Supplementary Material for:

Hundreds of Nuclear Genes and a New Synthesis of Fossil Evidence Shed Light on the Early Cretaceous Origin and Evolutionary History of Palms (Arecaceae)

Sidonie Bellot, Fabien L. Condamine, Kelly K.S. Matsunaga, Robert J. Morley, Santiago Ramirez-Barahona, Ángela Cano, Thomas L.P. Couvreur, Robyn Cowan, Wolf L. Eiserhardt, Benedikt G. Kuhnhäuser, Olivier Maurin, Michelle Siros, Felix Forest, Ilia J. Leitch, William J. Baker

The supplementary material provided below is available from GitHub at: https://github.com/sidonieB/Bellot_et_al_Palm_Early_Evolution_Supplementary_Material, together with the raw and clean sequence alignments, the gene trees and the species trees generated for this study.

The final version of the supplementary material will be deposited in a Dryad Digital Repository once the study has been peer-reviewed and accepted for publication.

Contents:

Supplementary Figure S1. Consistency of median age estimates across molecular dating analyses relying on node calibrations. For both strategies (left: "younger ages" and right: "older ages"), median age estimates are compared between analyses constraining or not the tree topology to reflect that of the species tree inferred with ASTRAL.

Supplementary Figure S2. Phylogenetic relationships among palm genera. Left: all-genes tree based on 802 nuclear genes classified as paralogous and 231 classified as orthologous; **Right:** orthologs-only tree based only on the 231 genes classified as orthologous. Labels above branches indicate local posterior probabilities. Stars indicate branches where gene tree discordance was significantly inconsistent with ILS and/or GTE alone, without (*) or with (**) Bonferroni correction. Numbers below branches indicate the effective number of gene trees in which that branch was recovered, while numbers above branches are local posterior probabilities. Pie charts represent quartet scores, with black representing the scores of the displayed topology while grey and white represent the scores of the alternative topologies.

Supplementary Figure S3. Relationship between effective number of genes, branch resolution and gene tree discordance. The identification of branches where gene tree discordance was significantly inconsistent with ILS and/or GTE alone was performed with (top) and without (bottom) Bonferroni correction.

Supplementary Figure S4. Variation in divergence time estimates between the "younger ages" and "older ages" strategies for the node calibration and FBD approaches. Median ages (bold numbers) and 95% highest posterior density intervals (bars) are indicated at nodes. Small numbers correspond to clades listed in Table S4. Nodes calibrated in the node calibration approach are indicated with red arrows and the fossil name. The insets show the relationships between branch length and support in the all-genes species tree shown in Figure 1 (Strong: local posterior probability $\geq$ 0.9; Weak: local posterior probability < 0.9) after excluding four branches that were supported by no more than 50 genes.

Supplementary Figure S5. Ancestral ranges estimated based on the younger ages tree including fossils and genus-level tip ranges coded following the 10-area scheme. Rectangles at the nodes represent the ancestral range with the highest probability – when the range encompasses multiple areas, multiple rectangles are printed on the same line. The

second and third most probable ranges are printed above and under the nodes for cases where the most probable range had a probability inferior to 0.66 and to twice the probability of the second most probable range. Pie charts indicate the probability of the most probable range (black), second most probable range (grey), third most probable range (light grey) and of all other ranges cumulated (white).

Supplementary Figure S6. Ancestral ranges estimated based on the younger ages tree pruned from fossils and genus-level tip ranges coded following the 10-area scheme. Rectangles at the nodes represent the ancestral range with the highest probability – when the range encompasses multiple areas, multiple rectangles are printed on the same line. The second and third most probable ranges are printed above and under the nodes for cases where the most probable range had a probability inferior to 0.66 and to twice the probability of the second most probable range. Pie charts indicate the probability of the most probable range (black), second most probable range (grey), third most probable range (light grey) and of all other ranges cumulated (white).

Supplementary Figure S7. Ancestral ranges estimated based on the younger ages tree including fossils and species-level tip ranges coded following the 10-area scheme. Rectangles at the nodes represent the ancestral range with the highest probability – when the range encompasses multiple areas, multiple rectangles are printed on the same line. The second and third most probable ranges are printed above and under the nodes for cases where the most probable range had a probability inferior to 0.66 and to twice the probability of the second most probable range. Pie charts indicate the probability of the most probable range (black), second most probable range (grey), third most probable range (light grey) and of all other ranges cumulated (white).

Supplementary Figure S8. Ancestral ranges estimated based on the younger ages tree including fossils and genus-level tip ranges coded following the 7-area scheme. Rectangles at the nodes represent the ancestral range with the highest probability – when the range encompasses multiple areas, multiple rectangles are printed on the same line. The second and third most probable ranges are printed above and under the nodes for cases where the most probable range had a probability inferior to 0.66 and to twice the probability of the second most probable range. Pie charts indicate the probability of the most probable range

(black), second most probable range (grey), third most probable range (light grey) and of all other ranges cumulated (white).

Supplementary Figure S9. Ancestral ranges estimated based on the older ages tree including fossils and genus-level tip ranges coded following the 10-area scheme. Rectangles at the nodes represent the ancestral range with the highest probability – when the range encompasses multiple areas, multiple rectangles are printed on the same line. The second and third most probable ranges are printed above and under the nodes for cases where the most probable range had a probability inferior to 0.66 and to twice the probability of the second most probable range. Pie charts indicate the probability of the most probable range (black), second most probable range (grey), third most probable range (light grey) and of all other ranges cumulated (white).

Supplementary Methods

Tissue sampling and DNA sequencing

All currently recognised palm genera were represented by one species. Additional species were sampled in two genera (*Areca*, *Oncosperma*) as we expected them to represent yet undescribed genera based on preliminary analyses. The genus *Dasypogon* (Dasypogonaceae) was included as an outgroup, following most recent monocot phylogenomic studies (Barrett et al., 2016; Li et al., 2021). Plant material was obtained from silica-dried leaf tissue or from herbarium specimens. Species names, sample origin and voucher information are provided in Table S1.

After DNA extraction, DNA longer than ca. 1,000 base pairs (bp) was fragmented using a M220 Focused-ultrasonicator™ and AFA Fiber Pre-Slit Snap-Cap microTUBES (Covaris, Woburn, MA, USA), with the following settings: peak power: 50; duty % factor: 20; cycles/burst: 200; power: 10; duration: 55 seconds; temperature: 20°C. Hybridisation lasted 24 hours and was conducted separately for the Angiosperms353 and PhyloPalm probe kits (Johnson et al., 2019; Loiseau et al., 2019), with a temperature of hybridisation of 65°C. It was followed by 12–16 PCR cycles as recommended in the manufacturer's protocol (http://www.arborbiosci.com/mybaits-manual). Libraries enriched in Angiosperms353 or PhyloPalm regions were pooled and sequenced separately, generating two sequence datasets for each sample, except for 23 samples (out of 187) for which only PhyloPalm data could be generated, and 2 for which only Angiosperms353 data could be generated, because there was not enough DNA library to perform two hybridisations (the type of data that was generated for each sample is indicated in Table S1). DNA sequencing was performed on an Illumina MiSeq with v2 or v3 chemistry at the Royal Botanic Gardens, Kew, or on an Illumina HiSeq X at Macrogen Inc. (Seoul, Korea), yielding 2 x 300 bp-long or 2 x 150 bp-long paired-end reads, respectively. For nine taxa (Table S1), the PhyloPalm dataset had been generated prior to this study and was downloaded from public repositories or provided by co-authors. The raw sequencing read data was deposited in GenBank (accession numbers in Table S1). Parts of the data that were newly generated as part of this study were used in two recent studies: Angiosperm353 regions from most genera were included in a study on the diversification of all angiosperms (Zuntini, Carruthers et al., 2024) and a small subset of the PhyloPalm regions (known as the Heyduk regions; Heyduk et al., 2016) from 48 samples was used in studies of the palms of New Caledonia (Pérez-Calle, Bellot et al., 2024) and Madagascar (Eiserhardt et al., 2022).

Genomic data cleaning and assembly

Sequence data quality was assessed using FASTQC v. 0.11.9 (Andrews, 2010), and Trimmomatic v. 0.39 (Bolger et al., 2014) was used to remove Illumina sequencing adapters using the paired-end palindrome mode with 1 seed mismatch allowed, palindrome and simple clips thresholds of 30 and 7 respectively, a minimum adapter length of 2 bp and keeping both reads. Bases with low quality at the end of the reads were then removed using a sliding window ("SLIDINGWINDOW" parameter) of 4 bp and a minimum Phred quality threshold

of 30, while bases of low quality at the beginning of the reads were removed using the "LEADING" parameter and the same minimum quality threshold. Reads shorter than 40 bp after the trimming were discarded.

Clean reads were analysed using HybPiper v. 1.3.1 (Johnson et al., 2016) to recover and assemble the regions targeted by both probe kits. The original reference sequences used to recover reads matching the target regions comprised the default Angiosperms353 reference sequences (https://github.com/mossmatters/Angiosperms353/Angiosperms353_targetSequences.fasta), and the sequences used to design the PhyloPalm probe kit, mostly derived from the *Elaeis guineensis* genome (K. Heyduk, com. pers.; M. Paris, com. pers.; Heyduk et al., 2016; Loiseau et al., 2019). Using blastn 2.5.0 (part of the BLAST+ suite; Camacho et al., 2009), we found that 68 Angiosperms353 target regions aligned with the PhyloPalm target regions over more than 50 bp, with an e-value < 0.05, so the Angiosperms353 duplicated regions were discarded from the original reference sequences so that the same region would not be recovered and analysed twice. To maximise gene recovery across all samples, a custom set of reference sequences was built that included all the original reference sequences as well as target sequences recovered by submitting five high-quality datasets to an initial HybPiper run using the original reference sequences. The five datasets comprised data from species representing the five palm subfamilies: *Attalea butyracea* (Arecoideae), *Korthalsia echinometra* (Calamoideae), *Phytelephas aequatorialis* (Ceroxyloideae), *Chuniophoenix nana* (Coryphoideae) and *Nypa fruticans* (Nypoideae). This final set of reference sequences comprises 11,348 sequences corresponding to 1,255 target regions and is available at https://github.com/sidonieB/Palm_phylogenomics_resources/blob/main/A353PP_noOverlap_285-970_wPalmRefs.fasta.

Classification of paralogous and orthologous regions

Previous studies have provided evidence for the occurrence of a whole genome duplication event in the ancestral lineage that gave rise to all current palm species (Barrett et al., 2019), suggesting that some or all the genetic regions targeted by our study may originally have been duplicated in the ancestral lineage, even if they now appear to be single copy regions. This issue has previously been addressed for subfamily Calamoideae by using published whole genome sequences of *Calamus* (Zhao et al., 2018) to identify genes with multiple copies. To mitigate for the risk of aligning and analysing paralogs instead of orthologs across the palm family, we expanded an approach first developed for Calamoideae (Kuhnhäuser, 2021) by classifying the regions into orthologous vs. paralogous regions based on their copy numbers in the annotated genomes of palm species representing the three main palm subfamilies Arecoideae, Calamoideae and Coryphoideae: *Elaeis guineensis* (Arecoideae; coding sequences from the EG5 genome version 3 downloaded from http://genomsawit.mpob.gov.my on 20 November 2020; Chan et al., 2017), *Calamus simplicifolius* (Calamoideae; coding sequences file downloaded from http://gigadb.org/dataset/101052 on 8 March 2020; Zhao et al., 2018) and *Phoenix dactylifera* (Coryphoideae; transcript variants downloaded from https://datepalmgenomehub.abudhabi.nyu.edu/?q=node/12; Hazzouri et al., 2019). Our target regions were aligned against all the coding sequences from the annotated genome and a

region was considered paralogous if the search identified at least two hits with the coding sequence, each with e-value scores < 10^{-21} and overlapping by more than 30% of the coding sequence length. This was conducted separately for each of the three annotated genomes, and a region identified as paralogous based on at least one of these searches was considered paralogous, while regions not identified as paralogous based on any of the three searches were considered orthologous. The latter is based on the rationale that even if these regions were duplicated in the ancestral lineage that gave rise to palms, the fact that they are found to be single copy regions in calamoids, coryphoids and arecoids suggests that reduction to a single copy occurred before these subfamilies diverged from each-other, and therefore before palms diversified, leading to the same copy being subsequently shared by all subfamilies and species. In addition, we classified as paralogous any region that raised a "paralog warning" in at least one sample during the HybPiper analysis. These warnings are raised when multiple overlapping, different contigs are assembled from the reads matching a single target region, which can indicate that the region is present in multiple copies in the sample being analysed. Classifying such regions as paralogous therefore allowed us to mitigate the risk of paralogy due to more recent genome/gene duplications that might have taken place throughout palm evolution. This paralog search resulted in classifying 951 out of 1,255 target regions as paralogous and 304 as orthologous

Cleaning of multiple sequence alignments

Multiple sequence alignments were generated for each region using MAFFT v. 7 (Katoh & Standley, 2013) with the "--genafpair" setting and 1,000 iterations. OptrimAl (https://github.com/keblat/bioinfo-utils/blob/master/docs/advice/scripts/optrimAl.txt) was then used to trim alignment columns so as to minimise gaps in alignments while keeping their informativeness high. Trimmed alignments were then further cleaned with CIAlign v. 1.1.0 (Tumescheit et al., 2022) to remove sequences that were very divergent from the others and therefore likely spurious. After trying different thresholds, sequences were removed if less than 85% of their nucleotides corresponded to the most common base in the alignment at that position. Care was taken to not discard the outgroup taxon sequence data during this process by using the option "--retain-str". TAPER v. 1.0 (Zhang et al., 2021) was then used on the resulting alignments to remove mis-aligned, highly divergent stretches of otherwise non-spurious sequences, using default settings. The resulting alignments were then trimmed again with OptrimAl. As a result of the cleaning process, 137 target regions were discarded because they contained less than four taxa or because the alignment entirely comprised sequences shorter than 250 bp or considered uninformative by OptrimAl.

Molecular dating analyses using the node calibration approach

Times of divergence between palm genera were estimated using BEAST 2.6 (Bouckaert et al., 2019). Because of computational limitations and potential conflicts between gene histories, we based the molecular dating analysis on the 22 orthologous regions that yielded the top 10% gene trees with the highest bipartition agreement to the species tree based on orthologous genes, as identified by (Smith et al., 2018). These 22 regions were concatenated using AMAS (Borowiec, 2016) and BEAST was run allowing each region to be analysed under a different nucleotide substitution model corresponding to the one identified

previously using IQ-TREE v. 1.6.12 (Minh et al., 2020), or to the closest model available in BEAST. Clocks and trees were linked across the partitions to enable the analyses to converge in a reasonable time. Substitution rate priors were set to lognormal distributions to facilitate chain mixing, and the tree generation prior was set according to the incomplete sampling birth-death model. To account for substitution rate heterogeneity between taxa, we used an optimised relaxed clock (ORC) model (Douglas et al., 2021). Analyses (see next section) were performed with and without constraining the tree topology based on the ASTRAL species tree obtained from summarizing all the genes. Topological constraints were implemented by modifying some parameters in the BEAST xml files following recommendations at https://www.beast2.org/2014/07/28/all-about-starting-trees and https://www.beast2.org/fix-starting-tree/. Specifically, we specified the species tree topology and we set the weights of the subtreeslide, narrowExchange, wideExchang, wilsonBalding and ORCAdaptableOperatorSampler_NER parameters to 0 directly in the xml configuration files. We used Tracer v. 1.7.2 (Rambaut et al., 2018) to monitor the convergence of the parameter estimates and to check that effective sampling sizes were mostly above 200. A run without using the data ("prior-only") was also performed for each analysis, in order to verify that the prior settings did not drive the posterior parameter estimations more than the data themselves. Number of runs, generations and effective sampling sizes are provided in Table S3 for each analysis. A maximum clade credibility tree was generated from the posterior tree distribution of each analysis using TreeAnnotator, which is part of BEAST, after removing the first 10% posterior trees generated (burn-in fraction). Median ages estimated for each clade across the post-burn-in posterior tree distribution were reported on this tree, together with the 95% highest posterior density age intervals.

Fossil selection and age calibrations used in the node calibration molecular dating approach

We used fossil records to calibrate the age of five nodes spread across the palm family. *Mauritiidites* spp. pollen records from the Maastrichtian to Campanian of Nigeria, Sudan, Egypt, Cameroon, Gabon and Angola (Edet & Nyong, 1993; Eisawi & Schrank, 2008; Rull, 1998; Salami, 1990; Salard-Cheboldaeff, 1990; Schrank, 1994; Table S2) were used to calibrate the age of the stem node of Mauritiinae. Pollen records of *Dicolpopollis cenomanicus* from the Cenomanian (Burger, 1990; Kuhnhäuser et al., 2025; Macphail & Jordan, 2015; Totterdell & Mitchell, 2009) were used to calibrate the stem node of the clade made of Metroxylinae, Plectocomiinae, Calaminae and Pigafettinae (Calamoideae, Calameae). The phylogenetic affinity and age of these pollen grains have been extensively discussed in a recent study (Kuhnhäuser et al., 2025). *Sabal bigbendense* Manch., Wheeler, & Lehman from the Campanian of Texas (ca. 77 Ma; (Cano, 2018; Manchester et al., 2010) was used to calibrate the stem node of *Sabal* (Coryphoideae, Sabaleae). *Hyphaeneocarpon indicum* Bande, Prakash, & Ambwani emend. Matsunaga, S.Y.Sm., Manch., Srivastava, & Kapgate (Bande et al., 1982; Matsunaga et al., 2019) from the Deccan Intertrappean Beds (64–67 Ma; Matsunaga et al., 2018) was used to calibrate the crown node of Hyphaeninae (Coryphoideae, Borasseae), and *Palmocarpon drypetoides* (Mehrotra, Prakash & Bande) Manch., Bonde, Nipunage, Srivastava, Mehrotra & S.Y.Sm. (Manchester et al., 2016), also from the Deccan Intertrappean Beds (64 – 67 Ma; Matsunaga et al., 2018), was used to calibrate the crown of Attaleinae (Arecoideae, Cocoseae). The placement of

these two fossils was supported by previous morphology-based phylogenetic inference (Matsunaga & Smith, 2021). Further details on the fossils are provided in Table S2.

Molecular dating analyses using the FBD approach

We jointly inferred divergence times and phylogenetic relationships between extant palm genera and 113 fossils by implementing the fossilized birth-death model (REF) in RevBayes (REF). To facilitate computation, the analysis was performed on the concatenated alignment of the five regions yielding gene trees most compatible with the ASTRAL species tree, as identified by SortADate. These trees were also among the most clock-like (i.e. showed low substitution rate heterogeneity), and GTR was inferred to be the best model for all but one, so GTR+G was used as the nucleotide substitution model for the concatenated alignment, which was unpartitioned to limit computational burden. After preliminary runs, topological constraints were selectively applied on relationships between extant palm genera so that the posterior tree would be consistent with our best knowledge of their relationships, i.e. the ASTRAL species tree obtained from summarizing all the genes. Additional constraints were applied on fossils based on their taxonomic affinities as reviewed in this study and specified in Table S2. The strategy followed for the calibration of the origin time is provided in the main text. The diversification rate was set as a function of the number of taxa (including fossils) and the maximum origin time. The sampling fraction was set to 0.074 based on the 2522 palm species accepted on 30 September 2024 (POWO). Moves between parameter values, and prior distributions and initial values on the turnover and psi parameters mainly followed previous studies (Ramírez-Barahona, 2024) on similarly-sized datasets, with some tweaks after preliminary trials. Scripts used to run the analyses including all the parameter values are available on GitHub at https://github.com/sidonieB/Bellot_et_al_Palm_Early_Evolution_Supplementary_Material_Version_2. For the younger ages analysis, two runs were conducted (each of 300 000 generations, sampling every 100 generation), and after checking that their parameter values converged, the second run was extended to 800 000 generations, and its posterior distribution was used to generate the "younger ages" maximum clade credibility tree, after removing a burn-in fraction of 40%. For the older ages analysis, three runs were conducted (two of 300 000 generations and one of 600 000 generations, sampling every 100 generation), and as the parameter values of the second and third run converged, the second run was extended to 750 000 generations, and its posterior distribution was used to generate the "older ages" maximum clade credibility tree after removing a burn-in fraction of 40%. Ages reported on the maximum clade credibility trees were median ages.

Fossil selection for the FBD approach

Although individual fossil reports were grouped as a single entry in Table S2 if they belonged to the same taxon, geological horizon and region, one fossil per entry was selected to be used in the FBD analysis, i.e. we did not include two fossils that had the same taxon name, and occurred in the same geological horizon and region. As we could only use a FBD model that cannot account for stratigraphic ranges of occurrence, fossils recorded through multiple geological periods in the same location were represented by the fossil present in the oldest demonstrated stratigraphic interval, which was then interpreted as an uncertainty

around the age of the fossil. Once this was applied, age intervals with same minimum and maximum age were expanded to the rounded down value of the minimum age and the rounded up value of the maximum age so that the interval width was 1 and not 0 (this only concerned the pollen records from Jaramillo et al., 2024). Finally, ranges expanding into the Cretaceous without further precision had their maximum age set to 100.5 as there was no evidence justifying ages older than the Late Cretaceous.

Fossils were selected for this analysis if they were clearly assigned to a clade below subfamily level, if their age and the uncertainty around it was clear (i.e. no mention of age being uncertain in Table S2), and if their placement could be constrained as follows. Multiple fossils were allowed to belong to a same clade (or to form a clade with an extant palm genus). However, fossils belonging to two nested clades were only allowed if the shallowest clade was only comprised of fossils and a single extant palm genus. For instance, we allowed to have one fossil placed in Attaleinae and one in *Attalea*, as the first fossil could still at least be placed sister to the second fossil + *Attalea*, but we did not allow to have a fossil placed in Cocoseae and a fossil placed in Attaleineae, because constraining the placement of the second fossil would prevent the first fossil to belong to Attaleineae. When a choice had to be made between two fossils because of their nested taxonomic placement, the solution accommodating fossils placed with the highest precision was preferred (for instance we selected Attaleineae and Bactridinae fossils instead of Cocoseae fossils).

Ancestral range inferences using DEC

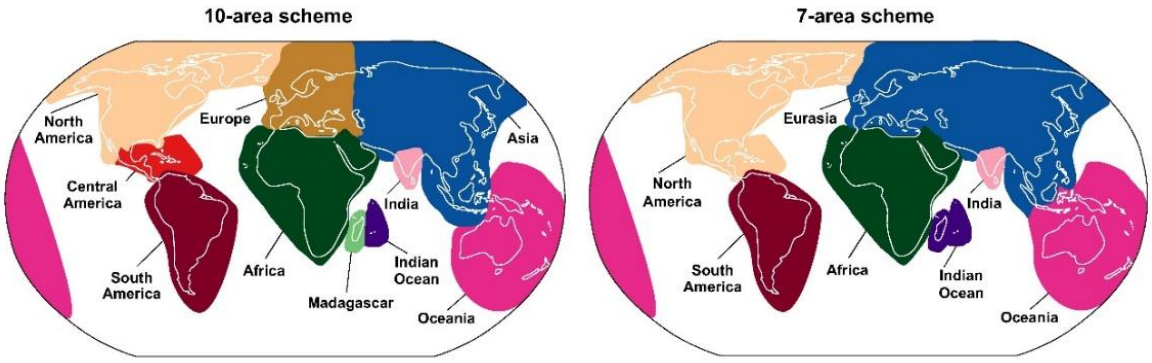
The evolution of ancestral ranges along the palm phylogeny was estimated using the Dispersal-Extinction-Cladogenesis (DEC) model (Ree & Sanmartín, 2018), as implemented in the DEC eXtended version (DECX, Beeravolu & Condamine, 2016; available at: https://github.com/champost/DECX). DECX requires a time-calibrated tree, the distribution of each taxon for a set of geographic areas, and a time-stratified geographic model that is represented by connectivity and dispersal multiplier matrices for specified time intervals spanning the entire evolutionary history of the group. DECX allows classical vicariance as a cladogenetic event by using temporally flexible constraints on the connectivity between any two given areas following the movement of landmasses and dispersal opportunity over time. However, DECX does not incorporate the founder-event speciation (+J parameter) because of concerns with statistical validity of model choice among DEC-derived models (Ree & Sanmartín, 2018). Also, founder-event speciation often leads to inferences that are decoupled from time, with null or extremely low extinction rates, an effect of the model favouring cladogenetic events over anagenetic events (Ree & Sanmartín, 2018), which makes it inadequate for reconstructing the history of ancient groups with widespread distributions. Analyses were performed on the younger and older trees obtained from the FBD analyses, with and without pruning the fossils from the trees. When fossils were kept, their distribution range was used to inform the analysis (see Table S2 and main Methods).

Area delimitation for the ancestral range inferences

Analyses were performed using two area delimitation schemes. The first scheme (10-area) comprised North America, Central America, South America, Europe, Asia, Africa,

Oceania, India, Indian Ocean, and Madagascar. The second scheme (7-area) comprised North America (including Central America from the 10-area scheme) South America, Africa, India, Eurasia (merging Europe and Asia from the 10-area scheme), Oceania and Indian Ocean (including Madagascar from the 10-area scheme). The delimitation between Asia and Oceania was different in each scheme (following Lidekker's line in the 10-area scheme, and Wallace's line in the 7-area scheme; see map below). Using both schemes enabled us to explore the impact of different rationales subtending area definitions, and to compare results to the main previous global palm biogeography study (Baker & Couvreur, 2013) that used the 7-area scheme.

Maps of both area schemes:



The table on the following page provides the precise definition and underlying rationale for each area for each scheme.

10-area scheme	**Delimitation**	**Rationale (based on geology during periods relevant for genera divergences, i. e. Oligocene and older)**
North America	Includes Florida, Bermuda, and northern Mexico down to Sinaloa, Durango, Zacatecas, Nuevo Leon, San Luis Potosi and Tamaulipas included	The rest of Mexico was not formed/emerged as early as the part included here
Central America	Antilles (but not Trinidad), and everything between Panama and the Mexico states excluded from North America	Trinidad is geologically part of South America. Central America as defined here had a geological history very different from North and South America.
South America	Includes Trinidad and everything south of Central America	See Central America
Europe	From the British isles until the Turgai strait	The Turgai strait was the main potential barrier across Eurasia
Asia	From the Turgai strait to everything north west of Lydekker line, including Wallacea, Nicobar island, Andaman islands.	In the past, Wallacea was physically closer to South East Asia than to New Guinea
Africa	Africa, the Arabian Peninsula, Canary Islands, Cape Verde	The Arabian Peninsula was closer to Africa than Asia for a long part of their geological history
Oceania	Includes Australia, New Guinea, New Zealand, and everything south east of Lydekker line, including the south Pacific islands and Hawaii.	See Asia. Assignment of south Pacific islands to Oceania was decided because a detailed study of the spread of palms across the Pacific can only be meaningfully done at the species level and by delimiting many new areas corresponding to each archipelago. Hawaii was included here as it was more parsimonious considering the distribution of *Pritchardia* species.
India	Includes Sri Lanka, Assam, and Bengladesh	These regions were either not emerged or part of India when it was not yet attached to Asia
Indian Ocean	Mascarenes, Seychelles	Fragments of areas which were much bigger in the past, and which were separated from India and Madagascar
Madagascar	Includes Comoros and other islands in the Mozambique Channel	Assignment of Comoros and nearby islands to Madagascar instead of Africa is arbitrary, they could be on their own in a more precise analysis but they concern very few palms so we chose to avoid adding an area just for these.
7-area scheme		
North America	Includes everything north of Panama (included) + Antilles (but not Trinidad)	Follows Baker & Couvreur 2013; Merges North America and Central America of Scheme 1
South America	Includes everything south of Panama, incl. Trinidad	Follows Baker & Couvreur 2013
Eurasia	From the British Isles to Wallace's line, including Andaman and Nicobar islands	Follows Baker & Couvreur 2013; Merges Europe and a slightly reduced Asia from Scheme 1 (excludes Wallacea)
Oceania	Everything East of Wallace's line, incl. Australia, New Guinea, New Zealand, and the Pacific islands	Follows Baker & Couvreur 2013; slightly expanded compared to Scheme 1 (includes Wallacea)
Africa	Includes the Arabian Peninsula, Canary Islands, Cape Verde	Follows Baker & Couvreur 2013
India	Includes Sri Lanka, Assam, Bengladesh	Follows Baker & Couvreur 2013
Indian Ocean	Mascarenes, Seychelles, Madagascar, Comoros	Follows Baker & Couvreur 2013; Merges Indian Ocean and Madagascar from Scheme 1

Connectivity information for the ancestral range inferences

Analyses were performed under a model where area connectivity was constrained based on how distances between areas changed through geological periods, after preliminary analyses showed that analyses performed without such constraints provided unrealistic results where ancestral species would be spread across many distant areas. Based on paleogeographic reconstructions (e.g.: Kocsis & Scotese, 2021; Seton et al., 2012), we built connectivity matrices to represent major changes in tectonic conditions that could have affected the distribution of palms. Constraints on area connectivity were specified in a matrix by coding 0 if any two areas are not connected or 1 if these areas are connected during a given period. The time interval was dissected into four time slices: from 0 to 36 Ma, 36 to 56 Ma, 56 to 85 Ma, and 85 Ma to the tree root age. Hence, a time-stratified geographic model was created in the form of binary matrices that consider paleogeographic changes through time with time slices indicating the possibility or not for a species to colonize a new area (Beeravolu & Condamine, 2016).

Antarctica was not included in the area schemes as no palm currently occurs there. However, an occurrence in Antarctica remains implicitly allowed by the model via the connectivity matrices allocating non-null probabilities to connections between Oceania, Africa and America during the Cretaceous, when these regions were connected through what is now Antarctica.

The tables on the following page show the connectivity matrices for each scheme. In these matrices, a ‘1’ at the intersection of two areas means that a taxon can occupy a range encompassing the two areas, while a ‘0’ means that it cannot occupy a range encompassing the two areas.

10-area scheme

200 - 86 Ma	NA	CA	SA	EU	AF	IN	AS	OC	IO	MA
North America	1									
Central America	1	1								
South America	1	1	1							
Europe	1	1	1	1						
Africa	1	1	1	1	1					
India	1	1	1	0	1	1				
Asia	1	1	1	1	0	0	1			
Oceania	1	1	1	0	1	1	0	1		
Indian Ocean	1	1	1	0	1	1	0	1	1	
Madagascar	1	1	1	0	1	1	0	1	1	1

56 - 37 Ma	NA	CA	SA	EU	AF	IN	AS	OC	IO	MA
North America	1									
Central America	1	1								
South America	0	1	1							
Europe	1	0	0	1						
Africa	0	0	0	1	1					
India	0	0	0	0	0	1				
Asia	1	0	0	1	0	1	1			
Oceania	0	0	1	0	0	0	0	1		
Indian Ocean	0	0	0	0	1	1	1	0	1	
Madagascar	0	0	0	0	1	1	0	0	1	1

85 - 57 Ma	NA	CA	SA	EU	AF	IN	AS	OC	IO	MA
North America	1									
Central America	1	1								
South America	0	1	1							
Europe	1	0	0	1						
Africa	0	0	0	1	1					
India	0	0	0	0	1	1				
Asia	1	0	0	1	0	0	1			
Oceania	0	0	1	0	0	0	0	1		
Indian Ocean	0	0	0	0	1	1	0	0	1	
Madagascar	0	0	0	0	1	1	0	0	1	1

36 - 0 Ma	NA	CA	SA	EU	AF	IN	AS	OC	IO	MA
North America	1									
Central America	1	1								
South America	1	1	1							
Europe	1	0	0	1						
Africa	0	0	0	1	1					
India	0	0	0	1	1	1				
Asia	1	0	0	1	1	1	1			
Oceania	0	0	0	0	0	0	1	1		
Indian Ocean	0	0	0	0	1	1	1	0	1	
Madagascar	0	0	0	0	1	1	0	0	1	1

7-area scheme

200 - 86 Ma	NA	SA	AF	IN	EU	OC	IO
North America	1						
South America	1	1					
Africa	1	1	1				
India	1	1	1	1			
Eurasia	1	1	1	0	1		
Oceania	1	1	1	1	0	1	
Indian Ocean	1	1	1	1	0	1	1

56 - 37 Ma	NA	SA	AF	IN	EU	OC	IO
North America	1						
South America	0	1					
Africa	0	0	1				
India	0	0	0	1			
Eurasia	1	0	1	1	1		
Oceania	0	1	0	0	1	1	
Indian Ocean	0	0	1	1	1	0	1

85 - 57 Ma	NA	SA	AF	IN	EU	OC	IO
North America	1						
South America	0	1					
Africa	0	0	1				
India	0	0	1	1			
Eurasia	1	0	1	0	1		
Oceania	0	1	0	0	0	1	
Indian Ocean	0	0	1	1	0	0	1

36 - 0 Ma	NA	SA	AF	IN	EU	OC	IO
North America	1						
South America	1	1					
Africa	0	0	1				
India	0	0	1	1			
Eurasia	1	0	1	1	1		
Oceania	0	0	0	0	1	1	
Indian Ocean	0	0	1	1	1	0	1

References

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (21 June 2024, date last accessed)*.

Baker, W. J., & Couvreur, T. L. P. (2013). Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical biogeography. *Journal of Biogeography*, *40*(2), 274–285. https://doi.org/10.1111/j.1365-2699.2012.02795.x

Bande, M. B., Prakash, U., & Ambwani, K. (1982). A fossil palm fruit hyphaeneocarpon indicum gen. et sp. nov. from the Deccan Intertrappean beds, India. *The Palaeobotanist*, *30*(3), 303–309.

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., Wm. Stevenson, D., Zomlefer, W. B., & Davis, J. I. (2016). Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytologist*, *209*(2), 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 Biology and Evolution*, *11*(5), 1501–1511. https://doi.org/10.1093/gbe/evz092

Beeravolu, R., & Condamine, F. (2016). An extended maximum likelihood inference of geographic range evolution by dispersal, local extinction and cladogenesis. *BioRxiv*.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 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*, *2016*(1). 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., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *15*(4). https://doi.org/10.1371/journal.pcbi.1006650

Burger, D. (1990). Early Cretaceous angiosperms from Queensland, Australia. In *Review of Palaeobotany and Palynology* (Vol. 65).

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*. https://doi.org/10.1186/1471-2105-10-421

Cano, A. (2018). *What can palm evolution in time and space say about the historical assembly of diversity in the Caribbean and Central America?* Université de Geneve.

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. *Biology Direct*, *12*(1). https://doi.org/10.1186/s13062-017-0191-4

Douglas, J., Zhang, R., & Bouckaert, R. (2021). Adaptive dating and fast proposals: Revisiting the phylogenetic relaxed clock model. *PLoS Computational Biology*, *17*(2). https://doi.org/10.1371/JOURNAL.PCBI.1008322

Edet, J. J., & Nyong, E. E. (1993). Depositional environments, sea–Level history and palaeobiogeography of the Late Campanian–Maastrichtian on the Calabar Flank, southeast Nigeria. . *Palaeogeography, Palaeoclimatology, Palaeoecology*, *102*, 161–175.

Eisawi, A., & Schrank, E. (2008). Upper Cretaceous to Neogene palynology of the Melut Basin, southeast Sudan. *Palynology*, *32*, 101–129.

Eiserhardt, W. L., Bellot, S., Cowan, R. S., Dransfield, J., Hansen, L. E. S. F., Heyduk, K., Rabarijaona, R. N., Rakotoarinivo, M., & Baker, W. J. (2022). Phylogenomics and generic limits of Dypsidinae (Arecaceae), the largest palm radiation in Madagascar. *Taxon*, *71*(6), 1170–1195. https://doi.org/10.1002/tax.12797

Hazzouri, K. M., Gros-Balthazard, M., Flowers, J. M., Copetti, D., Lemansour, A., Lebrun, M., Masmoudi, K., Ferrand, S., Dhar, M. I., Fresquez, Z. A., Rosas, U., Zhang, J., Talag, J., Lee, S., Kudrna, D., Powell, R. F., Leitch, I. J., Krueger, R. R., Wing, R. A., … Purugganan, M. D. (2019). Genome-wide association mapping of date palm fruit traits. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-12604-9

Heyduk, K., Trapnell, D. W., Barrett, C. F., & Leebens-Mack, J. (2016). Phylogenomic analyses of species relationships in the genus Sabal (Arecaceae) using targeted sequence capture. *Biological Journal of the Linnean Society*, *117*(1), 106–120. https://doi.org/10.1111/bij.12551

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. *Applications in Plant Sciences*, *4*(7).

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., Eiserhardt, W. L., Epitawalage, N., Forest, F., Kim, J. T., Leebens-Mack, J. H., Leitch, I. J., Maurin, O., Soltis, D. E., Soltis, P. S., Wong, G. K. S., Baker, W. J., & Wickett, N. J. (2019). A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any Flowering Plant Designed Using k-Medoids Clustering. *Systematic Biology*, *68*(4), 594–606. https://doi.org/10.1093/sysbio/syy086

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kocsis, Á. T., & Scotese, C. R. (2021). Mapping paleocoastlines and continental flooding during the Phanerozoic. In *Earth-Science Reviews* (Vol. 213). Elsevier B.V. https://doi.org/10.1016/j.earscirev.2020.103463

Kuhnhäuser, B. G. (2021). *Phylogenomics and biogeography of the calamoid palms, D.Phil. thesis*. University of Oxford.

Kuhnhäuser, B. G., Bates, C. D., Dransfield, J., Geri, C., Henderson, A., Julia, S., Ying Lim, J., Morley, R. J., Rustiami, H., Schley, R. J., Bellot, S., Chomicki, G., Eiserhardt, W. L., Hiscock, S. J., & Baker, W. J. (2025). Island geography drives evolution of rattan palms in tropical Asian rainforests. *Science*, *387*, 1204–1209. https://www.science.org

Li, H. T., Luo, Y., Gan, L., Ma, P. F., Gao, L. M., Yang, J. B., Cai, J., Gitzendanner, M. A., Fritsch, P. W., Zhang, T., Jin, J. J., Zeng, C. X., Wang, H., Yu, W. Bin, Zhang, R., van der Bank, M., Olmstead, R. G., Hollingsworth, P. M., Chase, M. W., … Li, D. Z. (2021). Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biology*, *19*(1). https://doi.org/10.1186/s12915-021-01166-2

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., … Salamin, N. (2019). Targeted capture of hundreds of nuclear genes unravels phylogenetic relationships of the diverse neotropical palm tribe geonomateae. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.00864

Macphail, M., & Jordan, G. (2015). Tropical palms and arums at near-polar latitudes: Fossil pollen evidence from the tamar and macquarie grabens, Northern Tasmania. In *Papers and Proceedings of the Royal Society of Tasmania* (Vol. 149, pp. 23–28). Royal Society of Tasmania. https://doi.org/10.26749/rstpp.149.23

Manchester, S. R., Bonde, S. D., Nipunage, D. S., Srivatava, R., Mehrotra, R. C., & Smith, S. Y. (2016). Trilocular palm fruits from the Deccan Intertrappean beds of India. *International Journal of Plant Sciences*, *177*(7), 633–641. https://doi.org/10.1086/687290

Manchester, S. R., Lehman, T. M., & Wheeler, E. A. (2010). Fossil palms (arecaceae, coryphoideae) associated with juvenile herbivorous dinosaurs in the Upper Cretaceous Aguja Formation, Big Bend National Park, TEXAS. *International Journal of Plant Sciences*, *171*(6), 679–689. https://doi.org/10.1086/653688

Matsunaga, K. K. S., Manchester, S. R., Srivastava, R., Kapgate, D. K., & Smith, S. Y. (2019). Fossil palm fruits from India indicate a Cretaceous origin of Arecaceae tribe Borasseae. In *Botanical Journal of the Linnean Society* (Vol. 190). https://academic.oup.com/botlinnean/article-abstract/190/3/260/5518348

Matsunaga, K. K. S., & Smith, S. Y. (2021). Fossil palm reading: using fruits to reveal the deep roots of palm diversity. *American Journal of Botany*, *108*(3), 472–494. https://doi.org/10.1002/ajb2.1616

Matsunaga, K. K. S., Smith, S. Y., Manchester, S. R., Kapgate, D., Ramteke, D., Garbout, A., & Villarraga-Gómez, H. (2018). Reinvestigating an enigmatic Late Cretaceous monocot: Morphology, taxonomy, and biogeography of Viracarpon. *PeerJ*, *2018*(4). https://doi.org/10.7717/peerj.4580

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., Lanfear, R., & Teeling, E. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Pérez-Calle, V., Bellot, S., Kuhnhäuser, B. G., Pillon, Y., Forest, F., Leitch, I. J., & Baker, W. J. (2024). Phylogeny, biogeography and ecological diversification of New Caledonian palms (Arecaceae). *Annals of Botany*. https://doi.org/10.1093/aob/mcae043

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Ramírez-Barahona, S. (2024). Incorporating fossils into the joint inference of phylogeny and biogeography of the tree fern order Cyatheales. *Evolution*, *78*(5), 919–933. https://doi.org/10.1093/evolut/qpae034

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. *Journal of Biogeography*, *45*(4), 741–749. https://doi.org/10.1111/jbi.13173

Rull, V. (1998). Biogeographical and evolutionary considerations of Mauritia (Arecaceae), based on palynological evidence. *Review of Palaeobotany and Palynology*, *100*, 122.

Salami, M. B. (1990). Palynomorph taxa from the Lower Coal Measures deposits (?Campanian–Maastrichtian) of Anambra Trough, southeastern Nigeria. *African Journal of Earth Sciences*, *11*, 135–150.

Salard-Cheboldaeff, M. (1990). Intertropical African palynostratigraphy from Cretaceous to late Quaternary times. *Journal of African Earth Sciences (and the Middle East)*, *11*, 1–6.

Schrank, E. (1994). Palynology of the Yesomma Formation in Northern Somalia: a study of pollen spores and associated phytoplankton from the late Cretaceous Palmae Province. *Palaeontographica*, *231*, 63–112.

Seton, M., Müller, R. D., Zahirovic, S., Gaina, C., Torsvik, T., Shephard, G., Talsma, A., Gurnis, M., Turner, M., Maus, S., & Chandler, M. (2012). Global continental and ocean basin reconstructions since 200Ma. In *Earth-Science Reviews* (Vol. 113, Issues 3–4, pp. 212–270). https://doi.org/10.1016/j.earscirev.2012.03.002

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*(5). https://doi.org/10.1371/journal.pone.0197433

Totterdell, J., & Mitchell, C. (2009). Bight Basin geological sampling and seepage survey. *Geoscience Australia Record*, *24*, 1–28.

Tumescheit, C., Firth, A. E., & Brown, K. (2022). CIAlign: A highly customisable command line tool to clean, interpret and visualise multiple sequence alignments. *PeerJ*. https://doi.org/10.7717/peerj.12983

Zhang, C., Zhao, Y., Braun, E. L., & Mirarab, S. (2021). TAPER: Pinpointing errors in multiple sequence alignments despite varying rates of evolution. *Methods in Ecology and Evolution*, *12*(11), 2145–2158. https://doi.org/10.1111/2041-210X.13696

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., … Jiang, Z. (2018). The chromosome-level genome assemblies of two rattans (Calamus simplicifolius and Daemonorops jenkinsiana). *GigaScience*, *7*(9). https://doi.org/10.1093/gigascience/giy097

Zuntini, A. R., Carruthers, T., Maurin, O., Bailey, P. C., Leempoel, K., Brewer, G. E., Epitawalage, N., Françoso, E., Gallego-Paramo, B., McGinnie, C., Negrão, R., Roy, S. R., Simpson, L., Toledo Romero, E., Barber, V. M. A., Botigué, L., Clarkson, J. J., Cowan, R. S., Dodsworth, S., … Baker, W. J. (2024). Phylogenomics and the rise of the angiosperms. *Nature*. https://doi.org/10.1038/s41586-024-07324-0

Figure S1

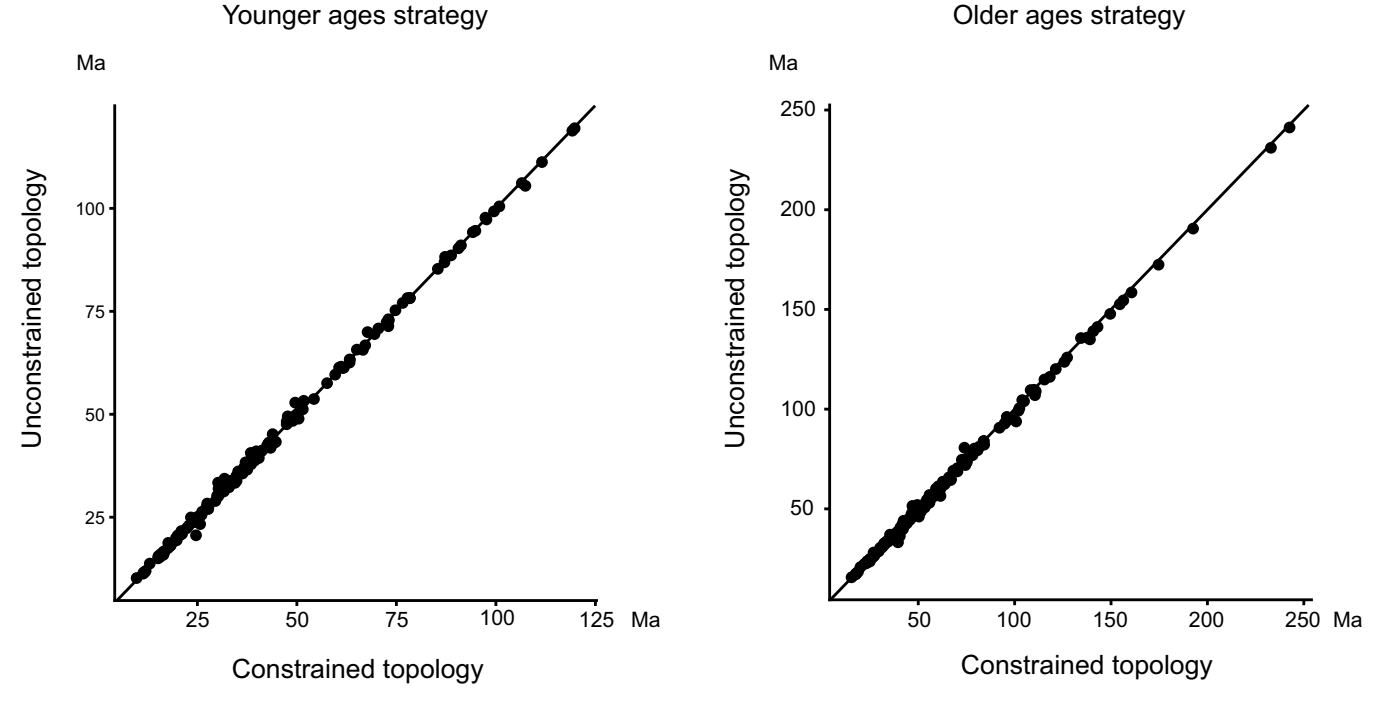

Younger ages strategy
Ma
Unconstrained topology
100
75
50
25
25
50
75
100
125
Ma
Constrained topology
Older ages strategy
Ma
Unconstrained topology
250
200
150
100
50
50
100
150
200
250
Ma
Constrained topology

Figure S2

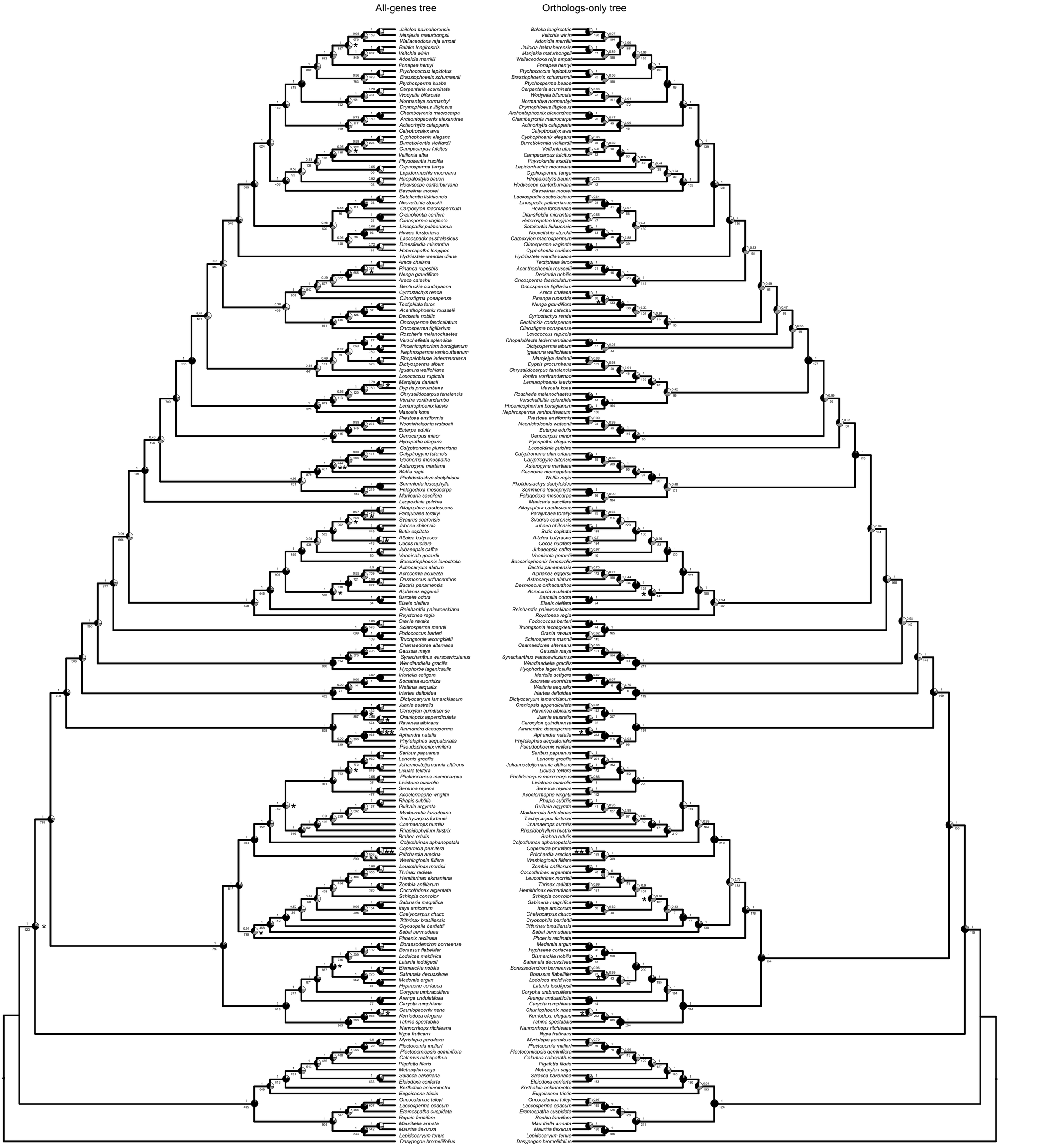

All-genes tree
Orthologs-only tree
Jailoloa halmaherensis
Manjekia maturbongsii
Wallaceodoxa raja ampat
Balaka longirostris
Veitchia winin
Adonidia merrillii
Ponapea hentyi
Ptychococcus lepidotus
Brassiophoenix schumannii
Ptychosperma buabe
Carpentaria acuminata
Wodyetia bifurcata
Normanbya normanbyi
Drymophloeus litigiosus
Chambeyronia macrocarpa
Archontophoenix alexandrae
Actinorhytis calapparia
Calyptrocalyx awa
Cyphophoenix elegans
Burretiokentia vieillardii
Campecarpus fulcitus
Veillonia alba
Physokentia insolita
Cyphosperma tanga
Lepidorrhachis mooreana
Rhopalostylis baueri
Hedyscepe canterburyana
Basselinia moorei
Satakentia liukiuensis
Neoveitchia storckii
Carpoxylon macrospermum
Cyphokentia cerifera
Clinosperma vaginata
Linospadix palmerianus
Howea forsteriana
Laccospadix australasicus
Dransfieldia micrantha
Heterospathe longipes
Hydriastele wendlandiana
Areca chaiana
Pinanga rupestris
Nenga grandiflora
Areca catechu
Bentinckia condapanna
Cyrtostachys renda
Clinostigma ponapense
Tectiphiala ferox
Acanthophoenix rousselii
Deckenia nobilis
Oncosperma fasciculatum
Oncosperma tigillarium
Roscheria melanochaetes
Verschaffeltia splendida
Phoenicophorium borsigianum
Nephrosperma vanhoutteanum
Rhopaloblaste ledermanniana
Dictyosperma album
Iguanura wallichiana
Loxococcus rupicola
Marojejya darianii
Dypsis procumbens
Chrysalidocarpus tanalensis
Vonitra vonitrandambo
Lemurophoenix laevis
Masoala kona
Prestoea ensiformis
Neonicholsonia watsonii
Euterpe edulis
Oenocarpus minor
Hyospathe elegans
Calyptronoma plumeriana
Calyptrogyne tutensis
Geonoma monospatha
Asterogyne martiana
Welfia regia
Pholidostachys dactyloides
Sommieria leucophylla
Pelagodoxa mesocarpa
Manicaria saccifera
Leopoldinia pulchra
Allagoptera caudescens
Parajubaea torallyi
Syagrus cearensis
Jubaea chilensis
Butia capitata
Attalea butyracea
Cocos nucifera
Jubaeopsis caffra
Voanioala gerardii
Beccariophoenix fenestralis
Astrocaryum alatum
Acrocomia aculeata
Desmoncus orthacanthos
Bactris panamensis
Aiphanes eggersii
Barcella odora
Elaeis oleifera
Reinhardtia paiewonskiana
Roystonea regia
Orania ravaka
Sclerosperma mannii
Podococcus barteri
Truongsonia lecongkietii
Chamaedorea alternans
Gaussia maya
Synechanthus warscewiczianus
Wendlandiella gracilis
Hyophorbe lagenicaulis
Iriartella setigera
Socratea exorrhiza
Wettinia aequalis
Iriartea deltoidea
Dictyocaryum lamarckianum
Juania australis
Ceroxylon quindiuense
Oraniopsis appendiculata
Ravenea albicans
Ammandra decasperma
Aphandra natalia
Phytelephas aequatorialis
Pseudophoenix vinifera
Saribus papuanus
Lanonia gracilis
Johannesteijsmannia altifrons
Licuala telifera
Pholidocarpus macrocarpus
Livistona australis
Serenoa repens
Acoelorrhaphe wrightii
Rhapis subtilis
Guihaia argyrata
Maxburretia furtadoana
Trachycarpus fortunei
Chamaerops humilis
Rhapidophyllum hystrix
Brahea edulis
Colpothrinax aphanopetala
Copernicia prunifera
Pritchardia arecina
Washingtonia filifera
Leucothrinax morrisii
Thrinax radiata
Hemithrinax ekmaniana
Zombia antillarum
Coccothrinax argentata
Schippia concolor
Sabinaria magnifica
Itaya amicorum
Chelyocarpus chuco
Trithrinax brasiliensis
Cryosophila bartlettii
Sabal bermudana
Phoenix reclinata
Borassodendron borneense
Borassus flabellifer
Lodoicea maldivica
Latania loddigesii
Bismarckia nobilis
Satranala decussilvae
Medemia argun
Hyphaene coriacea
Corypha umbraculifera
Arenga undulatifolia
Caryota rumphiana
Chuniophoenix nana
Kerriodoxa elegans
Tahina spectabilis
Nannorrhops ritchieana
Nypa fruticans
Myrialepis paradoxa
Plectocomia mulleri
Plectocomiopsis geminiflora
Calamus calospathus
Pigafetta filaris
Metroxylon sagu
Salacca bakeriana
Eleiodoxa conferta
Korthalsia echinometra
Eugeissona tristis
Oncocalamus tuleyi
Laccosperma opacum
Eremospatha cuspidata
Raphia farinifera
Mauritiella armata
Mauritia flexuosa
Lepidocaryum tenue
Dasypogon bromeliifolius

Figure S3

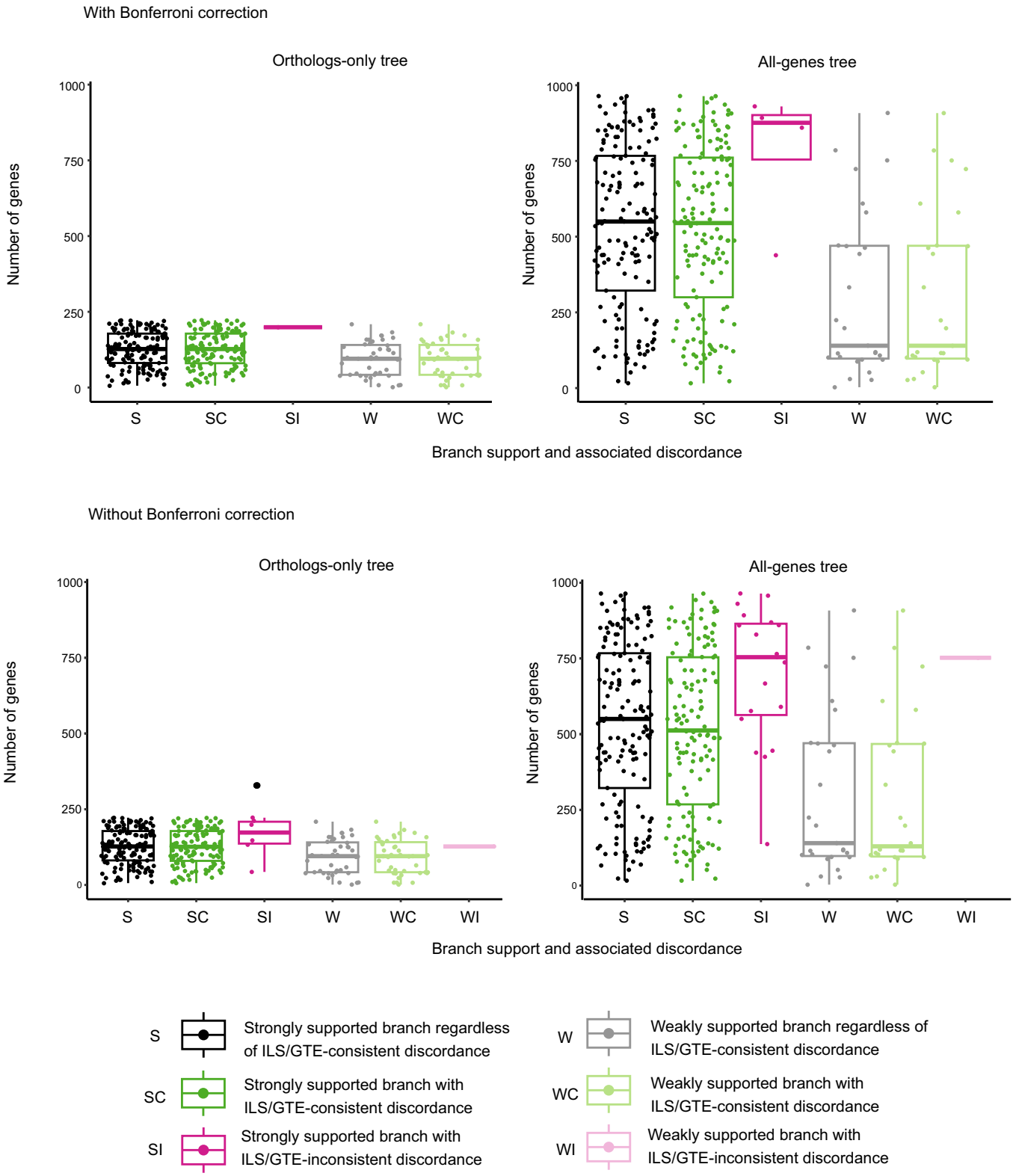

With Bonferroni correction
Orthologs-only tree
All-genes tree
Number of genes
1000
750
500
250
0
S
SC
SI
W
WC
Branch support and associated discordance
Without Bonferroni correction
Orthologs-only tree
All-genes tree
Number of genes
1000
750
500
250
0
S
SC
SI
W
WC
WI
Branch support and associated discordance
S
Strongly supported branch regardless of ILS/GTE-consistent discordance
SC
Strongly supported branch with ILS/GTE-consistent discordance
SI
Strongly supported branch with ILS/GTE-inconsistent discordance
W
Weakly supported branch regardless of ILS/GTE-consistent discordance
WC
Weakly supported branch with ILS/GTE-consistent discordance
WI
Weakly supported branch with ILS/GTE-inconsistent discordance

Figure S4

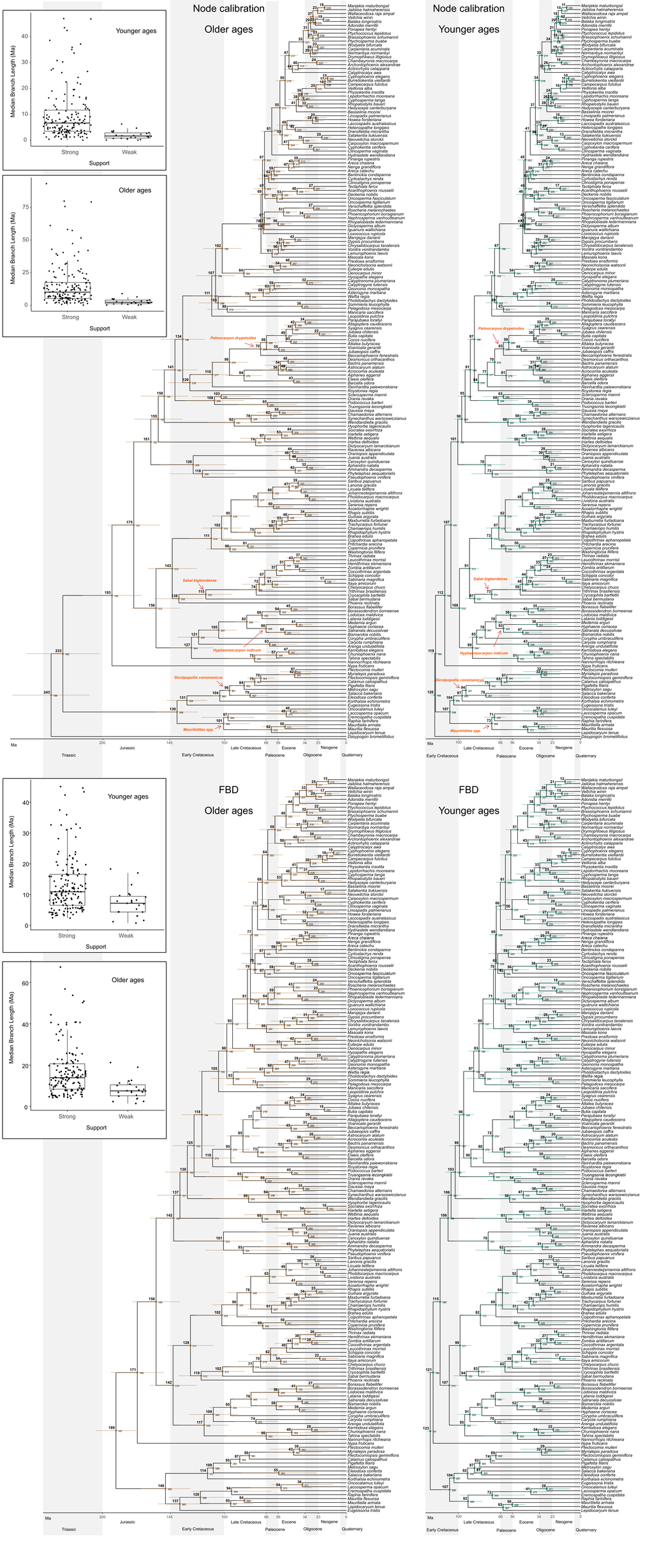

Node calibration
Older ages
Younger ages
Median Branch Length (Ma)
Strong
Weak
Support
Palmocarpon drygetoides
Sabal bigbendense
Hyphaenecarpon indicum
Dicolpopollis cenomanicus
Maurilitidites spp.
Ma
Triassic
Jurassic
Early Cretaceous
Late Cretaceous
Paleocene
Eocene
Oligocene
Neogene
Quaternary
Node calibration
Younger ages
Older ages
FBD
Younger ages
Older ages

Figure S5

Genus-level area coding, younger ages, 10-area scheme, with fossils

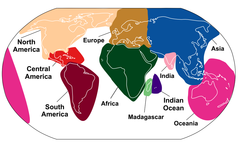


Figure S6

Genus-level area coding, younger ages, 10-area scheme, without fossils

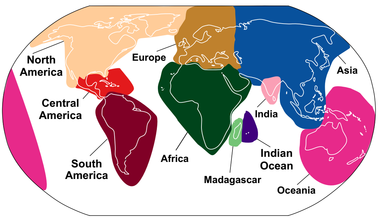


Figure S7

Species-level area coding, younger ages, 10-area scheme, with fossils

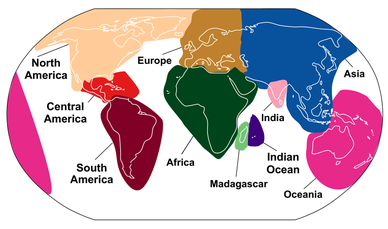
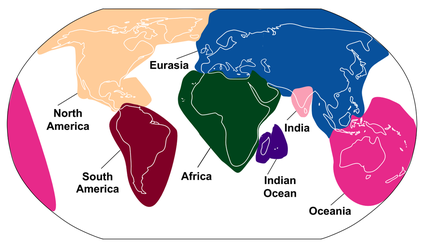


Figure S8

Genus-level area coding, younger ages, 7-area scheme, with fossils



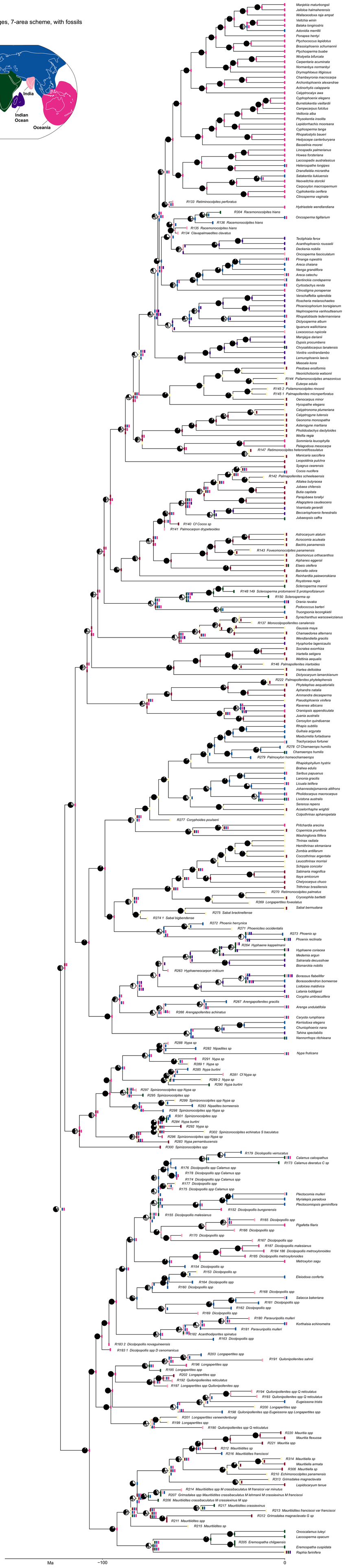


Figure S9

Genus-level area coding, older ages, 10-area scheme, with fossils