersal history of the most densely sampled major tropical rainforest lineage to date. Together, these findings allow unique new insights into the temporal and spatial dynamics of tropical rainforest evolution. There are many exciting avenues for research building on this highly detailed evolutionary and biogeographic framework, which may unlock even more secrets of the most species-rich terrestrial biome on earth.

# 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.

Albalat, R., Cañestro, C., 2016. Evolution by gene loss. Nat. Rev. Genet. 17, 379-391. https://doi.org/10.1038/nrg.2016.39.

Allen, L., Allen, E., Kunst, C., Sosebee, R., 1991. A diffusion model for dispersal of *Opuntia imbricata* (cholla) on rangeland. J. Ecol. 79, 1123-1135.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402. https://doi.org/10.1093/nar/25.17.3389.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed 22 July 2021).

Antonelli, A., Hettling, H., Condamine, F.L., Vos, K., Nilsson, R.H., Sanderson, M.J., Sauquet, H., Scharn, R., Silvestro, D., Topel, M., Bacon, C.D., Oxelman, B., Vos, R.A., 2017. Toward a self-updating platform for estimating rates of speciation and migration, ages, and relationships of taxa. Syst. Biol. 66, 152-166. https://doi.org/10.1093/sysbio/syw066.

Antonelli, A., Sanmartín, I., 2011. Mass extinction, gradual cooling, or rapid radiation? Reconstructing the spatiotemporal evolution of the ancient angiosperm genus *Hedyosmum* (Chloranthaceae) using empirical and simulated approaches. Syst. Biol. 60, 596-615. https://doi.org/10.1093/sysbio/syr062.

Asmussen, C.B., Baker, W.J., Dransfield, J., 2000. Phylogeny of the palm family (Arecaceae) based on *rps*16 intron and *trn*L-*trn*F plastid DNA sequences. In: Wilson, K. L., Morrison, D. A. (Eds.), Monocots: systematics and evolution. CSIRO, Melbourne, pp. 525-537.

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.

Ayres, D.L., Cummings, M.P., Baele, G., Darling, A.E., Lewis, P.O., Swofford, D.L., Huelsenbeck, J.P., Lemey, P., Rambaut, A., Suchard, M.A., 2019. BEAGLE 3: improved performance, scaling, and usability for a high-performance computing library for statistical phylogenetics. Syst. Biol. 68, 1052-1061. https://doi.org/10.1093/sysbio/syz020.

Baker, W.J., Dransfield, J., 2000. Towards a biogeographic explanation of the calamoid palms. Monocots: systematics and evolution. Collingwood, Australia: CSIRO, 545-553.

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., Asmussen, C.B., Barrow, S.C., Dransfield, J., Hedderson, T.A., 1999a. A phylogenetic study of the palm family (Palmae) based on chloroplast DNA sequences from the *trn*L-*trn*F region. Plant. Syst. Evol. 219, 111-126.

Baker, W.J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L.R., Brewer, G., Carruthers, T., Clarkson, J.J., Cook, J., Cowan, R.S., Dodsworth, S., Epitawalage, N., Françoso, E., Gallego, B., Johnson, M.G., Kim, J.T., Leempoel, K., Maurin, O., Mcginnie, C., Pokorny, L., Roy, S., Stone, M., Toledo, E., Wickett, N.J., Zuntini, A.R., Eiserhardt, W.L., Kersey, P.J., Leitch, I.J., Forest, F., 2021. A comprehensive phylogenomic platform for exploring the angiosperm Tree of Life. Syst. Biol. https://doi.org/10.1093/sysbio/syab035.

Baker, W.J., Coode, M.J.E., Dransfield, J., Dransfield, S., Harley, M.M., Hoffmann, P., Johns, R.J., 1998. Patterns of distribution of Malesian vascular plants. In: Hall, R., Holloway, J. (Eds.), Biogeography and geological evolution of SE Asia. Backhuys Publishers, Leiden.

Baker, W.J., Couvreur, T.L., 2012. Biogeography and distribution patterns of Southeast Asian palms. In: Gower, D., Johnson, K., Richardson, J., Rosen, B., Rüber, L., Williams, S. (Eds.), Biotic evolution and environmental change in Southeast Asia. Cambridge University Press, Cambridge, pp. 164-190.

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., 1999b. 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 *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., Hedderson, T.A., Dransfield, J., 2000c. 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., 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.

Balslev, H., Valencia, R., Paz-y-Mino-C, G., Christensen, H., Nielsen, I., 1998. Species count of vascular plants in one hectare of humid lowland forest in Amazonian Ecuador. In: Dallmeier, F., Comiskey, J. A. (Eds.), Forest Biodiversity in North, Central and South America, and the Caribbean. UNESCO, Paris, pp. 585-594.

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., 2016a. 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., 2016b. 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.

Beaulieu, J.M., O'Meara, B.C., 2015. Extinction can be estimated from moderately sized molecular phylogenies. Evolution 69, 1036-1043. https://doi.org/10.1111/evo.12614.

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.

Bogotá-Ángel, G., Huang, H., Jardine, P.E., Chazot, N., Salamanca, S., Banks, H., Pardo-Trujillo, A., Plata, A., Dueñas, H., Star, W., Langelaan, R., Eisawi, A., Umeji, O.P., Enuenwemba, L.O., Parmar, S., da Silveira, R.R., Lim, J.Y., Prasad, V., Morley, R.J., Bacon, C.D., Hoorn, C., 2021. Climate and geological change as drivers of Mauritiinae palm biogeography. J. Biogeog., 1-22. https://doi.org/10.1111/jbi.14098.

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.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., Drummond, A.J., 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15, e1006650. https://doi.org/10.1371/journal.pcbi.1006650.

Bromham, L., 2019. Six impossible things before breakfast: assumptions, models, and belief in molecular dating. Trends Ecol. Evol. 34, 474-486. https://doi.org/10.1016/j.tree.2019.01.017.

Bromham, L., Duchêne, S., Hua, X., Ritchie, A.M., Duchêne, D.A., Ho, S.Y.W., 2018. Bayesian molecular dating: opening up the black box. Biol. Rev. 93, 1165-1191. https://doi.org/10.1111/brv.12390.

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.

Brummitt, R.K., Pando, F., Hollis, S., Brummitt, N., 2001. World geographical scheme for recording plant distributions. International Working Group on Taxonomic Databases for Plant Sciences (TDWG).

Budd, G.E., Mann, R.P., Doyle, J.A., Coiro, M., Hilton, J., 2021. Fossil data do not support a long pre-Cretaceous history of flowering plants. bioRxiv, 2021.2002.2016.431478. https://doi.org/10.1101/2021.02.16.431478.

Burger, D., 1990. Early cretaceous angiosperms from Queensland, Australia. Rev. Palaeobot. Palynol. 65, 153-163. https://doi.org/10.1016/0034-6667(90)90066-R.

Burnham, R.J., Johnson, K.R., 2004. South American palaeobotany and the origins of neotropical rainforests. Philos. Trans. R. Soc. B. 359, 1595-1610. https://doi.org/10.1098/rstb.2004.1531.

Cámara-Leret, R., Frodin, D.G., Adema, F., Anderson, C., Appelhans, M.S., Argent, G., Arias Guerrero, S., Ashton, P., Baker, W.J., Barfod, A.S., Barrington, D., Borosova, R., Bramley, G.L.C., Briggs, M., Buerki, S., Cahen, D., Callmander, M.W., Cheek, M., Chen, C.-W., Conn, B.J., Coode, M.J.E., Darbyshire, I., Dawson, S., Dransfield, J., Drinkell, C., Duyfjes, B., Ebihara, A., Ezedin, Z., Fu, L.-F., Gideon, O., Girmansyah, D., Govaerts, R., Fortune-Hopkins, H., Hassemer, G., Hay, A., Heatubun, C.D., Hind, D.J.N., Hoch, P., Homot, P., Hovenkamp, P., Hughes, M., Jebb, M., Jennings, L., Jimbo, T., Kessler, M., Kiew, R., Knapp, S., Lamei, P., Lehnert, M., Lewis, G.P., Linder, H.P., Lindsay, S., Low, Y.W., Lucas, E., Mancera, J.P., Monro, A.K., Moore, A., Middleton, D.J., Nagamasu, H., Newman, M.F., Nic Lughadha, E., Melo, P.H.A., Ohlsen, D.J., Pannell, C.M., Parris, B., Pearce, L., Penneys, D.S., Perrie, L.R., Petoe, P., Poulsen, A.D., Prance, G.T., Quakenbush, J.P., Raes, N., Rodda, M., Rogers, Z.S., Schuiteman, A., Schwartsburd, P., Scotland, R.W., Simmons, M.P., Simpson, D.A., Stevens, P., Sundue, M., Testo, W., Trias-Blasi, A., Turner, I., Utteridge, T., Walsingham, L., Webber, B.L., Wei, R., Weiblen, G.D., Weigend, M., Weston, P., de Wilde, W., Wilkie, P., Wilmot-Dear, C.M., Wilson,

H.P., Wood, J.R.I., Zhang, L.-B., van Welzen, P.C., 2020. New Guinea has the world's richest island flora. Nature 584, 579-583. https://doi.org/10.1038/s41586-020-2549-5.

Cannon, C.H., Morley, R.J., Bush, A.B.G., 2009. The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. Proc. Natl. Acad. Sci. 106, 11188-11193. https://doi.org/10.1073/pnas.0809865106.

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.

Carvalho, M.R., Jaramillo, C., de la Parra, F., Caballero-Rodríguez, D., Herrera, F., Wing, S., Turner, B.L., D'Apolito, C., Romero-Báez, M., Narváez, P., Martínez, C., Gutierrez, M., Labandeira, C., Bayona, G., Rueda, M., Paez-Reyes, M., Cárdenas, D., Duque, Á., Crowley, J.L., Santos, C., Silvestro, D., 2021. Extinction at the end-Cretaceous and the origin of modern Neotropical rainforests. Science 372, 63-68. https://doi.org/10.1126/science.abf1969.

Chan, K.-L., Tatarinova, T.V., Rosli, R., Amiruddin, N., Azizi, N., Halim, M.A.A., Sanusi, N.S.N.M., Jayanthi, N., Ponomarenko, P., Triska, M., Solovyev, V., Firdaus-Raih, M., Sambanthamurthi, R., Murphy, D., Low, E.-T.L., 2017. Evidence-based gene models for structural and functional annotations of the oil palm genome. Biol. Direct 12, 21. https://doi.org/10.1186/s13062-017-0191-4.

Chase, M.W., Duvall, M.R., Hills, H.G., Conran, J.G., Cox, A.V., Eguiarte, L.E., Hartwell, J., Fay, M.F., Caddick, L.R., Cameron, K.M., Hoot, S., 1995. Molecular phylogenetics of Lilianae. Monocotyledons: systematics and evolution 1, 109-137.

Chase, M.W., Fay, M.F., Devey, D.S., Maurin, O., Rønsted, N., Davies, T.J., Pillon, Y., Peterson, G., Tamura, M.N., Asmussen, C.B., Hilu, K., Borsch, T., Davis, J.I., Stevenson, D.W., Pires, J.C., Givnish, T.J., Sytsma, K.J., McPherson, M.A., Graham, S.W., Rai, H.S., 2006. Multigene analyses of monocot relationships. Aliso 22, 63-75.

Chase, M.W., Soltis, D.E., Olmstead, R.G., Morgan, D., Les, D.H., Mishler, B.D., Duvall, M.R., Price, R.A., Hills, H.G., Qiu, Y.-L., Kron, K.A., Rettig, J.H., Conti, E., Palmer, J.D., Manhart, J.R., Sytsma, K.J., Michaels, H.J., Kress, W.J., Karol, K.G., Clark, W.D., Hedren, M., Gaut, B.S., Jansen, R.K., Kim, K.-J., Wimpee, C.F., Smith, J.F., Furnier, G.R., Strauss, S.H., Xiang, Q.-Y., Plunkett, G.M., Soltis, P.S., Swensen, S.M., Williams, S.E., Gadek, P.A., Quinn, C.J., Eguiarte, L.E., Golenberg, E., Learn, G.H., Graham, S.W., Spencer, C.H.B., Dayanandan, S., Albert, V.A.,

1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Mo. Bot. Gard. 80, 528-580. https://doi.org/10.2307/2399846.

Chase, M.W., Soltis, D.E., Soltis, P.S., Rudall, P.J., Fay, M.F., Hahn, W.H., Sullivan, S., Joseph, J., Molvray, M., Kores, P.J., Givnish, T.J., Sytsma, K.J., Pires, J.C., 2000. Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification. Monocots: Systematics and Evolution, 3-16.

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.

Condamine, F.L., Nagalingum, N.S., Marshall, C.R., Morlon, H., 2015. Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. BMC Evol. Biol. 15, 65. https://doi.org/10.1186/s12862-015-0347-8.

Cook, L.G., Crisp, M.D., 2005. Directional asymmetry of long-distance dispersal and colonization could mislead reconstructions of biogeography. J. Biogeog. 32, 741-754. https://doi.org/10.1111/j.1365-2699.2005.01261.x.

Corlett, R.T., Primack, R.B., 2011. Tropical rain forests: an ecological and biogeographical comparison. John Wiley & Sons.

Couvreur, T.L.P., 2015. Odd man out: why are there fewer plant species in African rain forests? Plant. Syst. Evol. 301, 1299-1313. https://doi.org/10.1007/s00606-014-1180-z.

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., Dauby, G., Blach-Overgaard, A., Deblauwe, V., Dessein, S., Droissart, V., Hardy, O.J., Harris, D.J., Janssens, S.B., Ley, A.C., Mackinder, B.A., Sonké, B., Sosef, M.S.M., Stévart, T., Svenning, J.-C., Wieringa, J.J., Faye, A., Missoup, A.D., Tolley, K.A., Nicolas, V., Ntie, S., Fluteau, F., Robin, C., Guillocheau, F., Barboni, D., Sepulchre, P., 2021. Tectonics, climate and the diversification of the tropical African terrestrial flora and fauna. Biol. Rev. 96, 16-51. https://doi.org/10.1111/brv.12644.

Couvreur, T.L.P., Forest, F., Baker, W.J., 2011a. 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.

Couvreur, T.L.P., Kissling, W.D., Condamine, F.L., Svenning, J.C., Rowe, N.P., Baker, W.J., 2015. Global diversification of a tropical plant growth form: environmental correlates and historical

contingencies in climbing palms. Front. Genet. 5, 1-18. https://doi.org/10.3389/fgene.2014.00452.

Couvreur, T.L.P., Pirie, M.D., Chatrou, L.W., Saunders, R.M.K., Su, Y.C.F., Richardson, J.E., Erkens, R.H.J., 2011b. Early evolutionary history of the flowering plant family Annonaceae: steady diversification and boreotropical geodispersal. J. Biogeog. 38, 664-680. https://doi.org/10.1111/j.1365-2699.2010.02434.x.

Cowell, C., Paton, A., Borrell, J.S., Williams, C., Wilkin, P., Antonelli, A., Baker, W.J., Buggs, R., Fay, M.F., Gargiulo, R., Grace, O.M., Kuhnhäuser, B.G., Woudstra, Y., Kersey, P.J., 2021. Uses and benefits of digital sequence information from plant genetic resources: lessons learnt from botanical collections. Plants, People, Planet, 11. https://doi.org/10.1002/ppp3.10216.

Crayn, D.M., Costion, C., Harrington, M.G., 2015. The Sahul–Sunda floristic exchange: dated molecular phylogenies document Cenozoic intercontinental dispersal dynamics. J. Biogeog. 42, 11-24. https://doi.org/10.1111/jbi.12405.

Crisp, M.D., Cook, L.G., 2011. Cenozoic extinctions account for the low diversity of extant gymnosperms compared with angiosperms. New Phytol. 192, 997-1009. https://doi.org/10.1111/j.1469-8137.2011.03862.x.

Crisp, M.D., Trewick, S.A., Cook, L.G., 2011. Hypothesis testing in biogeography. Trends Ecol. Evol. 26, 66-72. https://doi.org/10.1016/j.tree.2010.11.005.

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 Bruyn, M., Stelbrink, B., Morley, R.J., Hall, R., Carvalho, G.R., Cannon, C.H., van den Bergh, G., Meijaard, E., Metcalfe, I., Boitani, L., Maiorano, L., Shoup, R., von Rintelen, T., 2014. Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity. Syst. Biol. 63, 879-901. https://doi.org/10.1093/sysbio/syu047.

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.

de Queiroz, A., 2014. The monkey's voyage : how improbable journeys shaped the history of life. Basic Books, New York.

De Smet, R., Adams, K.L., Vandepoele, K., Van Montagu, M.C.E., Maere, S., Van de Peer, Y., 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc. Natl. Acad. Sci. 110, 2898-2903. https://doi.org/10.1073/pnas.1300127110.

Degnan, J.H., 2018. Modeling hybridization under the network multispecies coalescent. Syst. Biol. 67, 786-799. https://doi.org/10.1093/sysbio/syy040.

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.

Donoghue, M.J., Edwards, E.J., 2014. Biome shifts and niche evolution in plants. Annu. Rev. Ecol. Evol. S. 45, 547-572. https://doi.org/10.1146/annurev-ecolsys-120213-091905.

Donoghue, M.J., Sanderson, M.J., 2015. Confluence, synnovation, and depauperons in plant diversification. New Phytol. 207, 260-274. https://doi.org/10.1111/nph.13367.

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., 1981a. Palms and Wallace's line. In: Whitmore, T. (Ed.), Wallace's line and plate tectonics. Clarendon Press, Oxford, p. 91.

Dransfield, J., 1981b. 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., 1987. Bicentric distribution in Malesia as exemplified by palms. In: Whitmore, T. C. (Ed.), Biogeographical evolution of the Malay Archipelago. Clarendon Press, Oxford, pp. 60-72.

Dransfield, J., 1992. The rattans of Sarawak. Royal Botanic Gardens Kew and Sarawak Forest Department, Richmond and Kuching.

Dransfield, J., Mogea, J., 1981. A reassessment of the genus *Lophospatha* Burret. Principes 25, 178-180.

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.

Drummond, A.J., Bouckaert, R.R., 2015. Bayesian evolutionary analysis with BEAST. Cambridge University Press, Cambridge.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, 699-710. https://doi.org/10.1371/journal.pbio.0040088.

Dupin, J., Matzke, N.J., Särkinen, T., Knapp, S., Olmstead, R.G., Bohs, L., Smith, S.D., 2017. Bayesian estimation of the global biogeographical history of the Solanaceae. J. Biogeog. 44, 887-899. https://doi.org/10.1111/jbi.12898.

Durand, E.Y., Patterson, N., Reich, D., Slatkin, M., 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28, 2239-2252. https://doi.org/10.1093/molbev/msr048.

Edet, J.J., Nyong, E.E., 1993. Depositional environments, sea-level history and palaeobiogeography of the late Campanian-Maastrichtian on the Calabar flank, SE Nigeria. Palaeogeogr. Palaeocl. 102, 161-175. https://doi.org/10.1016/0031-0182(93)90010-G.

Eisawi, A., Schrank, E., 2008. Upper Cretaceous to Neogene palynology of the Melut Basin, Southeast Sudan. Palynology 32, 101-129. https://doi.org/10.2113/gspalynol.32.1.101.

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.

Eiserhardt, W.L., Svenning, J.-C., Kissling, W.D., Balslev, H., 2011. Geographical ecology of the palms (Arecaceae): determinants of diversity and distributions across spatial scales. Ann. Bot. 108, 1391-1416. https://doi.org/10.1093/aob/mcr146.

Erkens, R.H.J., Chatrou, L.W., Maas, J.W., van der Niet, T., Savolainen, V., 2007. A rapid diversification of rainforest trees (*Guatteria*; Annonaceae) following dispersal from Central into South America. Mol. Phylogenet. Evol. 44, 399-411. https://doi.org/10.1016/j.ympev.2007.02.017.

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.

Faye, A., Pintaud, J.C., Baker, W.J., Sonké, B., Couvreur, T.L.P., 2014. A plastid phylogeny of the African rattans (Ancistrophyllinae, Arecaceae). Syst. Bot. 39, 1099-1107. https://doi.org/10.1600/036364414X682634.

Fernández-Palacios, J.M., Otto, R., Borregaard, M.K., Kreft, H., Price, J.P., Steinbauer, M.J., Weigelt, P., Whittaker, R.J., 2021. Evolutionary winners are ecological losers among oceanic island plants. J. Biogeog 48, 2186-2198. https://doi.org/10.1111/jbi.14143.

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., Puad, A.S.A., Arifiani, D., Sahromi, Wickett, N.J., Zerega, N.J.C., 2020. Paralogs and off-target sequences improve phylogenetic resolution in a densely-sampled study of the breadfruit genus (*Artocarpus*, Moraceae). Syst. Biol. syaa073, 1-18. https://doi.org/10.1093/sysbio/syaa073.

Gaut, B.S., Morton, B.R., McCaig, B.C., Clegg, M.T., 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbc*L. Proc. Natl. Acad. Sci. 93, 10274-10279. https://doi.org/10.1073/pnas.93.19.10274.

Gentry, A.H., 1992. Tropical forest biodiversity: distributional patterns and their conservational significance. Oikos, 19-28.

Ghosh Dasgupta, M., Dev, S.A., Muneera Parveen, A.B., Sarath, P., Sreekumar, V.B., 2021. Draft genome of *Korthalsia laciniosa* (Griff.) Mart., a climbing rattan elucidates its phylogenetic position. Genomics 113, 2010-2022. https://doi.org/10.1016/j.ygeno.2021.04.023.

Givnish, T.J., Renner, S.S., 2004. Tropical intercontinental disjunctions: Gondwana breakup, immigration from the boreotropics, and transoceanic dispersal. Int. J. Plant Sci. 165, S1-S6. https://doi.org/10.1086/424022.

Givnish, T.J., Zuluaga, A., Spalink, D., Soto Gomez, M., Lam, V.K.Y., Saarela, J.M., Sass, C., Iles, W.J.D., de Sousa, D.J.L., Leebens-Mack, J., Chris Pires, J., Zomlefer, W.B., Gandolfo, M.A., Davis, J.I., Stevenson, D.W., dePamphilis, C., Specht, C.D., Graham, S.W., Barrett, C.F., Ané, C., 2018. 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. 105, 1888-1910. https://doi.org/10.1002/ajb2.1178.

Govaerts, R., Dransfield, J., 2005. World checklist of palms. Royal Botanic Gardens, Kew.

Govaerts, R., Dransfield, J., Zona, S.F., Hodel, D.R., Henderson, A., 2021. World checklist of Arecaceae. Facilitated by the Royal Botanic Gardens, Kew. Available at: http://wcsp.science.kew.org (accessed 22 July 2021).

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

Grudinski, M., Wanntorp, L., Pannell, C.M., Muellner-Riehl, A.N., 2014. West to east dispersal in a widespread animal-dispersed woody angiosperm genus (*Aglaia*, Meliaceae) across the Indo-Australian Archipelago. J. Biogeog. 41, 1149-1159. https://doi.org/10.1111/jbi.12280.

Hagen, O., Skeels, A., Onstein, R., Jetz, W., Pellissier, L., 2021. Earth history events shaped the evolution of uneven biodiversity across tropical moist forests. Proc. Natl. Acad. Sci. 118, e2026347118. https://doi.org/10.1073/pnas.2026347118.

Hahn, W.J., 2002a. 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.

Hahn, W.J., 2002b. A phylogenetic analysis of the arecoid line of palms based on plastid DNA sequence data. Mol. Phylogenet. Evol. 23, 189-204. https://doi.org/10.1016/S1055-7903(02)00022-2.

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.

Hall, R., 1998. The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. In: Hall, R., Holloway, J. (Eds.), Biogeography and geological evolution of SE Asia. Bukhuys Publishers, Amsterdam, pp. 99-131.

Hall, R., 2009. Southeast Asia's changing palaeogeography. Blumea 54, 148-161. https://doi.org/10.3767/000651909x475941.

Hall, R., 2011. Australia–SE Asia collision: plate tectonics and crustal flow. Geological Society, London, Special Publications 355, 75-109. https://doi.org/10.1144/sp355.5.

Hall, R., 2012a. Late Jurassic–Cenozoic reconstructions of the Indonesian region and the Indian Ocean. Tectonophysics 570-571, 1-41. https://doi.org/10.1016/j.tecto.2012.04.021.

Hall, R., 2012b. Sundaland and Wallacea: geology, plate tectonics and palaeogeography. In: Gower, D. J., Richardson, J. E., Rosen, B. R., Rüber, L., Williams, S. T. (Eds.), Biotic evolution and environmental change in Southeast Asia. Cambridge University Press, Cambridge, pp. 32-78. https://doi.org/10.1017/CBO9780511735882.005.

Hall, R., 2017. Southeast Asia: new views of the geology of the Malay Archipelago. Annu. Rev. Earth Pl. Sc. 45, 331-358. https://doi.org/10.1146/annurev-earth-063016-020633.

Hansen, L.E.S.F., Baker, W.J., Tietje, M., Eiserhardt, W.L., 2021. Testing tropical biogeographical regions using the palm family as a model clade. J. Biogeog 48, 2502– 2511. https://doi.org/10.1111/jbi.14216.

Harley, M.M., 2006. A summary of fossil records for Arecaceae. Bot. J. Linn. Soc. 151, 39-67. https://doi.org/10.1111/j.1095-8339.2006.00522.x.

Harley, M.M., Morley, R.J., 1995. Ultrastructural studies of some fossil and extant palm pollen, and the reconstruction of the biogeographical history of subtribes Iguanurinae and Calaminae. Rev. Palaeobot. Palynol. 85, 153-182. https://doi.org/10.1016/0034-6667(94)00133-5.

Harmon, L.J., Harrison, S., 2015. Species diversity is dynamic and unbounded at local and continental scales. Am. Nat. 185, 584-593.

Hart, M.L., Forrest, L.L., Nicholls, J.A., Kidner, C.A., 2016. Retrieval of hundreds of nuclear loci from herbarium specimens. Taxon 65, 1081-1092. https://doi.org/10.12705/655.9.

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.

Heled, J., Drummond, A.J., 2011. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Syst. Biol. 61, 138-149. https://doi.org/10.1093/sysbio/syr087.

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.

Hendriks, K.P., Mandáková, T., Hay, N.M., Ly, E., Hooft van Huysduynen, A., Tamrakar, R., Thomas, S.K., Toro-Núñez, O., Pires, J.C., Nikolov, L.A., Koch, M.A., Windham, M.D., Lysak, M.A., Forest, F., Mummenhoff, K., Baker, W.J., Lens, F., Bailey, C.D., 2021. The best of both worlds: combining lineage-specific and universal bait sets in target-enrichment hybridization reactions. Appl. Plant Sci. 9. https://doi.org/10.1002/aps3.11438.

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.

Hijmans, R.J., Williams, E., Vennes, C., 2019. R package 'geosphere: spherical trigonometry'.

Hill, J.L., Hill, R.A., 2001. Why are tropical rain forests so species rich? Classifying, reviewing and evaluating theories. Prog. Phys. Geogr. 25, 326-354. https://doi.org/10.1177/030913330102500302.

Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M., Crandall, K.A., Deng, J., Drew, B.T., Gazis, R., Gude, K., Hibbett, D.S., Katz, L.A., Laughinghouse, H.D., McTavish, E.J., Midford, P.E., Owen, C.L., Ree, R.H., Rees, J.A., Soltis, D.E., Williams, T., Cranston, K.A., 2015. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proc. Natl. Acad. Sci. 112, 12764-12769. https://doi.org/10.1073/pnas.1423041112.

Ho, S.Y.W., Phillips, M.J., 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst. Biol. 58, 367-380. https://doi.org/10.1093/sysbio/syp035.

Holland, S.M., 2016. The non-uniformity of fossil preservation. Philos. Trans. R. Soc. B 371, 20150130. https://doi.org/10.1098/rstb.2015.0130.

Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., Sanmartín, I., Sanchez-Meseguer, A., Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., Negri, F.R., Hooghiemstra, H., Lundberg, J., Stadler, T., Särkinen, T., Antonelli, A., 2010. Amazonia

through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science 330, 927-931. https://doi.org/10.1126/science.1194585.

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.

Huang, H., Morley, R., Licht, A., Dupont-Nivet, G., Grímsson, F., Zetter, R., Westerweel, J., Win, Z., Wa Aung, D., Hoorn, C., 2020. Eocene palms from central Myanmar in a South-East Asian and global perspective: evidence from the palynological record. Bot. J. Linn. Soc., 1-30. https://doi.org/10.1093/botlinnean/boaa038.

Hughes, C.E., Pennington, R.T., Antonelli, A., 2013. Neotropical plant evolution: assembling the big picture. Bot. J. Linn. Soc. 171, 1-18. https://doi.org/10.1111/boj.12006.

Huxley, T.H., 1868. On the classification and distribution of the Alectoromorphae and Heteromorphae. Proc. Zool. Soc. Lond. 294-319.

Iles, W.J.D., Smith, S.Y., Gandolfo, M.A., Graham, S.W., 2015. Monocot fossils suitable for molecular dating analyses. Bot. J. Linn. Soc. 178, 346-374. https://doi.org/10.1111/boj.12233.

Isnard, S., 2006. Biomechanics and development of rattans: what is special about *Plectocomia himalayana* Griff. (Calamoideae, Plectocomiinae)? Bot. J. Linn. Soc. 151, 83-91. https://doi.org/10.1111/j.1095-8339.2006.00524.x.

Isnard, S., Rowe, N.P., 2008. The climbing habit in palms: biomechanics of the cirrus and flagellum. Am. J. Bot. 95, 1538-1547. https://doi.org/10.3732/ajb.0700005.

Johnson, M., Pokorny, L., Dodsworth, S., Botigué, L., Cowan, R., Devault, A., Eiserhardt, W., Epitawalage, N., Forest, F., Kim, J., Leebens-Mack, J.H., Leitch, I., Maurin, O., Soltis, D., Soltis, P., Ka-Shu Wong, G., Baker, W., Wickett, N., 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. Syst. Biol. 68, 594–606. https://doi.org/10.1093/sysbio/syy086.

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.

Joyce, E.M., Pannell, C.M., Rossetto, M., Yap, J.-Y.S., Thiele, K.R., Wilson, P.D., Crayn, D.M., 2021. Molecular phylogeography reveals two geographically and temporally separated floristic

exchange tracks between Southeast Asia and northern Australia. J. Biogeog. 48, 1213-1227. https://doi.org/10.1111/jbi.14072.

Joyce, E.M., Thiele, K.R., Slik, F.J.W., Crayn, D.M., 2020a. Checklist of the vascular flora of the Sunda-Sahul Convergence Zone. Biodivers. Data J. 8. https://doi.org/10.3897/BDJ.8.e51094.

Joyce, E.M., Thiele, K.R., Slik, J.W.F., Crayn, D.M., 2020b. Plants will cross the lines: climate and available land mass are the major determinants of phytogeographical patterns in the Sunda–Sahul Convergence Zone. Biol. J. Linn. Soc. https://doi.org/10.1093/biolinnean/blaa194.

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.

Khan, A., 1976. Palynology of Tertiary sediments from Papua New Guinea. I. New form genera and species from upper Tertiary sediments. Aust. J. Bot. 24, 753-781. https://doi.org/10.1071/BT9760753.

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

Klaus, K.V., Matzke, N.J., 2019. Statistical comparison of trait-dependent biogeographical models indicates that Podocarpaceae dispersal is influenced by both seed cone traits and geographical distance. Syst. Biol. 69, 61-75. https://doi.org/10.1093/sysbio/syz034.

Koenen, E.J.M., Clarkson, J.J., Pennington, T.D., Chatrou, L.W., 2015. Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. New Phytol. 207, 327-339. https://doi.org/10.1111/nph.13490.

Kooyman, R.M., Morley, R.J., Crayn, D.M., Joyce, E.M., Rossetto, M., Slik, J.W.F., Strijk, J.S., Su, T., Yap, J.-Y.S., Wilf, P., 2019. Origins and assembly of Malesian rainforests. Annu. Rev. Ecol. Evol. S. 50, 119–143. https://doi.org/10.1146/annurev-ecolsys-110218-024737.

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

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. https://doi.org/10.5281/zenodo.4359280.

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., 2021. A robust phylogenomic framework for the calamoid palms. Mol. Phylogenet. Evol. 157, 107067. https://doi.org/10.1016/j.ympev.2020.107067.

Landis, M.J., Matzke, N.J., Moore, B.R., Huelsenbeck, J.P., 2013. Bayesian analysis of biogeography when the number of areas is large. Syst. Biol. 62, 789-804. https://doi.org/10.1093/sysbio/syt040.

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.

Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., Deyholos, M.K., Gitzendanner, M.A., Graham, S.W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., Porsch, M., Quint, M., Rensing, S.A., Soltis, D.E., Soltis, P.S., Stevenson, D.W., Ullrich, K.K., Wickett, N.J., DeGironimo, L., Edger, P.P., Jordon-Thaden, I.E., Joya, S., Liu, T., Melkonian, B., Miles, N.W., Pokorny, L., Quigley, C., Thomas, P., Villarreal, J.C., Augustin, M.M., Barrett, M.D., Baucom, R.S., Beerling, D.J., Benstein, R.M., Biffin, E., Brockington, S.F., Burge, D.O., Burris, J.N., Burris, K.P., Burtet-Sarramegna, V.,

Caicedo, A.L., Cannon, S.B., Çebi, Z., Chang, Y., Chater, C., Cheeseman, J.M., Chen, T., Clarke, N.D., Clayton, H., Covshoff, S., Crandall-Stotler, B.J., Cross, H., dePamphilis, C.W., Der, J.P., Determann, R., Dickson, R.C., Di Stilio, V.S., Ellis, S., Fast, E., Feja, N., Field, K.J., Filatov, D.A., Finnegan, P.M., Floyd, S.K., Fogliani, B., García, N., Gâteblé, G., Godden, G.T., Goh, F., Greiner, S., Harkess, A., Heaney, J.M., Helliwell, K.E., Heyduk, K., Hibberd, J.M., Hodel, R.G.J., Hollingsworth, P.M., Johnson, M.T.J., Jost, R., Joyce, B., Kapralov, M.V., Kazamia, E., Kellogg, E.A., Koch, M.A., Von Konrat, M., Könyves, K., Kutchan, T.M., Lam, V., Larsson, A., Leitch, A.R., Lentz, R., Li, F.-W., Lowe, A.J., Ludwig, M., Manos, P.S., Mavrodiev, E., McCormick, M.K., McKain, M., McLellan, T., McNeal, J.R., Miller, R.E., Nelson, M.N., Peng, Y., Ralph, P., Real, D., Riggins, C.W., Ruhsam, M., Sage, R.F., Sakai, A.K., Scascitella, M., Schilling, E.E., Schlösser, E.-M., Sederoff, H., Servick, S., Sessa, E.B., Shaw, A.J., Shaw, S.W., Sigel, E.M., Skema, C., Smith, A.G., Smithson, A., Stewart, C.N., Stinchcombe, J.R., Szövényi, P., Tate, J.A., Tiebel, H., Trapnell, D., Villegente, M., Wang, C.-N., Weller, S.G., Wenzel, M., Weststrand, S., Westwood, J.H., Whigham, D.F., Wu, S., Wulff, A.S., Yang, Y., Zhu, D., Zhuang, C., Zuidof, J., Chase, M.W., Pires, J.C., Rothfels, C.J., Yu, J., Chen, C., Chen, L., Cheng, S., Li, J., Li, R., Li, X., Lu, H., Ou, Y., Sun, X., Tan, X., Tang, J., Tian, Z., Wang, F., Wang, J., Wei, X., Xu, X., Yan, Z., Yang, F., Zhong, X., Zhou, F., Zhu, Y., Zhang, Y., Ayyampalayam, S., Barkman, T.J., Nguyen, N.-p., Matasci, N., Nelson, D.R., Sayyari, E., Wafula, E.K., Walls, R.L., Warnow, T., An, H., Arrigo, N., Baniaga, A.E., Galuska, S., Jorgensen, S.A., Kidder, T.I., Kong, H., Lu-Irving, P., Marx, H.E., Qi, X., Reardon, C.R., Sutherland, B.L., Tiley, G.P., Welles, S.R., Yu, R., Zhan, S., Gramzow, L., Theißen, G., Wong, G.K.-S., 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574, 679-685. https://doi.org/10.1038/s41586-019-1693-2.

Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards, S.V., Forest, F., Gilbert, M.T.P., Goldstein, M.M., Grigoriev, I.V., Hackett, K.J., Haussler, D., Jarvis, E.D., Johnson, W.E., Patrinos, A., Richards, S., Castilla-Rubio, J.C., van Sluys, M.-A., Soltis, P.S., Xu, X., Yang, H., Zhang, G., 2018. Earth BioGenome Project: sequencing life for the future of life. Proc. Natl. Acad. Sci. 115, 4325-4333. https://doi.org/10.1073/pnas.1720115115.

Lewis, C.E., Doyle, J.J., 2001. Phylogenetic utility of the nuclear gene malate synthase in the palm family (Arecaceae). Mol. Phylogenet. Evol. 19, 409-420. https://doi.org/10.1006/mpev.2001.0932.

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.

Lim, J.Y., Huang, H., Farnsworth, A., Lunt, D.J., Baker, W.J., Morley, R.J., Kissling, W.D., Hoorn, C., 2021. The Cenozoic history of palms: global diversification, biogeography and the decline of megathermal forests. Glob. Ecol. Biogeogr. 31, 425-439. https://doi.org/10.1111/geb.13436.

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.

Lomolino, M.V., Riddle, B.R., Whittaker, R.J., 2016. Biogeography: biological diversity across space and time. Sinauer Associates Inc., Sunderland.

Louca, S., Pennell, M.W., 2020. Extant timetrees are consistent with a myriad of diversification histories. Nature. https://doi.org/10.1038/s41586-020-2176-1.

Louca, S., Pennell, M.W., 2021. Why extinction estimates from extant phylogenies are so often zero. Curr. Biol. https://doi.org/10.1016/j.cub.2021.04.066.

Lydekker, R., 1896. A geographical history of mammals. Cambridge University Press, Cambridge.

MacArthur, R.H., Wilson, E.O., 1963. An equilibrium theory of insular zoogeography. Evolution 17, 373-387. https://doi.org/10.2307/2407089.

Macphail, M.K., Hill, R.S., 2018. What was the vegetation in northwest Australia during the Paleogene, 66–23 million years ago? Aust. J. Bot. 66, 556-574. https://doi.org/10.1071/BT18143.

Macphail, M.K., Jordan, G.J., 2015. Tropical palms and arums at near-polar latitudes: fossil pollen evidence from the Tamar and Macquarie grabens, northern Tasmania. Pap. Proc. R. Soc. Tasmania 149, 23-28. https://doi.org/10.26749/rstpp.149.23.

Madulid, D.A., 1981. A monograph of *Plectocomia* (Palmae: Lepidocaryoideae). Kalikasan 10, 1-94.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., Hernández-Hernández, T., 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytol. 207, 437-453. https://doi.org/10.1111/nph.13264.

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.

Martius, K.F.P.v., Mohl, H.v., Unger, F., 1823-50. Historia naturalis palmarum. T.O. Weigel, Munich.

Matzke, N.J., 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. PhD thesis, University of California.

Matzke, N.J., 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63, 951-970. https://doi.org/10.1093/sysbio/syu056.

Matzke, N.J., 2021. Statistical comparison of DEC and DEC+J is identical to comparison of two ClaSSE submodels, and is therefore valid. OSF Preprints. https://doi.org/10.31219/osf.io/vqm7r.

McClatchey, W.C., 1998. A revision of the genus *Metroxylon* section *Coelococcus* (Arecaceae). PhD thesis, University of Florida.

Merckx, V.S.F.T., Hendriks, K.P., Beentjes, K.K., Mennes, C.B., Becking, L.E., Peijnenburg, K.T.C.A., Afendy, A., Arumugam, N., de Boer, H., Biun, A., Buang, M.M., Chen, P.-P., Chung, A.Y.C., Dow, R., Feijen, F.A.A., Feijen, H., Soest, C.F.-v., Geml, J., Geurts, R., Gravendeel, B., Hovenkamp, P., Imbun, P., Ipor, I., Janssens, S.B., Jocqué, M., Kappes, H., Khoo, E., Koomen, P., Lens, F., Majapun, R.J., Morgado, L.N., Neupane, S., Nieser, N., Pereira, J.T., Rahman, H., Sabran, S., Sawang, A., Schwallier, R.M., Shim, P.-S., Smit, H., Sol, N., Spait, M., Stech, M., Stokvis, F., Sugau, J.B., Suleiman, M., Sumail, S., Thomas, D.C., van Tol, J., Tuh, F.Y.Y., Yahya, B.E., Nais, J., Repin, R., Lakim, M., Schilthuizen, M., 2015. Evolution of endemism on a young tropical mountain. Nature 524, 347. https://doi.org/10.1038/nature14949.

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.

Moore, H.E., 1973a. The major groups of palms and their distribution. Gentes Herbarum 11, 27-141.

Moore, H.E., 1973b. Palms in the tropical forest ecosystems of Africa and South America. In: Meggers, B., Ayensu, E., Duckworth, W. (Eds.), Tropical forest ecosystem of Africa and South America: a comparative review. Smithsonian Institute Press, Washington, DC, pp. 63-88.

Morley, R.J., 1998. Palynological evidence for Tertiary plant dispersals in the SE Asian region in relation to plate tectonics and climate. In: Hall, R., Holloway, J. D. (Eds.), Biogeography and geological evolution of SE Asia. Backhuys, Leiden, pp. 211-234.

Morley, R.J., 2000. Origin and evolution of tropical rain forests. John Wiley & Sons, Chichester, UK.

Morley, R.J., 2011. Cretaceous and Tertiary climate change and the past distribution of megathermal rainforests. In: Bush, M., Flenley, J., Gosling, W. (Eds.), Tropical rainforest responses to climatic change. Springer Berlin, Heidelberg, pp. 1-34. https://doi.org/10.1007/978-3-642-05383-2_1.

Morley, R.J., 2018. Assembly and division of the South and South-East Asian flora in relation to tectonics and climate change. J. Trop. Ecol. 34, 209-234. https://doi.org/10.1017/S0266467418000202.

Müller, R.D., Seton, M., Zahirovic, S., Williams, S.E., Matthews, K.J., Wright, N.M., Shephard, G.E., Maloney, K.T., Barnett-Moore, N., Hosseinpour, M., Bower, D.J., Cannon, J., 2016. Ocean basin evolution and global-scale plate reorganization events since Pangea breakup. Annu. Rev. Earth Pl. Sc. 44, 107-138. https://doi.org/10.1146/annurev-earth-060115-012211.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853-858. https://doi.org/10.1038/35002501.

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.

Nagalingum, N.S., Marshall, C.R., Quental, T.B., Rai, H.S., Little, D.P., Mathews, S., 2011. Recent synchronous radiation of a living fossil. Science 334, 796-799. https://doi.org/10.1126/science.1209926.

Nauheimer, L., Boyce, P.C., Renner, S.S., 2012. Giant taro and its relatives: a phylogeny of the large genus *Alocasia* (Araceae) sheds light on Miocene floristic exchange in the Malesian region. Mol. Phylogenet. Evol. 63, 43-51. https://doi.org/10.1016/j.ympev.2011.12.011.

Nugraha, A.M.S., Hall, R., 2018. Late Cenozoic palaeogeography of Sulawesi, Indonesia. Palaeogeogr. Palaeocl. 490, 191-209. https://doi.org/10.1016/j.palaeo.2017.10.033.

Onstein, R.E., 2019. Darwin's second 'abominable mystery': trait flexibility as the innovation leading to angiosperm diversity. New Phytol. https://doi.org/10.1111/nph.16294.

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. R. Soc. B. 359, 1455-1464. https://doi.org/10.1098/rstb.2004.1539.

Pennington, R.T., Hughes, M., Moonlight, P.W., 2015. The origins of tropical rainforest hyperdiversity. Trends Plant Sci. 20, 693-695. https://doi.org/10.1016/j.tplants.2015.10.005.

Pennington, R.T., Richardson, J.E., Lavin, M., 2006. Insights into the historical construction of species-rich biomes from dated plant phylogenies, neutral ecological theory and phylogenetic community structure. New Phytol. 172, 605-616. https://doi.org/10.1111/j.1469-8137.2006.01902.x.

Pérez-Escobar, O., Bellot, S., Przelomska, N.A.S., Flowers, J.M., Nesbitt, M., Ryan, P., Gutaker, R.M., Gros-Balthazard, M., Wells, T., Kuhnhäuser, B.G., Schley, R., Bogarín, D., Dodsworth, S., Diaz, R., Lehmann, M., Petoe, P., Eiserhardt, W.L., Preick, M., Hofreiter, M., Hajdas, I., Purugganan, M., Antonelli, A., Gravendeel, B., Leitch, I.J., Torres Jiménez, M.F., Papadopulos, A.S.T., Chomicki, G., Renner, S.S., Baker, W.J., 2021. Molecular clocks and archaeogenomics of a Late Period Egyptian date palm leaf reveal introgression from wild relatives and add timestamps on the domestication. Mol. Biol. Evol., msab188. https://doi.org/10.1093/molbev/msab188.

Petchey, O.L., Gaston, K.J., 2002. Functional diversity (FD), species richness and community composition. Ecol. Lett. 5, 402-411. https://doi.org/10.1046/j.1461-0248.2002.00339.x.

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

Purvis, A., 2008. Phylogenetic approaches to the study of extinction. Annu. Rev. Ecol. Evol. S. 39, 301-319. https://doi.org/10.1146/annurev-ecolsys-063008-102010.

Putz, F., 1990. Growth habits and trellis requirements of climbing palms (*Calamus* spp.) in North-Eastern Queensland. Aust. J. Bot. 38, 603-608. https://doi.org/10.1071/BT9900603.

Putz, F.E., Chai, P., 1987. Ecological studies of lianas in Lambir National Park, Sarawak, Malaysia. J. Ecol. 75, 523-531. https://doi.org/10.2307/2260431.

R Core Team, 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Rabosky, D.L., 2009. Ecological limits on clade diversification in higher taxa. Am. Nat. 173, 662-674. https://doi.org/10.1086/597378.

Rabosky, D.L., 2010. Extinction rates should not be estimated from molecular phylogenies. Evolution 64, 1816-1824. https://doi.org/10.1111/j.1558-5646.2009.00926.x.

Rabosky, D.L., 2016. Challenges in the estimation of extinction from molecular phylogenies: a response to Beaulieu and O'Meara. Evolution 70, 218-228. https://doi.org/10.1111/evo.12820.

Rabosky, D.L., Hurlbert, A.H., 2015. Species richness at continental scales is dominated by ecological limits. Am. Nat. 185, 572-583. https://doi.org/10.1086/680850.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901-904. https://doi.org/10.1093/sysbio/syy032.

Ramírez-Barahona, S., Sauquet, H., Magallón, S., 2020. The delayed and geographically heterogeneous diversification of flowering plant families. Nat. Ecol. Evol. 4, 1232-1238. https://doi.org/10.1038/s41559-020-1241-3.

Ree, R.H., Sanmartín, I., 2018. Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. J. Biogeog. 45, 741-749. https://doi.org/10.1111/jbi.13173.

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.

Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3, 217-223. https://doi.org/10.1111/j.2041-210X.2011.00169.x.

Richards, P.W., 1973. Africa, the 'odd man out'. In: Meggers, B., Ayensu, E., Duckworth, W. (Eds.), Tropical forest ecosystems in Africa and South America. Smithsonian Institute Press, Washington, DC, pp. 21-26.

Richardson, J.E., Pennington, R.T., 2016. Origin of tropical diversity: from clades to communities. Front. Genet. 7. https://doi.org/10.3389/fgene.2016.00186.

Richardson, J.E., Pennington, R.T., Pennington, T.D., Hollingsworth, P.M., 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. Science 293, 2242-2245. https://doi.org/10.1126/science.1061421.

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.

Salami, M.B., 1990. Palynomorph taxa from the "Lower Coal Measures" deposits (? Campanian-Maastrichtian) of Anambra trough, Southeastern Nigeria. J. Afr. Earth Sci. 11, 135-150. https://doi.org/10.1016/0899-5362(90)90084-R.

Salard-Cheboldaeff, M., 1990. Intertropical African palynostratigraphy from Cretaceous to late quaternary times. J. Afr. Earth Sci. 11, 1-24. https://doi.org/10.1016/0899-5362(90)90072-M.

Sauquet, H., 2013. A practical guide to molecular dating. CR Palevol. 12, 355-367. https://doi.org/10.1016/j.crpv.2013.07.003.

Sauquet, H., Ramírez-Barahona, S., Magallón, S., 2021. The age of flowering plants is unknown. EcoEvoRxiv. https://doi.org/10.32942/osf.io/n4v6b.

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.

Schuster, R., 1972. Continental movements,"Wallace's Line" and Indomalayan-Australasian dispersal of land plants: some eclectic concepts. Bot. Rev. 38, 3-86.

Sclater, W.L., Sclater, P.L., 1899. The geography of mammals. K. Paul, Trench, Trübner, London.

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.

Silvestro, D., Bacon, C.D., Ding, W., Zhang, Q., Donoghue, P.C.J., Antonelli, A., Xing, Y., 2021. Fossil data support a pre-Cretaceous origin of flowering plants. Nat. Ecol. Evol. 5, 449-457 https://doi.org/10.1038/s41559-020-01387-8.

Simpson, G.G., 1977. Too many lines; the limits of the Oriental and Australian zoogeographic regions. Proc. Am. Philos. Soc. 121, 107-120.

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.

Slik, J.W.F., Aiba, S.-I., Bastian, M., Brearley, F.Q., Cannon, C.H., Eichhorn, K.A.O., Fredriksson, G., Kartawinata, K., Laumonier, Y., Mansor, A., Marjokorpi, A., Meijaard, E., Morley, R.J., Nagamasu, H., Nilus, R., Nurtjahya, E., Payne, J., Permana, A., Poulsen, A.D., Raes, N., Riswan, S., van Schaik, C.P., Sheil, D., Sidiyasa, K., Suzuki, E., van Valkenburg, J.L.C.H., Webb, C.O., Wich, S., Yoneda, T., Zakaria, R., Zweifel, N., 2011. Soils on exposed Sunda Shelf shaped biogeographic patterns in the equatorial forests of Southeast Asia. Proc. Natl. Acad. Sci. 108, 12343-12347. https://doi.org/10.1073/pnas.1103353108.

Slik, J.W.F., Arroyo-Rodríguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L.F., Ashton, P., Balvanera, P., Bastian, M.L., Bellingham, P.J., van den Berg, E., Bernacci, L., da Conceição Bispo, P., Blanc, L., Böhning-Gaese, K., Boeckx, P., Bongers, F., Boyle, B., Bradford, M., Brearley, F.Q., Breuer-Ndoundou Hockemba, M., Bunyavejchewin, S., Calderado Leal Matos, D., Castillo-Santiago, M., Catharino, E.L.M., Chai, S.-L., Chen, Y., Colwell, R.K., Chazdon, R.L., Clark, C., Clark, D.B., Clark, D.A., Culmsee, H., Damas, K., Dattaraja, H.S., Dauby, G., Davidar, P., DeWalt, S.J., Doucet, J.-L., Duque, A., Durigan, G., Eichhorn, K.A.O., Eisenlohr, P.V., Eler, E., Ewango, C., Farwig, N., Feeley, K.J., Ferreira, L., Field, R., de Oliveira Filho, A.T., Fletcher, C., Forshed, O., Franco, G., Fredriksson, G., Gillespie, T., Gillet, J.-F., Amarnath, G., Griffith, D.M., Grogan, J., Gunatilleke, N., Harris, D., Harrison, R., Hector, A., Homeier, J., Imai, N., Itoh, A., Jansen, P.A., Joly, C.A., de Jong, B.H.J., Kartawinata, K., Kearsley, E., Kelly, D.L., Kenfack, D., Kessler, M., Kitayama, K., Kooyman, R., Larney, E., Laumonier, Y., Laurance, S., Laurance, W.F., Lawes, M.J., Amaral, I.L.d., Letcher, S.G., Lindsell, J., Lu, X., Mansor, A., Marjokorpi, A., Martin, E.H., Meilby, H., Melo, F.P.L., Metcalfe, D.J., Medjibe, V.P., Metzger, J.P., Millet, J., Mohandass, D., Montero, J.C., de Morisson Valeriano, M., Mugerwa, B., Nagamasu, H., Nilus, R., Ochoa-Gaona, S., Onrizal, Page, N., Parolin, P., Parren, M., Parthasarathy, N., Paudel, E., Permana, A., Piedade, M.T.F., Pitman, N.C.A., Poorter, L., Poulsen, A.D., Poulsen, J., Powers, J., Prasad, R.C., Puyravaud, J.-P., Razafimahaimodison, J.-C., Reitsma, J., dos Santos, J.R., Roberto Spironello, W., Romero-Saltos, H., Rovero, F., Rozak, A.H., Ruokolainen, K., Rutishauser, E.,

Saiter, F., Saner, P., Santos, B.A., Santos, F., Sarker, S.K., Satdichanh, M., Schmitt, C.B., Schöngart, J., Schulze, M., Suganuma, M.S., Sheil, D., da Silva Pinheiro, E., Sist, P., Stevart, T., Sukumar, R., Sun, I.-F., Sunderland, T., Suresh, H.S., Suzuki, E., Tabarelli, M., Tang, J., Targhetta, N., Theilade, I., Thomas, D.W., Tchouto, P., Hurtado, J., Valencia, R., van Valkenburg, J.L.C.H., Van Do, T., Vasquez, R., Verbeeck, H., Adekunle, V., Vieira, S.A., Webb, C.O., Whitfeld, T., Wich, S.A., Williams, J., Wittmann, F., Wöll, H., Yang, X., Adou Yao, C.Y., Yap, S.L., Yoneda, T., Zahawi, R.A., Zakaria, R., Zang, R., de Assis, R.L., Garcia Luize, B., Venticinque, E.M., 2015. An estimate of the number of tropical tree species. Proc. Natl. Acad. Sci. 112, 7472-7477. https://doi.org/10.1073/pnas.1423147112.

Smith, A.G., Smith, D.G., Funnell, B.M., 1994. Atlas of Mesozoic and Cenozoic coastlines. Cambridge University Press, Cambridge.

Smith, M.L., Hahn, M.W., 2021. New approaches for inferring phylogenies in the presence of paralogs. Trends Genet. 37, 174-187. https://doi.org/10.1016/j.tig.2020.08.012.

Smith, S.A., Brown, J.W., Walker, J.F., 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. Plos One 13, e0197433. https://doi.org/10.1371/journal.pone.0197433.

Sniderman, J.M.K., Jordan, G.J., 2011. Extent and timing of floristic exchange between Australian and Asian rain forests. J. Biogeog. 38, 1445-1455. https://doi.org/10.1111/j.1365-2699.2011.02519.x.

Solís-Lemus, C., Bastide, P., Ané, C., 2017. PhyloNetworks: a package for phylogenetic networks. Mol. Biol. Evol. 34, 3292-3298. https://doi.org/10.1093/molbev/msx235.

Stadler, T., Rabosky, D.L., Ricklefs, R.E., Bokma, F., 2014. On age and species richness of higher taxa. Am. Nat. 184, 447-455. https://doi.org/10.1086/677676.

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.

Stebbins, G.L., 1974. Flowering plants: evolution above the species level. Edward Arnold, London.

Steenwyk, J.L., Buida, T.J., III, Labella, A.L., Li, Y., Shen, X.-X., Rokas, A., 2021. PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. Bioinformatics. https://doi.org/10.1093/bioinformatics/btab096.

Sterling, J.A., Seberg, O., Humphries, C.J., Borchsenius, F., Dransfield, J., 2007. Priority areas for rattan conservation on Borneo. In: Curry, G. B., Humphries, C. J. (Eds.), Biodiversity databases - techniques, politics, and applications, pp. 129-180.

Strijk, J.S., Binh, H.T., Ngoc, N.V., Pereira, J.T., Slik, J.W.F., Sukri, R.S., Suyama, Y., Tagane, S., Wieringa, J.J., Yahara, T., Hinsinger, D.D., 2020. Museomics for reconstructing historical floristic exchanges: divergence of stone oaks across Wallacea. Plos One 15, e0232936. https://doi.org/10.1371/journal.pone.0232936.

Sunderland, T.C.H., 2012. A taxonomic revision of the rattans of Africa (Arecaceae: Calamoideae). Phytotaxa 51, 76. https://doi.org/10.11646/phytotaxa.51.1.1.

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 6 August 2020).

Thomas, D.C., Hughes, M., Phutthai, T., Ardi, W.H., Rajbhandary, S., Rubite, R., Twyford, A.D., Richardson, J.E., 2012. West to east dispersal and subsequent rapid diversification of the mega-diverse genus *Begonia* (Begoniaceae) in the Malesian archipelago. J. Biogeog. 39, 98-113. https://doi.org/10.1111/j.1365-2699.2011.02596.x.

Tomlinson, P.B., 2006. The uniqueness of palms. Bot. J. Linn. Soc. 151, 5-14. https://doi.org/10.1111/j.1095-8339.2006.00520.x.

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

Totterdell, J.M., Mitchell, C., 2009. Bight Basin geological sampling and seepage survey: R/V Southern Surveyor Survey SS01/2007. Record 2009/024. Geoscience Australia, Canberra.

Toussaint, E.F.A., Hall, R., Monaghan, M.T., Sagata, K., Ibalim, S., Shaverdo, H.V., Vogler, A.P., Pons, J., Balke, M., 2014. The towering orogeny of New Guinea as a trigger for arthropod megadiversity. Nat. Commun. 5, 4001. https://doi.org/10.1038/ncomms5001.

Trénel, P., Gustafsson, M.H.G., Baker, W.J., Asmussen-Lange, C.B., Dransfield, J., Borchsenius, F., 2007. Mid-Tertiary dispersal, not Gondwanan vicariance explains distribution patterns in the wax palm subfamily (Ceroxyloideae: Arecaceae). Mol. Phylogenet. Evol. 45, 272-288. https://doi.org/10.1016/j.ympev.2007.03.018.

Truswell, E., Owen, J., 1988. Eocene pollen from Bungonia, New South Wales. Mem. Assoc. Australas. Palaeontol. 5, 259–284.

Turner, H., Hovenkamp, P., van Welzen, P.C., 2001. Biogeography of Southeast Asia and the West Pacific. J. Biogeog. 28, 217-230. https://doi.org/10.1046/j.1365-2699.2001.00526.x.

Uhl, N.W., Dransfield, J., 1987. Genera Palmarum: a classification of palms based on the work of Harold E. Moore, Jr. Bailey Hortorium and the International Palm Society, Lawrence, Kansas.

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

Upchurch, G.R., Wolfe, J.A., 1987. Mid-Cretaceous to Early Tertiary vegetation and climate: evidence from fossil leaves and woods. In: Friis, E. M., Chaloner, W. G., Crane, P. H. (Eds.), The origins of angiosperms and their biological consequences. Cambridge University Press, Cambridge, pp. 75-105.

van Welzen, P.C., Parnell, J.A.N., Slik, J.W.F., 2011. Wallace's Line and plant distributions: two or three phytogeographical areas and where to group Java? Biol. J. Linn. Soc. 103, 531-545. https://doi.org/10.1111/j.1095-8312.2011.01647.x.

Voris, H.K., 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J. Biogeog. 27, 1153-1167. https://doi.org/10.1046/j.1365-2699.2000.00489.x.

Vorontsova, M.S., Clark, L.G., Dransfield, J., Govaerts, R., Baker, W.J., 2016a. World checklist of bamboos and rattans. International Network of Bamboo and Rattan & Royal Botanic Gardens, Kew, Bejing.

Vorontsova, M.S., Clark, L.G., Dransfield, J., Govaerts, R., Wilkinson, T., Baker, W.J., 2016b. World atlas of bamboos and rattans. International Network of Bamboo and Rattan & Royal Botanic Gardens, Kew, Bejing.

Walker, J.D., Geissman, J., Bowring, S., Babcock, L., 2018. Geologic Time Scale v. 5.0. Geological Society of America.

Wallace, A.R., 1860. On the zoological geography of the Malay archipelago. Zool. J. Linn. Soc. 4, 172-184. https://doi.org/10.1111/j.1096-3642.1860.tb00090.x.

Wallace, A.R., 1878. Tropical nature, and other essays. Macmillan and Company, London.

Wang, W., Ortiz, R.D., Jacques, F.M.B., Xiang, X.G., Li, H.L., Lin, L., Li, R.Q., Liu, Y., Soltis, P.S., Soltis, D.E., Chen, Z.D., 2012. Menispermaceae and the diversification of tropical rainforests near the Cretaceous-Paleogene boundary. New Phytol. 195, 470-478. https://doi.org/10.1111/j.1469-8137.2012.04158.x.

Warren, B.H., Simberloff, D., Ricklefs, R.E., Aguilée, R., Condamine, F.L., Gravel, D., Morlon, H., Mouquet, N., Rosindell, J., Casquet, J., Conti, E., Cornuault, J., Fernández-Palacios, J.M., Hengl, T., Norder, S.J., Rijsdijk, K.F., Sanmartín, I., Strasberg, D., Triantis, K.A., Valente, L.M., Whittaker, R.J., Gillespie, R.G., Emerson, B.C., Thébaud, C., 2015. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. Ecol. Lett. 18, 200-217. https://doi.org/10.1111/ele.12398.

Webb, C.O., Ackerly, D.D., McPeek, M.A., Donoghue, M.J., 2002. Phylogenies and community ecology. Annu. Rev. Ecol. Syst. 33, 475-505. https://doi.org/10.1146/annurev.ecolysis.33.010802.150448.

Weigelt, P., Steinbauer, M.J., Cabral, J.S., Kreft, H., 2016. Late Quaternary climate change shapes island biodiversity. Nature 532, 99. https://doi.org/10.1038/nature17443.

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, 1400042. https://doi.org/10.3732/apps.1400042.

Whitmore, T.C., 1982. Wallace's line: a result of plate tectonics. Ann. Mo. Bot. Gard. 69, 668-675. https://doi.org/10.2307/2399087.

Whitmore, T.C., 1998. An introduction to tropical rain forests. Clarendon Press, Oxford.

Wiens, J.J., 2011. The causes of species richness patterns across space, time, and clades and the role of “ecological limits”. Q. Rev. Biol. 86, 75-96. https://doi.org/10.1086/659883.

Wiens, J.J., Donoghue, M.J., 2004. Historical biogeography, ecology and species richness. Trends Ecol. Evol. 19, 639-644. https://doi.org/10.1016/j.tree.2004.09.011.

Wilson, E.O., MacArthur, R.H., 1967. The theory of island biogeography. Princeton University Press, Princeton.

Wing, S.L., Boucher, L.D., 1998. Ecological aspects of the Cretaceous flowering plant radiation. Annu. Rev. Earth Pl. Sc. 26, 379-421. https://doi.org/10.1146/annurev.earth.26.1.379.

Wing, S.L., Hickey, L.J., Swisher, C.C., 1993. Implications of an exceptional fossil flora for Late Cretaceous vegetation. Nature 363, 342. https://doi.org/10.1038/363342a0.

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.

Zhao, H., Wang, S., Wang, J., Chen, C., Hao, S., Chen, L., Fei, B., Han, K., Li, R., Shi, C., Sun, H., Wang, S., Xu, H., Yang, K., Xu, X., Shan, X., Shi, J., Feng, A., Fan, G., Liu, X., Zhao, S., Zhang, C., Gao, Q., Gao, Z., Jiang, Z., 2018. The chromosome-level genome assemblies of two rattans (*Calamus simplicifolius* and *Daemonorops jenkinsiana*). GigaScience 7, 1-11. https://doi.org/10.1093/gigascience/giy097.

Zona, S., Henderson, A., 1989. A review of animal-mediated seed dispersal of palms. Selbyana 11, 6-21.

# Appendix

## Appendix 1. Lanjak Entimau Wildlife Sanctuary – a palm hotspot in the heart of Borneo

Published journal article on following pages.

**Bibliographic information:**

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

# Lanjak Entimau Wildlife Sanctuary – a Palm Hotspot in the Heart of Borneo

Peter Petoe*[1,2]
Benedikt G. Kuhnhäuser*[1,3]
Connie Geri[4]
and
William J. Baker[1]

[1]*Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, UK*

[2]*Department of Biology, Aarhus University, 8000 Aarhus C, Denmark*

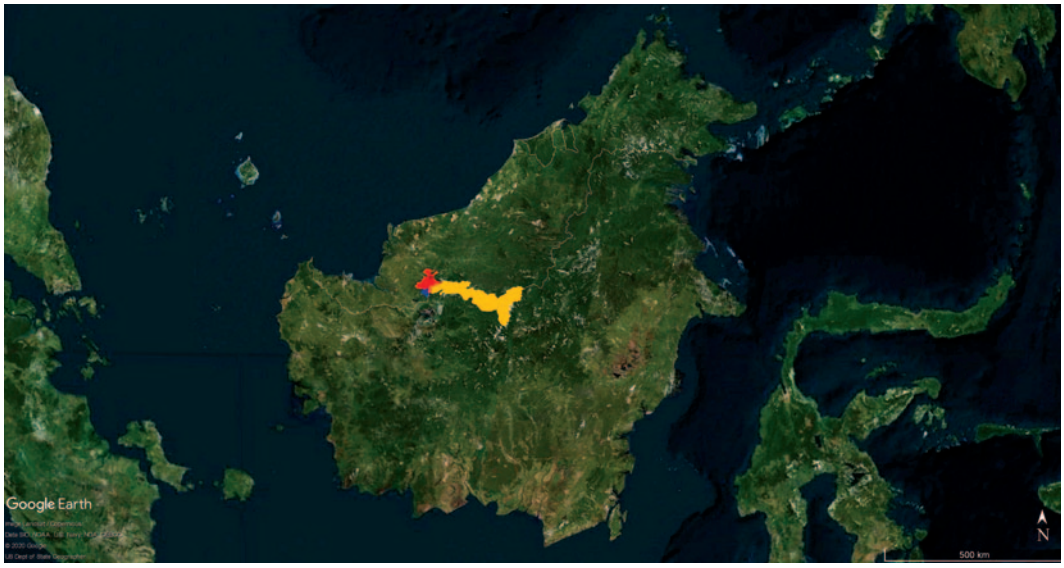
[3]*Department of Plant Sciences, University of Oxford, Oxford, OX1 3RB, UK*

[4]*Sarawak Forestry Corporation, Kuching, Sarawak, Malaysia*

**peter.petoe@bios.au.dk, benedikt.kuhnhaeuser@bnc.ox.ac.uk, equal contributors.*

**With over 300 species, Borneo has the richest palm flora of any Malesian island. Here, we describe an expedition to Lanjak Entimau Wildlife Sanctuary, a remote region of Sarawak in the heart of Borneo and provide a checklist of the 46 palm species that we found there.**

Our expedition started in Kuching, the state capital of Sarawak in western Malaysian Borneo. We used the city as a base from which to access nearby localities including Gunung Matang in Kubah National Park, which is known for its astounding palm diversity and for Beccari's efforts to describe it during the 19th century (Beccari 1904), and Bako National Park, home to several populations of the iconic palm *Johannesteijsmannia altifrons*. Although exploring and collecting within these popular sites was rewarding, we were particularly excited about what was yet to come: an expedition to the remote and incompletely known Lanjak Entimau Wildlife Sanctuary (LEWS). LEWS is the largest of Sarawak's 66 Totally Protected Areas and is only accessible to scientists and local people living within its immediate vicinity (Forest Department Sarawak 2020). It is probably most widely known for its significant role in the conservation of the critically endangered northwest Bornean orangutan (Ancrenaz et al. 2016). However, the sanctuary protects a large diversity of wildlife, including an estimated 2000 species of plants (Chai 2000). The sanctuary abuts the vast

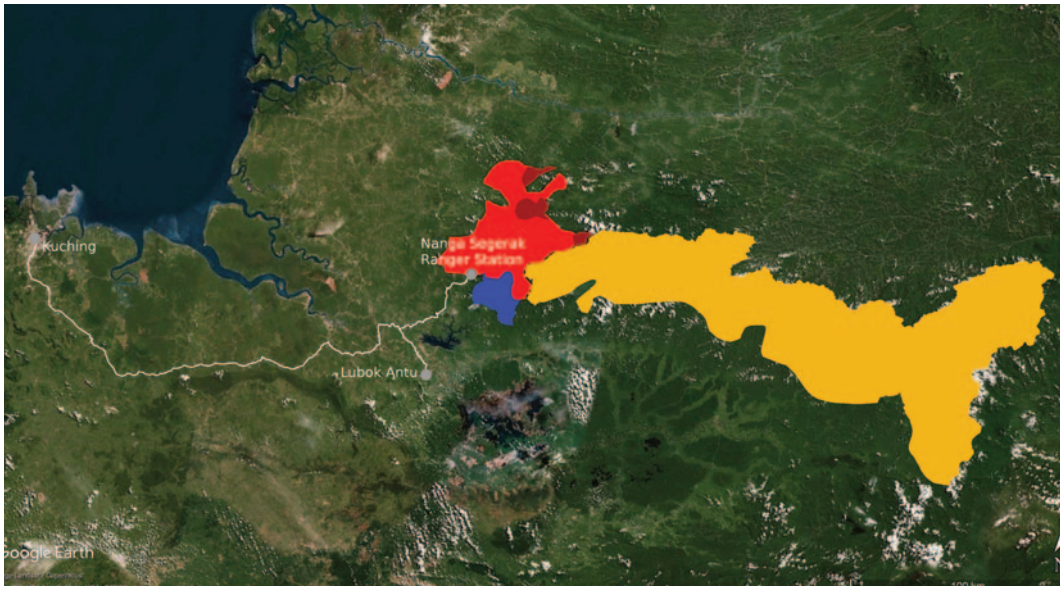



1. The island of Borneo showing Lanjak Entimau Wildlife Sanctuary (red) with extensions (dark red), Betung Kerihun National Park (yellow) and Batang Ai National Park (blue).

Betung Kerihun National Park across the border in Kalimantan, which is more than five times the size of LEWS. Together, the two reserves form the Transboundary Biodiversity Conservation Area (Fig. 1; Brouwer 2012), a core part of the Heart of Borneo initiative, which aims to conserve the remaining intact forests in the center of the island (Rautner et al. 2005). However, the biodiversity of LEWS remains insufficiently known. A checklist of the plants of LEWS has been published, which records over 100 palm species (Chai 2000), but many palm records are yet to be formally confirmed and further ground-truthing is required.

At 6:30 in the morning of 1 November 2018, we departed Kuching in a fully loaded pickup-truck with the objective of undertaking a rigorous inventory of the palm diversity around Nanga Segerak Ranger Station, which is located in Ulu Engkari, in the southwestern portion of

2. Our route from Kuching to Lanjak Entimau Wildlife Sanctuary. Protected areas color coded as in Figure 1.



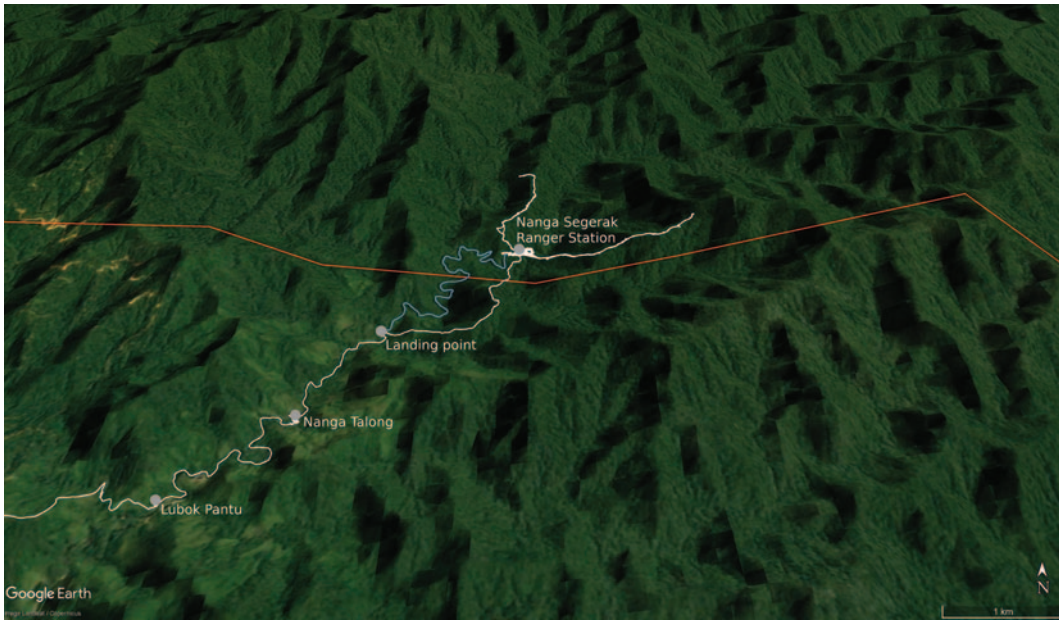


3. Palm cabbage of *Plectocomiopsis geminiflora* (left) and *Eugeissona utilis* (right) sold at local stall near Lubok Antu. Photo: B.G. Kuhnhäuser.

LEWS (Fig. 2). Sirukit anak Dubod from the Sarawak Forestry Corporation accompanied us and had already proven to be a versatile and indispensable companion during our fieldwork in Sarawak. We started out by driving for several hours along the main road to Lubok Antu. Upon reaching this small town, which is near the border with Indonesia, we transferred to a 4-WD truck. We had a brief stop on the roadside where palm cabbage, or *umbut* in Malay, of *Plectocomiopsis geminiflora* and *Eugeissona utilis* was being sold by local people (Fig. 3) and then branched off the main road onto a dirt track that had been constructed for logging purposes and would take us north of Batang Ai National Park toward LEWS. Some common palms, such as *Plectocomiopsis geminiflora* (Fig. 4) and *Eugeissona utilis* flourished in the disturbed vegetation along this stretch of the journey, which lasted for roughly 50 km. Eventually, we reached Lubok Pantu on the Sungai Engkari where we were greeted by our boatmen. The journey upstream would take us first to Nanga Talong and, the following day, on to Nanga Segerak Ranger Station (Fig. 5). The longboats were lined up and ready for departure on the riverbank opposite an idyllic Iban village adorned by *Arenga pinnata*, *Cocos nucifera* and *Areca catechu* (Fig. 6). After a short and pleasant journey up the river we arrived at Nanga Talong, the last longhouse on Sungai Engkari before LEWS. We were heartily welcomed and hosted by the Headman Bada anak

4. *Plectocomiopsis geminiflora* Photo: B.G. Kuhnhäuser.






5. Routes travelled in and around Lanjak Entimau Wildlife Sanctuary. The boundaries of the sanctuary are marked with a red line.

Chendai, who gladly accepted several of the *umbut* we had bought earlier in the day, and enjoyed them incorporated into our dinner.

Early on the following morning we took off again in two longboats, which sat low in the water due to the combined weight of us and our provisions. Upstream of Nanga Talong, the river soon became shallow, with emergent rocks and boulders, forcing us to disembark occasionally to push the boat upstream. Lush vegetation clothed both sides of the river with common riverside palms including *Arenga undulatifolia*, *Salacca vermicularis* (Fig. 7) and several species of *Calamus*. Our three boatmen, Mugu anak Sanggap, Dellie anak Medie and Medie

6. River view of Lubok Pantu showing *Arenga pinnata*, *Cocos nucifera* and *Areca catechu* in the background Photo: W.J. Baker.





7. *Salacca vermicularis* Photo: B.G. Kuhnhäuser.

anak Aloh, would be our local guides for the remainder of the expedition. These fearless masters of the river swapped deftly between outboard engine, paddle and punting stick to navigate the shallows, rapids and narrow bends of Sungai Engkari. They also possessed extensive knowledge of the area and the natural resources of the forest. Before the river became impassable for the heavily laden boats, we landed at a riverbank ("Landing point" in Fig. 5) to allow Mugu, Dellie and Medie to continue upstream with the now considerably lighter boats to deliver our provisions to the research station. After a short trek of about two kilometers through the forest we arrived at the station, which is located right inside the boundary of LEWS. Being equipped with electricity, a field laboratory and even internet, the ranger station provides basic but comfortable accommodation and facilities for field workers. In addition, a nearby suspension bridge gives immediate access to excellent collecting grounds for a multitude of palm species. For each of the five following days, we would explore the trails radiating from the research station from early in the morning until the rains set in, normally during the mid-afternoon, at which point we sought shelter back at base. We were able to make herbarium specimens of almost every palm species that we found. These were deposited at the Sarawak Forest Department herbarium in Kuching, with duplicates for the herbaria at the Royal Botanic Gardens, Kew and Aarhus University. As a result, we were able to confirm our field identifications by comparisons with expertly named specimens after the expedition. The palms we recorded are described in this article and enumerated in Appendix 1.

## Palm habitats in LEWS

The sanctuary is geologically complex with several different forest types including riparian, old secondary, lowland mixed dipterocarp and hill mixed dipterocarp forests (ITTO 2001). All of these forest types occur to varying extents within our inventoried area, within an elevational range from 320 m (Nanga Segerak Ranger Station) to 780 m. Although the mixed dipterocarp forest is known to be the most extensive and biologically diverse habitat within the sanctuary, we found the highest species richness and abundance of palms in wet valley bottoms of the riparian forest (Fig. 8). Palms remained conspicuous in the hill slopes and ridge tops of the adjacent mixed dipterocarp forest, although at lower densities and local species richness.

## Calamoideae

In the subfamily Calamoideae, we recorded 30 species belonging to the genera *Calamus* (18 spp.), *Eugeissona* (1 sp.), *Korthalsia* (5 spp.), *Plectocomia* (1 sp.), *Plectocomiopsis* (3 spp.), and *Salacca* (1 sp.). They displayed an extraordinary array of morphological and ecological diversity.

Asian calamoid palms are often associated with the climbing growth form. However, about 10% of Calamoideae are not climbing, but self-supporting or stemless



8. Riparian forest understory with *Pinanga* aff. *brevipes*, *Pinanga tomentella*, *Calamus sabalensis* and *Licuala petiolulata* (from front to back). Photo: W.J. Baker.



9. *Calamus ursinus* Photo: B.G. Kuhnhäuser.

(Couvreur et al. 2015). At LEWS, we encountered five of these species. The remarkable *Calamus ursinus* grew in large numbers on a steep riverbank near the research station (Fig. 9). It was formerly placed in the genus *Pogonotium* (now a synonym of *Calamus* [Baker 2015]), which is immediately distinguished by the long, erect auricles flanking the petiole. Having met with *Calamus ursinus* before on the top of Gunung Matang, we were surprised to find it here again in such a different environment. Indeed, it was previously thought to be constrained to submontane ridgetops (Dransfield 1992).

*Calamus collarifer*, a self-supporting, litter-trapping palm up to 3 m tall, was one of the most abundant components of the understory (Fig. 10). *Calamus sabalensis*, a slender, stemless species with the appearance of a young palm seedling (Fig. 11), would almost have gone unnoticed, had we not discovered an individual bearing an old inflorescence. Besides these species, we encountered the aforementioned *Salacca vermicularis* and the massive tree palm *Eugeissona utilis* (locally known as *pantu*), several specimens of which having been felled for their delectable palm hearts (Back Cover).

The aesthetic value of rattans is generally underappreciated! Near the river, we were attracted to the beautiful, bright orange flowers of *Calamus sparsiflorus*, which emitted a sweet fragrance of jasmine. In the specimen we collected, the six stamens were consistently arranged in groups of one, two, and three (Fig. 12). Nearby grew *Calamus pilosellus*, a neat rattan whose regularly arranged, closely adjacent leaflets were interlocked by bristly hairs on the leaflet margins (Fig. 13) and *Plectocomia mulleri*, a robust high-climbing palm with golden spines arranged like eyelashes (see p. 55).

Contrasting with the appeal of these species, several rattans at LEWS lived in close association with belligerent ants. Several species of *Korthalsia* provided living space in their inflated ocreas, i.e. papery elongations of the leaf sheath



10. *Calamus collarifer*. Photo: W.J. Baker.



11. *Calamus sabalensis*. Photo: W.J. Baker.



12. Inflorescence branch of *Calamus sparsiflorus*, with stamens of each flower arranged in groups of one, two and three. Photo: B.G. Kuhnhäuser.



13. *Calamus pilosellus*. Photo: B.G. Kuhnhäuser.

beyond the insertion of the petiole. *Korthalsia rostrata*, *K. cheb* (Fig. 14) and *K. echinometra* (Fig. 15) could be easily distinguished by the shape, size and armature of their ocreas. The long, divergent ocreas of *Korthalsia hispida* (Fig. 16) were inhabited by "noisy ants" producing a rattling sound upon disturbance before rushing out to attack. In *Calamus sabut*, interlocking whorls of spines formed so-called ant galleries (Fig. 17). These were connected by carton tunnels made of bitten off spines. Not all *Korthalsia* species accommodate ants; *Korthalsia debilis* bore a tightly sheathing, disintegrating ocrea (Fig. 18).

Other rattans displayed morphological extremes and curiosities. Whilst *Calamus corrugatus* was the most slender of all palms we encountered with a stem diameter of only 5 mm (Fig. 19), *Calamus comptus* had the longest inflorescence at over 7 m. In stark contrast to these rather slender rattans was the massive *Calamus ornatus*, which we encountered only as a seedling. *Calamus myriacanthus*, a usually

14. Inflated ocrea of *Korthalsia cheb* with entrance hole at the right. Photo: W.J. Baker.





15 (left). Inflated ocrea of *Korthalsia echinometra*. Photo: B.G. Kuhnhäuser. 16 (right). Divergent ocrea of *Korthalsia hispida*. Photo: W.J. Baker.

short-stemmed palm widespread throughout Sarawak, here climbed to 8 m despite possessing no specialized climbing organs. The diamond-shaped, discolorous leaflets of *Calamus hallierianus* were strongly reminiscent of *Korthalsia* (Figs. 20 & 21). Apart from this superficial resemblance, *Calamus hallierianus* was easily distinguished from *Korthalsia* by its extraordinary inflorescence, which was enclosed entirely within the pod-like prophyll, a character diagnostic for the group of species formerly included in the genus *Ceratolobus* (now a synonym of *Calamus* [Baker 2015]). The inflorescence is accessible to potential pollinators only via narrow openings at the prophyll tip. *Calamus pseudoulur* caught our interest as it possessed both a cirrus and a vestigial flagellum, whereas most rattans have either one or the other of these two highly specialized climbing organs . The fruits of *Calamus didymophyllus* were covered in red glossy resin, so-called "dragon's blood," and *Plectocomiopsis mira*, immediately distinguished by its truncate, orange ocrea, was the only rattan considered poisonous by our local guides.

Further calamoid palms we encountered at LEWS were *Calamus* cf. *periacanthus*, *C. javensis*, *C. korthalsii* with long, slender spines at the mouth of the sheath, *C. laevigatus var. mucronatus*, whose basal leaflets were reflexed across the stem, *C. maculatus*, *C. zonatus*, *Plectocomiopsis geminiflora*, and a potentially undescribed species of *Plectocomiopsis* with discolorous leaves and a triangular ocrea.

### Coryphoideae

Five species of coryphoid palm belonging to the genera *Arenga* (1 sp.), *Caryota* (1 sp.) and *Licuala* (3 spp.) were recorded. The relatively low diversity of *Licuala* was somewhat surprising given that the genus is particularly diverse in Borneo with 48 known species (Govaerts et al. 2020), but it is possible that the narrow endemism of many of the species causes high species-



17. Calamus sabut with interlocking whorls of spines forming ant galleries. Photo: B.G. Kuhnhäuser.



18. *Korthalsia debilis*. Photo: W.J. Baker.



19. *Calamus corrugatus*. Photo: B.G. Kuhnhäuser.



20 (left). *Calamus hallierianus*. Photo: W.J. Baker. 21 (right). Enclosed inflorescence of *Calamus hallierianus*. Photo: W.J. Baker.

turnover within the genus. Nevertheless, *Licuala petiolulata* made up for the dearth of species in its sheer numbers of individuals. It was one of the commonest palms that we found growing around the ranger station, where it occurs in the soggy riparian forest as well as on drier slopes and ridges. Fertile individuals of *L. petiolulata* were very rare at the time of collecting but we could easily identify the palms we saw on account of their conspicuously petiolulate terminal leaf segments (Fig. 22). *Licuala rheophytica* was also rather common but occurred only in the riparian forest and a bit further up the mountain slopes. In contrast to *L. petiolulata*, this is a dainty palm and all the individuals we saw had short 4-segmented leaves and spicate inflorescences (Fig. 23). Very occasionally we would come across mottled licuala leaves, which were later attributed to *Licuala* cf. *maculata*. This palm could not be identified unequivocally due to the lack of fertile material, but is likely to be *L. maculata*, which is known from nearby Batang Ai National Park.

The largest coryphoid palm that we collected was *Caryota mitis*, which grew in the riparian forest to about 12 meters in height. The only other caryotoid palm to be recorded was *Arenga undulatifolia*, which occurred frequently along the banks of Sungai Engkari. Both of these species have broad distributions that extend well beyond Borneo.

**Arecoideae**

We recorded 11 species of arecoid palm belonging to the genera *Areca* (1 sp.), *Iguanura* (1 sp.) *Oncosperma* (1 sp.) and *Pinanga* (8 spp.). The arecoid palm seen most frequently was *Pinanga* aff. *brevipes*, which was particularly abundant in the riparian forest but also occurred on higher slopes and ridges. The individuals we observed were all small, acaulescent understory palms with relatively short leaves that had three or four pairs of



22. *Licuala petiolulata* with conspicuously petiolulate terminal leaf segments. Photo: P. Petoe.



23. *Licuala rheophytica*. Photo: P. Petoe.

relatively broad pinnae inserted onto the rachis, and bifid inflorescences. Despite demonstrating clear affinities with the type of *Pinanga brevipes*, known from Kuching, the form that occurs around the ranger station seemed altogether smaller and with broader leaflets. Most strikingly, the specimens that we encountered produce their inflorescences buried in the leaf litter or even deeper in the soil, where they successfully flower and set fruit (Figs. 24 & 25).

*Areca chaiana*, *Oncosperma horridum*, *Pinanga tomentella*, *P. albescens* and *P.* sp. 1 all grew within a relatively short distance from the river. Of these palms, *Areca chaiana* was perhaps the most exciting discovery as it is the only species of *Areca* with a spicate inflorescence (Figs. 26 & 27). Having so far been regarded an anomalous member of the genus (Dransfield 1984), the placement of this species in the genus *Areca* may yet be challenged (Bellot pers. comm.). We found only a few, scattered individuals of *Pinanga tomentella* and *P. albescens*, which are both small single-stemmed species like *Pinanga* aff. *brevipes* but differ in having entire leaves. In addition, all the individuals of *P. tomentella* that we recorded appeared to be of an unusual form with conspicuously bullate leaves (Fig. 28). Though common elsewhere in Borneo, *Oncosperma horridum* was rare in this area, and we observed only a single, fertile individual of *Pinanga* sp. 1. Despite the existence of a good collection, the identity of *P.* sp. 1 with its sigmoid leaflets and bright red, spicate infructescence (Fig. 29) remains unresolved.

On higher slopes and ridges we collected *Pinanga* cf. *ligulata*, *P. lepidota* and *Iguanura chaiana* and recorded, but did not collect, *Pinanga salicifolia* and *P.* sp. 2. These species were all occasional except for *Iguanura chaiana* and *Pinanga* sp. 2 each of which were only represented by one individual. The latter species could not be identified due to a lack of fertile material, although the variegated,





24 (left). *Pinanga* aff. *brevipes*. Photo: W.J. Baker. 25 (right). Entirely subterranean infructescence of *Pinanga* aff. *brevipes*. Photo: P. Petoe.

pinnate leaves displayed by this species stood out among the surrounding vegetation (Fig. 30). Attaining 8 m in height, *Pinanga* cf. *ligulata* was the most imposing of the pinangs we saw in LEWS and was referred to by our guides as *pinang mureng*. *Iguanura chaiana* was an exciting find because few records exist of this rare species (Fig. 31). In contrast, *Pinanga lepidota* and *P. salicifolia* are widely distributed and rather common throughout much of Borneo.

26. *Areca chaiana*, note the spicate inflorescence. Photo: P. Petoe.



## Conclusion

After six intense days, we had to leave LEWS behind us. Having focused on taking an as complete as possible inventory of the palms of Nanga Segerak Ranger Station, we had met with an astonishing diversity of palms in its immediate surroundings. In a radius of less than 2 km around the research station, focusing mainly on easily accessible habitats such as riparian forest and ridge tops, we had recorded 46 palm species and collected herbarium vouchers for 42 of them, with accompanying leaf samples for ongoing research on palm evolution. We had also gained insights into the many uses of the palms of the



27. *Areca chaiana*. Photo: W.J. Baker.



28. *Pinanga tomentella*, form with bullate foliage. Photo: P. Petoe.



29. *Pinanga* sp. 1. Photo: W.J. Baker.



30. *Pinanga* sp. 2. Photo: P. Petoe.

wildlife sanctuary. We had eaten the freshly cooked *umbut* of two *Calamus* species for lunch while sitting on the leaves of *Licuala petiolulata*, had enjoyed a delicious vegetarian curried *pantu*, and had watched our guides collect canes of the slender and flexible *Calamus corrugatus* and *Calamus laevigatus var. mucronatus*.

Now, we had to surrender our precious collections to the mercy of the winding rapids of the Sungai Engkari (Fig. 32). The ride down the narrow river was both thrilling and terrifying, adrenalin peaking when one of our boatmen fell overboard. Drenched to the skin, we were glad to reach the calmer waters of Nanga Talong



31. *Iguanura chaiana*. Photo: P. Petoe.

safely with our collections intact. As our onward journey was planned for the next day, we took the opportunity to collect a specimen of *Arenga pinnata* growing next to the longhouse. Being well aware of the old tradition of headhunting practiced by the Iban people in the past, we couldn't resist the temptation to ascertain whether skulls from former times still existed at the longhouse. When we were presented with an unearthly assortment of nine skulls (Fig. 33), we were amazed to find that some of these were kept in harnesses made of woven rattan canes! On the next morning, we continued our journey back to Kuching, passing through a melancholy landscape of secondary forest necklaced with low-hanging clouds (Fig. 34). At Kuching, our specimens were pressed and dried at the Sarawak Forestry Corporation's Semenggoh Arboretum, and sorted into duplicates.

The richness of palms we documented around Nanga Segerak Ranger Station bears witness to the importance of LEWS as a refuge for Borneo's palm diversity. As indicated by the existing checklist of the flora of LEWS (Chai 2000), we feel confident that further fieldwork would bring to light many more palm treasures.

**Acknowledgments**

We thank the CEO of the Sarawak Forestry Corporation Haji Zolkipli Mohamad Aton for allocating resources and staff making this fieldwork possible. In particular, Julia Sang was instrumental with regards to planning and formalities, Sirukit anak Dubod took on a multitude of essential tasks including supporting our field collecting effort, cooking, and drying our specimens and Embahalbana anak Jangoh was our host and local contact at Nanga Segerak Ranger Station. In addition, the advice given to us by our three boatmen Mugu anak Sanggap, Dellie anak Medie and Medie anak Aloh was invaluable on the rapids of the river as well as in the forest collecting, and the inhabitants of Nanga Talong including the Headman Bada anak Chendai, who accommodated us in his home, all made us feel very welcome. We thank Dr John Dransfield, Dr Andrew Henderson and Dr Leng Guan




32. Boat journey from the ranger station to Nanga Talong. Photo: W.J. Baker.



33.



33. Nanga Talong longhouse resident with collection of headhunting skulls in rattan harnesses. Photo: B.G. Kuhnhäuser.

Saw for helping with determining specimens and for general advice on palms. Dr Paul Chai provided valuable information about the field site. Sarawak Forest Department kindly granted permission to collect (permits No. (207)JHS/NCCD/600-7/2/107 and No. WL107/2018) and export plant specimens (permit No. 18642). This work was supported logistically by the Royal Botanic Gardens, Kew. It was made possible thanks to generous funding from the Aarhus University Research foundation (grant AUFF-E-2017-7-19 to Wolf L. Eiserhardt), a Bentham-Moxon Trust grant (BMT1-2017) to WJB and BGK, a Society of Systematic Botanists Graduate Student Research Award to BGK, and a Brasenose College Student Support Funding grant to BGK.

## Literature Cited

Ancrenaz, M., M. Gumal, A.J. Marshall, E. Meijaard, S.A. Wich and S. Husson. 2016. *Pongo pygmaeus* ssp. *pygmaeus*. The IUCN Red List of Threatened Species 2016. https://www.iucnredlist.org.

Baker, W.J. 2015. A revised delimitation of the rattan genus *Calamus* (Arecaceae). Phytotaxa 197: 139–152.

Beccari, O. 1904. Wanderings in the Great Forests of Borneo; Travels and Researches of a Naturalist in Sarawak. Constable, London.

Brouwer, M. 2012. The Ecosystem Promise. Bunnik Meindert Brouwer Partner in Communicatie, Bunnik.

Chai, P.K. (ed.) 2000. Lanjak Entimau Wildlife Sanctuary, a check-list of flora, fauna, food and medicinal plants. Forest Department Sarawak and International Tropical Timber Organization (ITTO), Yokohama and Kuching.

Couvreur, T.L.P., W.D. Kissling, F.L. Condamine, J.C. Svenning, N.P. Rowe and W.J. Baker. 2015. Global diversification of a tropical plant growth form: environmental correlates and historical contingencies in climbing palms. Frontiers in Genetics 5: 1–18.

Dransfield, J. 1984. The genus *Areca* (Palmae: Arecoideae) in Borneo. Kew Bulletin 39: 1–22.

Dransfield, J. 1992. The rattans of Sarawak. Royal Botanic Gardens Kew and Sarawak Forest Department, Richmond and Kuching.

Forest Department Sarawak. 2020. Wild Life Sanctuaries. https://forestry.sarawak.gov.my.

Govaerts, R., J. Dransfield, S. Zona, D.R. Hodel and A. Henderson. 2020. World Checklist of Arecaceae. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; http://wcsp.science.kew.org.

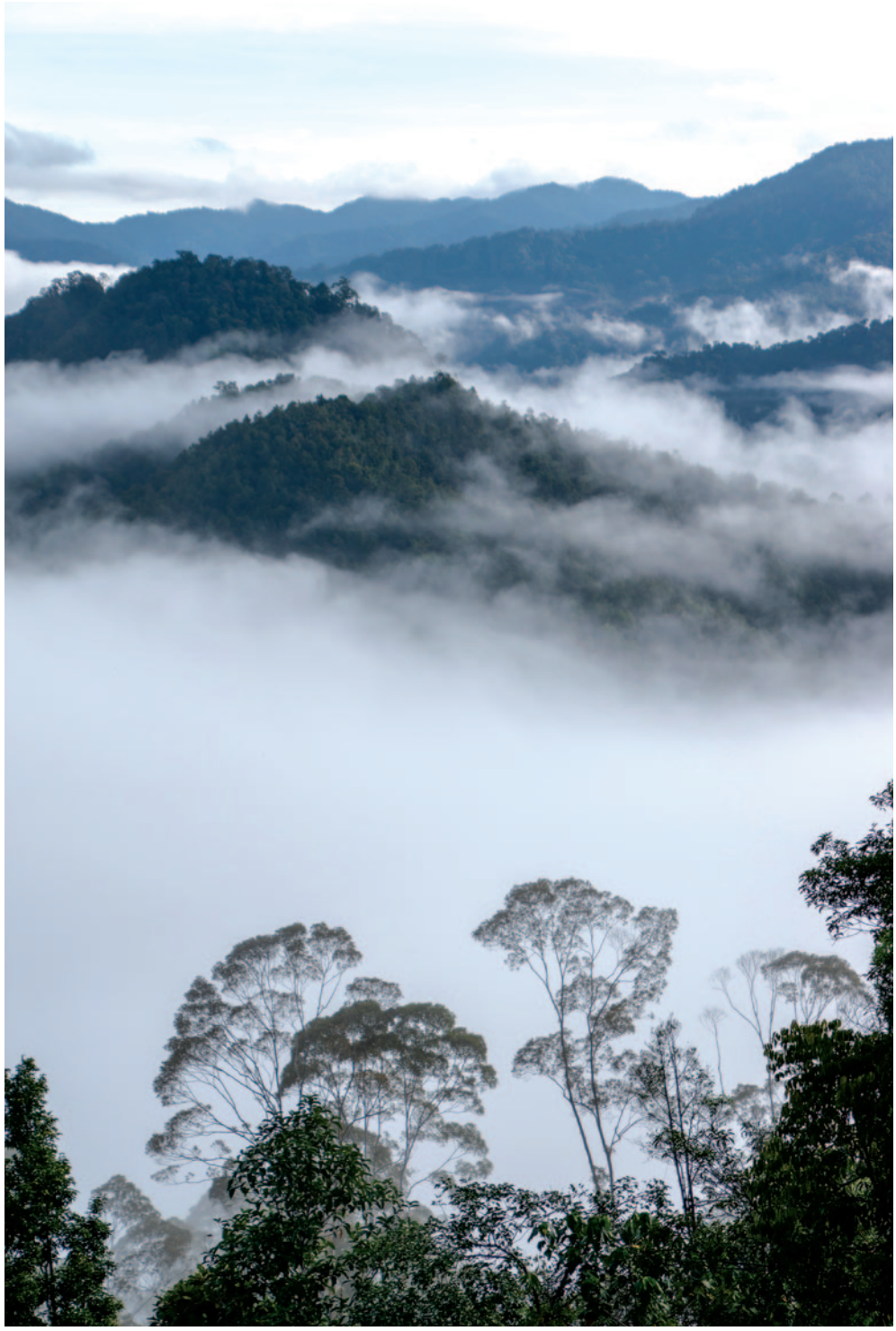
ITTO. 2001. Lanjak Entimau Wildlife Sanctuary: A Conservation Project Supported by International Tropical Timber Organization (ITTO). ITTO/Forest Department Sarawak.

Mohd-Azlan, J. and L. Engkamat. 2013. Camera trapping and conservation in

Lanjak Entimau Wildlife Sanctuary, Sarawak, Borneo. The Raffles Bulletin of Zoology 61: 397–405.

Rautner, M., M. Hardiono, R.J. Alfred and WWF Malaysia. 2005. Borneo: Treasure Island at Risk. WWF Germany.

34. Landscape near Lanjak Entimau Wildlife Sanctuary. Photo: W.J. Baker.



**Appendix 1. Palms recorded around Nanga Segerak Ranger Station, Lanjak Entimau Wildlife Sanctuary. Specimen codes represent herbarium collections: BK = Benedikt G. Kuhnhäuser, PP = Peter Petoe, X = no voucher specimen collected (sight record).**

| Subfamily | Genus | Specific epithet | Specimen |
|---|---|---|---|
| Arecoideae | *Areca* | *chaiana* | PP35, PP50 |
| | *Iguanura* | *chaiana* | PP45 |
| | *Oncosperma* | *horridum* | PP41 |
| | *Pinanga* | aff. *brevipes* | PP32, PP34 |
| | | *albescens* | PP39 |
| | | cf. *ligulata* | PP38 |
| | | *lepidota* | PP43 |
| | | *salicifolia* | X |
| | | sp. 1 | PP36 |
| | | sp. 2 | X |
| | | *tomentella* | PP37 |
| Calamoideae | *Calamus* | cf. *periacanthus* | X |
| | | *collarifer* | BK25 |
| | | *comptus* | BK27, BK46 |
| | | *corrugatus* | BK41 |
| | | *didymophyllus* | BK48 |
| | | *hallierianus* | BK37 |
| | | *javensis* | BK32 |
| | | *korthalsii* | BK33, PP33 |
| | | *maculatus* | BK39 |
| | | *myriacanthus* | BK28, BK30, BK45 |
| | | *ornatus* | PP44 |
| | | *pilosellus* | BK52, BK53 |
| | | *pseudoulur* | PP42 |
| | | *sabalensis* | BK31 |
| | | *sabut* | BK36 |
| | | *sparsiflorus* | BK51 |
| | | *ursinus* | BK26 |
| | | *zonatus* | BK43 |
| | *Eugeissona* | *utilis* | BK38 |
| | *Korthalsia* | *cheb* | PP40 |
| | | *debilis* | BK47, BK55 |
| | | *echinometra* | BK50 |
| | | *hispida* | BK40 |
| | | *rostrata* | BK34 |
| | *Plectocomia* | *mulleri* | BK54 |
| | *Plectocomiopsis* | *geminiflora* | BK49 |
| | | *mira* | BK44 |
| | | sp. | BK35 |
| | *Salacca* | *vermicularis* | BK56 |
| Coryphoideae | *Arenga* | *undulatifolia* | X |
| | *Caryota* | *mitis* | PP48 |
| | *Licuala* | cf. *maculata* | PP46 |
| | | *petiolulata* | PP47 |
| | | *rheophytica* | PP31, PP49, PP51 |

## Appendix 2. A robust phylogenomic framework for the calamoid palms

Published journal article on following pages.

**Bibliographic information:**

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., 2021. A robust phylogenomic framework for the calamoid palms. Molecular Phylogenetics and Evolution 157, 107067. https://doi.org/10.1016/j.ympev.2020.107067.

Molecular Phylogenetics and Evolution 157 (2021) 107067

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

# A robust phylogenomic framework for the calamoid palms

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 Animal and Plant Sciences, University of Sheffield, United Kingdom
[f] Department of Biology, Aarhus University, Denmark

ARTICLE INFO

*Keywords:*
Arecaceae
Calamoideae
Rattan
Phylogenomics
Systematics
Gene tree conflict

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 more than one fifth 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. 1a) and the presence of spines (Fig. 1b). They also exhibit an extraordinary variety of growth forms (Fig. 1c-e), the climbing habit (Fig. 1e) being predominant (Kissling et al., 2019).

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

https://doi.org/10.1016/j.ympev.2020.107067
Received 19 August 2020; Received in revised form 23 December 2020; Accepted 30 December 2020
Available online 5 January 2021
1055-7903/© 2021 Elsevier Inc. All rights reserved.

**Table 1**
**Classification, diversity, and distribution of the calamoid genera.** Classification follows Baker and Dransfield (2016). Species numbers according to Govaerts et al. (2021) 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<br>Subtribe<br>*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] |

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 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 (Dransfield et al., 2008; Baker and Dransfield, 2016). 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* (*Calospatha* Becc., *Ceratolobus* Blume ex Schult. & Schult.f., *Daemonorops* Blume, *Pogonotium* J.Dransf., and *Retispatha* J. Dransf.) have now been synonymized to render the genus monophyletic (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 123 of the 176 loci of Heyduk et al. (2015), 8 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 primarily for the phylogenetically divergent Arecoideae and Coryphoideae (Heyduk et al., 2015; de La Harpe 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 tribes, subtribes and genera, 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.




**Fig. 1. Diversity of calamoid palms. a**, **b**, Synapomorphies of the Calamoideae. **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.

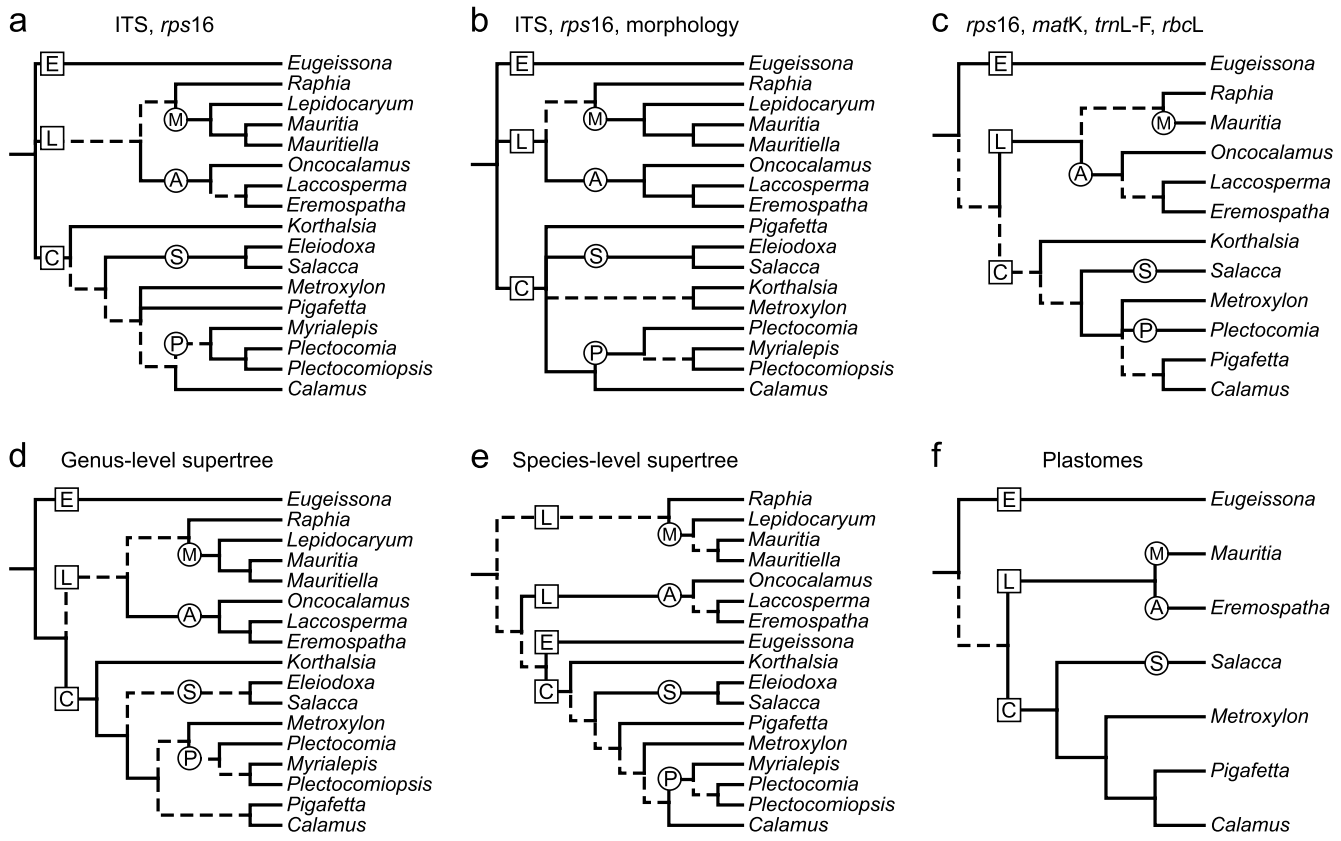



**Fig. 2. Previous phylogenetic inferences of the Calamoideae.** Selected previous inferences of relationships among calamoid tribes, subtribes and genera. Most topologies resolve the same generic relationships within the Lepidocaryeae, as well as identical placements of *Korthalsia* and the Salaccinae within the Calameae. In contrast, inferred relationships among the remaining genera of the Calameae and among the three tribes are often differing and generally 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.

### 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 × 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.

**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<br>Subtribe<br>*Genus*<br>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) |
| 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) |
| 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) |

(*continued on next page*)

**Table 2** (*continued*)

| **TRIBE**<br>**Subtribe**<br>*Genus*<br>Section | Species | Voucher |
|---|---|---|
| *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) |

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., 2020), 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 coalescent 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 reads on target, 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

**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 | 394,577 | 61,164 | 6.1 | 133 | 11 | 0 | 0 |
| | Q1 | 2,299,154 | 399,717 | 15.7 | 923 | 874 | 607 | 176 |
| | Median | 3,343,687 | 713,749 | 22.2 | 947 | 912 | 722 | 307 |
| | Q3 | 5,290,890 | 1,061,112 | 27.1 | 952 | 926 | 769 | 357 |
| | Maximum | 13,234,981 | 3,858,999 | 39.8 | 959 | 941 | 823 | 463 |
| Outgroup | Median | 5,675,654 | 1,340,924 | 31.6 | 963 | 955 | 917.5 | 698 |

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. 2a). Overarching methodologies (coalescent vs. concatenation) and datasets (exons vs. supercontigs) had the strongest effect on differences in tree topology, whereas substitution model testing in the coalescent analyses and partitioning of concatenated analyses had the smallest effects (Supplementary Fig. 2b). When only considering overarching methodologies and datasets, the coalescent analyses inferred highly similar relationships, whereas larger degrees of conflict existed between the other possible pairwise comparisons of species tree topologies (Supplementary Fig. 2b).

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. 2a). 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. 3a). 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. 3a). 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. 2a).

We found low gene tree conflict for all tribal and subtribal relationships, as well as higher-level relationships in Mauritiinae and Salaccinae (Fig. 4a,b,c,e,f). In contrast, substantial gene tree conflict was identified in Ancistrophyllinae and especially Plectocomiinae (Fig. 4d, 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. 2a), whereas substantial conflict was inferred for two of the other three investigated relationships in the backbone of *Calamus* (Fig. 4h).

## 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

**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 | 1,167,751 | 18.3 | 576,659 | 49.4 | 322,370 | 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 | 10,121 | 57.6 | 7359 | 87.3 | 4749 | 69.0 | 63.0 |
| | Concatenated | 79 | 1,943,739 | 28.0 | 1,202,869 | 61.9 | 664,320 | 34.2 | 41.2 |
| Splash zones | Concatenated | 79 | 775,988 | 42.6 | 626,210 | 80.7 | 341,950 | 44.1 | 29.4 |

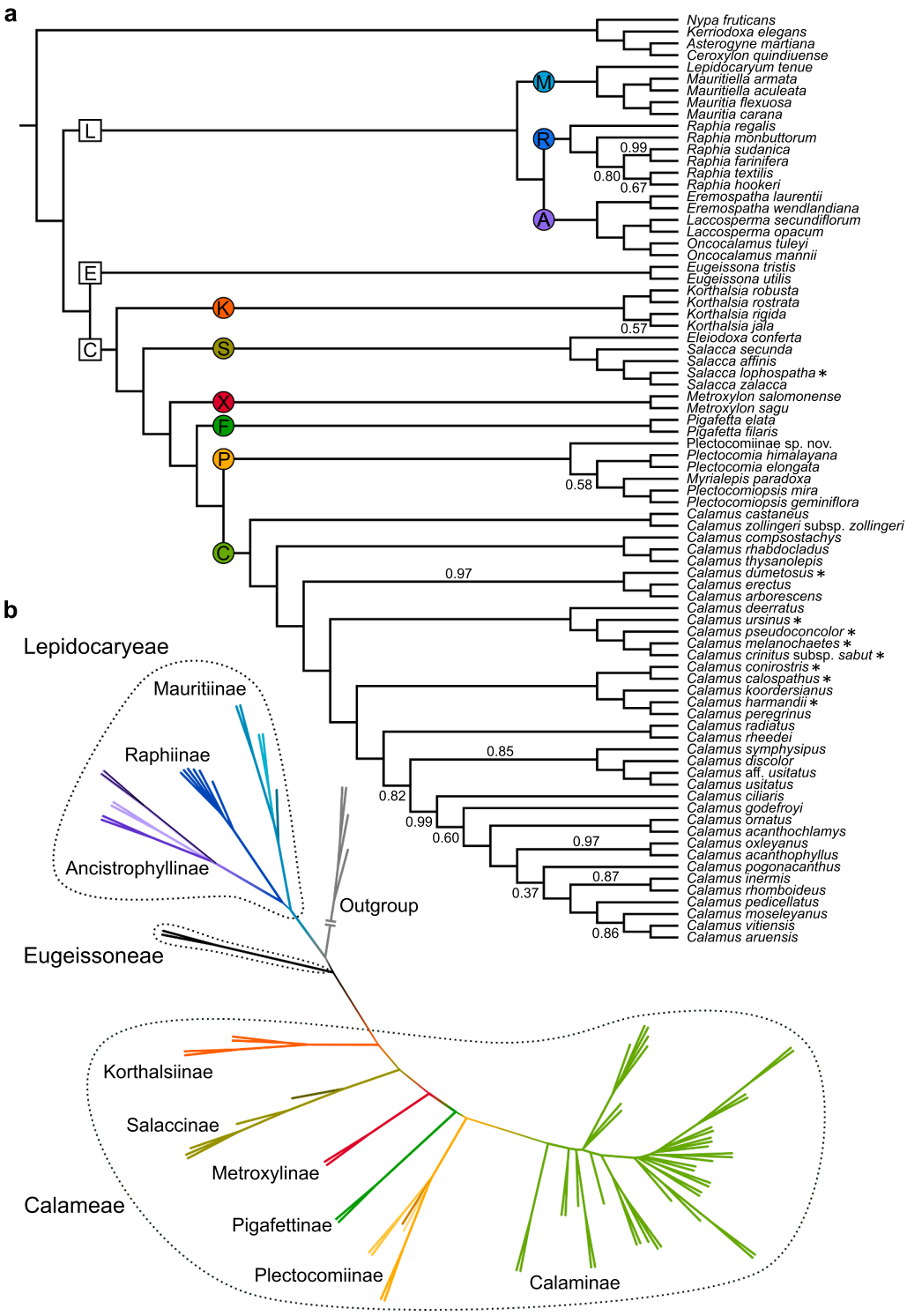



**Fig. 3. Phylogenomic inferences of the Calamoideae based on coalescent analysis of 945 supercontigs.** Phylogenetic trees of 75 species representing the diversity of the Calamoideae, plus four outgroup species (*Nypa fruticans*, *Kerriodoxa elegans*, *Asterogyne martiana*, *Ceroxylon quindiuense*) representing the remaining four palm subfamilies. Individual 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 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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. 2a). 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

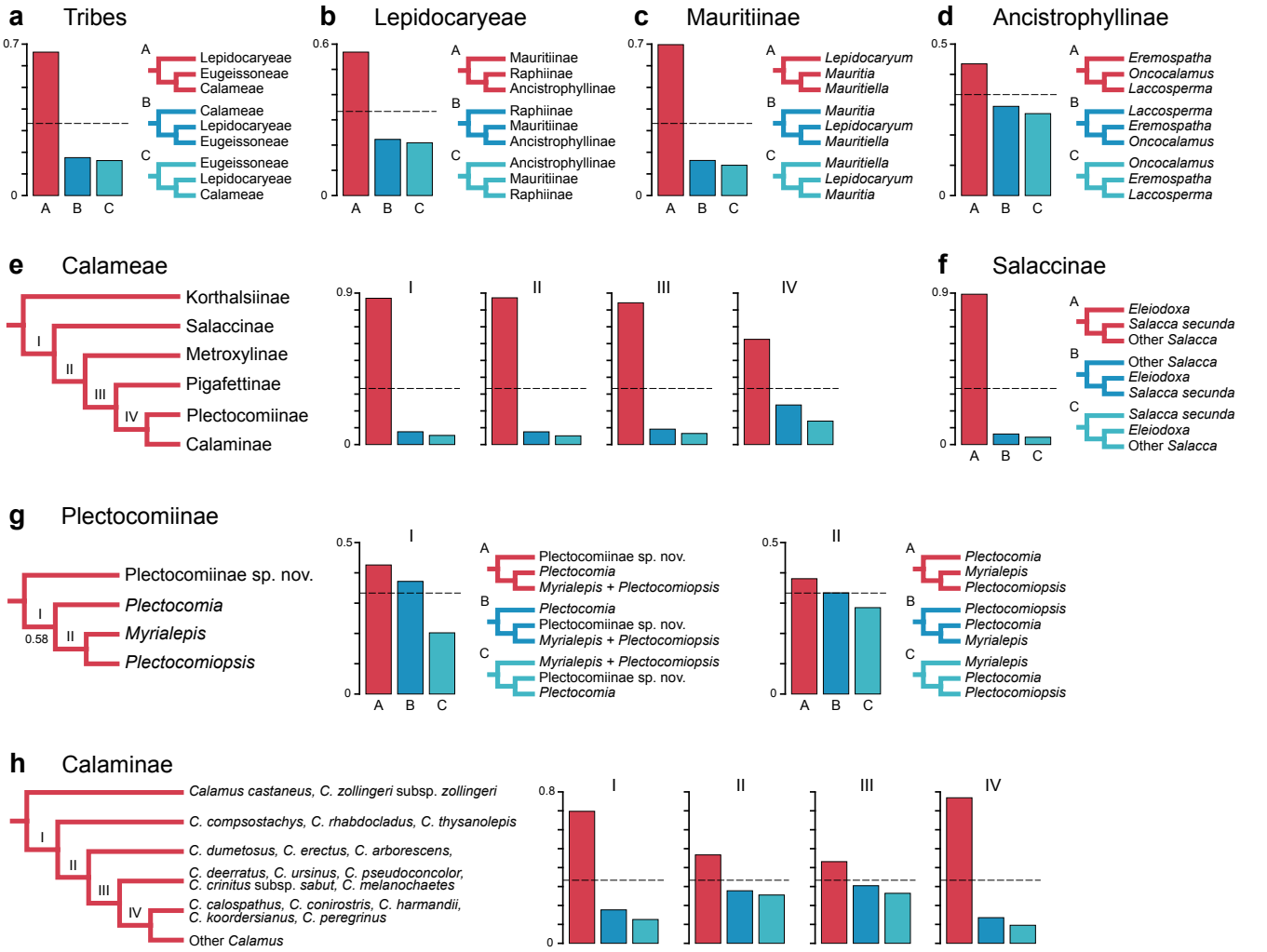



**Fig. 4. Gene-tree conflict in higher-level relationships of the Calamoideae.** Gene tree quartet frequencies for alternative topologies of selected higher-level relationships, based on coalescent analysis of 945 supercontigs (Fig. 3). Tribal and subtribal relationships, as well as relationships within 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 coalescent 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.

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. 4a) 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. 4b), 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 high degree of gene tree conflict in Ancistrophyllinae (Fig. 4d) 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. 4c). In Calameae, all subtribal relationships were resolved with maximal support and low gene tree conflict (Figs. 3, 4e), 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. 4g) mirrors the inference 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. 2a) 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. 4h, Supplementary Fig. 2a).

### 4.3. Potential causes of gene tree conflict

Substantial gene tree conflict was observed in Ancistrophyllinae, Plectocomiinae and Calaminae (Fig. 4d,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. 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. 4g), which appears morphologically intermediate between the three genera of Plectocomiinae. 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. 4f), 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. 4g) hints to a complex evolutionary history of the subtribe.

In Calaminae, the strongly supported nesting of all synonymized genera within *Calamus* (Fig. 3a) 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. 4h), 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. 3a).

## 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).

## 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 NERC Independent Research Fellowship (NE/S014470/1) to GC; 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.

## CRediT authorship contribution statement

**Benedikt G. Kuhnhäuser:** Conceptualization, Funding acquisition, Resources, Methodology, Investigation, Formal analysis, Validation, Visualization, Data curation, Project administration, Writing - original draft, Writing - review & editing. **Sidonie Bellot:** Funding acquisition, Resources, Methodology, Investigation, Supervision, Writing - review & editing. **Thomas L.P. Couvreur:** Resources, Methodology, Writing - review & editing. **John Dransfield:** Resources, Methodology, Validation, Writing - review & editing. **Andrew Henderson:** Resources, Methodology, Validation, Writing - review & editing. **Rowan Schley:** Methodology, Investigation, Writing - review & editing. **Guillaume Chomicki:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Wolf L. Eiserhardt:** Conceptualization, Funding acquisition, Resources, Methodology, Investigation, Supervision, Writing - review & editing. **Simon J. Hiscock:** Conceptualization, Methodology, Supervision, Writing - review & editing. **William J. Baker:** Conceptualization, Funding acquisition, Resources, Methodology, Validation, Project administration, Supervision, Writing - review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## 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.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2020.107067.

## 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.

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., 2016a. 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., 2016b. 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, C.C., Webb, C.O., Wurdack, K.J., Jaramillo, C.A., Donoghue, M.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.
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).
Gardner, E.M., Johnson, M.G., Pereira, J.T., Puad, A.S.A., Arifiani, D., Sahromi, Wickett, N.J., Zerega, N.J.C., 2020. Paralogs and off-target sequences improve phylogenetic resolution in a densely-sampled study of the breadfruit genus (Artocarpus, Moraceae). Syst. Biol. 073, 1–18. https://doi.org/10.1093/sysbio/syaa073.
Govaerts, R., Dransfield, J., Zona, S.F., Hodel, D.R., Henderson, A., 2021. World checklist of Arecaceae. Facilitated by the Royal Botanic Gardens, Kew. Available at: http://wcsp.science.kew.org (accessed 18 January 2021).
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., 2012a. 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., 2012b. 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.
Shahimi, S., 2018. Systematics and evolution of the rattan genus *Korthalsia* Bl. (Arecaceae) with special reference to domatia. Ph.D. thesis University of Reading.
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 Bioinf. 19, 153. https://doi.org/10.1186/s12859-018-2129-y.