

Colonization and diversification history of Madagascan palms with new phylogenomic evidence from the genus *Orania* (Arecaceae)

Maya Schrödl



AARHUS
UNIVERSITY

Colonization and diversification history of Madagascan palms with new phylogenomic evidence from the genus *Orania* (Arecaceae)

60 ECTS Master's Thesis

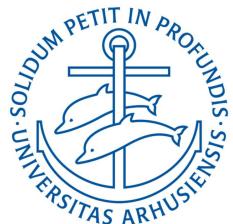
by

Maya Schrödl

Supervisor: Wolf Eiserhardt

International Master of
Biodiversity Ecology and Evolution ([IMABEE](#))

2020



Section for Ecoinformatics and Biodiversity
[Aarhus Universitet](#)
Nordre Ringgade 1
8000 Aarhus C

Front and back cover:
Madagascan palms on the Masoala peninsula ©W. J. Baker

COLONIZATION AND DIVERSIFICATION HISTORY OF MADAGASCAN PALMS WITH NEW PHYLOGENOMIC EVIDENCE FROM THE GENUS *Orania* (ARECACEAE)

Abstract

Biodiversity is unevenly distributed among locations and clades, and also varies over time. A major focus of biodiversity studies is which eco-evolutionary processes shaped and shape these patterns. In order to understand the diversification of a lineage in a specific region, it is crucial to determine when and wherefrom the lineage arrived to this region. Here, the focus was on the colonization times of Madagascan palms. Because of its long isolation from other landmasses, Madagascar builds an ideal model for the study of colonization histories.

After a general introduction on biogeography in a phylogenomic context, two major questions were addressed.

First, it was tested whether the uneven distribution of species richness among Madagascan palms could be explained by the earlier arrival of some lineages to the island, them thereby having had "more time to diversify". The results of our literature review based on available species-level phylogenies suggested that current species richness differences between Madagascan palms cannot be explained by different arrival times alone. Second, the main part of the thesis dealt with the disjunct distribution pattern of the genus *Orania*, which is mainly present in Southeast Asia with only three species endemic to Madagascar. A first dated species-level phylogeny was built using targeted sequencing and through ancestral range estimation. It was inferred that Madagascan *Orania* is monophyletic and arrived to the island via long-distance dispersal from Southeast Asia between the Oligocene and early Miocene.

To conclude, this thesis provides new insight and a phylogenomic key element for understanding colonization and diversification patterns in Madagascan palms.

Key-words: biogeography, disjunct distribution, palms, POS clade, species-level phylogeny

Qui peut être réellement de souche ?

Même les arbres viennent d'une graine amenée là par un oiseau.

Jean-Luc Raharimanana, Malagasy author, 2008

PREFACE

This Master's thesis was prepared during my ten months at the Section for Ecoinformatics and Biodiversity, Aarhus University and is the final work of my International Master of Biodiversity Ecology and Evolution (IMABEE).

This thesis is structured in three parts:

Part A is a general introduction to biodiversity genesis, biogeography, and phylogenomics. This will allow the reader to get familiar with the scientific background knowledge needed for the understanding of the other parts.

Part B consists of an additional project to the main project of the thesis.

Part C is the main part of the thesis. This part is a manuscript in preparation for submission to TAXON. For this part, I prepared the sequencing of samples during three months of labwork. Unfortunately, the COVID-19 pandemic and following lockdown prevented the sequencing of these samples. Hence, here are only represented the results based on sequences which were not sequenced by myself. Once the missing samples will have been sequenced, they will be included in the analyses and the final version of the manuscript.

Aarhus, 12/06/2020

A handwritten signature in blue ink, appearing to read "M. Scholl".

Epigraph quote: Raharimanana, J. L. 2008. Pp. 108-109 in: Collectif. *Il me sera difficile de venir te voir : correspondances littéraires sur les conséquences de la politique française d'immigration*. Paris : Éd. Vent d'ailleurs.

ACKNOWLEDGEMENTS

First of all, I am extremely grateful to Wolf Eiserhardt for being the best supervisor I could have hoped for. I would like to thank him for introducing me to the world of phylogenomics, and guiding and supporting me through the entire project. Many thanks for always staying positive, even in times of technical difficulties and helping me with any question I had (so many). Thanks to Wolf, I learnt a lot during this project in a very positive and welcoming environment.

The whole PEB Team made me feel welcomed from the beginning onwards and I would like to thank each one of the team for contributing to the good mood on our floor and for very helpful discussions. Special thanks to Miao who helped me out many times when I was lost in the complicated world of bioinformatics. I would also especially like to thank Camilla, Peter, and Pirada for introducing me with a lot of patience to all the laboratory techniques which were so new to me.

On a more technical note, I thank Thomas Couvreur and Sidonie Bellot for providing me the sequences without which this project would not have been possible. Thanks also to GenomeDK and Aarhus University for providing computational resources.

Many thanks also to my office mates Emil, Asger, Alexander, Ben, Sune, Mathias, and Sebastian for the many hours that we spent laughing and drinking coffee. Thanks for teaching me some Danish and for all the good jokes.

Thanks should also go to the Section for Ecoinformatics and Biodiversity at Aarhus University for weekly breakfasts, inspiring seminars, and most importantly: the coffee machine. I would like to thank The Social Scientists for all the fun activities that were organized.

I would very much like to thank the IMABEE program to have given me this wonderful opportunity to come to Denmark, particularly Cécile Le Lann, Phillippe Vanderkoornhuyse, and Jesper Sørensen. Thanks should also go to the ERASMUS program for their financial support.

The other IMABEE students Monique, Clémie and Suzon deserve a special thanks for supporting me all the way through my stay and project. These 10 months would not have been the same without you, and I am very grateful to have you as my friends. I would also especially like to thank Colin who cheered me up during the dark Danish winter and made the lockdown enjoyable. Finally, many thanks to Monique, Colin, Suzon, and Ralf for taking the time to proofread my thesis in detail.

I wish all the people involved in this project in one way or another good luck for the future. *Held og lykke.*

TABLE OF CONTENTS

Part A: General Background	9
1 Which processes shaped current biodiversity patterns?	10
2 Historical biogeography: understanding distributions through space and time	12
2.1 Ancestral range estimation with parametric biogeographic models	13
3 Phylogenomics to retrace the past	16
4 Madagascar as an "evolutionary laboratory" for plant biodiversity evolution	18
4.1 Madagascan palms	20
5 Aim of this thesis	20
Bibliography	21
Part B: Testing the 'Time-for-Speciation hypothesis'	26
1 Introduction	27
2 Materials and Methods	28
2.1 Definition of taxonomic palm groups on Madagascar and their Madagascan species richness	28
2.2 Intervals of possible arrival time to Madagascar	30
2.3 Summary of the arrival time intervals	30
2.4 Correlation between species richness and arrival time	31
3 Results	32
4 Discussion	32
5 Conclusion	36
Supplementary Material	37
Bibliography	38

Part C: Phylogeny and biogeography of the palm genus	
<i>Orania</i>	41
1 Introduction	43
2 Materials and Methods	45
2.1 Taxon Sampling	45
2.2 DNA Extraction, Library Preparation, Target Capture, and Se- quencing	45
2.3 Read Trimming and Filtering	46
2.4 Contig Assembly, Gene filtering, Read coverage trimming, and Alignment	48
2.5 Phylogenetic Inference	49
2.6 Molecular Dating and Calibration	49
2.7 Ancestral Range Estimation	51
3 Results	52
3.1 Target capture sequencing, assembly, and phylogenetic analyses	52
3.2 Phylogenetic relationships	52
3.3 Divergence time estimates	54
3.4 Ancestral ranges and colonization histories	56
4 Discussion	58
4.1 Systematic implications	58
4.2 Biogeographic explanations for the recent radiation in SE Asian <i>Orania</i>	60
4.3 Biogeographic pathways of the POS clade and origin of the Madagascan <i>Orania</i> clade	60
5 Conclusion	62
Supplementary Material	63
Bibliography	66

PART A

General Background

1 Which processes shaped current biodiversity patterns?

Biodiversity varies among locations, biological groups, and over time (Wiens, 2011). Humanity is rapidly reshaping these biodiversity patterns and now more than ever, it is crucial to understand how these differences emerged, which is a fundamental question of evolutionary biology. Which mechanisms shaped today's biodiversity has to be understood in order to predict future changes in a changing world.

One might naively think that current biodiversity patterns are mainly driven by recent ecological variables, like climate, soil composition, and biotic factors. However, already Alfred Russel Wallace, the 'Father of biogeography', noticed that "the present state of the earth and of the organisms now inhabiting it, is but the last stage of a long and uninterrupted series of changes which it has undergone" (Wallace, 1855). A key element towards the understanding of current biodiversity patterns is the study of past evolutionary processes: speciation, extinction, dispersal, and time itself (Ricklefs, 1987). The contribution of each of these factors to the number of species of a clade has been strongly discussed and comes down to three major hypotheses.

First, it has been hypothesised that only time determines species richness: the longer a clade has been present in a region, the more time it had to diversify (Stephens & Wiens, 2003). This so-called 'Time-for-Speciation hypothesis' assumes that net diversification rates (speciation minus extinction) are similar between regions and clades and are constant and positive (more speciation than extinction) over time (Figure 1a: I).

The second hypothesis also assumes that diversification rates are constant over time, but that these rates vary between clades (Rabosky, 2009a). Under this 'Rate hypothesis', only variation in diversification rates between clades would explain differences in species richness (Figure 1a: II).

The third hypothesis states that diversification rates are diversity-dependent, *i.e.* decrease with the increase of species richness over time until an equilibrium is reached and species richness remains constant (Rabosky, 2009a). This equilibrium state is called the "ecological limit", *i.e.* carrying capacity, for a clade in a given region. Under this 'Ecological limit hypothesis', higher species richness in different clades would be explained by higher ecological limits (Figure 1a: III).

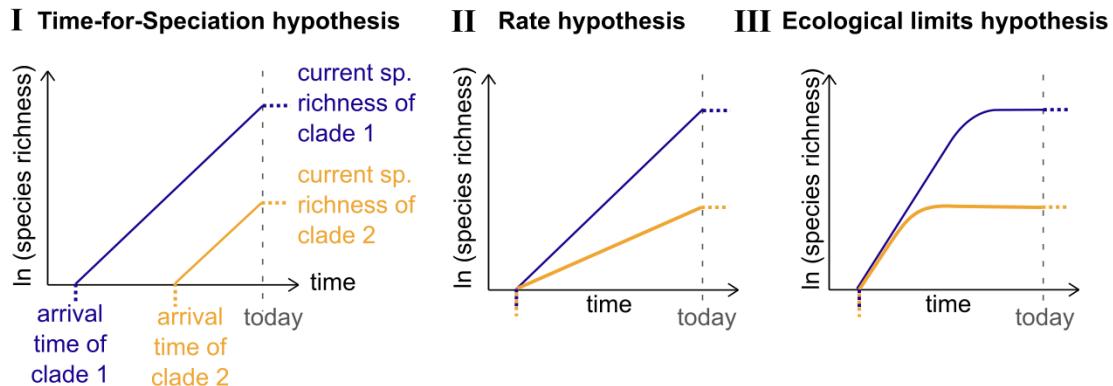
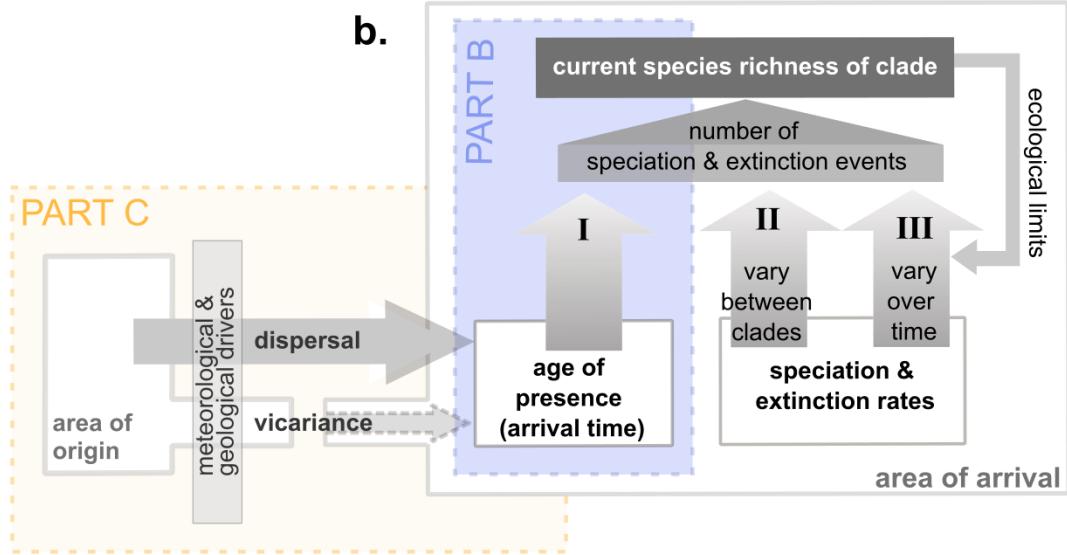
a.**b.**

Figure 1: Theoretical comparison of the three major hypotheses on the genesis of regional biodiversity and their underlying processes. (a) Theoretical evolution of species richness over time within one region under the three hypotheses. For visualization purpose, species richness is given as its natural logarithm. Blue represents one clade ("clade 1") and orange another clade ("clade 2"). The arrival time corresponds to the time the clade has been present in a certain region. For the graphs of hypothesis II and III, both clades arrived at the same time. The dashed horizontal lines always correspond to the current species (sp.) richness of the corresponding clade. These plots are based on the results of Rabosky (2009a). (b) Schema representing the biogeographic and evolutionary factors on which the three hypotheses are based. The current species richness of a clade within a region depends directly on the number of past speciation & extinction events. This number is defined by different factors, depending on the hypothesis. The three hypotheses are represented by the vertical arrows with roman numerals. Under the 'Rate hypothesis' (II) and 'Ecological limits hypothesis' (III), the number of speciation/extinction events depend on the variation of diversification rates, either between respectively different clades or over time. The variation over time (III) is linked to diversity dependent control by ecological limits. Under the 'Time-for-Speciation hypothesis' (I), the age of presence, *i.e.* arrival time, directly determines the number of speciation/extinction events. This hypothesis will be tested in PART B of this thesis (blue box). PART C (orange box) deals with the biogeographic processes preceding this time of arrival, *i.e.* transition from the area of origin to the area of arrival by dispersal or vicariance which are driven by meteorological factors, *e.g.* wind, or geological factors, *e.g.* continental drift. This schema is partly based on Eiserhardt & al. (2017).

Indirect evidence has been found for the 'Time-for-Speciation hypothesis' (references summarised in Li & Wiens, 2019), for the 'Rate hypothesis' (Scholl & Wiens, 2016; Harris & Davies, 2016) and for the 'Ecological limits hypothesis' (Rabosky, 2009b, 2012; Rabosky & Hurlbert, 2015). However, no consensus on these three hypotheses, or a possible combination of them, has yet been reached.

Testing these diversification scenarii within a geographical constraint area, as for example the tropical rain forest biome or regions within this biome, is of particular interest because of their intrinsic ecological and historical characteristics. In this context, it is a crucial point to determine the time of arrival of clades to this area, in order to understand the following regional diversification processes (Figure 1b).

2 Historical biogeography: understanding distributions through space and time

Past diversification dynamics of clades can only be understood under the light of the past distribution patterns of these clades and the mechanisms shaping these patterns. In order to understand when a lineage arrived to a certain area and following diversification processes, it has to be understood where the ancestors of today's species were present, and how and when the transition to the current distribution happened (cf. Figure 1b). This is the main focus of historical biogeography.

The timing and nature of this transition can be especially well studied in clades that present current disjunct distributions, *i.e.* their species occur in very distant places. Why some lineages with similar physiognomy occur in such distant places has already intrigued Alexander von Humboldt (1805). Since then, understanding these disjunct distributions has been one of the major questions in biogeography. For a long time, these distribution patterns were attributed to oceanic dispersal until the validation of the plate tectonics theory (cf. Nelson, 1978). This was followed by the popular explanation that disjunct distributions can rather be explained by vicariant theory (Cracraft, 1988) than by dispersal. Vicariance is the process of speciation induced by the movement of land resulting in the separation of populations through newly formed barriers like mountains or oceans (Croizat & al., 1974). However, with the recent flourishing of the study of phylogenies, vicariance has been excluded in some examples (Yoder & Nowak, 2006) and nowadays, vicariant and dispersalist theories are considered to be not mutually exclusive.

2.1 Ancestral range estimation with parametric biogeographic models

This emergence of large-scale disjunct distribution patterns and the biogeographic history of lineages in general, can be nowadays estimated based on time-scaled phylogenies, current distribution data, and parametric biogeographic models that link the latter. This is called "ancestral range estimation", where the ancestral range of a lineage are the area(s) it was present at a certain moment.

Most of these parametric biogeographic models (from now on simply called *biogeographic models* in the text) include parameters based on two different categories of biogeographic processes: those happening in between speciation events (called anagenetic processes) and those directly linked to speciation (called cladogenetic processes). The anagenetic processes are dispersal (range expansion) and extirpation (range contraction) (cf. Box 1). The cladogenetic processes included in biogeographic models are vicariance, "sympatric speciation", and "jump dispersal" (cf. Box 1, also for an explanation why these terms are put in quotes). Biogeographic models infer which combination of these processes can best explain the current species distribution given a phylogenetic tree.

The most commonly used biogeographic models of range evolution are: the dispersal–extinction–cladogenesis model (DEC; [Ree & Smith, 2008](#)), the dispersal–vicariance model (DIVA; [Ronquist, 1997](#)), and BayArea ([Landis & al., 2013](#)). Maximum Likelihood versions of these models are implemented in the R package BIOGEOBEARS ([Matzke, 2013a](#)) which allows model selection. The main difference between these three likelihood models, are that they consider different scales of vicariance and "sympatric speciation" ([Matzke, 2013a](#)).

For all three models, the anagenetic parameters (dispersal & extirpation) are free parameters that can vary over time, on the branches between nodes in the phylogeny. Cladogenetic parameters (vicariance, "sympatric speciation", and "jump dispersal") are constrained to occur on the nodes, where speciation is supposed to have happened. The "jump dispersal" parameter implemented in BIOGEOBEARS ([Matzke, 2014](#)) has received strong criticism due to conceptual and statistical flaws ([Ree & Sanmartín, 2018](#)) and its use should be avoided in model comparison. These BIOGEOBEARS models do not consider speciation or extinction rates. Other models can implement these rates estimated on phylogenies ([Goldberg & al., 2011](#)), but it has been recently criticised that the information of speciation and extinction rates cannot be extracted from phylogenies ([Louca & Pennell, 2020](#)).

In their simplest version, the BIOGEOBEARS models are only based on phylogenetic parameters, without taking into account the geographic and geological evolution between and within the areas. An additional option in BIOGEOBEARS allows to do this by fixing dispersal probabilities between areas over time. These dispersal probabilities can be estimated based on abiotic parameters (dis)favouring dispersal routes like distance between areas, availability of land bridges, stepping stones, marine currents, winds, and suitable climate conditions for species and their ancestors given conserved niches over time (Sanmartín, 2012; Matzke, 2013a).

Box 1: Biogeographic processes

The historical biogeographic processes that shaped and shape species distribution can be divided into two categories: cladogenetic processes, *i.e.* processes that result in speciation, and anagenetic processes, *i.e.* processes that happen between speciation events.

An example scenario is given to illustrate and explain the different biogeographic processes (Figure 2). (1) Let us assume an island with four distinct biogeographic areas and an omnipresent species A. (2) A movement of land, *i.e.* continental drift, isolates different populations of A which is followed by the emergence of two new species B and C due to allopatric speciation, *i.e.* speciation in geographic isolation. This cladogenetic process is called "vicariance" (Croizat & al., 1974). (3) The range of the species C is contracted to one area. This local extinction is also called "extirpation" to avoid confusion with species extinction. This is an example of an anagenetic process. (4) The species C undergoes a speciation within the same area resulting in species D and E. Despite the repeated use of "sympatric speciation" for this cladogenetic process (*e.g.* Ronquist, 1997; Matzke, 2013a), this speciation within the same region does not necessarily imply sympatric speciation, *i.e.* speciation without restrictions to gene flow (Ree & al., 2005). Therefore, this term was put in quotes. (5) As another anagenetic process, the range of B expands to another island through dispersal. (6) The species E disperses rapidly to a new island which is immediately followed by speciation into F. In contrast to simple dispersal, this "jump dispersal" occurs rapidly over long distance and immediately results in allopatric speciation (Matzke, 2014). What is called "jump dispersal" in Matzke (2013a) is in fact jump dispersal followed immediately by speciation and this term is therefore put in quotes.

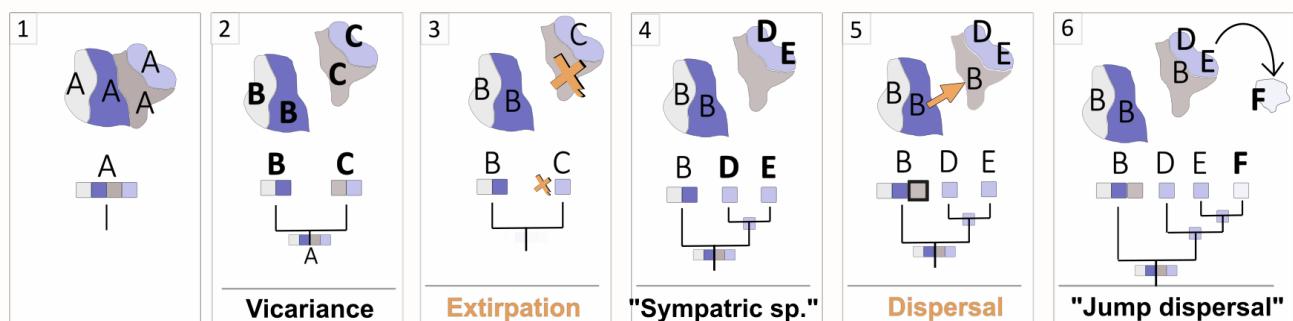


Figure 2: Example of range and species evolution representing the evolutionary processes assumed by different biogeographic models. A temporal sequence of theoretical events in a theoretical environment and their resulting phylogenies with ancestral ranges are depicted. Each colour represents a different area (defined prior to model application) and each letter stands for a different species. Anagenetic processes, *i.e.* processes happening in between speciation events, are represented in orange. Cladogenetic processes, *i.e.* speciation events, are represented in black. New species or ranges within a species are marked in bold. This schema is based on Matzke (2013a). Explanations for why "sympatric speciation" and "jump dispersal" were put in quotes can be found in the text of this box.

Limits and challenges of ancestral range estimations

When estimating the biogeographic history of a clade, one has to keep in mind that these estimations are based on statistics and do not present precise reconstructions. The only direct information about ancestral ranges are fossils (Lieberman, 2003). Fossil records can be implemented in ancestral range estimation (Matzke, 2013b) and should be implemented, whenever fossil records are available for the considered group. However, for many lineages, no fossils are available and in order to get a more direct insight into their biogeographic history, it is crucial to focus on finding new fossils.

A critical step in ancestral range estimation is the definition of the geographical areas on which the estimated ranges are based. Most studies define the size and shape of areas based on specific hypotheses (Ree & Smith, 2008; Ree & Sanmartín, 2009). For studies where no such delimitation can be made, the region of interest can be divided into areas of the same size on a grid (Landis & al., 2013). However, these different approaches can result in very different estimated ancestral ranges (Landis & al., 2013). Choosing the appropriate geographical resolution for a study can be challenging and hence the results of biogeographic estimations should be carefully interpreted under this light.

In connection with this is the assignment of each extant species to various areas, *i.e.* their current distribution, on which the ancestral range estimations rely extensively. However, the known distribution of a species is always only an approximation of the true distribution. This so-called "imperfect detection" could bias the ancestral range estimations when a species also occurs in a different area than the ones where data is available. In ecology on a population level, imperfect detection can be modelled on this small-scale (Warton & al., 2017). However, biogeographic analyses often focus on very large-scale areas like continents (Matzke, 2013b) which makes it difficult to model the scope of imperfect detection. Hence, an increase in sampling effort is crucial to approach the true distribution of extant species.

Lastly, the backbone of ancestral range estimations are phylogenies. Their topology and timing are assumed to be correct. However, phylogenies are only statistical estimates of the evolutionary history and can present great uncertainties in their topology (low node support) or timing (large credibility intervals of node ages). In order to estimate the biogeographic history of a lineage, it is crucial to build dated phylogenies which approach the true evolutionary history of a clade as much as possible.

3 Phylogenomics to retrace the past

A key element for reconstructing dispersal histories are dated phylogenies. Phylogenies represent the evolutionary relationships between different species, populations, or genes.

Already Charles Darwin tried to establish these evolutionary relationships based on the observation of morphological character differences between species (Darwin, 1859). These morphological differences have long been the sole source of data in phylogenetics (Wanninger, 2015). With the development of new laboratory techniques in the 1960s and 1970s, like precursors to the polymerase chain reaction (PCR) (*e.g.* Kleppe & al., 1971), more sensitive electrophoresis methods (Loening, 1967), and DNA sequencing (Sanger & al., 1977), a new additional source of data became available: molecular data. This has been the base for phylogenetic analyses ever since. The discipline of molecular phylogenetics relies on the theory that the reproductive isolation of two populations, followed by random mutational processes, results in an accumulation of different nucleotides (Rowe & al., 2017). The shared common ancestor between two diverged genomes, could therefore be determined through the amount of different nucleotides between these genomes (Rowe & al., 2017).

However, until recently these analyses were only based on few genes and the evolutionary relationships of one gene does not necessarily reflect the evolutionary relationships of the species. This can be explained by the following mechanisms (Degnan & Rosenberg, 2009):

- (1) incomplete lineage sorting, *i.e.* copies of the same gene in different species fail to coalesce at the time speciation happened;
- (2) horizontal gene transfer (HGT), *i.e.* a gene "jumped" from the genome of one lineage to a distant one;
- (3) hybridization: similar to HGT, genes are transferred from one ancestral species to another, but hybridization affects whole genomes and not only small segments like HGT;
- (4) gene duplication and loss, *i.e.* when a gene is duplicated and a copy is lost in a species, the species which haven't lost this copy, can appear more closely related.

Therefore, it is crucial to include as many genes as possible when trying to reconstruct the evolutionary history between species. However, it is important to focus on low-copy nuclear genes to get a maximum amount of phylogenetic informativeness whilst minimising sequencing effort (cf. *e.g.* Sang, 2002).

This has been made possible in recent years with the development of the target capture method (Mamanova & al., 2010; Weitemier & al., 2014), followed by high throughput sequencing, allowing the sequencing of thousands of genes at the same time (Johnson & al., 2019). For a specific large group of organisms, so-called "baits" (*e.g.* Angiosperms353 for flowering plants; Johnson & al., 2019) are designed to target low-copy regions of the genomes and can be used to "capture" these regions and only sequence them. This technique is rapidly becoming the golden standard for reconstructing phylogenies, especially in plant biology (Couvreur & al., 2019).

In order to not only represent the evolutionary relationships of species, but also place these divergences on a timeline, these reconstructed phylogenies can be dated (Drummond & Rambaut, 2007). The framework for this dating is built by the molecular clock hypothesis, which states that the rate at which nucleotides within a genome change, is relatively constant over time and between organisms (Kimura, 1968). Hence, the amount of genetic differences between two genomes would be directly correlated to the time these genomes diverged (Kumar, 2005). This "strict" molecular clock has been considered to be too simplistic and in the last two decades methods have been developed to "relax" the molecular clock in a Bayesian framework by allowing minimal variations in evolutionary rates between genomes (Kumar, 2005).

The molecular clock can only estimate relative timings of divergence events, but not assign absolute dates (Kumar, 2005). Therefore, the time-scaled tree needs to be calibrated by adding prior knowledge of dates. This can consist of dated fossils of common ancestors (*e.g.* Marshall, 2008), or dates taken from former phylogenies, also known as "secondary calibration" (*e.g.* Schenk, 2016). These calibrations result in a phylogenetic tree where each species split has an assigned date.

Despite all these new techniques for phylogeny reconstruction, the number of available phylogenies on a species level is far from complete. Constructing missing species-level phylogenies will not only bring research closer to the ultimate goal of constructing a complete species-level tree of life, but also provide a framework for understanding evolutionary processes on a species scale.

4 Madagascar as an "evolutionary laboratory" for plant biodiversity evolution

A major focus of biodiversity science is the tropical rain forest biome, because of its extraordinary high species richness. This biome contains most of the regions classified as "biodiversity hotspots": regions, that present high levels of endemism and are therefore the most threatened (Myers & al., 2000). Hence, unravelling the genesis of this hyperdiversity and hyperendemism within the biodiversity hotspots is a major challenge of biogeography.

In reason of this sheer diversity, it is important to focus on specific areas in biodiversity studies. The island of Madagascar is one of the biodiversity hotspots (Myers & al., 2000) and presents an excellent model for the study of biogeographic and evolutionary processes because of its long isolation from other landmasses (Vences & al., 2009). The island was once part of the supercontinent Gondwana and after its split-up, Madagascar has been isolated since the late Cretaceous (cf. box 2). Due to this long isolation, there has been a limited number of colonization events, and most colonizations have led to some degree of *in situ* diversification (Vences & al., 2009; Couvreur & al., 2011). This builds a good framework for testing whether the amount of *in situ* diversification following an event of arrival depends on the timing of that event (*i.e.* Time-for-Speciation hypothesis).

Madagascar also builds a good framework for testing the biogeographic processes underlying this event of arrival. Because of its peculiar geological history, it can be well distinguished when vicariance or dispersal were the dominant processes in shaping ancestral and current distributions of species (Yoder & Nowak, 2006; cf. Box 2).

Box 2: Geological history of Madagascar

Madagascar has once been part of the supercontinent Gondwana. At the beginning of the Cretaceous (166 Mya) the supercontinent began to split up in two parts: West Gondwana (South America and Africa) and East Gondwana (Madagascar, India, Antarctica, and Australia) (1. in Figure 3). During the split up, dispersal between Africa and Madagascar was possible. In the mid-Cretaceous (116 Mya - 83.5 Mya), India split from Madagascar (2. in Figure 3). Dispersal between Africa/Madagascar/India was possible. Subsequently, India drifted towards SE Asia until the first contact with SE Asia in the beginning of the Eocene (57 Mya) (3. in Figure 3). Dispersal between Africa/Madagascar and India became less and less likely. India definitely collided with SE Asia and reached its current position around 35 Mya (4. in Figure 3). The timing of these geological events and of dispersal routes were respectively taken from [Ali & Aitchison \(2008\)](#) and [Buerki & al. \(2013\)](#)

Current large-scale distribution patterns between Madagascan species and sisters of other continents, like Africa or Asia, can be either explained by vicariance or dispersal ([Yoder & Nowak, 2006](#)). When looking at divergence times on dated phylogenies, current distribution patterns can be explained by vicariance only until the break-up between India and Madagascar. Post-Gondwanan (< 86 Mya) divergence times indicate that dispersal was the predominant force shaping ancestral and current distribution patterns.

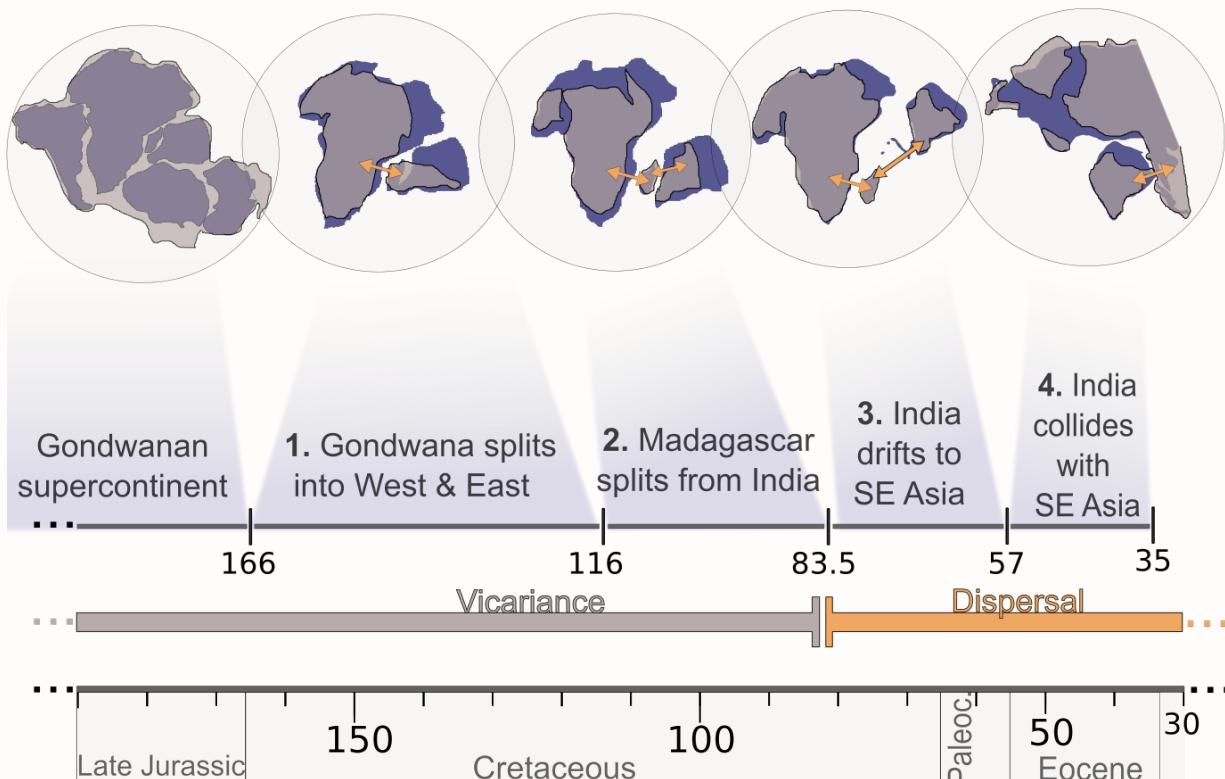


Figure 3: Timeline of the geological evolution of Madagascar and India from the Late Jurassic to the end of the Eocene, and related biogeographic mechanisms. All dates are given in Million years ago (Mya). Paleocene is abbreviated as Paleoc. The shapes of current continents and islands are represented in dark blue. The shapes of the continents and islands at the corresponding time are represented in dark grey. Only parts of the world are shown and only the relative position of continents/islands are represented, but not their absolute position on the world. The shapes, position of the continents/islands to each other, and the relative timing of events is taken from [Ali & Aitchison \(2008\)](#). Possible biotic dispersal routes are represented with orange arrows and are taken from [Buerki & al. \(2013\)](#). "Greater India" (India + a northern extension) is referred to as "India". Southeast Asia is abbreviated as SE Asia. The lower part shows until when Gondwanan vicariance was the predominant mechanism underlying the biogeographic history of Madagascan lineages and when dispersal got more important. This information was taken from [Yoder & Nowak \(2006\)](#).

4.1 Madagascan palms

To study biodiversity patterns in the tropical rain forests of Madagascar, palms are an especially good model. First, palms are ecologically representative of tropical rain forests since most palm species are present in this biome (Couvreur & Baker, 2013). Secondly, basic biodiversity data, like the taxonomy of palm species and occurrence data is very well known (Couvreur & Baker, 2013). Thirdly, palms in Madagascar are highly endemic, with only three non-endemics out of the 208 currently accepted native Madagascan palm species (William J. Baker, pers. comm., 2020).

Most species-level phylogenies of Madagascan palms have been inferred, at least to some extent. Moreover, Madagascan palms present large differences in species richness between clades. These phylogenies and different species richnesses, in combination with Madagascar's long isolation (see above), build an excellent model for testing the Time-for-Speciation hypothesis.

5 Aim of this thesis

This thesis aimed to provide an insight into the past eco-evolutionary mechanisms that might have shaped today's biodiversity patterns in Madagascan palms. Two different scales of biodiversity and biogeography were addressed: the differences in diversity between clades within a region, and secondly between regions within a clade.

The first question (part B) addressed whether the differences in diversity between Madagascan palm clades can be explained by the differences in how long these clades have been present in Madagascar.

The second study (part C) is the core part of the thesis. Its focus was on the disjunct distribution between Madagascar, Southeast Asia and Africa within a palm clade. The arrival time and origin was estimated and this study provided a phylogenomic key element of the biogeographic evolution of Madagascan palms.

Bibliography

- Ali, J.R. & Aitchison, J.C.** 2008. Gondwana to Asia: Plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166–35 Ma). *Earth-Science Reviews*, 88(3-4): 145–166. [doi:10.1016/j.earscirev.2008.01.007](https://doi.org/10.1016/j.earscirev.2008.01.007).
- Buerki, S., Devey, D.S., Callmander, M.W., Phillipson, P.B., & Forest, F.** 2013. Spatio-temporal history of the endemic genera of Madagascar: Evolution of the Endemic Madagascan Flora. *Botanical Journal of the Linnean Society*, 171(2): 304–329. [doi:10.1111/boj.12008](https://doi.org/10.1111/boj.12008).
- Couvreur, T.L. & Baker, W.J.** 2013. Tropical rain forest evolution: Palms as a model group. *BMC Biology*, 11(1): 48. [doi:10.1186/1741-7007-11-48](https://doi.org/10.1186/1741-7007-11-48).
- Couvreur, T.L., 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 Biology*, 9(1): 44. [doi:10.1186/1741-7007-9-44](https://doi.org/10.1186/1741-7007-9-44).
- Couvreur, T.L.P., Helmstetter, A.J., Koenen, E.J.M., Bethune, K., Brandão, R.D., Little, S.A., Sauquet, H., & Erkens, R.H.J.** 2019. Phylogenomics of the Major Tropical Plant Family Annonaceae Using Targeted Enrichment of Nuclear Genes. *Frontiers in Plant Science*, 9. [doi:10.3389/fpls.2018.01941](https://doi.org/10.3389/fpls.2018.01941).
- Cracraft, J.** 1988. Vicariance Biogeography: Theory, Methods, and Applications. *Systematic Biology*, 37(3): 219–220. [doi:10.1093/sysbio/37.3.219](https://doi.org/10.1093/sysbio/37.3.219).
- Croizat, L., Nelson, G., & Rosen, D.E.** 1974. Centers of Origin and Related Concepts. *Systematic Biology*, 23(2): 265–287. [doi:10.1093/sysbio/23.2.265](https://doi.org/10.1093/sysbio/23.2.265).
- Darwin, C.** 1859. *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*. London: John Murray.
- Degnan, J.H. & Rosenberg, N.A.** 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, 24(6): 332–340. [doi:10.1016/j.tree.2009.01.009](https://doi.org/10.1016/j.tree.2009.01.009).
- Drummond, A.J. & Rambaut, A.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7(1): 214. [doi:10.1186/1471-2148-7-214](https://doi.org/10.1186/1471-2148-7-214).
- 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 Phytologist*, 214(4): 1408–1422. [doi:10.1111/nph.14516](https://doi.org/10.1111/nph.14516).
- Goldberg, E.E., Lancaster, L.T., & Ree, R.H.** 2011. Phylogenetic Inference of Reciprocal Effects between Geographic Range Evolution and Diversification. *Systematic Biology*, 60(4): 451–465. [doi:10.1093/sysbio/syr046](https://doi.org/10.1093/sysbio/syr046).

- Harris, L.W. & Davies, T.J.** 2016. A Complete Fossil-Calibrated Phylogeny of Seed Plant Families as a Tool for Comparative Analyses: Testing the "Time for Speciation" Hypothesis. *PLOS ONE*, 11(10): e0162907. doi:[10.1371/journal.pone.0162907](https://doi.org/10.1371/journal.pone.0162907).
- 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. doi:[10.1093/sysbio/syy086](https://doi.org/10.1093/sysbio/syy086).
- Kimura, M.** 1968. Evolutionary Rate at the Molecular Level. *Nature*, 217(5129): 624--626. doi:[10.1038/217624a0](https://doi.org/10.1038/217624a0).
- Kleppe, K., Ohtsuka, E., Kleppe, R., Molineux, I., & Khorana, H.G.** 1971. Studies on polynucleotides: XCVI. Repair replication of short synthetic DNA's as catalyzed by DNA polymerases. *Journal of Molecular Biology*, 56(2): 341--361. doi:[10.1016/0022-2836\(71\)90469-4](https://doi.org/10.1016/0022-2836(71)90469-4).
- Kumar, S.** 2005. Molecular clocks: Four decades of evolution. *Nature Reviews Genetics*, 6(8): 654--662. doi:[10.1038/nrg1659](https://doi.org/10.1038/nrg1659).
- Landis, M.J., Matzke, N.J., Moore, B.R., & Huelsenbeck, J.P.** 2013. Bayesian Analysis of Biogeography when the Number of Areas is Large. *Systematic Biology*, 62(6): 789--804. doi:[10.1093/sysbio/syt040](https://doi.org/10.1093/sysbio/syt040).
- Li, H. & Wiens, J.J.** 2019. Time Explains Regional Richness Patterns within Clades More Often than Diversification Rates or Area. *The American Naturalist*, 193(4): 514--529. doi:[10.1086/702253](https://doi.org/10.1086/702253).
- Lieberman, B.S.** 2003. Paleobiogeography: The Relevance of Fossils to Biogeography. *Annual Review of Ecology, Evolution, and Systematics*, 34(1): 51--69. doi:[10.1146/annurev.ecolsys.34.121101.153549](https://doi.org/10.1146/annurev.ecolsys.34.121101.153549).
- Loening, U.E.** 1967. The fractionation of high-molecular-weight ribonucleic acid by polyacrylamide-gel electrophoresis. *The Biochemical Journal*, 102(1): 251--257. doi:[10.1042/bj1020251](https://doi.org/10.1042/bj1020251).
- Louca, S. & Pennell, M.W.** 2020. Extant timetrees are consistent with a myriad of diversification histories. *Nature*, 580(7804): 502--505. doi:[10.1038/s41586-020-2176-1](https://doi.org/10.1038/s41586-020-2176-1).
- Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., & Turner, D.J.** 2010. Target-enrichment strategies for next-generation sequencing. *Nature Methods*, 7(2): 111--118. doi:[10.1038/nmeth.1419](https://doi.org/10.1038/nmeth.1419).

- Marshall, C.R.** 2008. A Simple Method for Bracketing Absolute Divergence Times on Molecular Phylogenies Using Multiple Fossil Calibration Points. *The American Naturalist*, 171(6): 726--742. doi:[10.1086/587523](https://doi.org/10.1086/587523).
- Matzke, N.J.** 2013a. Probabilistic historical biogeography: New models for founder- event speciation, imperfect detection, and fossils allow im- proved accuracy and model-testing. *Frontiers of Biogeography*, (5): 242--248. doi:[10.21425/F5FBG19694](https://doi.org/10.21425/F5FBG19694).
- Matzke, N.J.** 2013b. *Probabilistic Historical Biogeography: New Models for Founder-Event Speciation, Imperfect Detection, and Fossils Allow Improved Accuracy and Model-Testing*. Ph.D. thesis, Department Integrative Biology and Designated Emphasis in Computational and Genomic Biology, University of California, Berkeley.
- Matzke, N.J.** 2014. Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades. *Systematic Biology*, 63(6): 951--970. doi:[10.1093/sysbio/syu056](https://doi.org/10.1093/sysbio/syu056).
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., & Kent, J.** 2000. Biodiversity hotspots for conservation priorities. *Nature*, 403(6772): 853--858. doi:[10.1038/35002501](https://doi.org/10.1038/35002501).
- Nelson, G.** 1978. From Candolle to Croizat: Comments on the history of biogeography. *Journal of the History of Biology*, 11(2): 269--305. doi:[10.1007/BF00389302](https://doi.org/10.1007/BF00389302).
- Rabosky, D.L.** 2009a. Ecological limits and diversification rate: Alternative paradigms to explain the variation in species richness among clades and regions. *Ecology Letters*, 12(8): 735--743. doi:[10.1111/j.1461-0248.2009.01333.x](https://doi.org/10.1111/j.1461-0248.2009.01333.x).
- Rabosky, D.L.** 2009b. Ecological Limits on Clade Diversification in Higher Taxa. *The American Naturalist*, 173(5): 662--674. doi:[10.1086/597378](https://doi.org/10.1086/597378).
- Rabosky, D.L.** 2012. Testing the time-for-speciation effect in the assembly of regional biotas. *Methods in Ecology and Evolution*, 3(2): 224--233. doi:[10.1111/j.2041-210X.2011.00166.x](https://doi.org/10.1111/j.2041-210X.2011.00166.x).
- Rabosky, D.L. & Hurlbert, A.H.** 2015. Species Richness at Continental Scales Is Dominated by Ecological Limits. *The American Naturalist*, 185(5): 572--583. doi:[10.1086/680850](https://doi.org/10.1086/680850).
- Ree, R.H., Moore, B.R., Webb, C.O., & Donoghue, M.J.** 2005. A Like- lihood Framework for Inferring the Evolution of Geographic Range on Phyloge- netic Trees. *Evolution*, 59(11): 2299--2311. doi:[10.1111/j.0014-3820.2005.tb00940.x](https://doi.org/10.1111/j.0014-3820.2005.tb00940.x).
- Ree, R.H. & Sanmartín, I.** 2009. Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography*, 36(7): 1211--1220. doi:[10.1111/j.1365-2699.2008.02068.x](https://doi.org/10.1111/j.1365-2699.2008.02068.x).

- 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. doi:10.1111/jbi.13173.
- Ree, R.H. & Smith, S.A.** 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, 57(1): 4--14. doi:10.1080/10635150701883881.
- Ricklefs, R.E.** 1987. Community diversity: Relative roles of local and regional processes. *Science (New York, N.Y.)*, 235(4785): 167--171. doi:10.1126/science.235.4785.167.
- Ronquist, F.** 1997. Dispersal-Vicariance Analysis: A New Approach to the Quantification of Historical Biogeography. *Systematic Biology*, 46: 9.
- Rowe, G., Sweet, M., & Beebee, T.J.C.** 2017. *An Introduction to Molecular Ecology*. Oxford University Press.
- Sang, T.** 2002. Utility of Low-Copy Nuclear Gene Sequences in Plant Phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology*, 37(3): 121--147. doi:10.1080/10409230290771474.
- Sanger, F., Nicklen, S., & Coulson, A.R.** 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, 74(12): 5463--5467.
- Sanmartín, I.** 2012. Historical Biogeography: Evolution in Time and Space. *Evolution: Education and Outreach*, 5(4): 555--568. doi:10.1007/s12052-012-0421-2.
- Schenk, J.J.** 2016. Consequences of Secondary Calibrations on Divergence Time Estimates. *PLoS ONE*, 11(1). doi:10.1371/journal.pone.0148228.
- Scholl, J.P. & Wiens, J.J.** 2016. Diversification rates and species richness across the Tree of Life. *Proceedings of the Royal Society B: Biological Sciences*, 283(1838). doi:10.1098/rspb.2016.1334.
- Stephens, P.R. & Wiens, J.J.** 2003. Explaining Species Richness from Continents to Communities: The Time-for-Speciation Effect in Emydid Turtles. *The American Naturalist*, 161(1): 112--128. doi:10.1086/345091.
- Vences, M., Wollenberg, K.C., Vieites, D.R., & Lees, D.C.** 2009. Madagascar as a model region of species diversification. *Trends in Ecology & Evolution*, 24(8): 456--465. doi:10.1016/j.tree.2009.03.011.
- von Humboldt, A. & Bonpland, A.** 1805. *Essai Sur La Géographie Des Plantes: Accompagné d'un Tableau Physique Des Régions Équinoxiales, Fondé Sur Des Mesures Exécutées, Depuis Le Dixième Degré de Latitude Boréale Jusqu'au Dixième Degré de Latitude Australe, Pendant Les Années 1799, 1800, 1801, 1802 et*

- 1803 /Par Al. de Humboldt et A. Bonpland ; Rédigée Par Al. de Humboldt. Paris: Chez Levrault, Schoell et compagnie, libraires. doi:[10.5962/bhl.title.9309](https://doi.org/10.5962/bhl.title.9309).
- Wallace, A.R.** 1855. On the law which has regulated the introduction of new species. *Annals and Magazine of Natural History*, 16(2): 184--196.
- Wanninger, A.** 2015. Morphology is dead – long live morphology! Integrating MorphoEvoDevo into molecular EvoDevo and phylogenomics. *Frontiers in Ecology and Evolution*, 3. doi:[10.3389/fevo.2015.00054](https://doi.org/10.3389/fevo.2015.00054).
- Warton, D.I., Stoklosa, J., Guillera-Arroita, G., MacKenzie, D.I., & Welsh, A.H.** 2017. Graphical diagnostics for occupancy models with imperfect detection. *Methods in Ecology and Evolution*, 8(4): 408–419. doi:[10.1111/2041-210X.12761](https://doi.org/10.1111/2041-210X.12761).
- 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 phylogenomics1. *Applications in Plant Sciences*, 2(9). doi:[10.3732/apps.1400042](https://doi.org/10.3732/apps.1400042).
- Wiens, J.J.** 2011. The Causes Of Species Richness Patterns Across Space, Time, And Clades And The Role Of "Ecological Limits". *The Quarterly Review of Biology*, 86(2): 75--96. doi:[10.1086/659883](https://doi.org/10.1086/659883).
- Yoder, A.D. & Nowak, M.D.** 2006. Has Vicariance or Dispersal Been the Predominant Biogeographic Force in Madagascar? Only Time Will Tell. *Annual Review of Ecology, Evolution, and Systematics*, 37(1): 405--431. doi:[10.1146/annurev.ecolsys.37.091305.110239](https://doi.org/10.1146/annurev.ecolsys.37.091305.110239).

PART B

Can the differences in species richness in Madagascan palm lineages be explained by the 'Time-for-Speciation hypothesis'?

1 Introduction

Species richness is unevenly distributed among the different groups of the tree of life (Macarthur, 1965). Explaining which evolutionary mechanisms drive this variation is a major goal in evolutionary biology. It is now widely accepted that the processes that shape diversity are speciation, extinction, dispersal, and time (Ricklefs, 1987). However, the interplay and relative importance of each factor are still under discussion. Within one region, different net diversification rates (speciation minus extinction) between clades or over time (Scholl & Wiens, 2016; Rabosky, 2009), but also different arrival times to a certain area could have determined today's differences in species richness. The so-called 'Time-for-Speciation hypothesis' (Stephens & Wiens, 2003) claims that the longer a clade has been present in a certain region, the more time it will have had to diversify and hence have a higher species richness today. This can only be hypothesised assuming constant and positive diversification rates over time and related to this, diversity-independent processes (Rabosky, 2012).

The Time-for-Speciation hypothesis has been tested in a great variety of groups, and has been either rejected (Rabosky & al., 2012; Scholl & Wiens, 2016) or accepted (for a summary of sources see Li & Wiens, 2019). Nevertheless, most of these studies do not focus on a specific region. Diversification rates of different groups of one lineage are expected to be more similar within one region, because of more similar ecological variables (Li & Wiens, 2019). Therefore, focusing on a specific region might reduce the noise of varying diversification rates, and approach an ideal context for testing of the Time-for-Speciation hypothesis.

The island of Madagascar is a good model region for this testing. First, the time of arrival of each group can be defined rather well, because of its long isolation from other continents, and hence its clear definition as a region (Vences & al., 2009). Second, most of these first arrivals have led to some degree of *in situ* diversification (Vences & al., 2009; Couvreur & al., 2011).

Madagascan palm groups present a great variation in current species richness (Dransfield & Beentje, 1995). For example the subtribe Dypsidinae contains 172 Madagascar-endemic species, whereas the genus *Orania* only has 3 endemic species (WCSP). Moreover, palms might present a constant diversification rate for most of their history (Couvreur & al., 2011). For these two reasons, Madagascan palms present a good model to test the Time-for-Speciation hypothesis on a regional scale.

In this study, I conducted a literature search of time-calibrated species-level phy-

logenies in order to estimate the possible arrival time to Madagascar of each Madagascan palm group. It was tested whether a positive correlation between species richness and arrival time can be detected in Madagascan palms, as expected under the Time-for-Speciation hypothesis. I used a "clade-based scenario" as in [Rabosky \(2012\)](#), meaning that each group of lineages corresponds to one colonization event.

2 Materials and Methods

2.1 Definition of taxonomic palm groups on Madagascar and their Madagascan species richness

In this study, I only considered the native Madagascan palm species which are organised into 17 genera ([Dransfield & Beentje, 1995](#); for *Tahina*: [Dransfield & al., 2008](#)).

These genera and species were consequently grouped into 8 groups of lineages (or 9, depending on colonization scenario for *Borassus*). This definition of groups was based on dated molecular species-level palm phylogenies and ancestral range estimations ([Bellot & al., *in press*](#); [Baker & Couvreur, 2013](#); [Trénel & al., 2007](#); [Meerow & al., 2015](#)). Thus, there was one group for each Madagascan colonization event as in the clade-based scenario of [Rabosky \(2012\)](#).

For each group, the number of Madagascan species and their endemicity were based on the World Checklist of Selected Plant Families ([WCSP](#)).

One group consisted of the clade formed by the two monotypic endemic genera *Bismarckia* and *Satranala*, which arrived once to Madagascar ([Bellot & al., *in press*](#); [Baker & Couvreur, 2013](#)).

In the closely related genus *Hyphaene*, the non-endemic species *Hyphaene coriacea* formed another group ([Bellot & al., *in press*](#)). The endemic monotypic genus *Tahina* formed another.

The case of the genus *Borassus* is more complex. Two species are native to Madagascar, an endemic one, *B. madagascariensis* and a non-endemic one, *B. aethiopum*. The former and the Madagascan ecotype of the latter do not form a monophyletic group on their own, but they are monophyletic with the African *B. aethiopum* ecotypes and the African species *B. akeassii* ([Bellot & al., *in press*](#)). [Bellot & al. \(*in press*\)](#) could not determine the direction of the colonization (Africa - Madagascar or *vice versa*), resulting in two possible Madagascan colonization scenarios. The first scenario (here: *B1*) consists of one colonization event to Madagascar

and then colonization to Africa. The second scenario (here: *B2*) consists of first, one colonization event to Africa, and subsequently two colonization events to Madagascar. For the definition of groups regarding *Borassus*, I considered both scenarii. In scenario *B1*, one group was defined by *B. madagascariensis*, *B. aethiopium*, and *B. akeassii*, which I will call the Madagascar/Africa clade. The non-Madagascan species of this clade (*i.e.* *B. akeassii*) was not included when calculating species richness for this group. In scenario *B2*, two groups were defined: first, *B. madagascariensis*, and second, the Madagascan ecotype of *B. aethiopium*.

The widespread genus *Phoenix* contains only one Madagascan species, the non-endemic *Phoenix reclinata*. In reason of the lack of a species-level phylogeny for this genus, I excluded it from this study.

The genus *Ravenea* formed one group. It is a monophyletic group that is almost completely endemic to Madagascar, except two Comoros endemic species (Trénel & al., 2007). Trénel & al. (2007) suggested a single colonization event to Madagascar.

The relationship between the monotypic endemic genus *Voanioala* and the endemic genus *Beccariophoenix* is still debated. Either they form a clade with the South-African genus *Jubaeopsis* (Meerow & al., 2015), or not (Baker & Couvreur, 2013). Because a new genus-level palm phylogeny (Sidonie Bellot, pers. comm., March 2020) confirmed the clade *Jubaeopsis* - *Voanioala* - *Beccariophoenix*, I considered *Voanioala* and *Beccariophoenix* as one single group, with one colonization event to Madagascar, as proposed by Meerow & al. (2015). The calculation of species richness of this clade excluded the non-Madagascan monotypic *Jubaeopsis* genus.

The subtribe Dypsidinae consists of 4 genera (*Dypsis*, *Lemurophoenix*, *Marojejya*, and *Masoala*), and is near-endemic to Madagascar (Dransfield & Uhl, 2008). Its monophyly is controversial (Baker & al., 2009; Baker & Couvreur, 2013), but a new genus-level palm phylogeny (Sidonie Bellot, pers. comm., March 2020) confirms its monophyly. Therefore, I considered the Dypsidinae as one group.

The genus *Orania* is mainly present in South-East-Asia and contains three Madagascar endemic species (Dransfield & Uhl, 2008). In reason of the lack of a species-level phylogeny for this genus, I excluded it from this study.

2.2 Intervals of possible arrival time to Madagascar

For each group, an interval of the possible arrival time to Madagascar was estimated through a literature review of dated molecular species-level phylogenetic trees of these groups and ancestral range estimations. For each group, the dates were based on phylogenie(s) that go until the group's level. For the detailed dates and references for each group, see Table 2.

For the monotypic groups (*Hyphaene coriacea*, *Tahina*, *Borassus madagascariensis*, Madagascan *Borassus aethiopum*), the earliest possible arrival time was defined as the split between the considered group and its sister group, and the latest time defined as today for each phylogeny (cf. Table 2).

For the group *Bismarckia* & *Satranala*, the *Borassus* Madagascar/Africa group, *Ravenea*, the group *Voanioala* & Beccariophoenix, and Dypsidinae, the earliest possible arrival time was considered to be the stem age of the group, and the latest to be the crown age (cf. Table 2).

For *Ravenea*, Trénel & al. (2007) proposed various divergence time estimates, resulting of different models, and I chose the largest possible interval between any of the suggested stem and crown ages.

For the Dypsidinae group, I took the Bayesian posterior distribution of phylogenies created by Faurby & al. (2016), constructed a maximum clade credibility tree with confidence intervals using TreeAnnotator, a tool integrated in BEAST2 v2.6.1 (Bouckaert & al., 2019). In the resulting tree, Dypsidinae was monophyletic and the arrival time interval could be deducted.

2.3 Summary of the arrival time intervals

In order to summarise these dates, I first calculated the absolute interval for each group over all sources, taking as the absolute earliest possible arrival time the highest value over all sources including the 95% HPD (highest posterior density) interval limits, and as the absolute latest possible arrival time the lowest value (Approach A).

As a second approach, I did likewise but for selected sources (Approach S). For the coryphoids *Bismarckia* & *Satranala*, and *Tahina*, Bellot & al. (*in press.*) added two new fossil calibration points (Manchester & al., 2010; Matsunaga & al., 2019) to the coryphoid fossils of Baker & Couvreur (2013). I therefore chose to only consider

the dates from Bellot & al. (*in press.*) for the groups *Bismarckia* & *Satrana*, and *Tahina* in this second approach.

For each approach, two different scenarii were possible, depending on the colonization path of *Borassus*: one single colonization event to Madagascar (*B1*), and therefore one group (*Borassus* Madagascar/Africa clade) in this study; or two independent colonization events (*B2*), and therefore two groups for *Borassus* (*Borassus madagascariensis* and Madagascan *Borassus aethiopum*).

In total, there were four different summarised datasets (Table 1).

Table 1: Description of the four different datasets of arrival times according to which sources and which groups for *Borassus* were used. Selected sources mean that only the dates of Bellot & al. (*in press.*) were used for the groups where Bellot & al. (*in press.*) and Baker & Couvreur (2013) were available, because Bellot & al. (*in press.*) provided more accurate fossil calibrations. The *Borassus* groups which were included depend on the chosen colonization scenario of *Borassus*: either one Madagascan colonization (one *Borassus* group) or two Madagascan colonization (two *Borassus* groups).

Dataset name	Included sources	Included <i>Borassus</i> group(s)
A-B1	all	<i>Borassus</i> Madagascar/Africa clade
A-B2	all	<i>Borassus</i> Madagascar/Africa clade
S-B1	selected	<i>Borassus madagascariensis</i> and Madagascan <i>Borassus aethiopum</i>
S-B2	selected	<i>Borassus madagascariensis</i> and Madagascan <i>Borassus aethiopum</i>

2.4 Correlation between species richness and arrival time

I wanted to test the correlation between the arrival time and the natural logarithm of Madagascan species richness, assuming that species richness increases exponentially over time, based on a diversity-independent birth-death model (Rabosky, 2012).

Since only intervals and no exact arrival times were available, I could not directly test the correlation, but could estimate how probable a significant correlation would be according to these data. Thus, for each group, an arrival time was randomly sampled on the corresponding arrival time interval, considering a uniform probability distribution. This assumption that every time on the interval is equally probable was made, because of the lack of probability density functions of arrival times. This

sampling was repeated 100,000 times. For each series, the correlation between the natural logarithm of Madagascan species richness and the arrival time was tested. Finally, the proportion of significant correlations and significant positive correlations over all 100,000 series was calculated. This was repeated for all four datasets (cf. Table 1).

The calculations were performed using R v3.6.2 (R Core Team, 2013).

3 Results

The estimated intervals of arrival time for each source are represented in Table 2. The summary of intervals over all sources, and over selected sources can be found in Table 3.

The relation between the species richness and the arrival time intervals for the dataset S-B1 is represented in Figure 1a. All intervals overlapped with all other intervals. For this dataset, the proportion of significant correlations ($p < 0.05$) was 2% in this sampling and the proportion of positive significant correlations ($r^2 > 0$ and $p < 0.05$) was 0% (cf. Figure 1b).

For the other datasets, the figures looked similar to the one of the dataset S-B1 (Figure 1a), and the proportions of significant correlation were almost as low as for the dataset S-B1 (S-B2: 0.05%; A-B1: 2%; A-B2: 0.05%), and the proportion of positive significant correlation was also 0% for all the other datasets (cf. Figures S1, S2, S3).

The earliest possible arrival time of palms to Madagascar was 66 Mya (*Bismarckia & Satranala; Bellot & al., in press.*).

4 Discussion

No positive correlation between species richness and arrival time could be found for Madagascan palm lineages when sampling over all intervals of possible arrival times. This means that given the available data, no combination of arrival times can be chosen that would lead to a significant Time-for-Speciation effect in Madagascan palms. The only possibility to detect Time-for-Speciation effect in Madagascan palms would be if future studies move or expand the intervals.

Table 2: Summary of species richness and interval of possible arrival times to Madagascar according to different time-calibrated phylogenies for each clade. *Borassus* M/A stands for the *Borassus* Madagascar/Africa group. $n_{Madag.}$ corresponds to the number of Madagascan species of the corresponding group. *Endemism* indicates whether the Madagascan species of the group are endemic to Madagascar or not. n_{tot} corresponds to the total number of species of the corresponding group. Scenario B1 corresponds to the scenario with one colonization event of *Borassus* to Madagascar and scenario B2 to the one with two colonization events. Earliest and Latest correspond to the earliest and latest possible arrival time according to the corresponding source. These dates are given in Mya and 95% highest posterior densities are given in brackets.

Defined group	$n_{Madag.}$	Endemism	n_{tot}	Scenario	Phylogeny 1			Phylogeny 2		
					Source	Earliest	Latest	Source	Earliest	Latest
<i>Bismarckia & Satranala</i>	2	endemic	2	-	Bellot & al. (in press.)	65 (64-66)	45 (25-60)	Baker & Couvreur (2013)	20.7 (13-26.7)	15.5 (7-23)
<i>Hyphaene coriacea</i>	1	non-endemic	1	-	Bellot & al. (in press.)	23 (12-35)	0	-	-	-
<i>Tahina</i>	1	endemic	1	-	Bellot & al. (in press.)	41 (19-63)	0	Baker & Couvreur (2013)	8.9 (1.7-17.7)	0
<i>Borassus</i> M/A	2	endemic	5	B1	Bellot & al. (in press.)	24 (15-34)	15 (9-23)	-	-	-
<i>Borassus madagascariensis</i>	1	endemic	1	B2	Bellot & al. (in press.)	3 (1-7)	0	-	-	-
<i>Borassus aethiopium</i> (M)	1	non-endemic	1	B2	Bellot & al. (in press.)	9 (3-16)	0	-	-	-
<i>Ravenea</i>	19	endemic	21	-	Trénel & al. (2007)	41 (37-48)	16 (15-18)	Baker & Couvreur (2013)	13.4 (3.7-24.7)	0
<i>Phoenix</i>	1	non-endemic	15	-	-	-	-	-	-	-
<i>Voanioala & Beccarioiphoenix</i>	4	endemic	4	-	Meerow & al. (2015)	37.9 (29-48)	30.9 (29-42)	-	-	-
Dypsidinae	172	endemic	178	-	Faurby & al. (2016)	27 (0-28)	21 (0-24)	-	-	-
<i>Orania</i>	3	endemic	27	-	-	-	-	-	-	-

Table 3: Absolute summary of intervals of possible arrival dates to Madagascar for each group. *All* means that the dates were calculated over all available sources for this group (Approach *A*). *Selected* corresponds to *selected sources* and means that only the dates of Bellot & al. (*in press.*) were used for the groups where Bellot & al. (*in press.*) and (Baker & Couvreur, 2013) were available, because Bellot & al. (*in press.*) provided more accurate fossil calibrations (Approach *S*). The absolute earliest date of arrival corresponds to the highest possible value over all/selected sources, including HPD (highest posterior density) interval values, and the absolute latest date of arrival corresponds to the lowest possible value.

Defined group	Absolute earliest (all)	Absolute latest (all)	Absolute earliest (selected)	Absolute latest (selected)
<i>Bismarckia</i> & <i>Satranala</i>	66	7	66	25
<i>Hyphaene coriacea</i>	35	0	35	0
<i>Tahina</i>	63	0	63	0
<i>Borassus</i> Madagascar/Africa clade	34	9	34	9
<i>Borassus madagascariensis</i>	7	0	7	0
<i>Borassus aethiopium</i> (Madagascar)	16	0	16	0
<i>Ravenea</i>	48	0	48	0
<i>Phoenix</i>	-	-	-	-
<i>Voanioala</i> & <i>Beccariophoenix</i>	48	29	48	29
Dypsidinae	28	0	28	0
<i>Orania</i>	-	-	-	-

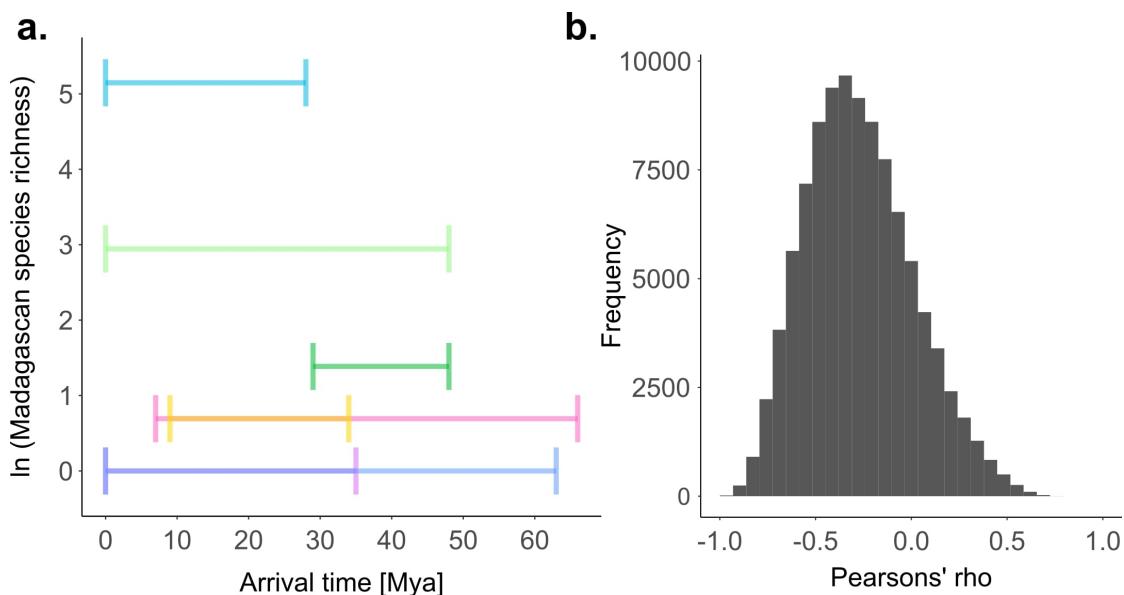


Figure 1: **a)** Representation of the natural logarithm of Madagascan species richness as a function of arrival time to Madagascar. Each interval corresponds to the interval of possible arrival time (lowest 95% HPD interval border of crown age - highest 95% HPD interval border of stem age; or present - lowest 95% HPD interval border of crown age) estimated from time-calibrated phylogenies, for each taxonomic group. HPD: highest posterior density. Light blue: Dypsidinae; light green: *Ravenea*; green: *Voanioala* & *Beccariophoenix*; pink: *Bismarckia* & *Satranala*; yellow: *Borassus* Madagascar/Africa clade; blue: *Tahina*; violet: *Hyphaene coriacea*. Here are represented only the results for the dataset S-B1 (selected sources & only one Madagascan colonization event for *Borassus*), but the figures for all other datasets look similar. **b)** Histogram representing the frequencies of different standardized slopes (Pearson's ρ) of the correlation between the natural logarithm of Madagascan species richness and the arrival time to Madagascar when randomly sampling ($n = 100,000$) over each uniform interval (panel (a) of this figure). 2% of the samples showed a significant negative correlation and 0% a significant positive correlation. This histogram and the percentage of significant (positive) correlations were similar for all other datasets.

In this study, I considered the whole continuous interval of possible arrival times for each group. Usually, single data points are used to test the Time-for-Speciation hypothesis. These arrival dates are arbitrarily chosen as being the crown-group age or the middle between the stem-age and crown age (for monotypic groups: middle of the terminal branch) (*e.g.* Li & Wiens, 2019; Harris & Davies, 2016). Li & Wiens (2019) argue that this major approximation can be done because they are not interested in exact ages, but only in relative ages of colonization events. This approximation might work for large-scale studies where intervals of possible colonization events do not overlap, but intuitively, such an approximation is not compatible with overlapping intervals, like in the present study. There is no way of knowing which clade colonized the region before the other when their intervals of possible arrival time overlap. Furthermore, Miller & Wiens (2017) state that deviations from their assumption that colonization for monotypic groups occurred mid-terminal-branch are unlikely to affect their conclusions, because "monotypic invasions are relatively young, reducing the range of possible ages". Here, the approximately 41 Million year old group *Tahina* consists of a clear counterexample. Hence, I argue that the mid-terminal-branch assumption cannot be made and the whole interval has to be taken into account.

Like in the present study, results of absence of Time-for-Speciation effect have been found in large-scale studies (*e.g.* Rabosky & al., 2012; Scholl & Wiens, 2016). Alternative explanations for differences in species richness between clades have been proposed. These say that these differences can be explained rather by variation in diversification rates between clades or over time than by differences in arrival time (Rabosky, 2009). Indeed, in Madagascan palms, it is possible that differences in species richness can be attributed to different diversification rates between clades. For example the tribe Dypsidinae could be so species rich because of an early rate shift in the Areceae (Couvreur & al., 2015) rather than because of a Madagascar-related process. However, to understand these complex dynamics of palm diversification in Madagascar, more precise estimates of net diversification are needed in all clades.

Another result worth mentioning is that, according to the available phylogenies, the absolute earliest possible arrival time of palms to Madagascar is at the beginning of the Paleocene (66 Mya) with the ancestor of the *Bismarckia/Satranala* clade. Even when taking into account the stem ages of the tribes corresponding to the clades that were not included in this study (*i.e.* Phoeniceae: 49 Mya [33-65 95% HPD]; Oranieae: 33 Mya [15-52 95% HPD]; cf. Baker & Couvreur, 2013), the

oldest arrival time was 66 Mya. During this time, Madagascar had already been isolated from other continents for at least 17 Mya (Indio-Madagascar split happened latest 83.5 Mya; [Ali & Aitchison, 2008](#)). Hence, I suggest that a vicariant origin of Madagascan palms can be ruled out and palms have arrived to Madagascar via dispersal, as it is the case in many other plant groups ([Yoder & Nowak, 2006](#)).

5 Conclusion

To conclude, the large intervals of arrival times and the absence of the Time-for-Speciation effect shows that palm colonization and diversification histories are still poorly known in Madagascar. More studies on the phylogenetic relationships, their arrival time and following diversification processes on the island are needed to explain current differences in species richness between clades. Most urgently, the focus should first lie the genus *Orania* which is the only genus with Madagascan endemics for which no relationships are known on a species-level.

Supplementary Material

All scripts and the following supplementary material can be found at:

https://github.com/pebgroup/pos_phylo_biogeo.

Figure S1: Same as Figure 1, but for the dataset S-B2 (selected sources & two Madagascan colonization events for *Borassus*).

Figure S2: Same as Figure 1, but for the dataset A-B1 (all sources & only one Madagascan colonization event for *Borassus*).

Figure S3: Same as Figure 1, but for the dataset A-B2 (all sources & two Madagascan colonization events for *Borassus*).

Bibliography

- Ali, J.R. & Aitchison, J.C.** 2008. Gondwana to Asia: Plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166–35 Ma). *Earth-Science Reviews*, 88(3-4): 145–166. doi:[10.1016/j.earscirev.2008.01.007](https://doi.org/10.1016/j.earscirev.2008.01.007).
- 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. doi:[10.1111/j.1365-2699.2012.02795.x](https://doi.org/10.1111/j.1365-2699.2012.02795.x).
- 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. *Systematic Biology*, 58(2): 240–256. doi:[10.1093/sysbio/syp021](https://doi.org/10.1093/sysbio/syp021).
- Bellot, S., Bayton, R.P., Couvreur, T.L.P., Dodsworth, S., Eiserhardt, W.L., Pritchard, H., Roberts, L., Toorop, P., & Baker, W.J.** *in press*. On the origin of giant seeds: Insights from the double coconut (*Lodoicea maldivica*) and close relatives (Borasseae, Arecaceae). *New Phytologist*.
- Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., Maio, N.D., 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 Computational Biology*, 15(4): e1006650. doi:[10.1371/journal.pcbi.1006650](https://doi.org/10.1371/journal.pcbi.1006650).
- Couvreur, T.L., 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 Biology*, 9(1): 44. doi:[10.1186/1741-7007-9-44](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. *Frontiers in Genetics*, 5. doi:[10.3389/fgene.2014.00452](https://doi.org/10.3389/fgene.2014.00452).
- Dransfield, J. & Beentje, H.** 1995. *The Palms of Madagascar*. Royal Botanic Gardens.
- Dransfield, J., Rakotoarinivo, M., Baker, W.J., Bayton, R.P., Fisher, J.B., Horn, J.W., Leroy, B., & Metz, X.** 2008. A new Coryphoid palm genus from Madagascar. *Botanical Journal of the Linnean Society*, 156(1): 79–91. doi:[10.1111/j.1095-8339.2007.00742.x](https://doi.org/10.1111/j.1095-8339.2007.00742.x).

- Dransfield, J. & Uhl, N.W.** (eds.) 2008. *Genera Palmarum: The Evolution and Classification of Palms*, Kew Plants People Possibilities, new edition. Kew Publ. Royal Botanical Garden.
- Faurby, S., Eiserhardt, W.L., Baker, W.J., & Svensson, J.C.** 2016. An all-evidence species-level supertree for the palms (Arecaceae). *Molecular Phylogenetics and Evolution*, 100: 57–69. doi:[10.1016/j.ympev.2016.03.002](https://doi.org/10.1016/j.ympev.2016.03.002).
- Harris, L.W. & Davies, T.J.** 2016. A Complete Fossil-Calibrated Phylogeny of Seed Plant Families as a Tool for Comparative Analyses: Testing the "Time for Speciation" Hypothesis. *PLOS ONE*, 11(10): e0162907. doi:[10.1371/journal.pone.0162907](https://doi.org/10.1371/journal.pone.0162907).
- Li, H. & Wiens, J.J.** 2019. Time Explains Regional Richness Patterns within Clades More Often than Diversification Rates or Area. *The American Naturalist*, 193(4): 514–529. doi:[10.1086/702253](https://doi.org/10.1086/702253).
- MacArthur, R.H.** 1965. Patterns of Species Diversity. *Biological Reviews*, 40(4): 510–533. doi:[10.1111/j.1469-185X.1965.tb00815.x](https://doi.org/10.1111/j.1469-185X.1965.tb00815.x).
- 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. doi:[10.1086/653688](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. *Botanical Journal of the Linnean Society*, 190(3): 260–280. doi:[10.1093/botlinnean/boz019](https://doi.org/10.1093/botlinnean/boz019).
- Meerow, A.W., Noblick, L., Salas-Leiva, D.E., Sanchez, V., Francisco-Ortega, J., Jestrow, B., & Nakamura, K.** 2015. Phylogeny and historical biogeography of the cocosoid palms (Arecaceae, Arecoideae, Cocoseae) inferred from sequences of six WRKY gene family loci. *Cladistics*, 31(5): 509–534. doi:[10.1111/cla.12100](https://doi.org/10.1111/cla.12100).
- Miller, E.C. & Wiens, J.J.** 2017. Extinction and time help drive the marine-terrestrial biodiversity gradient: Is the ocean a deathtrap? *Ecology Letters*, 20(7): 911–921. doi:[10.1111/ele.12783](https://doi.org/10.1111/ele.12783).
- R Core Team** 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria.
- Rabosky, D.L.** 2009. Ecological limits and diversification rate: Alternative paradigms to explain the variation in species richness among clades and regions. *Ecology Letters*, 12(8): 735–743. doi:[10.1111/j.1461-0248.2009.01333.x](https://doi.org/10.1111/j.1461-0248.2009.01333.x).
- Rabosky, D.L.** 2012. Testing the time-for-speciation effect in the assembly of regional biotas. *Methods in Ecology and Evolution*, 3(2): 224–233. doi:[10.1111/j.2041-210X.2011.00166.x](https://doi.org/10.1111/j.2041-210X.2011.00166.x).

- Rabosky, D.L., Slater, G.J., & Alfaro, M.E.** 2012. Clade Age and Species Richness Are Decoupled Across the Eukaryotic Tree of Life. *PLOS Biology*, 10(8): e1001381. doi:10.1371/journal.pbio.1001381.
- Ricklefs, R.E.** 1987. Community diversity: Relative roles of local and regional processes. *Science (New York, N.Y.)*, 235(4785): 167--171. doi:10.1126/science.235.4785.167.
- Scholl, J.P. & Wiens, J.J.** 2016. Diversification rates and species richness across the Tree of Life. *Proceedings of the Royal Society B: Biological Sciences*, 283(1838). doi:10.1098/rspb.2016.1334.
- Stephens, P.R. & Wiens, J.J.** 2003. Explaining Species Richness from Continents to Communities: The Time-for-Speciation Effect in Emydid Turtles. *The American Naturalist*, 161(1): 112--128. doi:10.1086/345091.
- Trénel, P., Gustafsson, M.H., 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). *Molecular Phylogenetics and Evolution*, 45(1): 272--288. doi:10.1016/j.ympev.2007.03.018.
- Vences, M., Wollenberg, K.C., Vieites, D.R., & Lees, D.C.** 2009. Madagascar as a model region of species diversification. *Trends in Ecology & Evolution*, 24(8): 456--465. doi:10.1016/j.tree.2009.03.011.
- WCSP**. World Checklist of Selected Plant Families. Royal Botanic Gardens, Kew. [Online]. <http://wcsp.science.kew.org/>.
- Yoder, A.D. & Nowak, M.D.** 2006. Has Vicariance or Dispersal Been the Predominant Biogeographic Force in Madagascar? Only Time Will Tell. *Annual Review of Ecology, Evolution, and Systematics*, 37(1): 405--431. doi:10.1146/annurev.ecolsys.37.091305.110239.

PART C

Phylogeny and biogeography of the
palm genus *Orania*

Understanding disjunct distribution patterns and historical biogeography in the palm genus *Orania* (Arecaceae) using a phylogenomic approach

Running title: Phylogeny and biogeography of the palm genus *Orania*

Maya Schrödl^{1*} and Wolf Eiserhardt

¹Aarhus University, Department of Bioscience, Ny Munkegade 114, 8000 Aarhus C, DK

*Author for correspondence: maya.schroedl@bios.au.dk

Abstract. Madagascar is an extraordinarily species-rich island with very high levels of endemism. However, the genesis of this biodiversity is still debated and understanding the biogeographic history of the Madagascan biota is a major step towards the resolution of this mystery. One particular focus is the study of lineages that are not only present in Madagascar, but also overseas in Southeast Asia and/or Africa. Understanding how these disjunct distributions were put in place is a major question in biogeography.

A genus that presents such a disjunct distribution is the palm genus *Orania*. This genus is mainly present in Southeast Asia with only three species endemic to Madagascar. The aim of this study was to understand when and wherefrom this genus arrived to Madagascar in a phylogenomic context. A first dated species-level phylogeny was built using targeted sequencing on silica and herbarium material. Through ancestral range estimation, it was inferred that Madagascan *Orania* is monophyletic and arrived to the island via long-distance dispersal from Southeast Asia between the Oligocene and early Miocene.

To conclude, this study provided one of the last species-level phylogenies of Madagascan palms and highlighted the importance of long-distance dispersal in shaping disjunct biogeography patterns.

Key-words: Ancestral range estimation, Madagascar, POS clade, species-level phylogeny, target capture

1 Introduction

The island of Madagascar is known to harbour an extraordinary biodiversity with high levels of species endemism. Madagascar contains 11,000 species of vascular plants, of which about 83% are endemics (Callmander & al., 2011) and is therefore considered to be one of Earth's most important biodiversity hotspots (Myers & al., 2000). However, the past eco-evolutionary processes shaping the island's current biodiversity are still poorly known.

One major step towards the understanding of the genesis of this plant hyperdiversity and hyperendemism, is the study of how Madagascan lineages are related to their relatives from other continents. This would allow to grasp their origin and time of arrival to Madagascar and following diversification processes.

Many Madagascan species groups present either strong affinities with Africa and/or Southeast Asia, including India, resulting in a disjunct distribution (Warren & al., 2010). Either the creation of a barrier or the overcoming of existing barriers can result in such distributions, due to two major eco-evolutionary forces: vicariance and dispersal (Myers & Giller, 1988). These two have been long discussed in the context of Madagascar (Yoder & Nowak, 2006).

However, with the rising amount of available dated phylogenies, the vicariance hypothesis has been rejected for most groups, because almost all of them presented post-Gondwanan split divergence times (Yoder & Nowak, 2006). Nowadays, it is thought that the current biodiversity patterns of Madagascan groups are rather due to recent colonization events in the Cenozoic period (66 Mya - 0 Mya) with vicariance only being a minor relict force (*e.g.* Yoder & Nowak, 2006; Warren & al., 2010; Agnarsson & Kuntner, 2012; Bacon & al., 2016).

Some lineages arrived to Madagascar from Africa, the closest continent, over the Mozambique Channel (*e.g.* Ali & Huber, 2010; Zhou & al., 2012). Other lineages arrived to Madagascar from Asia through long-distance dispersal (LDD) over the Indian Ocean (Warren & al., 2010). This could have been facilitated by "stepping stone" islands in the Indian Ocean, and/or marine currents and winds favouring east-to-west dispersal to Madagascar (Warren & al., 2010; Federman & al., 2015). Whether long-distance-dispersal from Asia or Africa was the foundation for Madagascan species remains untested in many taxa, especially plants, due to missing phylogenies.

One major plant group that is representative of the Madagascan flora and is relatively well-studied are palms (Arecaceae) (Couvreur & Baker, 2013). The origin and arrival time of Madagascan endemic palm species has been studied to some extent in a phylogenetic framework for all Madagascan palm groups (Bellot & al., *in press*; Baker & Couvreur, 2013a; Trénel & al., 2007; Meerow & al., 2015), except for two Arecoid groups: the hyperdiverse subtribe Dypsidinae, and the genus *Orania*.

The palm genus *Orania* Zipp. (Arecoids) is of particular interest because of its peculiar disjunct distribution. Three species are endemic to Madagascar, and the majority of its species (27, WCSP) are only present in Southeast Asia, from Malesia to Papuasia (Keim & Dransfield, 2012). *Orania*'s closest relatives are the West African genera *Podococcus* and *Sclerosperma*. All species of the three genera are mainly present in tropical rainforests (Dransfield & Uhl, 2008). The three corresponding monogeneric tribes form together the monophyletic Podococceae–Oranieae–Sclerospermeae (POS) clade (Dransfield & Uhl, 2008), which is sister to all Core Arecoids (Baker & Couvreur, 2013a). The relationships between the three tribes is still debated: either *Orania* & *Sclerosperma* form a clade (Baker & Couvreur, 2013a; Comer & al., 2016), or *Podococcus* & *Sclerosperma* (Lewis & Doyle, 2002; Comer & al., 2015).

Different hypotheses on the biogeographic history of the POS clade might explain the contemporary disjunct distribution of *Orania* with either arriving to Madagascar from Africa or from Southeast Asia. Baker & Couvreur (2013a) suggest that *Orania* and *Sclerosperma* split respectively in Eurasia and Africa, with *Orania* subsequently dispersing towards Madagascar and the rest of Southeast Asia. However, their genus-level phylogeny did not give a more detailed insight into the dispersal within *Orania*. After the basal split of *Orania* from either *Sclerosperma* or *Sclerosperma* & *Podococcus*, two different scenarii are possible:

- (1) Dransfield & Uhl (2008) suggested based on Morley (2003) a route from Africa over Madagascar towards Southeast Asia, possibly through India-Madagascan vicariance, a hypothesis also supported by (Comer & al., 2016). Also long distance dispersal from Madagascar to Southeast Asia could have been possible as in Bacon & al. (2016) and Zhou & al. (2012).
- (2) *Orania* dispersed from Southeast Asia to Madagascar in one or more long-distance colonization events over the Indian Ocean, as in Warren & al. (2010).

No species-level phylogeny of the POS clade is currently available and whether

Madagascan *Orania* species arrived to Madagascar from Africa or Asia, whether they descend from one colonist, and when they arrived, remain open questions.

In this study, we aim to answer these questions through phylogenomic inference using targeted sequencing methods, molecular dating, and ancestral range estimation with different biogeographic models. A first species-level phylogeny of the POS clade was constructed. Furthermore, the monophyly of each genus of the POS clade and of the three Madagascan *Orania* species was tested and the relationships of *Orania* to the other genera of the POS clade were established. Based on a dated phylogeny, the different biogeographic hypotheses on the origin of the Madagascan *Orania* species were tested. We provided one of the last species-level phylogenies to complete the Madagascan palm phylogeny puzzle.

2 Materials and Methods

2.1 Taxon Sampling

A total of 41 samples were obtained from mostly herbarium fragments or silica-dried leaves. For the genus *Orania* 23 species (77%) out of 30 currently recognized species were sampled, for the genus *Sclerosperma* all 3 species (100%) of the 3 currently recognized species, and for the genus *Podococcus* both species (100 %) of the 2 currently recognized species ([WCSP](#); Table 1). The outgroup consisted of one species chosen among the Core Arecoids (*Dypsis mananjarensis*).

2.2 DNA Extraction, Library Preparation, Target Capture, and Sequencing

For the samples that were sequenced in the Couvreur lab (see Table 1), the same sequencing preparation and sequencing methods as in [Couvreur & al. \(2019\)](#) were applied, using the custom nuclear exon bait kit designed by [Heyduk & al. \(2015\)](#), which targets 176 single-copy genes.

For the remaining samples, DNA was extracted using a modified cetyl trimethylammonium bromide (CTAB) protocol ([Doyle & Doyle, 1987](#)). The DNA extracts were quantified using a Qubit® Fluorometer v2.2 (Thermo Fisher Scientific, Waltham, MA, United States, USA) and average fragment sizes were assessed on an Agilent Technologies 4200 TapeStation System with High Sensitivity D1000 ScreenTapes (Santa Clara, CA, USA). The samples which had an average fragment size over 350 bp were sheared to 350 bp using a M220 Focused-ultrasonicator™ (Covaris, Woburn,

MA, USA) following the manufacturer's protocol with varied shearing times depending on the DNA fragment size.

When possible, 400 ng of DNA per sample were used for library preparation. Dual-indexed libraries for Illumina® sequencing were prepared using the DNA NEBNext® Ultra™II Library Prep Kit and the NEBNext® Multiplex Oligos for Illumina® (Dual Index Primers Set 1 and 2) from New England BioLabs® (Ipswich, MA, USA) following the manufacturer's instructions, size-selecting, if needed, for 300 - 400 bp. The quality of the libraries was assessed on the 4200 TapeStation and quantified with the Qubit® Fluorometer. The final average library size including the adapters was 350 bp or lower.

After pooling of the libraries, target capture was performed. For the samples sequenced by Bellot/Eiserhardt/Couvreur (see Table 1), the custom nuclear exon bait kit designed by Heyduk & al. (2015), which targets 176 single-copy genes, was used. For the samples sequenced by Schrödl, the nuclear exon bait kit *PhyloPalm* was used, which contains the baits of Heyduk & al. (2015) and targets 795 additional single-copy genes (Loiseau & al., 2019). Target capture was conducted following the myBaits® Custom Target Capture Kits protocol (v4.0, 2018; <http://www.arborbiosci.com/mybaits-manual>), with 24 h incubation time at 65°C and post-capture amplification of 6-10 PCR cycles. PCR products were quantified with the Qubit® Fluorometer and run on the 4200 TapeStation for quality assessment. For the samples sequenced by Bellot and Eiserhardt (see Table 1), the pooled target capture reactions were sequenced on an Illumina MiSeq sequencer with 600 cycles as 2 x 300 bp paired-end reads. The samples whose sequencing was prepared by Schrödl, could not be sequenced yet due to the COVID-19 pandemic and resulting lockdown and are not included in the downstream analyses.

2.3 Read Trimming and Filtering

The reads were adapter trimmed and filtered, keeping the reads with a mean Phred quality score (Q) > 30 and a length > 35 bp. This was done using cutadapt v1.2.1 (Martin, 2011) with the default parameters and a custom script (Couvreur & al., 2019) for the samples sequenced in the Couvreur lab (see Table 1). For the samples sequenced by Bellot/Eiserhardt, Trimmomatic v0.39 (Bolger & al., 2014) was used with additionally trimming low quality read endings ($Q > 30$) and removing bases that belonged to a 4 bp window of average quality < 30 , as in Brewer & al. (2019). For quality control, FastQC v0.11.9 (Andrews, 2010) was used before and after the trimming and filtering.

Table 1: Details on the specimens of this study. The upper part of samples was sequenced and included in the analyses, whereas the lower part was not due to the COVID-19 pandemic. The sequencing was prepared either by Thomas Couvreur, Sidonie Bellot, Wolf Eiserhardt, or Maya Schrödl.

Taxon name [individual number]	Country	Collector (Herbarium)	Silica/ Herbarium	Sequenced by
<i>Orania archboldiana</i> Burret	Papua New Guinea	Baker, W.J. 1104 (K)	silica	Couvreur
<i>Orania dafonsoroensis</i> A.P.Keim & J.Dransf.	Indonesia	Heatubun, C. 278 (K)	silica	Couvreur
<i>Orania deflexa</i> A.P.Keim & J.Dransf.	Papua New Guinea	Baker, W.J. 1319 (K)	silica	Couvreur
<i>Orania grandiflora</i> A.P.Keim & J.Dransf.	Indonesia	Heatubun, C. 179 (K)	silica	Couvreur
<i>Orania lauterbachiana</i> Becc. [1]	Papua New Guinea	Baker, W.J. 1099 (K)	silica	Couvreur
<i>Orania lauterbachiana</i> Becc. [2]	Papua New Guinea	Baker, W.J. 1107 (K)	silica	Couvreur
<i>Orania lauterbachiana</i> Becc. [3]	Papua New Guinea	Baker, W.J. 1114 (K)	silica	Couvreur
<i>Orania longisquama</i> (Jum.) J.Dransf. & N.W.Uhl	Madagascar	Dransfield, J. 7746 (K)	silica	Couvreur
<i>Orania macropetala</i> K.Schum. & Lauterb.	Papua New Guinea	Banka, R. 2014 (K)	silica	Couvreur
<i>Orania palindan</i> Merr. [1]	Indonesia	Baker, W.J. 1049 (K)	silica	Couvreur
<i>Orania palindan</i> Merr. [2]	Papua New Guinea	Baker, W.J. 1140 (K)	silica	Couvreur
<i>Orania ravaka</i> Beentje [1]	Madagascar	Dransfield, J. 7731 (K)	silica	Couvreur
<i>Orania ravaka</i> Beentje [2]	Madagascar	Rakotoarinivo, M. 338 (K)	silica	Couvreur
<i>Orania ravaka</i> Beentje [3]	Madagascar	Dransfield, J. 7731 (K)	silica	Bellot
<i>Orania sylvicola</i> (Griff.) H.E.Moore [1]	Singapore	Singapore Botanic Garden 20011403*B (K)	silica	Couvreur
<i>Orania sylvicola</i> (Griff.) H.E.Moore [2]	Kuala Lumpur	FRIM 80015 (KEP)	silica	Couvreur
<i>Orania tabubilensis</i> A.P.Keim & J.Dransf.	Papua New Guinea	Baker, W.J. 1129 (K)	silica	Couvreur
<i>Orania timikae</i> A.P.Keim & J.Dransf.	Indonesia	Heatubun, C. 173 (K)	silica	Couvreur
<i>Orania trispatha</i> (J.Dransf. & N.W.Uhl) Beentje & J.Dransf.	Madagascar	Rakotoarinivo, M. 704 (K)	silica	Couvreur
<i>Podococcus acaulis</i> Hua	Gabon	Couvreur 556 (WAG)	herbarium	Couvreur
<i>Podococcus barteri</i> Mann & H.Wendl. [1]	Nigeria	Tuley 600 (K)	herbarium	Couvreur
<i>Podococcus barteri</i> Mann & H.Wendl. [2]	Cameroon	Couvreur 699 (WAG)	herbarium	Couvreur
<i>Sclerosperma mannii</i> H.Wendl. [1]	Republic of Congo	Couvreur 816 (WAG)	herbarium	Couvreur
<i>Sclerosperma mannii</i> H.Wendl. [2]	Equatorial Guinea	Sunderland 1794 (K)	silica	Bellot
<i>Sclerosperma walkeri</i> A.Chev. [1]	Gabon	Couvreur 1076 (WAG)	herbarium	Couvreur
<i>Sclerosperma walkeri</i> A.Chev. [2]	Gabon	Couvreur 1112 (WAG)	herbarium	Couvreur
<i>Dypsis mananjarensis</i> (Jum. & H. Perrier) Beentje & J. Dransf.	Madagascar	Rakotoarinivo, M. 354 (K)	silica	Eiserhardt
<i>Orania bakeri</i> A.P.Keim & J.Dransf.	Papua New Guinea	Baker 581 (K)	herbarium	Schrödl (pending)
<i>Orania decipiens</i> Becc.	Philippines	Ebalo 975 (K)	herbarium	Schrödl (pending)
<i>Orania disticha</i> Burret	Indonesia	Maturbongs 703 (K)	herbarium	Schrödl (pending)
<i>Orania ferruginea</i> A.P.Keim & J.Dransf.	Indonesia	Keim 41 (K)	herbarium	Schrödl (pending)
<i>Orania littoralis</i> A.P.Keim & J.Dransf.	Papua New Guinea	Barfod 456 (K)	herbarium	Schrödl (pending)
<i>Orania longistaminodia</i> A.P.Keim & J.Dransf.	Papua New Guinea	Barfod 374 (K)	herbarium	Schrödl (pending)
<i>Orania macropetala</i> K.Schum. & Lauterb.	Papua New Guinea	Clemens 10894 (K)	herbarium	Schrödl (pending)
<i>Orania micrantha</i> Becc.	Papua New Guinea	Clemens 11363 (K)	herbarium	Schrödl (pending)
<i>Orania palindan</i> Merr. [4]	Indonesia	de Vogel 4402 (K)	herbarium	Schrödl (pending)
<i>Orania palindan</i> Merr. [5]	Philippines	Stone (PPI) 6708 (K)	herbarium	Schrödl (pending)
<i>Orania paraguensis</i> Becc.	Philippines	Podzorski 765 (K)	herbarium	Schrödl (pending)
<i>Orania subdisticha</i> A.P.Keim & J.Dransf.	Papua New Guinea	Baker 1116 (K)	herbarium	Schrödl (pending)
<i>Orania zonae</i> A.P.Keim & J.Dransf.	Indonesia	Zona 674 (K)	herbarium	Schrödl (pending)
<i>Sclerosperma profizianum</i> Valk. & Sunderl.	Ghana	Stauffer + Ouatarra 107 (G)	silica	Schrödl (pending)

2.4 Contig Assembly, Gene filtering, Read coverage trimming, and Alignment

Trimmed and filtered paired and unpaired reads were then processed in the pipeline HybPiper v1.3.1 (Johnson & al., 2016) with default settings to recover mapped and assembled reads. HybPiper was set to map the reads to the 176 target genes using BWA (Li & Durbin, 2009). Subsequently, the reads were assembled *de novo* into contigs using SPAdes (Bankevich & al., 2012). Exonerate (Slater & Birney, 2005) was executed to align the assembled contigs to the corresponding target sequence. The HybPiper `intronerate.py` script was then run to generate supercontigs (targeted exons + off-target flanking regions). All downstream analyses are based on the supercontigs, and not only on the targeted exon regions, in order to include a maximum amount of phylogenetic information. Target recovery success was assessed with the HybPiper scripts `hybpiper_stats.py` and `gene_recovery_heatmap.R` (Johnson & al., 2016).

As a next step, the target and recovered sequence lengths were generated with the HybPiper script `get_seq_lengths.py`. These were used to identify the exons which had 75% of their length reconstructed in 75% of all individuals. Only the 63 selected genes and the corresponding supercontigs were used for downstream analyses. This gene filtering step is necessary in order to avoid gene tree errors due to fragmented sequences and to limit the amount of missing data (Couvreur & al., 2019).

Subsequently, we masked the regions of the supercontigs presenting low read coverage in order to minimize the probability of sequencing errors. Each of the sample's reads (paired and unpaired reads) were mapped on the corresponding supercontigs using BWA (Li & Durbin, 2009) and the read coverage at each base pair was computed with SAMtools (Li & al., 2009). Like Gardner & al. (2019), we chose to set the read coverage threshold to 2. We hard-masked every position having a coverage lower than the threshold and trimmed the masked bases from the supercontig endings with a custom Python script.

The masked and trimmed supercontigs were then for each gene aligned with MAFFT v7.4 (Katoh & Standley, 2013) (with the "-auto" option) and cleaned with GBLOCKS v0.91b (Castresana, 2000) using the default parameters and allowing all gap positions ("b5 = a"). In order to remove poorly aligned fragments, the alignments were also manually cleaned using AliView v1.26 (Larsson, 2014).

2.5 Phylogenetic Inference

We used a coalescent approach for phylogenetic inference. Individual gene trees were constructed using RAxML-ng v0.9 (Kozlov & al., 2019) with 200 bootstrap replicated per gene and the GTR+G model. No model selection was undertaken, because parameter-rich models such as GTR+G have been shown to lead to similar inferences as other models, making model selection unnecessary (Hoff & al., 2016; Abadi & al., 2019). Branches with bootstrap support < 10 and also branches with a branch length under 10^{-5} were collapsed using Newick utilities (Junier & Zdobnov, 2010) in order to improve the accuracy of inferred species trees (Zhang & al., 2018). The resulting gene trees were used to estimate with ASTRAL-III v5.6.3 (Zhang & al., 2018) a tree of individuals ("extended species tree") and subsequently the species tree where individuals were combined to species with the ASTRAL-multi method (option "-a") (Rabiee & al., 2018). For the individuals of a same species which were not monophyletic in the tree of individuals, their monophyly was tested by constraining them to monophyly (option "-j") and re-running ASTRAL, as in Rabiee & Mirarab (2020). For all analyses in ASTRAL, branch support was computed as local posterior probabilities. As an additional step, PhyParts (Smith & al., 2015) with a bootstrap filter of 50% was used to assess concordance and discordance among gene trees for each bipartition of the ASTRAL tree of individuals. First, the gene trees and the species tree were rerooted with the R package ape v5.3 (Paradis & Schliep, 2019). Subsequently, PhyParts was run and the script phypartspiecharts.py (<https://github.com/mossmatters/phyloscripts/tree/master/phypartspiecharts>) was used to visualize the results on the species tree.

2.6 Molecular Dating and Calibration

Molecular dating was conducted with a subset of the samples. For the species where more than one individual was available, one individual was chosen for dating or two for those where the species were not resolved monophyletic in the tree of individuals. The individuals with the least gaps in their alignments were chosen. Twenty individuals were retained. The not-selected tips were dropped in the ASTRAL individual tree with ape v5.3 (Paradis & Schliep, 2019) and this reduced tree (referred to as "species tree" from now on) was used for all downstream analyses.

Two datasets were used: the first containing all 63 genes, and for the second "gene shopping", *i.e.* selection of a subset of genes, was conducted in order to

avoid possible bias in branch lengths through topological and rate heterogeneity across genes (Smith & al., 2018). We first "shop" for gene trees that have the least topological conflict with the species tree, as in Smith & al. (2018). In addition to this paper, we took into account node support and chose the genes that had no well-supported nodes ($BS > 75\%$) disagreeing with the species tree using a custom script. This selected 23 genes. In order to account for rate heterogeneity, we chose the most clock-like gene of these 23 genes (based on root-tip-variance, Smith & al., 2018). We tested for clock-likeness of this gene with BEAST2 v2.6.1 (Bouckaert & al., 2019) with the same parameters as below, using the nested sampling algorithm and model selection based on the Bayes factor. Since no clock-likeness was found, all 23 topology-selected genes were used in the reduced dataset.

The alignments of the selected genes were concatenated using phyx v3.0 (Brown & al., 2017) for each dataset. The dating method consisted in running BEAST2 v2.6.1 (Bouckaert & al., 2019) on these concatenated supercontigs whilst applying a multi monophyletic constraint, which was set to the rooted species tree. Due to the lack of fossils in the POS clade, a secondary calibration was applied using the crown age of the split POS clade and Core Arecoideae clade (57 Ma; 95% highest probability density [HPD] 45--69) from Baker & Couvreur (2013a) as MRCA prior with normal distribution.

The BEAST .xml input file was prepared using the *R* package **babette** v2.1.2 (Bilderbeek & Etienne, 2018), which is more user-friendly and reproducible than the BEAST2 default GUI BEAUTi. The multi monophyletic constraint was added manually to the .xml file. A starting tree was generated with the *chronopl* function of **ape** based on the species tree and the calibration point. BEAST was run with relaxed lognormal molecular clock and a Yule tree prior. For the first dataset with all genes, BEAST was run for 50,000,000 MCMC generations, and for the dataset of selected genes for 100,000,000 MCMC generations, logging every 1,000 generations for both datasets. Convergence to stable values was checked with Tracer v1.7.1 (Rambaut & al., 2018), *i.e.* whether the effective sample size (ESS) was above 200 for each parameter. The dated maximum clade credibility (MCC) tree with node heights and 95% HPD intervals were obtained with the BEAST2-integrated TreeAnnotator, with a 15% burnin for the first dataset and a 50% burnin for the second dataset. The trees were visualized with FigTree v1.4.4 (Rambaut, 2009).

2.7 Ancestral Range Estimation

We defined three large biogeographic areas: Madagascar, Southeast Asia (= Malesia & Papuasia), and Africa. The corresponding distribution matrix was constructed for the POS clade based on Keim & Dransfield (2012) for *Orania* and based on Dransfield & Uhl (2008) for *Podococcus* and *Sclerosperma*, and verified with distribution data from the Botanical Information and Ecology Network (BIEN v4.1.1; Enquist2009).

To infer the biogeographic history of the POS clade, an ancestral range estimation with the *R* package BioGeoBEARS v1.1.1 (Matzke, 2013) was performed. The analysis was done on the dated BEAST2 tree with selected genes. The maximum number of areas any species may occupy was set to two.

BioGeoBEARS implements three commonly used models of range evolution: dispersal-extinction-cladogenesis (DEC) (Ree & Smith, 2008), a likelihood version of DIVA (Ronquist, 1997) (DIVALIKE), and a likelihood version of BayArea (Landis & al., 2013) (BAYAREALIKE). We tested all three models with the maximum likelihood default parameters. BioGeoBEARS also allows to implement a jump parameter +J, commonly used, but because it presents conceptual and statistical flaws (Ree & Sanmartín, 2018), we decided against its implementation.

For each range evolution model, three dispersal probability models were tested. Firstly, a biogeographic "null" model with equal probability of dispersal among all biogeographic areas through time (M_0). Secondly, a symmetric dispersal constraint model based on the model of Baker & Couvreur (2013a) (M_B). Thirdly, an asymmetric dispersal constraint model based on the "terrestrial + marine" model of Federman & al. (2015) (M_F), which in addition to Baker & Couvreur (2013a) also implements marine currents in the Indian Ocean. The dispersal probability matrices summarized to our defined areas can be found in the Text S1.

The nine models in total (three range evolution models x three dispersal probability models) were compared and evaluated using the sample-size corrected Akaike information criterion ($AICc$; Hurvich & Tsai, 1989).

All *R* analyses were conducted in *R* v3.6.2 (R Core Team, 2013).

3 Results

3.1 Target capture sequencing, assembly, and phylogenetic analyses

Due to the COVID-19 pandemic and resulting lockdown, only 27 out of the 41 samples could be sequenced (Table 1). Hence, the following [Results](#) and [Discussion](#) only include 13 *Orania* species (43% of all accepted species), 2 *Sclerosperma* species (66%), and 2 *Podococcus* species (100%), and the outgroup.

The percentage of recovery, *i.e.* the number of nucleotides recovered divided by the mean length of the targeted gene (Figure S1), was high with an average of 73% over all species and genes. No potential paralogs were detected (Table S1). The average read depth on the supercontigs was 20x. The read coverage trimming hard-masked 8% of all nucleotides with coverage below 2 reads.

After alignment cleaning and filtering, the average length of the alignments varied between 509 bp and 3804 bp over all genes, with an average length of 1719 bp. All alignments together contained 5% of missing data. The alignments and gene trees were deposited online (respectively Data S1 and Data S2).

3.2 Phylogenetic relationships

The genus *Orania* (node II) was monophyletic with high support (local posterior probability [LPP] = 1) (Figure 1a,b).

Both *Sclerosperma* and *Podococcus* were also each a monophyletic genus with high support (both: LPP = 1) (Figure 1a,b).

Sclerosperma and *Podococcus* formed together a clade (node I) with low local posterior probability support (LPP = 0.88), and only 26 out of 61 informative gene trees supported this topology (Figure 1a,b; Figure S2). The top alternative bipartition at this node had *Podococcus* as a sister clade to *Sclerosperma* and *Orania*, which was supported by 17 out of 61 informative gene trees (Figures S3; node 1).

Within the genus *Orania*, the three Madagascan *Orania* species formed a clade (node III) with high support (LPP = 1) and also the SE Asian *Orania* species were monophyletic (node IV) with high support (LPP = 1) (Figure 1a,b).

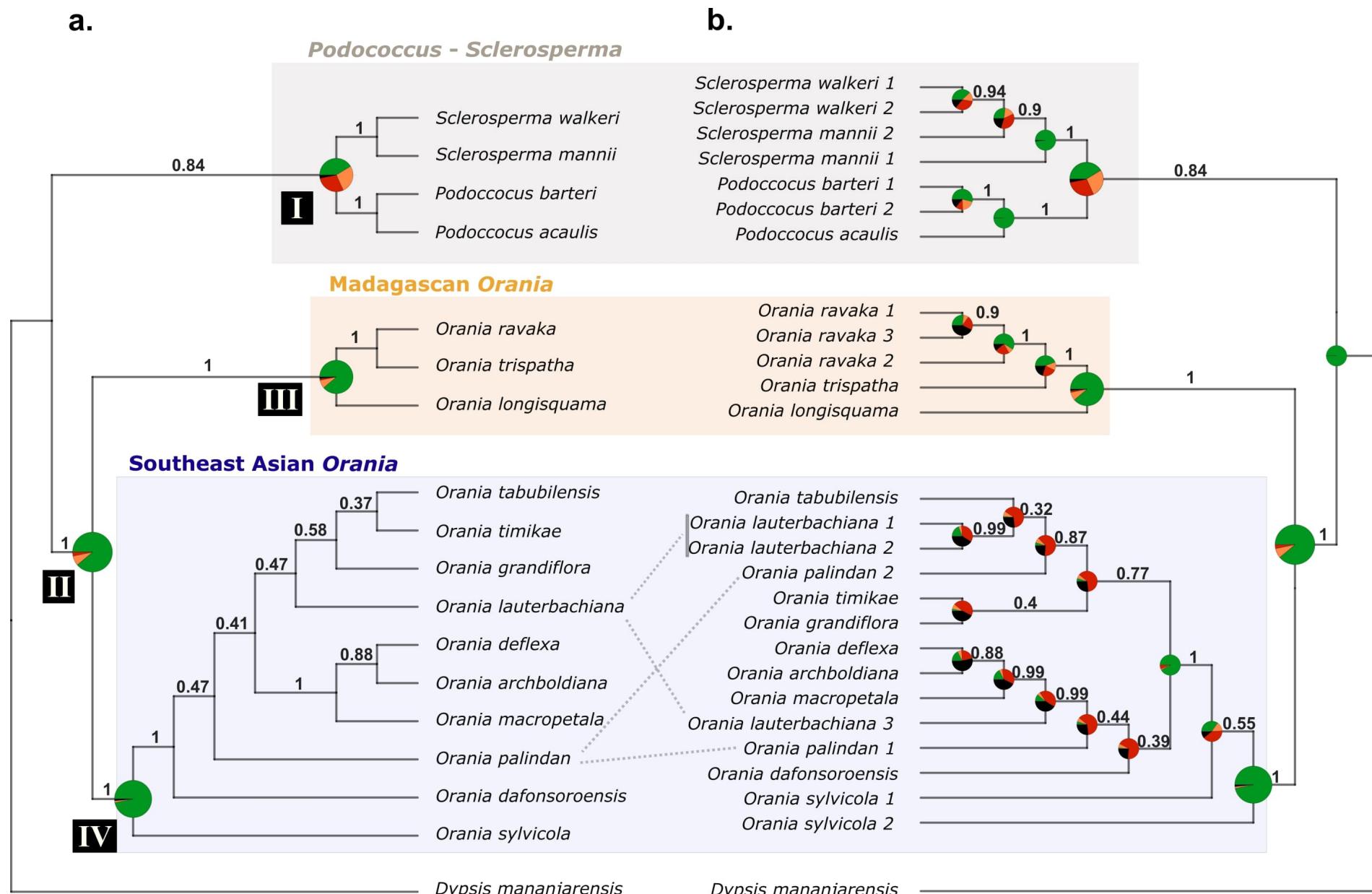


Figure 1: Cladograms of the POS clade using ASTRAL-III based on 63 gene trees generated from supercontigs. The nodes of interest are labelled with Roman numerals. Support values are given as local posterior probabilities on top of each branch. The pie charts are based on results from PhyParts and show the proportion of gene trees that are concordant with the species tree topology at the corresponding node (green), that support the main alternative topology (orange), that support other alternative topologies (red), and the proportion of gene trees that are uninformative for that node (black). The different sizes of the pie charts are only for visualization purpose. (a) Species tree, where individuals were combined to species using ASTRAL-multi; (b) Tree of individuals. Dashed lines between the tip labels of both trees show the different placement of the individuals of the same species which are not monophyletic (*Orania lauterbachiana* and *Orania palindan*). Values for pie charts were generated for the tree of individuals (b) with PhyParts, and transferred to the species tree (a) for the nodes of interest where the corresponding clades were monophyletic in both trees.

Within the SE Asian *Orania* clade, *Orania sylvicola* was sister to the rest with high support ($LPP = 1$) (Figure 1a,b). Within the rest of the SE Asian *Orania* clade, only the clade *O. matropetala*-*O. archboldiana*-*O. deflexa* was monophyletic with high support ($LPP = 1$ in Figure 1a; $LPP = 1$ in Figure 1b). Besides that, most relationships within the SE Asian *Orania* clade were uncertain, having very low LPP (< 60) in the species tree (Figure 1a). Also the proportion of gene trees supporting the topology (*i.e.* green part in Figure 1b) was very low (< 20%) for all nodes within the SE Asian *Orania* clade (exempting *O. sylvicola*). For most nodes, the proportion of the dominating alternative (*i.e.* orange part in Figure 1b) was very low and most gene trees had very different topologies (*i.e.* red part in Figure 1b) or gave no information at all for the corresponding node (*i.e.* black part in Figure 1b).

The tree of individuals (Figure 1b) had the same topology as the species tree (Figure 1a), except within the SE Asian clade (excluding *O. sylvicola*). In this clade, the individuals of *O. lauterbachiana* did not form one monophyletic group as expected, and neither did the individuals of *O. palindan*. These individuals were polyphyletic. When forcing these individuals to be monophyletic for each species with ASTRAL, the clade of *O. palindan* has very low support ($LPP = 0.14$) (cf. Figure S5) and the clade of *O. lauterbachiana* is not supported at all ($LPP = 0$) (cf. Figure S4). Furthermore, the position of the individuals of *O. sylvicola* is also uncertain in the tree of individuals ($LPP = 0.55$): either they form a clade or not. However, this would not change anything in the topology of the species tree.

3.3 Divergence time estimates

No difference in divergence times was found between the reduced dataset with 23 genes (Figure 2) and when dating on all 63 genes (Figure S7), since all credibility intervals of the estimated dates overlapped. All the following results are based on the reduced dataset.

After the split between the POS clade and the Core Arecoids (represented by our outgroup) between the Paleocene and first half of the Eocene (node VII including 95% HPD, calibration point), *Orania* split from *Podococcus-Sclerosperma* clade during the Eocene (node VI -- 95% HPD; Table 2). The genera *Podococcus* and *Sclerosperma* have been separated since the Eocene/early Oligocene (node I including 95% HPD; Table 2).

Within *Orania*, the Madagascan clade has been separated from the rest of the genus since the Oligocene/early Miocene (node II including 95% HPD; Table 2). This clade had its crown age in the late Miocene (node III including 95% HPD; Table 2). *Orania sylvicola* split from the rest of the SE Asian *Orania* species during the late Miocene (node IV including 95% HPD; Table 2). The clade consisting of all other SE Asian *Orania* species had its crown age in the Pliocene (node V including 95% HPD; Table 2). This clade is also the most diverse *Orania* clade.

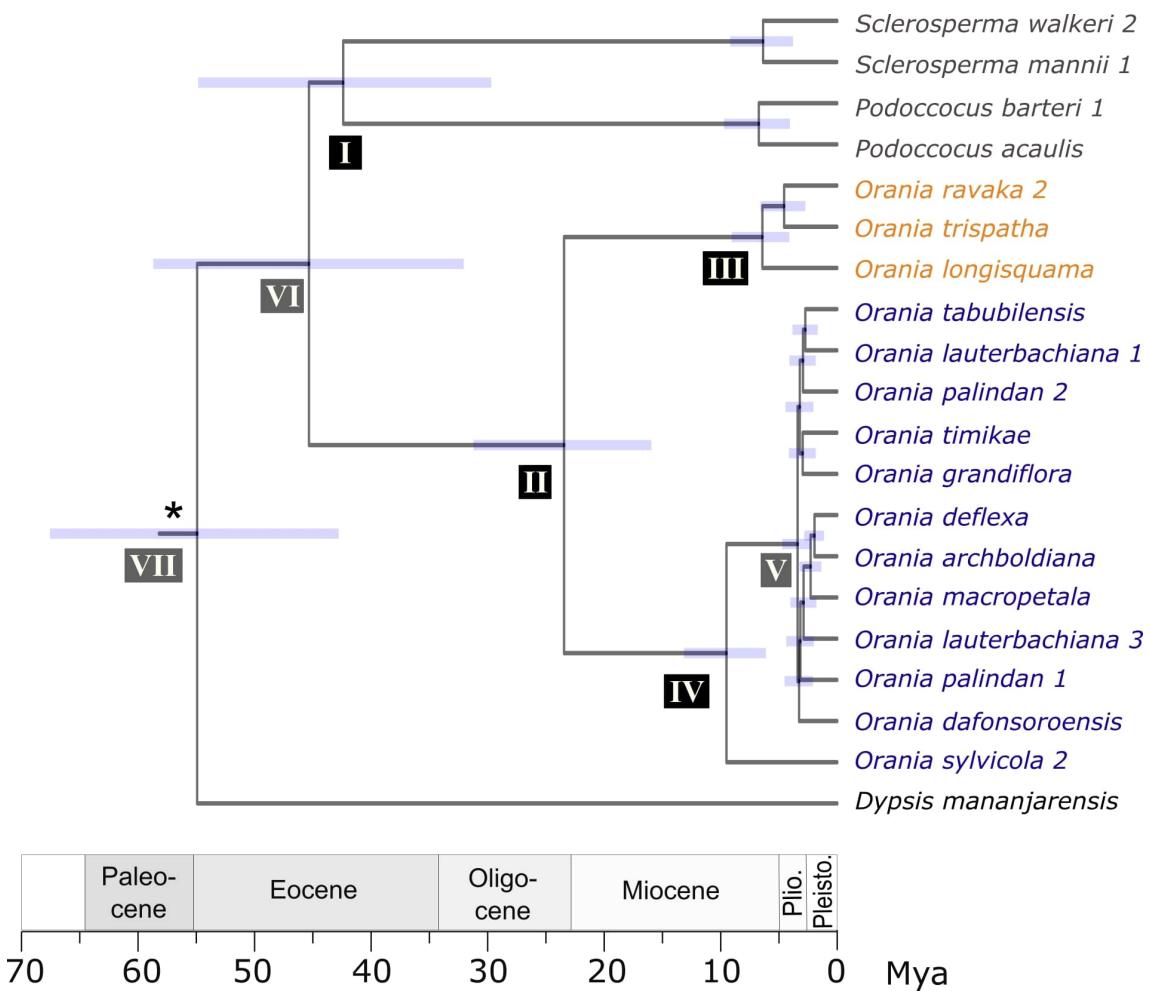


Figure 2: Maximum clade credibility chronogram of the POS clade using BEAST based on 23 genes. The blue bars indicate the 95% highest posterior density intervals. The different colours of the tip labels correspond to the different clades of interest, based on their geography: African *Sclerosperma* and *Podococcus* (grey); Madagascan *Orania* (orange); SE Asian *Orania* (blue). The nodes of interest are labelled with Roman numerals and grey boxes correspond to the numbers added to those in the cladogram (Figure 1). The asterisk (*) indicates the secondary calibration point (node VII). The Pliocene is abbreviated as "Plio." and the Pleistocene as "Pleisto."

Table 2: Divergence time estimates for the major clades based on 23 genes. The node numbers correspond to those in Figure 1. Age with 95% HPD interval in Mya. HPD: highest posterior density. The asterisk (*) indicates the secondary calibration point. The divergence time estimates for all other nodes can be found in Figure S6.

Node number	Node name	Age (95% HPD)
I	<i>Podococcus</i> - <i>Sclerosperma</i> clade crown	42.39 (29.83--54.69)
II	<i>Orania</i> crown	23.45 (16.10--31.07)
III	Madagascan <i>Orania</i> crown	6.42 (4.26--8.92)
IV	Southeast Asian <i>Orania</i> crown	9.51 (6.27--13.01)
V	Southeast Asian <i>Orania</i> [except <i>O. sylvicola</i>] crown	3.4 (2.36--4.58)
VI	POS clade crown	45.33 (32.20--58.55)
VII*	POS clade stem	54.93 (42.93--67.42)

3.4 Ancestral ranges and colonization histories

The best-fitting model was the DIVALIKE model without dispersal constraints (M_0) (Table S2). The most recent common ancestor of the POS clade was inferred to have been present in either Africa and Southeast Asia or in Africa and Madagascar (cf. node VI, Figure S8). Due to this ambiguity, we decided to take into account prior knowledge and constrain the split between the outgroup and the POS clade (node VII) to Southeast Asia, and an additional region, continental Eurasia, based on the biogeographic inference for this node by Baker & Couvreur (2013a) (region "F"). All following results are based on this constraint model.

For this new node-constraint model, the best-fitting model was also the DIVA-LIKE model without dispersal constraints (M_0) (Table S3). The ancestral range probabilities were 1 at all nodes (Figure 3). It was inferred that after the split with the Core Arecoids in SE Asia or continental Eurasia (as constrained based on Baker & Couvreur, 2013a), the ancestors of the POS clade dispersed into Africa in the early Eocene (between 55 and 45 Mya \pm 95% HPD (node VII -- node VI; Table 2)). The most recent common ancestor of the POS clade was then present in SE Asia and Africa. Subsequently, *Sclerosperma* and *Podococcus* would have emerged in Africa (cf. Figure 3, node I).

Orania was inferred to have arrived in one colonization event via long-distance dispersal from Southeast Asia to Madagascar earliest 58 Mya (cf. Figure 3 and Table 2, node VI, upper 95% HPD limit) and latest 16 Mya (cf. Figure 3 and Table 2, node VI, lower 95% HPD limit).

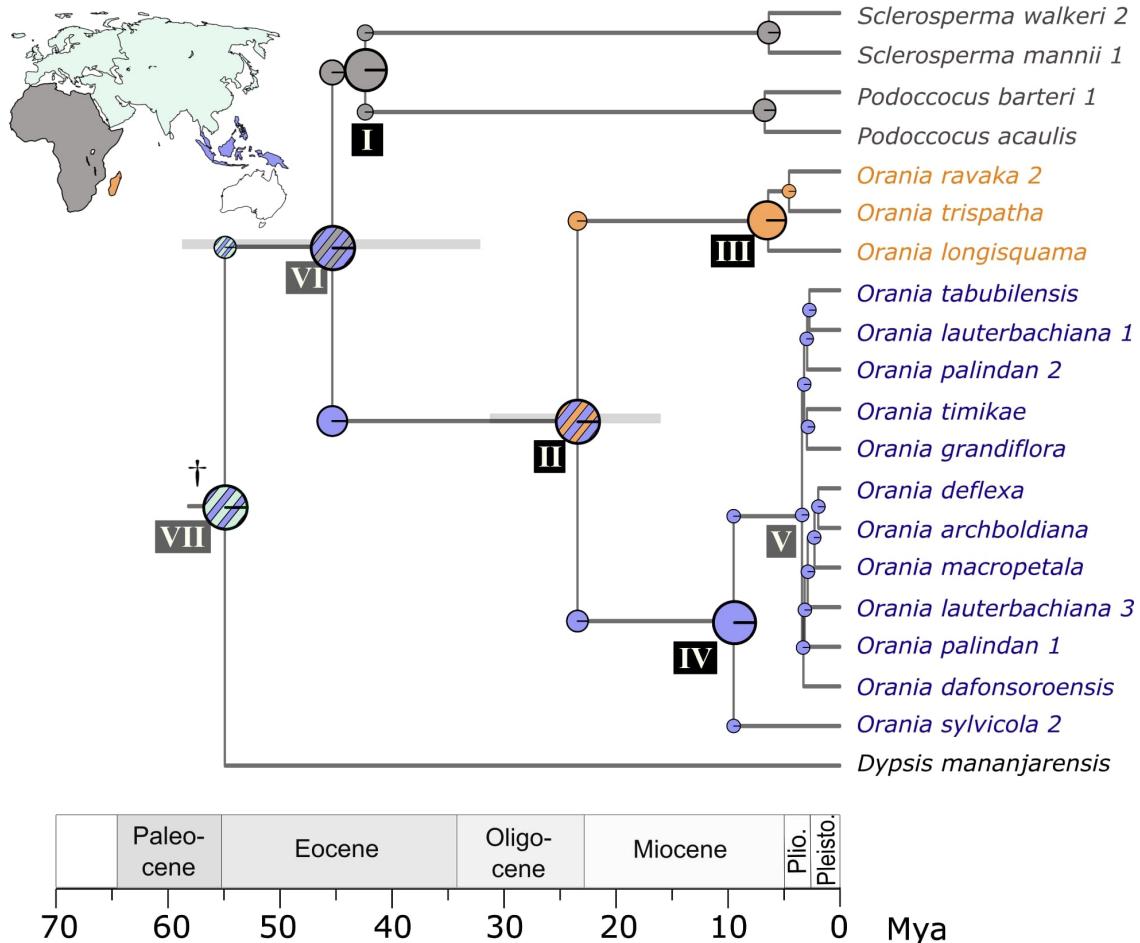


Figure 3: Biogeographic ancestral range estimates using BioGeoBears on the chronogram of Figure 2. At the top left is represented which colour corresponds to which geographical area: grey (Africa); orange (Madagascar); blue (Southeast Asia = Malesia + Papua); light green (continental Eurasia). The colours of the tip labels correspond to the current range of each species. The range of *Dypsis mananajarensis* was not taken into account and this tip label is therefore represented in black. The colours within the pie charts correspond to the range probabilities at the corresponding node, which are all equal to 1 in this case. The striped parts indicate the presence in multiple areas. The different sizes of the pie charts are only for visualization purposes. Pie charts on the stem nodes of the tips are not represented; the range probability is 1 for all corresponding tip ranges at these nodes. The nodes of interest are labelled with Roman numerals and grey boxes correspond to the numbers added to those in the cladogram (Figure 1). For node VI and node II, the credibility intervals of their age are represented for the discussion of uncertainty of the possible arrival time of *Orania* to Madagascar. The dagger (†) indicates the node which was constrained to the range of Southeast Asia (blue) and continental Eurasia (light green) based on region "F" of Baker & Couvreur (2013a) (node VII). The Pliocene is abbreviated as "Plio." and the Pleistocene as "Pleisto."

4 Discussion

4.1 Systematic implications

Our first species-level phylogeny confirms the monophyly of the genus *Orania*, as already suggested by Lewis & Doyle (2002) and Asmussen & al. (2006), who included only one Madagascan and one Southeast Asian species.

Furthermore, the Madagascan *Orania* form a clade that split a long time ago from the Southeast Asian (SE Asian) *Orania* clade (23 Mya [16--31, 95% HPD]). The existence of these two clades is in agreement with the morphological differences between these species: in contrast to their SE Asian sisters, the Madagascan *Orania* species present ramenta, and have more stamens and staminodes (Keim & Dransfield, 2012). Due to these morphological differences, two of the three Madagascan *Orania* species were formerly placed in different genera (Keim & Dransfield, 2012) (*O. longisquama* was *Sindroa longisquama* Jum., and *O. trispatha* was *Halmoorea trispatha* J.Dransf. N.W.Uhl). However, the recognition of these two monotypic genera would make *Orania* polyphyletic because of the position of *O. ravaka*. The three Madagascan species could be recognized as *Sindroa spp.*, since this clade has long been separated from the rest of *Orania*. However, this would imply two new combinations (*i.e.* "*Sindroa trispatha*" and "*Sindroa ravaka*"), which would violate the principle of nomenclatural stability and is therefore rejected.

In the SE Asian *Orania* clade, *O. sylvicola* was resolved as sister to all other SE Asian *Orania* species. Relative to the other splits within this last clade, *O. sylvicola* split a long time ago from this clade (10 Mya [6--13, 95% HPD]). This is concordant with the hypothesis of Essig (1980), that *O. sylvicola* is the "most primitive member of the genus", while only SE Asian *Orania* species were in this genus (The Madagascan species were placed/discovered after 1980: *O. longisquama* (Dransfield & Uhl, 1984); *O. trispatha* (Dransfield & Beentje, 1995); *O. ravaka* (Dransfield & Beentje, 1995)). This was based on morphological characters like well-developed pistillodes and staminodes.

Apart from *O. sylvicola*, the topology within the SE Asian clade remains unclear, since most nodes were poorly supported, ie. had low posterior probability values and many gene trees conflicting with the species tree and each other. A great part of gene trees gave no information on the corresponding node. These low-supported nodes can be mainly explained by consequences of the observed very short terminal branches in this clade, which indicates a recent rapid radiation. The

high amount of uninformative gene trees for the nodes, *i.e.* highly unresolved gene trees, might be explained by alignments that do not present enough variation. This would be the consequence of a too low number of substitutions during this rapid radiation to detect branching events (Smith & al., 2015). The high proportion of low-frequency topologies at most nodes might be due to incomplete lineage sorting, *i.e.* polymorphisms did not have enough time to sort within a lineage (Degnan & Rosenberg, 2009).

In order to resolve this, the inclusion of more genes rather than of more Southeast Asian species would be needed. If newly included species are placed within the rapid radiation clade, as would be expected, the already short terminal branches would become even shorter and probably lead to even lower support. Including more genes however, would increase the possibility of finding the dominant topology, and increase accuracy as in Rabiee & al. (2018). Our samples which have not yet been sequenced include 795 more targeted genes than the samples presented here, and twelve new Southeast Asian species. Because of the higher number of genes, but also species, we cannot speculate on whether these samples will improve accuracy within the rapidly radiating clade or not.

Within this Southeast Asian clade, especially the positions of *O. lauterbachiana* and *O. palindan* are uncertain. Different individual sequences of the same species are polyphyletic with high support. One might suggest that these individuals are hence different species. However, this conclusion cannot be made because it is possible that these individuals are not resolved as a clade due to weakly resolved gene trees within this radiating clade and/or homoplastic similarities with other species of the SE Asian clade. Species monophyly might be resolved when including our new samples which might improve general resolution (see above). Especially two additional individual sequences of *O. palindan* of our samples might resolve this species' polyphyly. If the individuals are still not monophyletic, the species definition should be questioned (Rabiee & al., 2018) and molecular approaches for species delimitation will be needed (Yang & Rannala, 2010; *e.g.* Smith & al., 2014).

Another uncertainty were the intergeneric relationships. Whether *Orania* is sister to *Sclerosperma*, or sister to *Sclerosperma* and *Podococcus*, could not be resolved. Our phylogenetic inference presents the latter topology $[(O, (S, P))]$, as in Lewis & Doyle (2002) and Comer & al. (2015), but only with low support in LPP and only one third of non-conflicting gene trees. The top alternative topology at this node places *Orania* as sister to *Sclerosperma* $[((O, S), P)]$, as in Baker & Couvreur

(2013a) and Comer & al. (2016). This topology was concordant with one fourth of the gene trees. One possible reason for having these two competing topologies could be ancient incomplete lineage sorting. For this process to occur, at least one short internal branch is needed (Degnan & Rosenberg, 2009), which is concordant with the short branch between the stem (node VI) and crown node (node I) of clade *Podococcus-Sclerosperma* (3 My without 95% HPD). Another possible scenario for these competing topologies would be a non-coalescent-neutral processes, as *e.g.* a hybridization event between *Sclerosperma* and *Orania*, or *Sclerosperma* and *Podococcus*.

4.2 Biogeographic explanations for the recent radiation in SE Asian *Orania*

Orania has experienced a recent burst of diversification within SE Asia since approximately 5 Mya. This recent radiation is surprising since we inferred that *Orania* and its ancestors have been present in SE Asia since at least the early Eocene. The inferred timing of the radiation coincides with the Pliocene fragmentation of the Sundaland block due to sea-level fluctuations in the Plio-Pleistocene (Zhong & al., 2004 *in* Bendiksby & al., 2010; de Bruyn & al., 2014). This might have induced a chain of vicariant speciation events resulting in speciation as already suggested in Bendiksby & al. (2010). The reconstruction of the biogeographic history of *Orania* within SE Asia is needed to confirm this hypothesis. Our study did not allow a reconstruction on such a level because the phylogeny was not well resolved within the SE Asian clade. The inclusion of not-yet sequenced samples might resolve the biogeographic history. An additional approach would be including the phylogenetic uncertainties into the ancestral range estimation through estimation of the ancestral range on a set of trees, as in *e.g.* Peter Linder & al. (2013).

4.3 Biogeographic pathways of the POS clade and origin of the Madagascan *Orania* clade

Orania was inferred to have arrived to Madagascar from Southeast Asia between 58 Mya and 16 Mya in a single colonization event. However, this interval can be narrowed down. The ancestor of *Orania* was unlikely to have persisted with a widespread SE Asian-Madagascan distribution for a very long time because sustained gene flow between the populations of these two regions was nearly impossible because of the sheer distance between SE Asia and Madagascar acting as a barrier. Hence, the colonization event of Madagascar should have happened very close to the

timing of the split between the Madagascan *Orania* and the SE Asian *Orania*, *i.e.* during the Oligocene or early Miocene (between 31 Mya and 16 Mya; node II including 95% HPD). This timing is consistent with westward marine surface currents and wind patterns during this time (Federman & al., 2015).

This inferred long-distance dispersal (LDD) event over the Indian Ocean to Madagascar implies two possibilities of dispersal mechanisms: dispersal by flying animals or dispersal by water. Dispersal mechanisms in *Orania* are little known, but it has been proposed that besides from wild pigs and cassowaries (flightless since at least 60 Mya; Maderspacher 2017), the fruits could be dispersed by flying foxes (especially *Pteropus* sp.) (Keim & Dransfield, 2012). However, these flying foxes dispersed only very recently during the Quaternary from Southeast Asia towards the Western Indian Ocean (Tsang & al., 2020) and could not have been involved in the ocean crossing of *Orania*. An argument against zochory is that *Orania* might be poisonous (Burkill & al., 1935), but this has not been confirmed yet and is according to Keim & Dransfield (2012) likely to be a myth. Also dispersal by water has been proposed (Keim & Dransfield, 2012; Dransfield & Rakotoarinivo, 2011), since adult individuals are sometimes found along water-courses. Fruits of *Orania* could have drifted over the Indian Ocean via westward ocean currents. It has not been tested whether fruits of the most recent common ancestor of *Orania* were capable of surviving such LDD. However, instead of floating in sea water, rafting would have been another possibility for the ocean crossing. Rafting as a possible explanation for LDD has already been proposed by *e.g.* Thiel & Gutow (2004) and is in our case the most likely dispersal mode for *Orania* over the Indian Ocean.

Our result of westward long-distance dispersal event is in part based on the biogeographic constraint of the split between the Core Arecoids and the POS clade to Southeast Asia and continental Eurasia, derived from Baker & Couvreur (2013a). Instead of taking into account the sister clade of the POS clade to determine the ancestral range of this split, this "secondary range calibration" was applied, because the sister of the POS clade is the enormously large Core Arecoids clade (> 1000 species). This estimated ancestral range is prone to change once a new species-level tree of all Arecoids is available, which could also influence the direction of the here-inferred dispersal event.

The result of dispersal over the Indian Ocean is in contrast with Comer & al. (2016), who defended the vicariance hypothesis claiming that *Orania* reached Southeast Asia from Madagascar using Greater India as a "raft" after its split from

Madagascar. However, their study presents a few conceptual and technical flaws concerning this conclusion on the biogeographic history of *Orania*. First, the split between Madagascar and India happened at between 116 Mya and 84 Mya (Ali & Aitchison, 2008), long time before the emergence of the POS clade (57 Mya [45–69, 95% HPD], Baker & Couvreur, 2013a, our calibration point), and vicariance is hence ruled out, as already suggested by Baker & Couvreur (2013b). Second, they extrapolate the results of the ancestral range estimation of Comer & al. (2015) onto their phylogeny, without having the same topology within the POS clade. Third, this ancestral range estimation of Comer & al. (2015) is erroneous, because the stem range of each POS tribe was estimated to be, besides Africa, South and North America which is impossible to have been inferred based on the current ranges of the tribes. Lastly and maybe most importantly, their conclusions on *Orania*'s biogeography were only based on one *Orania* species (the Asian *Orania palindan*) and hence are not valid for the inference of the biogeographic origin of Madagascan *Orania* species.

5 Conclusion

This study contributed a key element to the species-level phylogeny of Madagascan palms. Moreover, the arrival of *Orania* from Southeast Asia to Madagascar highlights the importance of long-distance dispersal events over the Indian Ocean. To conclude, this study is another step towards understanding colonization and following diversification patterns in Madagascan palms.

Supplementary Material

All scripts and the following supplementary material can be found at:

https://github.com/pebgroup/pos_phylo_biogeo.

Data S1: Alignments. 63 fasta files of alignments generated with MAFFT, trimmed with GBlocks and manually cleaned. Each fasta file corresponds to the alignment of one gene. The faster headers correspond to the individual names in "TAG notation". A text file is given for the conversion of the TAG notation to the species name for each individual.

Data S2: Gene trees. 63 newick files of gene trees generated with RAxMLng. Each newick file corresponds to the gene tree of one gene. The faster headers correspond to the individual names in "TAG notation". A text file is given for the conversion of the TAG notation to the species name for each individual. A pdf file is given showing all visualized gene trees.

Table S1: HybPiper statistics. Summary of target enrichment and gene recovery efficiency for all 20 samples, output by HybPiper. The columns correspond respectively to: Individual name, Number of reads, Number of reads on target, Percent reads on target, Number of genes with reads, Number of genes with contigs, Number of genes with sequences, Number of genes with sequences > 25% of the target length, Number of genes with sequences > 50% of the target length, Number of genes with sequences > 75% of the target length, Number of genes with sequences > 150% of the target length, Number of genes with paralog warnings, as described in the HybPiper Wiki (<https://github.com/moosmatters/HybPiper/wiki>).

Table S2: Biogeographic model fit statistics from BioGeoBEARS for the estimation without node constraint. The first part of the model name corresponds to the range evolution model: dec (DEC), dl (DIVALIKE), b (BAYAREALIKE); and the second part of the model name to the dispersal probability models: no dispersal probability constraints (M_0), dispersal probability constraints as in Baker & Couvreur (2013a) (B: M_B), dispersal probability constraints as in Federman & al. (2015) (F: M_F). LnL: Log likelihood; numparams: number of parameters in the model; d: estimated dispersal rate; e: estimated extinction (=extirpation) rate; AICc: AICc score; AICc_wt: AICc weight.

Table S3: Biogeographic model fit statistics from BioGeoBEARS for the estimation with node constraint. The node between the outgroup (core Arecooids) and the POS clade was constraint to the range Southeast Asia/Eurasia, based on Baker & Couvreur (2013a). The first part of the model name corresponds to the range evolution model: dec (DEC), dl (DIVALIKE), b (BAYAREALIKE); and the second part of the model name to the dispersal probability models: no dispersal probability constraints (M_0), dispersal probability constraints as in Baker & Couvreur (2013a) (B: M_B), dispersal probability constraints as in Federman & al. (2015) (F: M_F). LnL: Log likelihood; numparams: number of parameters in the model; d: estimated dispersal rate; e: estimated extinction (=extirpation) rate; AICc: AICc score; AICc_wt: AICc weight.

Figure S1: HybPiper heatmap. Visualization of recovery efficiency, figure created by HybPiper. Each row shows a sample, and each column is a gene. The amount of shading in each box corresponds to the length of the gene recovered for that sample, relative to the length of the target coding sequence (cf. HybPiper Wiki <https://github.com/mossmatters/HybPiper/wiki>).

Figure S2: ASTRAL cladogram of individuals with PhyParts as pie charts with node numbers and gene tree support values. The pie charts are based on results from PhyParts and show the proportion of gene trees that are concordant with the species tree topology at the corresponding node (green), that support the main alternative topology (orange), that support other alternative topologies (red), and the proportion of gene trees that are uninformative for that node (black). The values above and below the branches represent respectively the number of concordant and conflicting gene trees. The numbers on grey represent the node numbers.

Figures S3: Each figure corresponds to one node: Top alternative topology at corresponding node generated with PhyParts. Node numbers correspond to the node numbers on the phylogeny in Figure S2. The number of agreeing gene trees with this alternative topology corresponds to the orange part of the pie chart in Figure 1b and Figure S2.

Figure S4: Cladogram of individuals when constraining the individuals of *O. lauterbachiana* to be monophyletic with the ASTRAL option "-j". Support values are given as local posterior probabilities on top of each branch. The red box indicates the *O. lauterbachiana* clade.

Figure S5: Cladogram of individuals when constraining the individuals of *O. palindan* to be monophyletic with the ASTRAL option "-j". Support values are given as local posterior probabilities on top of each branch. The red box indicates the *O. palindan* clade.

Figure S6: Maximum clade credibility chronogram of the POS clade using BEAST based on 23 selected genes. These genes were selected based on the least topology conflict of the gene trees with the species tree. The blue bars indicate the 95% highest posterior density intervals. The number above each node indicates the node age and the number below the node indicates the lowest and highest 95% highest posterior density.

Figure S7: Maximum clade credibility chronogram of the POS clade using BEAST based on all 63 genes. The blue bars indicate the 95% highest posterior density intervals. The number above each node indicates the node age and the number below the node indicates the lowest and highest 95% highest posterior density.

Figure S8: Biogeographic ancestral range estimates using BioGeoBears without biogeographic node constraint on the chronogram of Figure 2 without the outgroup (*i.e.* *Dypsis mananjarensis*). At the top left is represented which colour corresponds to which geographical area: grey (Africa); orange (Madagascar); blue (Southeast Asia = Malesia + Papua). The colours of the tip labels correspond to the current range of each species. The parts within the pie charts correspond to the range probabilities at the corresponding node. The striped parts indicate the presence in multiple areas. The different sizes of the pie charts are only for visualization purposes. Pie charts on the stem nodes of the tips are not represented; the range probability is 1 for all corresponding tip ranges at these nodes. The nodes of interest are labelled with Roman numerals and grey boxes correspond to the numbers added to those in the cladogram (Figure 1). The Pliocene is abbreviated as "Plio." and the Pleistocene as "Pleisto."

Text S1: Dispersal probability matrices for ancestral range estimation. Text explaining how dispersal probability matrices were obtained for the two models: M_B (matrices adapted from Baker & Couvreur (2013a)) and M_F (matrices adapted from Federman & al. (2015)); and the resulting dispersal probability matrices for each model.

Bibliography

- Abadi, S., Azouri, D., Pupko, T., & Mayrose, I.** 2019. Model selection may not be a mandatory step for phylogeny reconstruction. *Nature Communications*, 10(1): 1–11. doi:10.1038/s41467-019-08822-w.
- Agnarsson, I. & Kuntner, M.** 2012. The Generation of a Biodiversity Hotspot: Biogeography and Phylogeography of the Western Indian Ocean Islands. Pp. 33–84 in: Anamthawat-Jnsson, K. (ed.), *Current Topics in Phylogenetics and Phylogeography of Terrestrial and Aquatic Systems*. InTech. doi:10.5772/38958.
- Ali, J.R. & Aitchison, J.C.** 2008. Gondwana to Asia: Plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166–35 Ma). *Earth-Science Reviews*, 88(3-4): 145–166. doi:10.1016/j.earscirev.2008.01.007.
- Ali, J.R. & Huber, M.** 2010. Mammalian biodiversity on Madagascar controlled by ocean currents. *Nature*, 463(7281): 653–656. doi:10.1038/nature08706.
- Andrews, S.** 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. [Online]. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- 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. *Botanical Journal of the Linnean Society*, 151(1): 15–38. doi:10.1111/j.1095-8339.2006.00521.x.
- Bacon, C.D., Simmons, M.P., Archer, R.H., Zhao, L.C., & Andriantiana, J.** 2016. Biogeography of the Malagasy Celastraceae: Multiple independent origins followed by widespread dispersal of genera from Madagascar. *Molecular Phylogenetics and Evolution*, 94: 365–382. doi:10.1016/j.ympev.2015.09.013.
- 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. *Journal of Biogeography*, 40(2): 274–285. doi: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. *Journal of Biogeography*, 40(2): 286–298. doi:10.1111/j.1365-2699.2012.02794.x.
- 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., Pyshkin, A.V., Sirotnik, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., & Pevzner, P.A.** 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, 19(5): 455–477. doi:10.1089/cmb.2012.0021.

- Bellot, S., Bayton, R.P., Couvreur, T.L.P., Dodsworth, S., Eiserhardt, W.L., Pritchard, H., Roberts, L., Toorop, P., & Baker, W.J. *in press*. On the origin of giant seeds: Insights from the double coconut (*Lodoicea maldivica*) and close relatives (Borasseae, Arecaceae). *New Phytologist*.
- Bendiksby, M., Schumacher, T., Gussarova, G., Nais, J., Mat-Salleh, K., Sofiyanti, N., Madulid, D., Smith, S.A., & Barkman, T. 2010. Elucidating the evolutionary history of the Southeast Asian, holoparasitic, giant-flowered Rafflesiaceae: Pliocene vicariance, morphological convergence and character displacement. *Molecular Phylogenetics and Evolution*, 57(2): 620--633. doi:10.1016/j.ympev.2010.08.005.
- Bilderbeek, R.J.C. & Etienne, R.S. 2018. Babette: BEAUti 2, BEAST2 and Tracer for R. *Methods in Ecology and Evolution*, 9(9): 2034--2040. doi:10.1111/2041-210X.13032.
- Bolger, A.M., Lohse, M., & Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15): 2114--2120. doi:10.1093/bioinformatics/btu170.
- Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., Maio, N.D., 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 Computational Biology*, 15(4): e1006650. doi:10.1371/journal.pcbi.1006650.
- Brewer, G.E., Clarkson, J.J., Maurin, O., Zuntini, A.R., Barber, V., Bellot, S., Biggs, N., Cowan, R.S., Davies, N.M.J., Dodsworth, S., Edwards, S.L., Eiserhardt, W.L., Epitawalage, N., Frisby, S., Grall, A., Kersey, P.J., Pokorny, L., Leitch, I.J., Forest, F., & Baker, W.J. 2019. Factors Affecting Targeted Sequencing of 353 Nuclear Genes From Herbarium Specimens Spanning the Diversity of Angiosperms. *Frontiers in Plant Science*, 10. doi:10.3389/fpls.2019.01102.
- Brown, J.W., Walker, J.F., & Smith, S.A. 2017. Phyx: Phylogenetic tools for unix. *Bioinformatics (Oxford, England)*, 33(12): 1886--1888. doi:10.1093/bioinformatics/btx063.
- Burkill, I.H., Watson, J.G., Scrivenor, J.B., Foxworthy, F.W., & Birtwistle, W. 1935. *A Dictionary of the Economic Products of the Malay Peninsula*. London: Published on behalf of the governments of the Straits settlements and Federated Malay states by the Crown agents for the colonies.
- Callmander, M.W., Phillipson, P.B., Schatz, G.E., Andriambololonera, S., Rabarimanarivo, M., Rakotonirina, N., Raharimampionona, J., Chatelein, C., Gautier, L., & Lowry II, P.P. 2011. The endemic and non-endemic

- vascular flora of Madagascar updated. *Plant Ecology and Evolution*, 144(2): 121-125. doi:[10.5091/plecevo.2011.513](https://doi.org/10.5091/plecevo.2011.513).
- Castresana, J.** 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*, 17(4): 540--552. doi:[10.1093/oxfordjournals.molbev.a026334](https://doi.org/10.1093/oxfordjournals.molbev.a026334).
- Comer, J.R., Zomlefer, W.B., Barrett, C.F., Davis, J.I., Stevenson, D.W., Heyduk, K., & Leebens-Mack, J.H.** 2015. Resolving relationships within the palm subfamily Arecoideae (Arecaceae) using plastid sequences derived from next-generation sequencing. *American Journal of Botany*, 102(6): 888--899. doi:[10.3732/ajb.1500057](https://doi.org/10.3732/ajb.1500057).
- 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). *Molecular Phylogenetics and Evolution*, 97: 32--42. doi:[10.1016/j.ympev.2015.12.015](https://doi.org/10.1016/j.ympev.2015.12.015).
- Couvreur, T.L. & Baker, W.J.** 2013. Tropical rain forest evolution: Palms as a model group. *BMC Biology*, 11(1): 48. doi:[10.1186/1741-7007-11-48](https://doi.org/10.1186/1741-7007-11-48).
- Couvreur, T.L.P., Helmstetter, A.J., Koenen, E.J.M., Bethune, K., Brandão, R.D., Little, S.A., Sauquet, H., & Erkens, R.H.J.** 2019. Phylogenomics of the Major Tropical Plant Family Annonaceae Using Targeted Enrichment of Nuclear Genes. *Frontiers in Plant Science*, 9. doi:[10.3389/fpls.2018.01941](https://doi.org/10.3389/fpls.2018.01941).
- 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. *Systematic Biology*, 63(6): 879--901. doi:[10.1093/sysbio/syu047](https://doi.org/10.1093/sysbio/syu047).
- Degnan, J.H. & Rosenberg, N.A.** 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, 24(6): 332--340. doi:[10.1016/j.tree.2009.01.009](https://doi.org/10.1016/j.tree.2009.01.009).
- Doyle, J.J. & Doyle, J.L.** 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Technical report.
- Dransfield, J. & Beentje, H.** 1995. *The Palms of Madagascar*. Royal Botanic Gardens.
- Dransfield, J. & Rakotoarinivo, M.** 2011. The biogeography of Madagascar palms. Pp. 179--196 in: Bramwell, D. & Caujape-castells, J. (eds.), *The Biology of Island Floras*. Cambridge: Cambridge University Press. doi:[10.1017/CBO9780511844270.008](https://doi.org/10.1017/CBO9780511844270.008).
- Dransfield, J. & Uhl, N.** 1984. *Halmoorea*, a new genus from Madagascar, with notes on *Sindroa* and *Orania*. *Principes (USA)*.

- Dransfield, J. & Uhl, N.W.** (eds.) 2008. *Genera Palmarum: The Evolution and Classification of Palms*, Kew Plants People Possibilities, new edition. Kew Publ. Royal Botanical Garden.
- Essig, F.B.** 1980. The Genus *Orania* Zipp (Arecaceae) in New Guinea. *Lyonia*, 1(5): 211--233.
- Federman, S., Dornburg, A., Downie, A., Richard, A.F., Daly, D.C., & Donoghue, M.J.** 2015. The biogeographic origin of a radiation of trees in Madagascar: Implications for the assembly of a tropical forest biome. *BMC Evolutionary Biology*, 15(1): 216. [doi:10.1186/s12862-015-0483-1](https://doi.org/10.1186/s12862-015-0483-1).
- Gardner, E.M., Johnson, M.G., Pereira, J.T., Puad, A.S.A., Arifiani, D., Sahromi, Wickett, N.J., & Zerega, N.J.C.** 2019. Paralogs and off-target sequences improve phylogenetic resolution in a densely-sampled study of the breadfruit genus (*Artocarpus*, Moraceae). *bioRxiv*, P. 854232. [doi:10.1101/854232](https://doi.org/10.1101/854232).
- 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. *Biological Journal of the Linnean Society*, 117(1): 106--120. [doi:10.1111/bij.12551](https://doi.org/10.1111/bij.12551).
- Hoff, M., Orf, S., Riehm, B., Darriba, D., & Stamatakis, A.** 2016. Does the choice of nucleotide substitution models matter topologically? *BMC Bioinformatics*, 17(1): 143. [doi:10.1186/s12859-016-0985-x](https://doi.org/10.1186/s12859-016-0985-x).
- Hurvich, C.M. & Tsai, C.L.** 1989. Regression and time series model selection in small samples. *Biometrika*, 76(2): 297--307. [doi:10.1093/biomet/76.2.297](https://doi.org/10.1093/biomet/76.2.297).
- 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 enrichment1. *Applications in Plant Sciences*, 4(7). [doi:10.3732/apps.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(13): 1669--1670. [doi:10.1093/bioinformatics/btq243](https://doi.org/10.1093/bioinformatics/btq243).
- 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. [doi:10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).
- Keim, A.P. & Dransfield, J.** 2012. A monograph of the genus *Orania* (Arecaceae: Oranieae). *Kew Bulletin*, 67(2): 127--190. [doi:10.1007/s12225-012-9356-6](https://doi.org/10.1007/s12225-012-9356-6).
- Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A.** 2019. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21): 4453--4455. [doi:10.1093/bioinformatics/btz305](https://doi.org/10.1093/bioinformatics/btz305).

- Landis, M.J., Matzke, N.J., Moore, B.R., & Huelsenbeck, J.P.** 2013. Bayesian Analysis of Biogeography when the Number of Areas is Large. *Systematic Biology*, 62(6): 789--804. doi:[10.1093/sysbio/syt040](https://doi.org/10.1093/sysbio/syt040).
- Larsson, A.** 2014. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22): 3276--3278. doi:[10.1093/bioinformatics/btu531](https://doi.org/10.1093/bioinformatics/btu531).
- Lewis, C.E. & Doyle, J.J.** 2002. A phylogenetic analysis of tribe Areceae (Arecaceae) using two low-copy nuclear genes. *Plant Systematics and Evolution*, 236(1): 1--17. doi:[10.1007/s00606-002-0205-1](https://doi.org/10.1007/s00606-002-0205-1).
- Li, H. & Durbin, R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14): 1754--1760. doi:[10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324).
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup** 2009. The Sequence Alignment/Map format and SAM-tools. *Bioinformatics (Oxford, England)*, 25(16): 2078--2079. doi:[10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352).
- 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. *Frontiers in Plant Science*, 10. doi:[10.3389/fpls.2019.00864](https://doi.org/10.3389/fpls.2019.00864).
- Maderspacher, F.** 2017. Evolution: Flight of the Ratites. *Current Biology*, 27(3): R110-R113. doi:[10.1016/j.cub.2016.12.023](https://doi.org/10.1016/j.cub.2016.12.023).
- Martin, M.** 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1): 10--12. doi:[10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200).
- Matzke, N.J.** 2013. Probabilistic historical biogeography: New models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography*, (5): 242--248. doi:[10.21425/F5FBG19694](https://doi.org/10.21425/F5FBG19694).
- Meerow, A.W., Noblick, L., Salas-Leiva, D.E., Sanchez, V., Francisco-Ortega, J., Jestrow, B., & Nakamura, K.** 2015. Phylogeny and historical biogeography of the cocosoid palms (Arecaceae, Arecoideae, Cocoseae) inferred from sequences of six WRKY gene family loci. *Cladistics*, 31(5): 509--534. doi:[10.1111/cla.12100](https://doi.org/10.1111/cla.12100).

- Morley, R.J.** 2003. Interplate dispersal paths for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics*, 6(1): 5--20. doi: 10.1078/1433-8319-00039.
- Myers, A.A. & Giller, P.** 1988. *Analytical Biogeography: An Integrated Approach to the Study of Animal and Plant Distributions*. Springer Science & Business Media.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., & Kent, J.** 2000. Biodiversity hotspots for conservation priorities. *Nature*, 403(6772): 853--858. doi:10.1038/35002501.
- Paradis, E. & Schliep, K.** 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35: 526--528.
- Peter Linder, H., Antonelli, A., Humphreys, A.M., Pirie, M.D., & Wüest, R.O.** 2013. What determines biogeographical ranges? Historical wanderings and ecological constraints in the danthonioid grasses. *Journal of Biogeography*, 40(5): 821--834. doi:10.1111/jbi.12070.
- R Core Team** 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria.
- Rabiee, M. & Mirarab, S.** 2020. Forcing external constraints on tree inference using ASTRAL. *BMC Genomics*, 21(Suppl 2). doi:10.1186/s12864-020-6607-z.
- Rabiee, M., Sayyari, E., & Mirarab, S.** 2018. Multi-allele species reconstruction using ASTRAL. *bioRxiv*, P. 439489. doi:10.1101/439489.
- Rambaut, A.** 2009. FigTree v1.4.4. [Online]. <http://tree.bio.ed.ac.uk/software/figtree/>.
- 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. doi:10.1093/sysbio/syy032.
- 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. doi:10.1111/jbi.13173.
- Ree, R.H. & Smith, S.A.** 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, 57(1): 4--14. doi:10.1080/10635150701883881.
- Ronquist, F.** 1997. Dispersal-Vicariance Analysis: A New Approach to the Quantification of Historical Biogeography. *Systematic Biology*, 46: 9.

- Slater, G.S.C. & Birney, E.** 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6(1): 31. doi:10.1186/1471-2105-6-31.
- Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C., & Brumfield, R.T.** 2014. Target Capture and Massively Parallel Sequencing of Ultraconserved Elements for Comparative Studies at Shallow Evolutionary Time Scales. *Systematic Biology*, 63(1): 83–95. doi:10.1093/sysbio/syt061.
- 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): e0197433. doi:10.1371/journal.pone.0197433.
- Smith, S.A., Moore, M.J., Brown, J.W., & Yang, Y.** 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology*, 15(1): 150. doi:10.1186/s12862-015-0423-0.
- Thiel, M. & Gutow, L.** 2004. The Ecology of Rafting in the Marine Environment. I. the Floating Substrata. Pp. 181–263 in: Gibson, R., Atkinson, R., & Gordon, J. (eds.), *Oceanography and Marine Biology*, volume 20042243. CRC Press. doi:10.1201/9780203507810.ch6.
- Trénel, P., Gustafsson, M.H., 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). *Molecular Phylogenetics and Evolution*, 45(1): 272–288. doi:10.1016/j.ympev.2007.03.018.
- Tsang, S.M., Wiantoro, S., Veluz, M.J., Sugita, N., Nguyen, Y.L., Simmons, N.B., & Lohman, D.J.** 2020. Dispersal out of Wallacea spurs diversification of Pteropus flying foxes, the world's largest bats (Mammalia: Chiroptera). *Journal of Biogeography*, 47(2): 527–537. doi:10.1111/jbi.13750.
- Warren, B.H., Strasberg, D., Bruggemann, J.H., Prys-Jones, R.P., & Thébaud, C.** 2010. Why does the biota of the Madagascar region have such a strong Asiatic flavour? *Cladistics*, 26(5): 526–538. doi:10.1111/j.1096-0031.2009.00300.x.
- WCSP**. World Checklist of Selected Plant Families. Royal Botanic Gardens, Kew. [Online]. <http://wcsp.science.kew.org/>.
- Yang, Z. & Rannala, B.** 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, 107(20): 9264–9269. doi:10.1073/pnas.0913022107.
- Yoder, A.D. & Nowak, M.D.** 2006. Has Vicariance or Dispersal Been the Predominant Biogeographic Force in Madagascar? Only Time Will Tell. *Annual Review of Ecology, Evolution, and Systematics*, 37(1): 405–431. doi:10.1146/annurev.ecolsys.37.091305.110239.

- Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S.** 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19(6): 153. doi:[10.1186/s12859-018-2129-y](https://doi.org/10.1186/s12859-018-2129-y).
- Zhong, G., Geng, J., Wong, H.K., Ma, Z., & Wu, N.** 2004. A semi-quantitative method for the reconstruction of eustatic sea level history from seismic profiles and its application to the southern South China Sea. *Earth and Planetary Science Letters*, 223(3-4): 443–459. doi:[10.1016/j.epsl.2004.04.039](https://doi.org/10.1016/j.epsl.2004.04.039).
- Zhou, L., Su, Y.C.F., Thomas, D.C., & Saunders, R.M.K.** 2012. ‘Out-of-Africa’ dispersal of tropical floras during the Miocene climatic optimum: Evidence from Uvaria (Annonaceae): Out-of-Africa dispersal during Miocene climatic optimum. *Journal of Biogeography*, 39(2): 322–335. doi:[10.1111/j.1365-2699.2011.02598.x](https://doi.org/10.1111/j.1365-2699.2011.02598.x).

