Ancestral Area Inference in Calamoideae Palms (Arecaceae) Using Biogeographic Models

Written by Thales de Lima

20/08/2022

Introduction

The palm family (Arecaceae) has more than 2,500 species, which are mainly concentrated in the tropics, but also occur in subtropical areas (Dransfield et al., 2008). The family is highly abundant in the tropics, for example it is estimated that 7 of the 20 most abundant tree species in the Amazon, including the most abundant of all, are palms (Ter Steege et al., 2013). The Arecaceae family is subdivided into five subfamilies: Calamoideae, Ceroxyloideae, Nypoideae, Coryphoideae and Arecoideae. The subfamily Calamoideae, together with Arecoideae, contains most of the species in the family.

The subfamily Calamoideae contains 17 genera and approximately 650 species (Dransfield et al., 2008), about 400 of which are contained in the genus Calamus (POW, 2022), the largest genus of the entire Arecaceae family. The species of this subfamily can be quickly distinguished from other palms by the reflex and overlapping scales present on the pericarp. Calamoideae is one of the best phylogenetically resolved subfamilies within Arecaceae (Baker et al., 2000b), being supported as monophyletic by many independent studies (Baker et al., 2000b; Baker et al., 2000a; Asmussen et al., 2006). Most Calamoideae species are distributed in the Old World, especially in Southeast Asia. Nevertheless, the subfamily is pantropical, occupying all the four tropical biogeographic regions (*sensu* Olson et al., 2001): Neotropic, Afrotropic, Indo-Malay and Australasia (see Figure 1).

Given this broad distribution, I have selected this group to perform a biogeographic study, more specifically, an inference of ancestral geographic range. Ancestral inference, also known as ancestral reconstruction, relies on current data about the focal characteristic mapped onto a phylogeny, an evolutionary model, and some kind of method (i.e, an algorithm) that establishes rules through each the extrapolation is made (Joy et al., 2016). Here I use a phylogeny of Calamoideae palms, their ranges in terms of biogeographic regions occupied, and a biogeographic model to infer the geographic ranges of these palms' ancestors.

Material and Methods

Performing an ancestral range inference for a clade requires a phylogeny describing the evolutionary relation among the species of such clade. The phylogeny used here is based on the one constructed by Baker et al. (2000b) and Baker et al. (2000a) to Calamoideae. The phylogeny was redone, including three new species whose sequences were available at GenBank (https://www.ncbi.nlm.nih.gov/genbank/). I have inferred the phylogeny using an Bayesian approach outside R during a course on phylogenetic systematics at the University of Campinas, in 2021. The methods description is not reported here, but the files needed to redone the analysis (i.e., the aligned sequencing data and the MrBayes script) and details on the generation of the tree can be found at `docs/tree_inference_documentation.pdf`. Be as it may, the final tree file is also available in the Git repository at `final_consensus_tree_run1.nex`. The final tree file just after inference has two problems that must be fixed before the ancestral range inference: 1) the tree is unrooted and 2) the tree is not ultrametric. Using the script `02_importing_n_checking_tree.R`. I rooted the tree using the outgroup taxa and coerced it into an ultrametric tree using a penalized likelihood method.
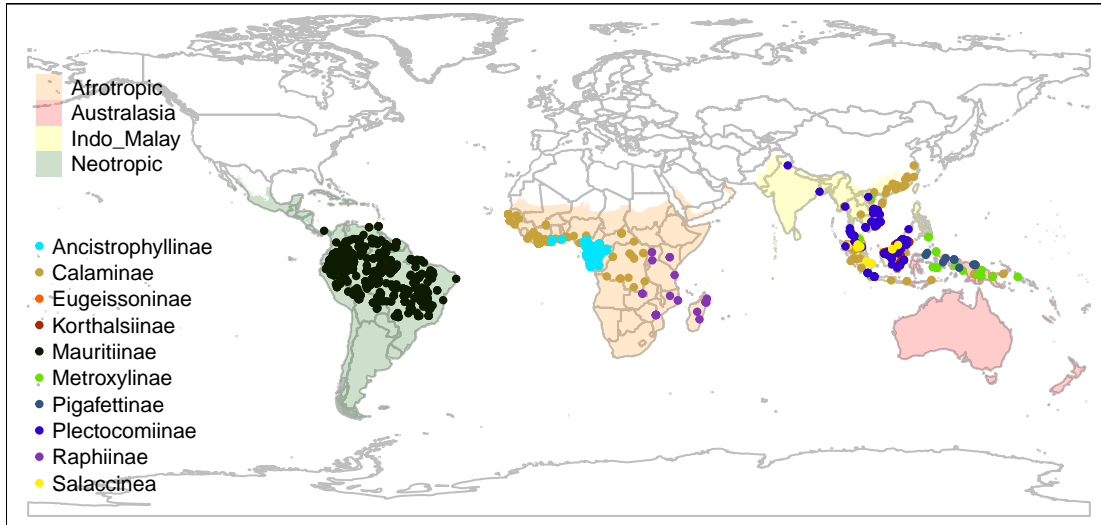
Figure 1: Map showing the occurrence points for the subfamily Calamoideae. The points are colored by subtribe, whose names are in the legend. The four biogeographic regions where the subfamily occurs are also shown.

The next step was determining which of the world's biogeographic regions does each species of Calamoideae palms inhabits. To do so first I imported the data for each species from GBIF (https://www.gbif.org/). From the GBIF data the geographic coordinates and the country of each occurrence was extracted. The ones with coordinates were further cleaned using the package Coordinate Cleaner 2.0-20 (Zizka et al., 2019). The ones without coordinates - only 3 species - had their country of occurrence inspected manually and any occurrence outside the known species range according to the Kew Botanical Gardens database (POW, 2022) were excluded. This is implemented in the script `03_GBIF_data_import_clean.R`. After that, to obtain the geographic regions inhabited by each species, the occurrences were transformed into spatial objects and projected onto a modified version of the WWF ecoregion shapefile that contained the biogeographic regions of the world (Olson et al., 2001).For those species without georeferenced occurrences their countries of occurrence were projected onto the map instead. The biogeographic information was complied in a presence/absence matrix. This is implemented in the scripts `04_spatial_objects.R` and `05_spp_biogeo_obtention.R`.

Finally, the presence/absence matrix and the phylogeny were combined using the DIVALIKE algorithm implemented in the package BioGeoBEARS 1.1.2 (Matzke, 2013). DIVALIKE is a adapted likelihood version of the maximum parsimony DIVA (dispersion and vicariance) of Ronquist (1997). The original DIVA considers speciation only through vicariance, and range change by dispersal and extinction. Here I implement the DIVALIKE model considering these three events and the DIVALIKE+J model, simalar to the DIVALIKE model, but considering also speciation via dispersion (i.e., founder-event speciation). I then compared these two models using AIC criterion. This is implemented in the script `06_ancestral_biogeo_model.R`.

Results and Discussion

Between the two models the one that performed better in terms of AIC values was the DIVALIKE+J one, that takes into account the evolutionary events modeled also by standard DIVALIKE model plus speciation by founder-event (Table 1).

Table 1: Ancestral range inference models statistics. K = number of parameters, LnL = negative log-likelihood, AIC = Akaike Information Criterium

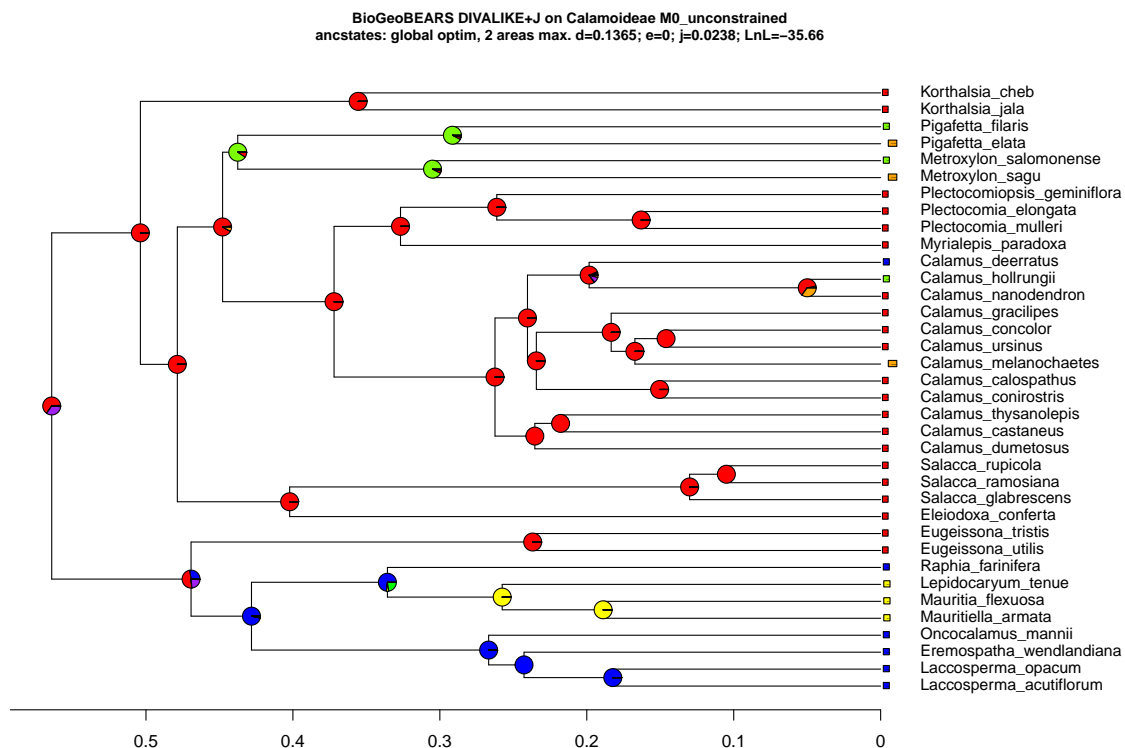|	K	LnL	AIC	delta.AIC
DIVALIKEj	3	-35.66	77.31	0.00
DIVALIKE	2	-38.82	81.63	4.32

Figure 2: Ancestral range inference of Calamoideae palms using the DIVALIKEJ model. Red = Indo-Maly, yellow = Netropic, blue = Afrotropic, green = Indo-Maly + Australasia, light green = Neotropic + Afrotropic, purple = Indo-Malay + Afrotropic.

The DIVALIKE+J model, hereafter just model, inferred the Indo-Malay area as the most likely ancestral area of the Calamoideae subfamily (Figure 2). Most species of Calamoideae are found in this same biogeographic region. From the most ancient node the tree splits into a mainly Indo-Malay and a mainly Afrotropic-Neotropic lineage. The colonization of the Neotropic happend from the Afrotropic region. In the mainly Indo-Malay clade two things are particularly interesting. First, the largest genus *Calamus*, although mainly Indo-Malay, is the only genus to occur in three different biogeographic regions. Moreover, the Afrotropic species *Calamus deerratus* is sister to the Australasian *Calamus hollrungii*, and their ancestor

is reconstructed as Indo-Malay. This interesting biogeographical event deserves more study. Second, the clade *Metroxylon + Pigafetta* is the only one mainly Australasian. Their ancestor is inferred as Indo-Malay, therefore, this group crossed the Wallace's line from Indo-Malay to Australasia and then recolonized the Indo-Malay area.

In conclusion, the Calamoideae subfamily has an interesting biogeographic history that deserves more attention. Future directions are the use of fossil records to date the phylogeny and the refinement of the model by including geological history information.

References

Asmussen, C. B., J. Dransfield, V. Deickmann, A. S. Barfod, J. C. Pintaud, and W. J. Baker. 2006. A new subfamily classification of the palm family (Arecaceae): Evidence from plastid DNA phylogeny. *Botanical Journal of the Linnean Society* 151: 15–38.

Baker, W. J., J. Dransfield, and T. A. Hedderson. 2000a. Phylogeny , Character Evolution , and a New Classification of the Calamoid Palms. *American Society of Plant Taxonomists* 25: 297–322.

Baker, W. J., T. A. Hedderson, and J. Dransfield. 2000b. Molecular phylogenetics of subfamily Calamoideae (Palmae) based on nrDNA ITS and cpDNA rps16 intron sequence data. *Molecular Phylogenetics and Evolution* 14: 195–217.

Dransfield, J., N. W. Uhl, C. B. Asmussen, W. J. Baker, M. M. Harley, and C. E. Lewis. 2008. Genera Palmarum. International Palm Society, Londres.

Joy, J. B., R. H. Liang, R. M. McCloskey, T. Nguyen, and A. F. Y. Poon. 2016. Ancestral Reconstruction. *PLoS Computational Biology* 12: 1–20.

Matzke, M. N. J. 2013. Package 'BioGeoBEARS'.

Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. N. Powell, E. C. Underwood, J. A. D'Amico, et al. 2001. Terrestrial ecoregions of the world: A new map of life on Earth. *BioScience* 51: 933–938.

POW. 2022. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew.

Ronquist, F. 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology* 46: 195–203.

Ter Steege, H., N. C. A. Pitman, D. Sabatier, C. Baraloto, R. P. Salomão, J. E. Guevara, O. L. Phillips, et al. 2013. Hyperdominance in the Amazonian tree flora. *Science* 342.

Zizka, A., D. Silvestro, T. Andermann, J. Azevedo, C. Duarte Ritter, D. Edler, H. Farooq, et al. 2019. CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* 10: 744–751.