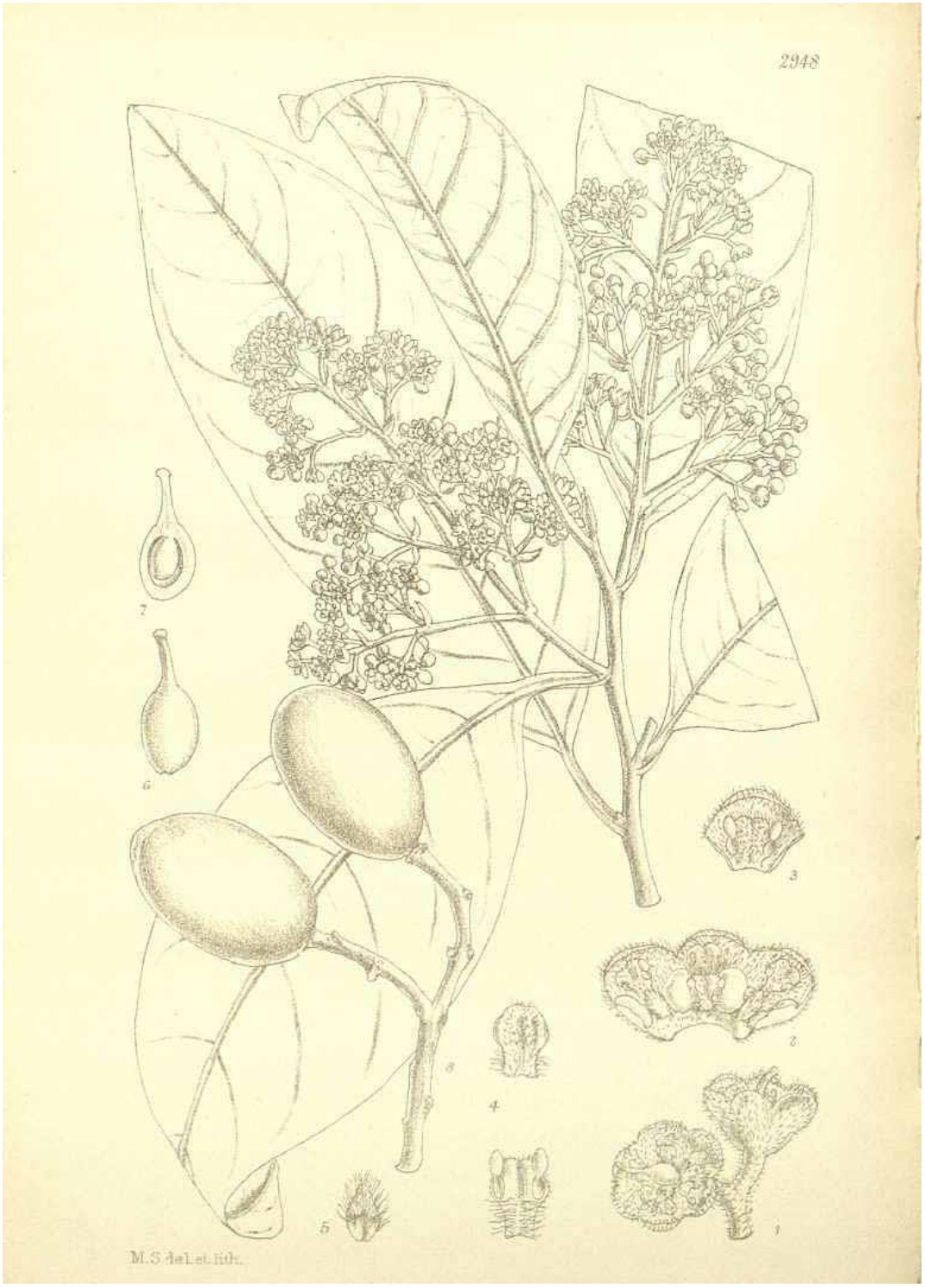
# SYSTEMATICS AND BIOGEOGRAPHY OF THE CRYPTOCARYA GROUP (LAURACEAE)

Laura Kragh Frederiksen



# Systematics and biogeography of the Cryptocarya group (Lauraceae)

60 ECTS MSc Thesis

by

Laura Kragh Frederiksen
Student ID: 201805319

Main supervisor: Wolf L. Eiserhardt
Co-supervisor: Jens G. Rohwer

Master of Biology

2023

Section for Ecoinformatics and Biodiversity
Aarhus University
Nordre Ringgade 1
8000 Aarhus C

Front cover:
Drawing of *Beilschmiedia ugandensis* Rendle by Matilda Smith (1911)

AARHUS UNIVERSITET

# ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Wolf Eiserhardt, for introducing me to the field of biogeography, which gave me the opportunity to develop a new skillset. I am thankful for his invaluable guidance, encouragement, and support throughout the entire process. Particularly I am grateful for Wolfs positive and solution-oriented approach to problems I encountered during the process.

I am also deeply grateful to my co-supervisor Jens Rohwer for introducing me to Lauraceae, and for providing very helpful feedback and suggestions along the way, given his expert knowledge on Lauraceae. Moreover, for providing the DNA extractions, which made this project possible.

I would also like to extend my thanks to the members of the PEB group, who all from day one provided a very welcoming environment and helpful discussions. I would like to give a special thanks to Paola for always being willing to help with the pipeline. Additionally, my office mates, Oscar, Lars Emil, and Josephine, who ensured an always good mood in the office. A special thanks to Oscar for introducing me to gwf and happily helping me troubleshoot and to Lars Emil for helping me 'tame the BEAST'. In this context I would also like to thank Victor Calle. Also, Camilla for her prompt preparation of DNA libraries in the lab.

Furthermore, I would like to thank all the people at Section for Ecoinformatics and Biodiversity I have come across during my project, who have all in some way contributed to my project and to the good atmosphere in the halls and at the weekly breakfast meetings. Here I would like to extend my deepest gratitude to Anders Barfod, who has been an invaluable support throughout my studies. If it hadn't been for Anders, I might never have developed a deep fascination for plants, but might still have believed that fish physiology was the way forward... Also, a special thank you for supporting my application to study at RBG Kew. Moreover, I would like to thank Henrik Balslev for many good talks and support. Most of all for believing in me and supporting my ambition to undertake a PhD.

I would also like to thank all my friends. Particularly, I would like to thank my study group, Amalie, Karen, Linette, and Mira. Without you these five years at Aarhus University would not have been the same. I am deeply grateful that you are a part of my life. An equally big thank you goes to Henriette, who always supports me and whose friendship means the world to me.

Lastly, I would like to express my deepest appreciation to my family for their unwavering love, understanding, and patience. Especially to my boyfriend René for always being my biggest support, and for always being willing to put his own needs aside, to be there for me. Also, to my mum for always providing me with endless love, and support. She made me see the positive and find solutions when things looked difficult. Without them this thesis would not have been possible. I love you.

# TABLE OF CONTENTS

# Part A: General introduction

## 1. Historical biogeography: What lives where and why?

A fundamental question in historical biogeography concerns how and when the current distribution and diversity of species was shaped (Sanmartín, 2012). In general, closely related species tend to occupy similar ranges, which reflects their shared evolutionary history (Carta et al., 2022; Donoghue, 2008). Yet, an intriguing aspect of this dates to Humboldt and Bonpland (1805) who questioned how closely related species can exhibit disjunct distributions (Lomolino et al., 2010). To understand this, it is fundamental to estimate the divergence time and ancestral ranges of species to uncover the evolutionary processes that might have shaped the current distribution. Two processes, long-distance dispersal (LDD) and vicariance (see Figure 1), have historically been recognised as two competing key processes, in shaping species distributions. Dispersal theory, which predominated until the recognition of plate tectonic theory (Nelson & Platnick, 1982), presupposes that species reach new ranges through movements, such as long-distance jump dispersals, to overcome a pre-existing geographic barrier. In contrast, the development of cladistic theory (Hennig, 1966) and the subsequent emergence of vicariance (Croizat et al., 1958 as cited in Morrone & Crisci, 1995) led to a paradigm shift within the field of biogeography. In the context of vicariance theory disjunct species distributions are instead attributed to separation of a formerly continuous population due to the formation of a barrier, such as geographical changes in the environment e.g., the opening of oceans or formation of mountains (Cox & Moore, 2010).

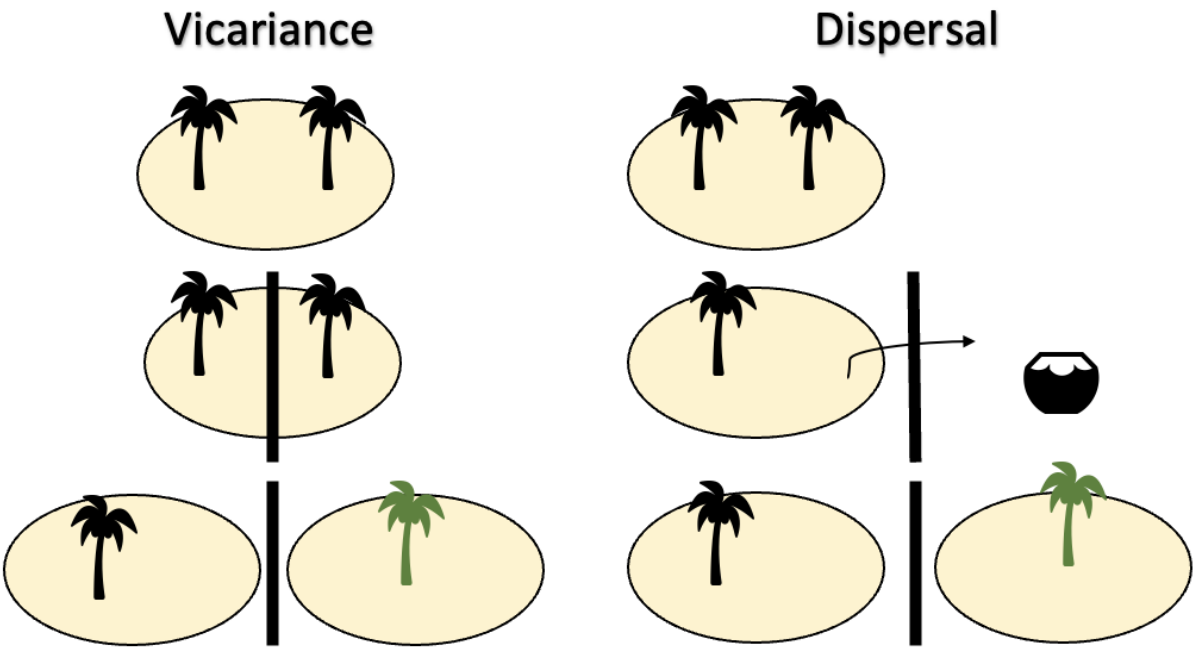



**Figure 1** Adapted from Humphries et al. (2017). Figure displaying the mechanism behind vicariance (left) and dispersal (right). See text for more information.

Currently advancements in biogeographical reconstruction methods have enabled us to assess the potential and challenges of both vicariance and dispersal. This can be achieved by examining divergence times, ancestral distributions, and the geographic patterns of related species. The current goal is therefore to understand the relative contributions of each process in shaping distributions (Bessa-Silva et al., 2020) .

### 1.1 Case: pantropical plant lineages

Terrestrial tropical biomes, specifically tropical rainforests (TRFs), are renowned for their extraordinary species richness and biodiversity, supporting close to half of the world's plant species on just c. 7% of the land surface (Hill & Hill, 2016; Lomolino et al., 2010). Presently, in the Holocene TRF's are distributed across the equatorial zone, spanning the Americas, Africa, tropical Asia, and Australia. Yet, many plant lineages (William J. Baker et al., 2013; Bardon et al., 2016; Bartish et al., 2011; Bechteler et al., 2017; Clayton et al., 2009; Conceição Oliveira et al., 2021; Couvreur, Forest, et al., 2011; Couvreur, Pirie, et al., 2011; Givnish & Renner, 2004; Lee et al., 2020; Nie et al., 2013; Schneider et al., 2022; Wang et al., 2012; Yang et al., 2018; Zerega et al., 2005; Zhao et al., 2022) are pantropical in spite of large intercontinental disjunctions. So, one may raise the question: How have these plant lineages overcome massive geographical barriers to achieve a pantropical distribution? To answer this question, we need to fast-backwards in time, to comprehend what the world looked like and how the climate was many million years ago (see Figure 2.). Based on this understanding of the Earths development through time as response to tectonic movements and periods of increasing and decreasing temperatures, several hypotheses that built upon dispersal and vicariance have been developed, to explain how these intercontinental transoceanic disjunctions might have developed. These hypotheses include Gondwanan vicariance, which refers to the breakup of the Gondwanan supercontinent (Axelrod, 1970); geodispersal, involving migration across land bridges such as the Northern Atlantic land bridge (NALB) during favourable climatic conditions (Tiffney, 1985), continental rafting, where land masses e.g., India, acts as rafts during tectonic plate movements (Conti et al., 2002), and LDD facilitated by different dispersal vectors such as wind (Muñoz et al., 2004), water (Harries & Clement, 2014), or birds (Onstein et al., 2019). Thus, if we understand the evolution of pantropical plant lineages, such as the TRF restricted palms (William J. Baker et al., 2013), we may be able to gain important insight into which processes have shaped todays rainforest.

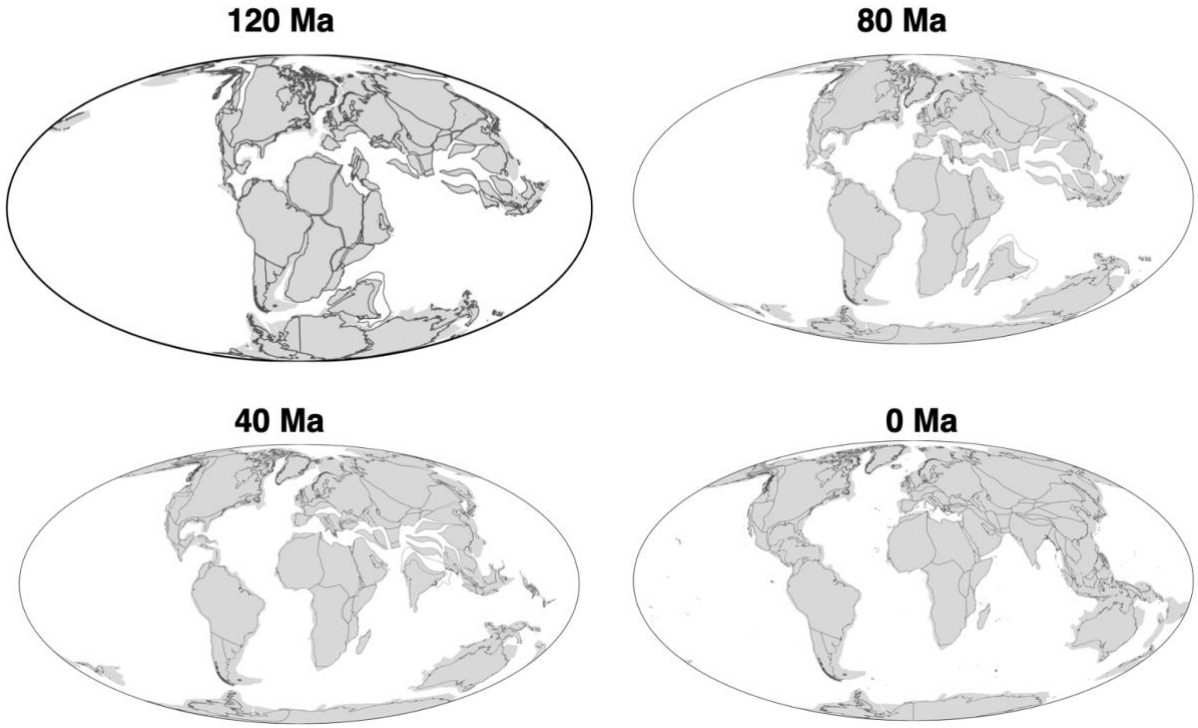



**Figure 2** Global continental maps showing continental trajectories, from Early Cretaceous (120 Ma) to present, at four different time points: 120 Ma, 80 Ma, 40 Ma and 20 Ma.

## 2. Ancestral range estimation

'Ancestral range estimation' is an important methodology for exploring the formation of species distributions and diversity. It provides insights into the relative contributions of evolutionary processes, specifically vicariance and dispersal (Díaz Gómez, 2011; Lomolino et al., 2010). This method relies on access to dated phylogenies, current distribution data for mapping biogeographical ranges, and the use of 'parametric' biogeographical models (Ree & Sanmartín, 2009). The term 'parametric' covers several models that use probabilistic approaches and geographical parameters, including divergence time estimates, past climate conditions, and the fossil record, to infer biogeographical patterns (Sanmartín, 2012). These 'parametric' biogeographical models enable the modelling of range evolution as a stochastic process, with parameters representing key biogeographical processes such as extinction and range expansion. In unison these methods have become fundamental tools in the field of historical biogeography, as evident by their widespread use in peer-reviewed studies (Matzke, 2022).

### 2.1 Parametric biogeographical models as implemented in '**BioGeoBEARS**'

Within the framework of parametric biogeographical models, three widely used approaches are the likelihood-based Dispersal-Extinction Cladogenesis model (DEC; Ree & Smith, 2008), the parsimony-based Dispersal-Vicariance Analysis (DIVA; Ronquist, 1997), and the Bayesian founded BayArea model (Landis et al., 2013). These models are implemented in the R package BioGeoBEARS (Matzke, 2013a) in a Maximum Likelihood framework, enabling model selection where the DIVA model is called DIVALIKE and BayArea is called BAYAREALIKE (Matzke, 2013b). The option to use model selection, rapidly made this package popular because *a priori* selection of biogeographic models was no longer needed. Consequently, this feature facilitated objective model choice, minimising the bias towards preconceived assumptions about the relative importance of different evolutionary processes (Matzke, 2014).

There is a set of processes, assumed by the abovementioned biogeographical models, that can either be anagenetic (happening between speciation events) or cladogenetic (happening during speciation events). Anagenetic processes include dispersal and extirpation (local extinction), which do not in themselves presuppose speciation, but can lead to it. As example the range expansion of a species from one island to another, is an example of dispersal. Cladogenetic processes are vicariance, sympatric speciation, and jump dispersal. Cladogenesis describes the origin of a new lineage through branching speciation (speciation events), understood as the splitting of an ancestral lineage into two or more distinct lineages (Emerson & Patiño, 2018). A classic example of vicariance is the paper on disjunct distribution of species within the genus *Nothafagus* Blume (Nothofagaceae Kuprian.) across the Southern Hemisphere, that were believed to split due to a historical breakup of land masses over time, due to plate tectonic movement (Nixon, 2001). However, this interpretation of the case has since been challenged by Knapp et al. (2005).

These anagenetic and cladogenetic processes can be modelled via the biogeographic models from following parameters: The DEC, DIVALIKE and BAYAREALIKE model each have two free anagenetic parameters, dispersal (d) and extinction (e), controlling anagenetic range expansion and range contraction, respectively, along branches of a phylogeny. Cladogenetic parameters are, on the contrary, associated with nodes where speciation events occur, and each parameter corresponds to a specific process modelled by the different biogeographical models. These processes are vicariance (v) in DIVALIKE or DEC, sympatry with narrow distribution (y) in all models, widespread sympatry (y) in BAYAREALIKE and subset sympatry (s) in DEC (Matzke, 2013). Additionally, a free cladogenetic parameter *j*, representing 'jump-

dispersal', was implemented in BioGeoBEARS, enabling the effects of founder event speciation to be tested (Matzke, 2013a). Based on simulations carried out in the paper by Matzke (2014), the DEC+*j* model generally outperformed the DEC model in several scenarios including island, and continental ranges. These BioGeoBEARS models including jump-dispersal, particularly the DEC+*j* model, have however been strongly critiqued.

Ree and Sanmartín (2018) argue that the conceptual and statistical framework underlying the BioGeoBEARS models is wrong. As example they remark that in the BioGeoBEARS Maximum Likelihood framework the 'cost' of cladogenetic processes is 'cheaper' in comparison to anagenetic processes. Accordingly, this pattern becomes additionally inflated in the DEC+*j* model due to the time-dependent processes being ignored in the DEC+*j* model. For these reasons they also state that DEC and DEC+*j* models should not be compared using model selection. Instead, they suggest that the ClaSSE framework (Goldberg & Igic, 2012) provides a better alternative to DEC+*j* to model phylogenetic range evolution with jump dispersal.

Following, the critique points of Ree and Sanmartín (2018) have been cited in many biogeographical studies and model selection has henceforth largely been ignored, according to Matzke (2022). As a response to the critique and argumentation of Ree and Sanmartín (2018) the critique was addressed and disentangled in the paper by Matzke (2022). Here, the critique was found to be flawed and given on a thin basis, i.e., tiny example datasets, thus disproving the claim that DEC+*j* is a degenerate model. In fact, 'the comparison of DEC and DEC+*j* is statistically valid according to their own criteria' (Matzke, 2022). As a result of the exchange of views regarding the validity of the DEC+*j* model it can be argued that the usage of the DEC+*j* model and model selection should be resumed, as considered a requirement by Matzke (2022). Nevertheless, provided that the results are evaluated in conjunction with arguments based on empirical biological or geographic considerations as recommended by Lamm and Redelings (2009) and Ree and Sanmartín (2018) to assess the performance of the ancestral range estimation model.

*Attention points regarding ancestral range estimation analyses and methods*

While parametric biogeographical models provide an exciting quantitative approach for estimating ancestral ranges (Lamm & Redelings, 2009), various challenges and limitations entail that require attention. One limitation to ancestral range estimation analyses is that the biogeographical models are somewhat error prone because the underlying model assumptions may not depict the evolutionary history of the lineages in question. The DEC model is for example subject to the Yule assumption of no extinction (Yule, 1925), which is bound to be incorrect for most clades. In continuation of this a phylogeny with incomplete sampling of the study group will suffer from similar problems because some cladogenetic events occurring in the phylogenetic tree will remain unobserved (Marshall, 2017).

It is furthermore important to consider limitations associated with ancestral range estimations, as they rely heavily on time-calibrated phylogenies that constitute the backbone of these analyses. The accuracy of these estimations thus depends upon the correctness of the phylogenetic topology and hence the evolutionary information. Despite these prerequisites, most phylogenies will inevitably contain unresolved nodes with low node support and imprecise age estimates. Correspondingly, biogeographic model estimates of ancestral ranges should not be interpreted as definitive but as a best estimate, since they rely on statistics and are not perfect reconstructions. The only direct source of evidence of ancestral ranges of extinct lineages can be derived from the fossil record (Meseguer et al., 2015). Hence, it is important

to incorporate fossils into biogeographic models, as achievable in BioGeoBEARS, to the extent possible, provided that a reliable fossil record is available (Matzke, 2013b).

A challenge associated with ancestral range estimation includes defining meaningful biogeographical regions, as the model states and parameters are built upon these (Ree & Sanmartín, 2009). Accordingly, care should be taken to circumscribe these regions, as ancestral range estimation can generate very different outputs, owing to the chosen circumscription (Landis et al., 2013). Consequently, these regions should not be based upon political boundaries but instead aim to capture main features of the biogeographical history, such as ecoregions (Matzke, 2023) and be relevant for the specific hypothesis in question (Ree & Sanmartín, 2009). It is therefore worth thinking about which geographical scale is meaningful for one's study group. As example the phytogeographical delineation by Carta et al. (2022) could provide a good starting point for defining meaningful phytogeographical regions at the global scale, whereas at a smaller geographical scale, such as Thailand, the seven floristic provinces of Thailand described by Smitinand (1958) would be of more use Anyhow, it is important not to define more areas than necessary, to maintain computational feasibility and phylogenetic signal (Ree & Sanmartín, 2009).

In continuation, allocating extant species to biogeographical areas is of equal importance because it forms the basis for where extinct species could have existed. Thus, tools like Global Biodiversity Information Facility (GBIF; GBIF.org, 2023) and the World Checklist of Vascular Plants (WCVP; Govaerts et al., 2021) that contain distribution data are of great importance in this process. In addition, coordinate data may be helpful as allocation of species to biogeographical regions can be troublesome if the area is loosely defined. Nevertheless, allocating species will constitute a major challenge despite the sampling effort made, because the known distribution will always be limited to an approximation of the true distribution.

To sum up these ancestral range estimation models cannot become better than the underlying data. Hence, one should aim to define appropriate biogeographical regions, increase taxon sampling, and construct the best possible time-calibrated phylogeny that approaches the true evolutionary history. Best approaches and pitfalls will be discussed in the following sections.

### 2.2 Phylogenomics

A key element in reconstructing evolutionary history and dispersal patterns are phylogenies (Ree & Sanmartín, 2009). A phylogeny is a 'tree' containing 'branches' and 'nodes', where branches represent the persistence of an evolutionary lineage and nodes represent speciation events.

The first phylogenies were reconstructed from morphological characters alone, but since the introduction of Sanger sequencing in the late 1970's (Sanger et al., 1977) and PCR in 1985 (Mullis et al., 1986), genetic information was enabled to be incorporated, presenting the field of molecular phylogenetics. This field aims to reconstruct the evolutionary history of organisms based on molecular data, such as nucleotide sequences (e.g., DNA) and relies on the principle of synapomorphy, that shared homologous characters indicate common descent. Thus, as an example, the most recent common ancestor (MRCA) of two species or taxa can be inferred based on differences in their nucleotide sequence, compared to other potentially related taxa (Kapli et al., 2020).

One way of inferring phylogenetic trees is by using statistical approaches, such as maximum-likelihood-based phylogenetics, as implemented in e.g., IQtree (Minh et al., 2020) that can estimate the most likely tree given a set of molecular data. This method compares molecular data, for example aligned DNA sequences, to all substitution models implemented or manually selected in the specific program (e.g., JC; Jukes & Cantor, 1969), HKY (Hasegawa et al., 1985) and GTR (Tavaré, 1986), along with model parameters (rate heterogeneity across sites), possible tree topologies and branch lengths. For each scenario the likelihood is calculated, and the model that maximises the likelihood of the selected criteria, such as the Akaike information criterion (AIC; Akaike, 1974), is selected as the most likely evolutionary tree for the given data (Young & Gillung, 2019).

Advances in sequencing technology have been developing swiftly since the introduction of Sanger sequencing (Metzker, 2005). In the mid-2000s the target capture method and High-throughput Sequencing (HTS) was introduced (Young & Gillung, 2019), which enabled thousands of genes to be targeted and sequenced in parallel at a low cost (Mardis, 2011). Thus, bringing the field of phylogenetics into the genomic era. Target capture functions by applying custom RNA probes, based on known sequences of a related species, which are designed to hybridise with the complementary DNA region (e.g., low copy nuclear genes) of the species of interest. Following, the hybridised DNA regions are captured and typically amplified using PCR, before sent to sequencing (Andermann et al., 2020). The probes can be designed for different scales, e.g., Angiosperms-353 for all flowering plants (Johnson et al., 2019), that is directed at genes across higher taxonomic ranks, or study group specific probes, e.g., PhyloPalm for palms enhancing resolution at the family level (Loiseau et al., 2019). This approach leads to increased sequencing depth of the targeted genes, which enables sequencing of possibly degraded DNA, such as DNA from herbarium and museum specimens, unlocking an extraordinary resource of previously underutilised data (Brewer et al., 2019; Dodsworth et al., 2019).

The magnitude of increase in data made possible by HTS, among others, means that phylogenetic tree inference is no longer constrained by the availability of only a few loci that potentially result in poorly supported phylogenetic trees (Young & Gillung, 2019). But more importantly the increase in data led to the realisation that different genes often have very different evolutionary histories. Hence, gene tree assembled from one or a few genes do not necessarily reflect the 'true' species tree. In order to give a good approximation of the evolutionary history of a species it is therefore crucial to include as many genes as possible to avoid generating a species tree reflecting incorrect genealogical histories (Degnan & Rosenberg, 2009). However, it is not without significance what kind of gene one uses. The plastome is known to evolve as one gene, hence representing just one gene tree, which may not correspond to the 'true' species tree (Doyle, 1992). However, low copy nuclear genes are well recognised for their importance in reconstructing the phylogenetical history of plants and are essential in understanding gene tree discordance resulting from differences in genealogical histories that may arise from the following processes (Degnan & Rosenberg, 2009; Morales-Briones et al., 2021):

1. Incomplete lineage sorting (ILS) i.e., failure of lineages to coalesce before the point of speciation.

2. Horizontal gene transfer (HGT) i.e., a gene 'jumps' from the genome of one lineage to a more distant lineage.

3. Hybridization, i.e., genes are transferred from one ancestral lineage to another. Like horizontal gene transfer but at the genome level.

4. Gene duplication and loss i.e., following a gene duplication event and subsequent gene loss in species, species that have not lost the gene can appear more closely related.

### 2.3 Molecular clock dating

Phylogenies provide a mean of retracing evolutionary history and additionally provide a framework to place species divergence events on a timeline (dos Reis et al., 2016). Applications of molecular dating include biogeography, since reconstructing biogeographic history typically relies on dated trees (Sauquet, 2013).

More than six decades ago Zuckerkandl and Pauling (1965) published two seminal papers introducing the molecular clock hypothesis, that soon after received theoretical backing by the 'neutral theory' biologist, Kimura (1968). This instantly revolutionised the field of molecular evolution (dos Reis et al., 2016; Lemey & Posada, 2009), where divergence time estimation had hitherto been limited to fossil records, i.e., paleontological dating (Laurin, 2012). This concept was centred around a molecular clock and the resulting underlying condition that the rate at which nucleotides or amino acids substitute within a genome remains relatively constant between species and over time (Kimura, 1968). Accordingly, divergence time between lineages could be estimated by measuring the genetic difference between two molecular sequences (Drummond et al., 2006; Sauquet, 2013). This 'strict' molecular clock was the first attempt of using substitutions to track evolutionary time, which was later criticised for being too simplistic. Several studies in fact found that evolutionary rates often vary markedly between distantly related lineages, violating the assumption of the strict molecular clock (dos Reis et al., 2016; Sauquet, 2013). With the introduction of Bayesian methods in the late 1990's and early 2000's (Rutschmann, 2006; Sanderson et al., 2004), the relaxed molecular clock models were developed. These methods allowed molecular dates to be estimated in the presence of rate heterogeneity among lineages (Sauquet et al., 2012). Within the broad research community today, this is generally accepted as the preferential molecular dating method (Sauquet, 2013).

Although the molecular clock was meant as an independent and improved alternative to the fossil record, molecular dating in combination with fossil calibration is now considered the definitive standard (Couvreur et al., 2018). Indeed, species divergence without the use of the fossil record merely provides dating in a relative time scale, whereas calibration from a secondary source, e.g., the fossil record is necessary to obtain absolute divergence times (Kumar, 2005; Sauquet, 2013). In continuation of this, a key element in any molecular dating, is calibration with prior knowledge in the form of dated fossils of common ancestors, or secondary calibration, which means calibrations from ages estimated in previous molecular dating studies (e.g., Schenk, 2016).

A Bayesian framework, as implemented in BEAST (Drummond & Rambaut, 2007), is accepted as the preferred method to conduct molecular dating analyses by the scientific community (dos Reis et al., 2016; Forest, 2009; Parham et al., 2012). The reason for this is, among others, that the Bayesian framework provides a way to include prior knowledge (priors), and incorporate uncertainty, e.g., time-priors and fossil-priors (dos Reis et al., 2016; Drummond & Rambaut, 2007; Parham et al., 2012).

The first step in molecular dating is typically deciding which and how many genes should form the basis for divergence-time estimation. SortaDate (Smith et al., 2018) provides a convenient

way of determining genes to base the divergence-time estimation on. This means choosing genes with minimal topological conflict to the species tree, thus reducing estimation error. In connection the common approach has been to base divergence-time estimates on concatenated alignment of multiple genes and a single species tree topology. Yet, Carruthers et al. (2022) have found that topological conflict between gene trees and the species tree has significant negative impact on divergence time estimation, due to influence on substitutions. Accordingly, they suggested that selecting fewer but more congruent genes would be preferable. The next step involves considering priors. Yule-process and Birth-death models are typical tree priors used in plant evolutionary studies above the population level (e.g., Kondraskov et al., 2015; Li et al., 2020). The birth-death model assumes an initial species that will give rise to a new species at a given rate and likewise subsequent extinction at a given rate (Gernhard, 2008). In comparison the Yule model assumes a pure birth process at an unknown rate, thus none of the species included are assumed to go extinct (Yule, 1925). Even though the specification of the most fitting tree priors is of concern, Couvreur et al. (2010) and Ritchie et al. (2017) argue that misspecification does not have a strong impact on Bayesian molecular dating results. Nevertheless, Ritchie et al. (2017) found that in studies containing a mix of inter- and intraspecific sampling, the birth-death model provided more accurate dating. In such sampling scenarios choosing a birth-death model may be preferential.

Assigning appropriate calibration points and related priors is a heavily debated topic, that in many instances has been an untransparent process (Parham et al., 2012). The debate includes correct placement of fossils, balancing between underestimating and overestimating divergence times (Pennington et al., 2004; Renner, 2005), as fossil taxa with a synapomorphy for any extant clade may be placed anywhere between the crown node and the stem node (Sauquet, 2013). As an example, Renner (2005) suggests placing fossils on the stem node, if no certain placement is given, whereas e.g., Pennington et al. (2004) argues that this approach will likely lead to underestimation of fossil age. Accordingly, the *a priori* selection and justification of fossils is of great importance (Parham et al., 2012). Nevertheless, it is in any case wrong to assume that the oldest fossil equals actual clade ages. For example, the age of a clade in rodents was pushed back to the double, after the discovery of a new fossil record (Kimura et al., 2015). Thus, calibration using fossils should be treated with caution and understood as a best estimate at a given time (Parham et al., 2012).

Additionally, the debate concerns fossil priors that among others include uniform, lognormal and normal priors. Of these, uniform priors constitute the most conservative but also least informative approach, because they assume that nothing about maximum fossil age can be known with certainty (Warnock et al., 2015).Thus, it is common practice and essential to impose a hard minimum age constraint based on the fossil record, as exemplified in Massoni et al. (2015), and to not assign a maximum age constraint, allowing it to approach infinity (Chazot et al., 2019). It is recommended to assign a maximum constraint to at least one calibration point, as the uniform prior is relatively vague (Chazot et al., 2019). Otherwise, the analyses may not converge or may send the maximum age estimate far back in time (Sauquet et al., 2012, 2013). Log normal priors instead incorporate a soft maximum age estimate that has the benefit of incorporating uncertainty linked to the age estimate of the fossil, accounting for a potential sampling gap (Sauquet et al., 2012). A limitation to the use of non-uniform priors, e.g., lognormal, is however the risk of assigning an arbitrary distributional shape, that may entail an erroneous estimate of uncertainty linked to fossil age (Warnock et al., 2012). The normal distribution in contrast to e.g., the lognormal and uniform prior assumes that the probability of a node is equally likely to be younger or older than the mean, based on a set standard deviation. These properties entail that this type of prior is appropriate to use for

secondary calibrations (calibrations based on the results of previous molecular dating studies; Schenk, 2016). This is because uncertainty linked to age estimates from e.g., a former molecular study is equally distributed around the mean, given the calibration point is taken from a Bayesian study (Forest, 2009).

Secondary calibration is generally seen as a last resort, as it may lead to biased age estimates, due to the assumptions made in the previous study. Therefore, they are most often used when the group of interest has an incomplete fossil record (Forest, 2009; Schenk, 2016). One risk of using secondary calibration is, e.g., the risk of passing on age estimates based on misplaced fossils or poorly assigned priors (Parham et al., 2012). Nevertheless, it is common practice to use secondary calibrations, alone or in combination with fossils (Sauquet et al., 2012; Shaul & Graur, 2002), because fossils can only provide minimum age constraints, but at least one maximum age constraint is needed, and secondary calibration is a convenient way of doing that (Sauquet, 2013). However, in the paper by Sauquet et al. (2012) it is argued that including outgroup calibration points constitute a more robust approach, than applying a secondary calibration point within the group of interest. Additionally, they argue that using a combination of ingroup and outgroup calibration points provide more accurate age estimates, than using the otherwise preferred ingroup calibrations alone. Altogether multiple fossil calibration points should be included in molecular dating analyses to avoid reliance on a single calibration point (Shaul & Graur, 2002).

### 3. Lauraceae, hereunder the Cryptocarya group

Lauraceae Juss., commonly known as the laurel family, is the largest family within the order Laurales, consisting of approximately 2500 species distributed among the 56 accepted genera (POWO). The family is pantropical, with two main centres of diversity in Tropical America and Tropical Asia. Due to their significant species richness and wide pantropical distribution, Lauraceae species are considered characteristic elements of tropical and subtropical regions (Chanderbali et al., 2001; Song et al., 2019). In addition, they are generally among the ten most frequent families in tropical rainforests worldwide (TRF; Gentry et al., 1988).

Although significant efforts have been made to resolve the systematics within Lauraceae, considerable uncertainty regarding the delimitation of genera and intraspecific relationships remains (Chanderbali et al., 2001; Rohwer, 2000; Rohwer & Rudolph, 2005; Song et al., 2019). However, there is consensus about the monotypic genus *Hypodaphnis* Stapf being sister to the remaining eight superior groupings: the Cryptocarya group, *Cassytha* Osbeck, *Neocinnamomum* H.Liu, *Caryodaphnopsis* Airy Shaw, the Mezilaurus group, the Persea group, tribe Laureae and Cinnamomeae (Figure 2).

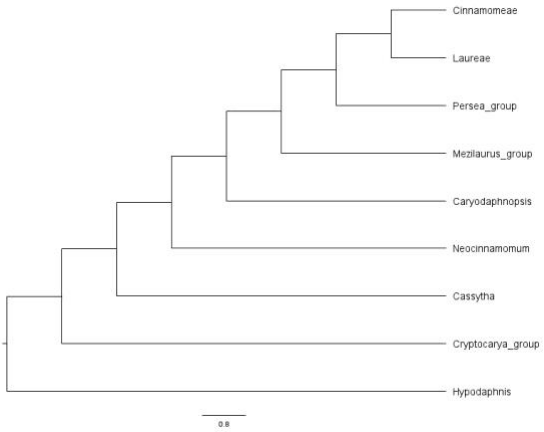


**Figure 2** Main evolutionary lineages in Lauraceae, based on Chanderbali et al. (2001), Rohwer (2000) , Rohwer and Rudolph (2005) and Song et al. (2019).

This thesis focuses on the pantropical Cryptocarya group, which comprises several genera. As of Rohwer et al. (2014), this group includes *Eusideroxylon* Teijsm. & Binn., *Potoxylon* Kosterm., *Aspidostemon* Rohwer & H.G.Richt., *Dahlgrenodendron* van der Merwe & van Wyk, *Cryptocarya* R.Br., and the Beilschmiedia complex. Additionally, Li et al. (2020) state that the Beilschmiedia complex is monophyletic and consists of *Beilschmiedia* Nees, *Endiandra* R.Br. (including *Brassiodendron* C.K.Allen and *Triadodaphne* Kosterm.), *Hexapora* Hook.f., *Potameia* Thouars, *Sinopora* J.Li, N.H.Xia & H.W.Li, *Syndiclis* Hook.f. and *Yasunia* van der Werff. The Cryptocarya group encompasses 826 species distributed across South America, Tropical Africa, Madagascar, Tropical Asia, Malesia, and Australia (POWO), which are mostly restricted to TRF (81%; Govaerts, 2022; van der Merwe et al., 2016). The two largest genera within the group are *Beilschmiedia* and *Cryptocarya*, which include 262 and 362 species, respectively (POWO).

The Cryptocarya group is characterised by some diagnostic features, including wood anatomy (Richter, 1981), inflorescence type (van der Werff & Richter, 1996), and molecular phylogenetic data (Chanderbali et al., 2001; Li et al., 2020; Rohwer et al., 2014; Rohwer & Rudolph, 2005). The inflorescence type in this group is to a greater or lesser extent paniculate-cymose with the ultimate branches not strictly opposite, and the wood can be recognised by characteristics such as marginal parenchyma and non-septate fibres with conspicuously bordered pits (van der Werff & Richter, 1996). Beyond that, there are few common morphological characteristics shared by the group.

One prominent characteristic of the Cryptocarya group is its ovary type, which has inspired the group's name. The name *Cryptocarya* is derived from Greek and can be interpreted as 'hidden fruit' or 'concealed nut', in resemblance to the ovary type found in the genera *Aspidostemon*, *Dahlgrenodendron*, and *Cryptocarya*, that is an inferior ovary immersed in a deeply tubular hypanthium, which after flowering forms a 'hidden' fruit (Rohwer, 1993; Rohwer et al., 2014). However, the monotypic genera *Eusideroxylon* and *Potoxylon* have a semi-inferior ovary and within the Beilschmiedia complex, the ovary is superior and thus not or only partially enclosed by receptacular tissue. Additionally, *Eusideroxylon* and *Potoxylon* have four separate pollen sacs, whereas all other members of the group have four pollen sacs, that are fused to a greater or lesser degree with two flaps, that makes the number of pollen sacs appear as to. Other inconsistencies within the group include leaf position as *Aspidostemon* and *Dahgrenodendron* have opposite leaves, whereas the rest of the group exhibits alternate leaf arrangement, with the exception of a few *Beilschmiedia* species with opposite leaves (Rohwer et al., 2014).
The inference of relationships within the group using taxonomic characters is however equally or even more difficult as recognised in the paper by Li et al. (2020). Previously, the size and distribution of terminal buds, was used to delimit subgroups within *Beilschmiedia,* which fall into the two categories: small and pubescent buds versus large and glabrous, as recognized by Hooker (1886). But previous studies e.g., Liu et al. (2013) have shown that only large buds but not small, may have taxonomic utility for delimiting major clades of *Beilschmiedia*. Similarly, *Endiandra* has commonly been delimited by number of stamens, typically three. However, several species have exceptional stamen numbers, such as two or six, making taxonomic delimitation challenging in these cases.

Despite substantial progress, since the first phylogeny of Lauraceae by Rohwer (2000), several relationships within the Cryptocarya group remain poorly understood. For example, the

circumscription of the polyphyletic *Beilschmiedia* is unsatisfactory, but taxonomic delimitation is hindered by incomplete understanding of the phylogenetic relationships (Li et al., 2020; Rohwer et al., 2014). Furthermore, that no single person has sufficient knowledge about the morphology of the species in different regions (J. G. Rohwer, personal communication, June 9, 2023).

### 3.1 Significance of the Lauraceae: economically and as a model group

Lauraceae is a family of great economic significance due to the many species with use potential within this group. The economic value is among other due to the characteristics of the family e.g., the aromatic bark utilised as spices and essential oils (e.g., *Cinnamomum aromaticum* Nees, *C. verum* J. Presl (cinnamon), *C. camphora* (L.) J.Presl (camphor) and *Laurus nobilis* L. (bay laurel), owing to nutritious fruit, *Persea americana* Mill. (avocado), and as a source of high-quality timber e.g., *Actinodaphne henryi* Gamble, *Cryptocarya obovata* R.Br., *Eusideroxylon zwageri* Teijsm. & Binn., *Persea indica* (L.) Spreng. and *Ocotea bullata* (Burch.) Baill. Unfortunately, these features act as a contributing factor that have led to overexploitation of several Lauraceae species and for some entailed status as threatened i.e., vulnerable, endangered, or critically endangered on the IUCN Red List (Li et al., 2008; Marques, 2001; Schroeder, 1976).

The family is additionally of great importance due to its ecological role that can be appropriated to its diversity, global distribution and great abundance driving physiognomic importance in tropical and subtropical ecosystems worldwide (Reis-Avila & Oliveira, 2017) e.g., in oak-laurel forests in the tropical montane zone of South-East Asia (Sri-Ngernyuang et al., 2003), Mixed Ombrophilous Forest (MOF) in high altitude areas of Brazil, Paraguay, and northern Argentina (Kieras et al., 2020) and in mixed broadleaved evergreen forests of the sub-tropical-humid zone (Laurisilva; Pignatti et al., 2015). The ecological properties of Lauraceae also entail that the group has great potential to serve as a model group in various biological fields, i.e., dendrochronology (Reis-Avila & Oliveira, 2017), possible drivers of diversity (Eiserhardt et al., 2018) and to explore biogeography in an early diverging angiosperm lineage (Chanderbali et al., 2001; Knight & Wilf, 2013).

Lauraceae constitute a good potential model group to study TRF evolution as they meet c. three of the four criteria set up by Couvreur and Baker (2013) for identifying suitable TRF model groups. The criteria include that the lineage should be ecologically representative of the biome under investigation, species taxonomy and distributions of the lineage should be well-documented, availability of comprehensive phylogenetic hypotheses, and the historical timeframe of the lineage should coincide with that of the biome (as indicated by the fossil record, for example). Lauraceae meets the first criterion by predominantly inhabiting the tropics and subtropics (Song et al., 2019), in addition to being one of the most encountered tree families in the tropical rainforest biome (TRF; Balslev et al., 1998; Gentry et al., 1988). It also satisfies the second criterion, because it is a sizeable family of > 3500 (POWO, 2023), like e.g., the TRF model groups Arecaceae (Couvreur & Baker, 2013) and Annonaceae (Couvreur, Pirie, et al., 2011), and has a relatively well-resolved taxonomy (Song et al., 2019), including moderately well-documented distributional data summarised in the World Checklist of Vascular Plants (Govaerts, 2022). Nevertheless, some species complexes like the Beilschmiedia complex require further studying (Li et al., 2020; Rohwer et al., 2014). Regarding the fourth criterion, the historic timeframe of Lauraceae dates back to the Early Cretaceous (Chanderbali et al., 2001; Massoni et al., 2015) as supported by the relatively well supported fossil record e.g., (Ramirez-Barahona et al., 2020), aligning with previous estimates

of TRF biome age (W. J. Baker et al., 2013). However, the third criterion of having a comprehensive phylogenetic hypothesis has yet to be accomplished and a genus-level chronogram has not been provided since Chanderbali et al. (2001). Therefore, to meet all four criteria and hence unlock the Lauraceae family's potential as a model group, resolving generic and specific relationships within Lauraceae further, to contribute to species-level mapping in Lauraceae, is crucial.

## 4. Aim of this thesis

The aim of this thesis was to significantly contribute to the understanding of the origins and distribution patterns of the pantropical Cryptocarya group, while also gaining a deeper understanding of the origin, assembly, and diversification of tropical rainforests (TRFs).

In Part B of this thesis, I wrote an article manuscript presenting my findings to this question. Here I provided a robust phylogenetic framework for the pantropical Cryptocarya group, incorporating molecular dating analyses and ancestral range estimations. By addressing key questions regarding the group's origin, distribution, and TRF evolution, this manuscript formed a piece of the puzzle in advancing our understanding of how pantropical groups have achieved a pantropical distribution and when characteristic elements of modern TRF originated across the globe.

## 5. References


Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716-723. https://doi.org/10.1109/TAC.1974.1100705

Andermann, T., Torres Jiménez, M. F., Matos-Maraví, P., Batista, R., Blanco-Pastor, J. L., Gustafsson, A. L. S., Kistler, L., Liberal, I. M., Oxelman, B., Bacon, C. D., & Antonelli, A. (2020). A Guide to Carrying Out a Phylogenomic Target Sequence Capture Project [Review]. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.01407

Axelrod, D. I. (1970). Mesozoic Paleogeography and Early Angiosperm History. *Botanical Review*, *36*(3), 277-319. http://www.jstor.org/stable/4353795

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

Baker, W. J., Couvreur, T. L. P., & Kissling, W. D. (2013). 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. https://doi.org/10.1111/j.1365-2699.2012.02794.x

Balslev, H. V., Renato; Paz-y-Mino-C, Guillermo; Christensen, H., Nielsen, I. (1998). Species Count of Vascular Plants in 1-hectare of humid lowland forest in Amazonian Ecuador. In F. C. Dallmeier, J. (Ed.), *Forest Biodiversity in North, Central and South America, and the Caribbean: Research and Monitoring* (Vol. 21, pp. 585-594). CRC Press.

Bardon, L., Sothers, C., Prance, G. T., Male, P. J., Xi, Z., Davis, C. C., Murienne, J., Garcia-Villacorta, R., Coissac, E., Lavergne, S., & Chave, J. (2016). Unraveling the biogeographical history of Chrysobalanaceae from plastid genomes. *Am J Bot*, *103*(6), 1089-1102. https://doi.org/10.3732/ajb.1500463

Bartish, I. V., Antonelli, A., Richardson, J. E., & Swenson, U. (2011). Vicariance or long-distance dispersal: historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). *Journal of Biogeography*, *38*(1), 177-190. https://doi.org/10.1111/j.1365-2699.2010.02389.x

Bechteler, J., Schafer-Verwimp, A., Lee, G. E., Feldberg, K., Perez-Escobar, O. A., Pocs, T., Peralta, D. F., Renner, M. A., & Heinrichs, J. (2017). Geographical structure, narrow species ranges, and Cenozoic diversification in a pantropical clade of epiphyllous leafy liverworts. *Ecology and Evolution*, *7*(2), 638-653. https://doi.org/10.1002/ece3.2656

Bessa-Silva, A., Vallinoto, M., Sampaio, I., Flores-Villela, O. A., Smith, E. N., & Sequeira, F. (2020). The roles of vicariance and dispersal in the differentiation of two species of the *Rhinella marina* species complex. *Molecular phylogenetics and evolution*, *145*, 106723.

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*, 1102. https://doi.org/10.3389/fpls.2019.01102

Carruthers, T., Sun, M., Baker, W. J., Smith, S. A., de Vos, J. M., & Eiserhardt, W. L. (2022). The Implications of Incongruence between Gene Tree and Species Tree Topologies

for Divergence Time Estimation. *Systematic Biology*, *71*(5), 1124-1146. https://doi.org/10.1093/sysbio/syac012

Carta, A., Peruzzi, L., & Ramirez-Barahona, S. (2022). A global phylogenetic regionalization of vascular plants reveals a deep split between Gondwanan and Laurasian biotas. *New Phytologist*, *233*(3), 1494-1504. https://doi.org/10.1111/nph.17844

Chanderbali, A. S., van der Werff, H., & Renner, S. S. (2001). Phylogeny and Historical Biogeography of Lauraceae: Evidence from the Chloroplast and Nuclear Genomes. *Annals of the Missouri Botanical Garden*, *88*(1). https://doi.org/10.2307/2666133

Chazot, N., Wahlberg, N., Freitas, A. V. L., Mitter, C., Labandeira, C., Sohn, J. C., Sahoo, R. K., Seraphim, N., de Jong, R., & Heikkila, M. (2019). Priors and Posteriors in Bayesian Timing of Divergence Analyses: The Age of Butterflies Revisited. *Systematic Biology*, *68*(5), 797-813. https://doi.org/10.1093/sysbio/syz002

Clayton, J. W., Soltis, P. S., & Soltis, D. E. (2009). Recent long-distance dispersal overshadows ancient biogeographical patterns in a pantropical angiosperm family (Simaroubaceae, Sapindales). *Systematic Biology*, *58*(4), 395-410. https://doi.org/10.1093/sysbio/syp041

Conceição Oliveira, L., Picanço Rodrigues, D., Fortune Hopkins, H. C., Peter Lewis, G., & John Gilbert Hopkins, M. (2021). Phylogeny and historical biogeography of the pantropical genus *Parkia* (Leguminosae, Caesalpinioideae, mimosoid clade). *Molecular phylogenetics and evolution*, *163*, 107219. https://doi.org/10.1016/j.ympev.2021.107219

Conti, E., Eriksson, T., Schönenberger, J., Sytsma, K. J., & Baum, D. A. (2002). Early Tertiary out-of-India Dispersal of Crypteroniaceae: Evidence from Phylogeny and Molecular Dating. *Evolution*, *56*(10), 1931-1942. https://doi.org/10.1111/j.0014-3820.2002.tb00119.x

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

Couvreur, T. L. P., Forest, F., & Baker, W. J. (2011). Origin and global diversification patterns of tropical rain forests: inferences from a complete genus-level phylogeny of palms. *BMC Biology*, *9*, 44. https://doi.org/10.1186/1741-7007-9-44

Couvreur, T. L. P., Franzke, A., Al-Shehbaz, I. A., Bakker, F. T., Koch, M. A., & Mummenhoff, K. (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Molecular Biology and Evolution*, *27*(1), 55-71. https://doi.org/10.1093/molbev/msp202

Couvreur, T. L. P., Helmstetter, A. J., Koenen, E. J. M., Bethune, K., Brandao, R. D., Little, S. A., Sauquet, H., & Erkens, R. H. J. (2018). Phylogenomics of the Major Tropical Plant Family Annonaceae Using Targeted Enrichment of Nuclear Genes. *Frontiers in Plant Science*, *9*, 1941. https://doi.org/10.3389/fpls.2018.01941

Couvreur, T. L. P., Pirie, M. D., Chatrou, L. W., Saunders, R. M. K., Su, Y. C. F., Richardson, J. E., & Erkens, R. H. J. (2011). Early evolutionary history of the flowering plant family Annonaceae: steady diversification and boreotropical geodispersal. *Journal of Biogeography*, *38*(4), 664-680. https://doi.org/10.1111/j.1365-2699.2010.02434.x

Cox, C. B., & Moore, P. D. (2010). *Biogeography: An Ecological and Evolutionary Approach*. John Wiley & Sons. https://books.google.gl/books?id=GP5HeCwkV2IC

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, *24*(6), 332-340. https://doi.org/10.1016/j.tree.2009.01.009

Díaz Gómez, J. M. (2011). Estimating ancestral ranges: testing methods with a clade of neotropical lizards (Iguania: Iiolaemidae). *PLOS ONE*, *6*(10), e26412. https://doi.org/10.1371/journal.pone.0026412

Dodsworth, S., Pokorny, L., Johnson, M. G., Kim, J. T., Maurin, O., Wickett, N. J., Forest, F., & Baker, W. J. (2019). Hyb-Seq for Flowering Plant Systematics. *Trends in Plant Science*, *24*(10), 887-891. https://doi.org/10.1016/j.tplants.2019.07.011

Donoghue, M. J. (2008). Colloquium paper: a phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *105 Suppl 1*(Suppl 1), 11549-11555. https://doi.org/10.1073/pnas.0801962105

dos Reis, M., Donoghue, P. C., & Yang, Z. (2016). Bayesian molecular clock dating of species divergences in the genomics era. *Nature Reviews Genetics*, *17*(2), 71-80. https://doi.org/10.1038/nrg.2015.8

Doyle, J. J. (1992). Gene Trees and Species Trees: Molecular Systematics as One-Character Taxonomy. *Systematic Botany*, *17*(1). https://doi.org/10.2307/2419070

Drummond, A. J., Ho, S. Y., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, *4*(5), e88. https://doi.org/10.1371/journal.pbio.0040088

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, *7*, 214. https://doi.org/10.1186/1471-2148-7-214

Eiserhardt, W. L., Antonelli, A., Bennett, D. J., Botigue, L. R., Burleigh, J. G., Dodsworth, S., Enquist, B. J., Forest, F., Kim, J. T., Kozlov, A. M., Leitch, I. J., Maitner, B. S., Mirarab, S., Piel, W. H., Perez-Escobar, O. A., Pokorny, L., Rahbek, C., Sandel, B., Smith, S. A., . . . Baker, W. J. (2018). A roadmap for global synthesis of the plant tree of life. *American Journal of Botany*, *105*(3), 614-622. https://doi.org/10.1002/ajb2.1041

Emerson, B. C., & Patiño, J. (2018). Anagenesis, Cladogenesis, and Speciation on Islands. *Trends in Ecology & Evolution*, *33*(7), 488-491. https://doi.org/10.1016/j.tree.2018.04.006

Forest, F. (2009). Calibrating the Tree of Life: fossils, molecules and evolutionary timescales. *Annals of Botany*, *104*(5), 789-794. https://doi.org/10.1093/aob/mcp192

GBIF.org. (2023). *GBIF Home Page*. https://www.gbif.org

Gentry, A. H. (1988). Changes in Plant Community Diversity and Floristic Composition on Environmental and Geographical Gradients. *Annals of the Missouri Botanical Garden*, *75*(1). https://doi.org/10.2307/2399464

Gernhard, T. (2008). The conditioned reconstructed process. *Journal of Theoretical Biology*, *253*(4), 769-778. https://doi.org/10.1016/j.jtbi.2008.04.005

Givnish, T. J., & Renner, S. S. (2004). Tropical Intercontinental Disjunctions: Gondwana Breakup, Immigration from the Boreotropics, and Transoceanic Dispersal. *International Journal of Plant Sciences*, *165*(S4), S1-S6. https://doi.org/10.1086/424022

Goldberg, E. E., & Igic, B. (2012). Tempo and mode in plant breeding system evolution. *Evolution*, *66*(12), 3701-3709. https://doi.org/10.1111/j.1558-5646.2012.01730.x

Govaerts, R. (2022). *The World Checklist of Vascular Plants (WCVP)* GBIF.org. https://doi.org/https://doi.org/10.15468/6h8ucr

Govaerts, R., Nic Lughadha, E., Black, N., Turner, R., & Paton, A. (2021). The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Scientific Data*, *8*(1), 215. https://doi.org/10.1038/s41597-021-00997-6

Harries, H. C., & Clement, C. R. (2014). Long-distance dispersal of the coconut palm by migration within the coral atoll ecosystem. *Annals of Botany*, *113*(4), 565-570. https://doi.org/10.1093/aob/mct293

Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*(2), 160-174. https://doi.org/10.1007/BF02101694

Hennig, W. (1966). *Phylogenetic Systematics*. University of Illinois Press.

Hill, J. L., & Hill, R. A. (2016). Why are tropical rain forests so species rich? Classifying, reviewing and evaluating theories. *Progress in Physical Geography: Earth and Environment*, *25*(3), 326-354. https://doi.org/10.1177/030913330102500302

Hooker, J. D. (1886). *Flora Of British India* (Vol. 5). L. Reeve.

Humboldt, A. v., & Bonpland, A. (1805). Essai sur la géographie des plantes. *Chez Levrault, Scoell et Campagnie, Libraries, Paris*.

Humphries, C. J., Escudero, M., & Martín-Bravo, S. (2017). Vicariance Biogeography. In *Reference Module in Life Sciences*. Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-809633-8.02178-6

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigue, 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., Baker, W. J., & Wickett, N. J. (2019). A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any Flowering Plant Designed Using k-Medoids Clustering. *Systematic Biology*, *68*(4), 594-606. https://doi.org/10.1093/sysbio/syy086

Jukes, T. H., & Cantor, C. R. (1969). CHAPTER 24 - Evolution of Protein Molecules. In H. N. Munro (Ed.), *Mammalian Protein Metabolism* (pp. 21-132). Academic Press. https://doi.org/https://doi.org/10.1016/B978-1-4832-3211-9.50009-7

Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, *21*(7), 428-444. https://doi.org/10.1038/s41576-020-0233-0

Kieras, W., Machado, S., Pelissari, A., Cysneiros, V., & Alves da Silva, S. (2020). Influence of the Lauraceae family on the dynamics of a mixed ombrophilous forest remnant. *Floresta*, *51*, 164-173. https://doi.org/10.5380/rf.v51i1.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, *217*(5129), 624-626. https://doi.org/10.1038/217624a0

Kimura, Y., Hawkins, M. T., McDonough, M. M., Jacobs, L. L., & Flynn, L. J. (2015). Corrected placement of Mus-Rattus fossil calibration forces precision in the molecular tree of rodents. *Scientific Reports*, *5*, 14444. https://doi.org/10.1038/srep14444

Knapp, M., Stockler, K., Havell, D., Delsuc, F., Sebastiani, F., & Lockhart, P. J. (2005). Relaxed molecular clock provides evidence for long-distance dispersal of *Nothofagus* (southern beech). *PLoS Biology*, *3*(1), e14. https://doi.org/10.1371/journal.pbio.0030014

Knight, C. L., & Wilf, P. D. (2013). Rare leaf fossils of Monimiaceae and Atherospermataceae (Laurales) from Eocene Patagonian rainforests and their biogeographic significance. *Palaeontologia Electronica*. https://doi.org/10.26879/386

Kondraskov, P., Schutz, N., Schussler, C., de Sequeira, M. M., Guerra, A. S., Caujape-Castells, J., Jaen-Molina, R., Marrero-Rodriguez, A., Koch, M. A., Linder, P., Kovar-Eder, J., & Thiv, M. (2015). Biogeography of Mediterranean Hotspot Biodiversity: Re-Evaluating the 'Tertiary Relict' Hypothesis of Macaronesian Laurel Forests. *PLOS ONE*, *10*(7), e0132091. https://doi.org/10.1371/journal.pone.0132091

Kumar, S. (2005). Molecular clocks: four decades of evolution. *Nature Reviews Genetics*, *6*(8), 654-662. https://doi.org/10.1038/nrg1659

Lamm, K. S., & Redelings, B. D. (2009). Reconstructing ancestral ranges in historical biogeography: properties and prospects. *Journal of Systematics and Evolution*, *47*(5), 369-382. https://doi.org/10.1111/j.1759-6831.2009.00042.x

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. https://doi.org/10.1093/sysbio/syt040

Laurin, M. (2012). Recent progress in paleontological methods for dating the Tree of Life. *Frontiers in Genetics*, *3*, 130. https://doi.org/10.3389/fgene.2012.00130

Lee, G. E., Condamine, F. L., Bechteler, J., Perez-Escobar, O. A., Scheben, A., Schafer-Verwimp, A., Pocs, T., & Heinrichs, J. (2020). An ancient tropical origin, dispersals via land bridges and Miocene diversification explain the subcosmopolitan disjunctions of the liverwort genus *Lejeunea*. *Scientific Reports*, *10*(1), 14123. https://doi.org/10.1038/s41598-020-71039-1

Lemey, P., & Posada, D. (2009). Molecular clock analysis. In A.-M. Vandamme, M. Salemi, & P. Lemey (Eds.), *The Phylogenetic Handbook* (2 ed., pp. 362-380). Cambridge University Press. https://doi.org/10.1017/cbo9780511819049.013

Li, H., Liu, B., Davis, C. C., & Yang, Y. (2020). Plastome phylogenomics, systematics, and divergence time estimation of the Beilschmiedia group (Lauraceae). *Molecular phylogenetics and evolution*, *151*, 106901. https://doi.org/10.1016/j.ympev.2020.106901

Li, S., Li, X., Li, J., Huang, P., Wei, F., Cui, H., & van der Werff, H. (2008). Flora of China 7. Lauraceae. In (Vol. 7, pp. 102). http://www.efloras.org/florataxon.aspx?flora_id=2&taxon_id=10479

Liu, B., Yang, Y., Xie, L., Zeng, G., & Ma, K. (2013). *Beilschmiedia turbinata*: a newly recognized but dying species of Lauraceae from tropical Asia based on morphological and molecular data. *PLOS ONE*, *8*(6), e67636. https://doi.org/10.1371/journal.pone.0067636

Loiseau, O., Olivares, I., Paris, M., de La Harpe, M., Weigand, A., Koubinova, D., Rolland, J., Bacon, C. D., Balslev, H., Borchsenius, F., Cano, A., Couvreur, T. L. P., Delnatte, C., Fardin, F., Gayot, M., Mejia, F., Mota-Machado, T., Perret, M., Roncal, J., . . . Salamin, N. (2019). Targeted Capture of Hundreds of Nuclear Genes Unravels Phylogenetic Relationships of the Diverse Neotropical Palm Tribe Geonomateae. *Frontiers in Plant Science*, *10*, 864. https://doi.org/10.3389/fpls.2019.00864

Lomolino, M. V., Riddle, B. R., Whittaker, R. J., & Brown, J. H. (2010). *Biogeography* (4 ed.). Sinauer Associates.

Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. *Nature*, *470*(7333), 198-203. https://doi.org/10.1038/nature09796

Marques, C. (2001). Importância econômica da família Lauraceae Lindl. *Floresta e Ambiente*, *8*, 195-206.

Marshall, C. R. (2017). Five palaeobiological laws needed to understand the evolution of the living biota. *Nature Ecology and Evolution*, *1*(6), 165. https://doi.org/10.1038/s41559-017-0165

Massoni, J., Couvreur, T. L., & Sauquet, H. (2015). Five major shifts of diversification through the long evolutionary history of Magnoliidae (angiosperms). *BMC Evolutionary Biology*, *15*, 49. https://doi.org/10.1186/s12862-015-0320-6

Matzke, N. J. (2013a). *BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts*. In (Version 1.1.2) University of California, Berkeley. https://github.com/nmatzke/BioGeoBEARS

Matzke, N. J. (2013b). *Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model*

*testing* University of California]. https://escholarship.org/content/qt44j7n141/qt44j7n141_noSplash_a25fcfcd2e4c86c599d6f7da5ee1d7f8.pdf?t=pga0we

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. https://doi.org/10.1093/sysbio/syu056

Matzke, N. J. (2022). Statistical comparison of DEC and DEC+J is identical to comparison of two ClaSSE submodels, and is therefore valid. *Journal of Biogeography*, *49*(10), 1805-1824. https://doi.org/10.1111/jbi.14346

Matzke, N. J. (2023). *BioGeoBEARS PhyloWiki*. Retrieved 4 June from http://phylo.wikidot.com/biogeobears#files

Meseguer, A. S., Lobo, J. M., Ree, R., Beerling, D. J., & Sanmartin, I. (2015). Integrating fossils, phylogenies, and niche models into biogeography to reveal ancient evolutionary history: the case of *Hypericum* (hypericaceae). *Systematic Biology*, *64*(2), 215-232. https://doi.org/10.1093/sysbio/syu088

Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome Research*, *15*(12), 1767-1776. https://doi.org/10.1101/gr.3770505

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

Morales-Briones, D. F., Kadereit, G., Tefarikis, D. T., Moore, M. J., Smith, S. A., Brockington, S. F., Timoneda, A., Yim, W. C., Cushman, J. C., & Yang, Y. (2021). Disentangling Sources of Gene Tree Discordance in Phylogenomic Data Sets: Testing Ancient Hybridizations in Amaranthaceae s.l. *Systematic Biology*, *70*(2), 219-235. https://doi.org/10.1093/sysbio/syaa066

Morrone, J. J., & Crisci, J. V. (1995). HISTORICAL BIOGEOGRAPHY: Introduction to Methods. *Annual Review of Ecology and Systematics*, *26*(1), 373-401. https://doi.org/10.1146/annurev.es.26.110195.002105

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, *51 Pt 1*, 263-273. https://doi.org/10.1101/sqb.1986.051.01.032

Muñoz, J., Felicísimo, Á. M., Cabezas, F., Burgaz, A. R., & Martínez, I. (2004). Wind as a Long-Distance Dispersal Vehicle in the Southern Hemisphere. *Science*, *304*(5674), 1144-1147. http://www.jstor.org/stable/3836873

Nelson, G., & Platnick, N. (1982). *Systematics and Biogeography: Cladistics and Vicariance* (Vol. 31). Columbia Univ. Press. https://doi.org/10.2307/2413040

Nie, Z. L., Deng, T., Meng, Y., Sun, H., & Wen, J. (2013). Post-Boreotropical dispersals explain the pantropical disjunction in *Paederia* (Rubiaceae). *Annals of Botany*, *111*(5), 873-886. https://doi.org/10.1093/aob/mct053

Nixon, K. C. (2001). Phylogeny. In S. A. Levin (Ed.), *Encyclopedia of Biodiversity (Second Edition)* (pp. 16-23). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-384719-5.00108-8

Onstein, R. E., Kissling, W. D., Chatrou, L. W., Couvreur, T. L. P., Morlon, H., & Sauquet, H. (2019). Which frugivory-related traits facilitated historical long-distance dispersal in the custard apple family (Annonaceae)? *Journal of Biogeography*, *46*(8), 1874-1888. https://doi.org/10.1111/jbi.13552

Parham, J. F., Donoghue, P. C., Bell, C. J., Calway, T. D., Head, J. J., Holroyd, P. A., Inoue, J. G., Irmis, R. B., Joyce, W. G., Ksepka, D. T., Patane, J. S., Smith, N. D., Tarver, J.

E., van Tuinen, M., Yang, Z., Angielczyk, K. D., Greenwood, J. M., Hipsley, C. A., Jacobs, L., . . . Benton, M. J. (2012). Best practices for justifying fossil calibrations. *Systematic Biology*, *61*(2), 346-359. https://doi.org/10.1093/sysbio/syr107

Pennington, R. T., Cronk, Q. C., & Richardson, J. A. (2004). Introduction and synthesis: Plant phylogeny and the origin of major biomes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *359*(1450), 1455-1464. https://doi.org/10.1098/rstb.2004.1539

Pignatti, E., Pignatti, S., D'Angeli, D., De Nicola, C., Maffei, L., Testi, A., & Tinelli, A. (2015). The Laurisilva as a cultural heritage: proposal for the protection of the relict of laurel forest near Ponte Renaro. *Rendiconti Lincei*, *26*(S3), 643-649. https://doi.org/10.1007/s12210-015-0389-2

POWO. (2023). *Plants of the World Online* http://www.plantsoftheworldonline.org/

Ramirez-Barahona, S., Sauquet, H., & Magallon, S. (2020). The delayed and geographically heterogeneous diversification of flowering plant families. *Nature Ecology and Evolution*, *4*(9), 1232-1238. https://doi.org/10.1038/s41559-020-1241-3

Ree, R. H., & Sanmartín, I. (2009). Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography*, *36*(7), 1211-1220. 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. https://doi.org/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. https://doi.org/10.1080/10635150701883881

Reis-Avila, G., & Oliveira, J. M. (2017). Lauraceae: A promising family for the advance of neotropical dendrochronology. *Dendrochronologia*, *44*, 103-116. https://doi.org/10.1016/j.dendro.2017.04.002

Renner, S. S. (2005). Relaxed molecular clocks for dating historical plant dispersal events. *Trends in Plant Science*, *10*(11), 550-558. https://doi.org/10.1016/j.tplants.2005.09.010

Richter, H. G. (1981). Anatomie des sekundaren Xylems und der Rinde der Lauraceae. *Sonderbande des Naturwiss*.

Ritchie, A. M., Lo, N., & Ho, S. Y. W. (2017). The Impact of the Tree Prior on Molecular Dating of Data Sets Containing a Mixture of Inter- and Intraspecies Sampling. *Systematic Biology*, *66*(3), 413-425. https://doi.org/10.1093/sysbio/syw095

Rohwer, J. G. (1993). Lauraceae. In K. Kubitzki, J. G. Rohwer, & V. Bittrich (Eds.), *Flowering Plants · Dicotyledons: Magnoliid, Hamamelid and Caryophyllid Families* (pp. 366-391). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-02899-5_46

Rohwer, J. G. (2000). Toward a Phylogenetic Classification of the Lauraceae: Evidence from matK Sequences. *Systematic Botany*, *25*(1). https://doi.org/10.2307/2666673

Rohwer, J. G., De Moraes, P. L. R., Rudolph, B., & Werff, H. V. D. (2014). A phylogenetic analysis of the Cryptocarya group (Lauraceae), and relationships of *Dahlgrenodendron*, *Sinopora*, *Triadodaphne*, and *Yasunia*. *Phytotaxa*, *158*(2). https://doi.org/10.11646/phytotaxa.158.2.1

Rohwer, J. G., & Rudolph, B. (2005). Jumping Genera: The Phylogenetic Positions of *Cassytha*, *Hypodaphnis*, and *Neocinnamomum* (Lauraceae) Based on Different Analyses of trnK Intron Sequences. *Annals of the Missouri Botanical Garden*, *92*(2), 153-178. http://www.jstor.org/stable/3298512

Ronquist, F. (1997). Dispersal-Vicariance Analysis: A New Approach to the Quantification of Historical Biogeography. *Systematic Biology*, *46*(1). https://doi.org/10.2307/2413643

Rutschmann, F. (2006). Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. *Diversity and Distributions*, *12*(1), 35-48. https://doi.org/10.1111/j.1366-9516.2006.00210.x

Sanderson, M. J., Thorne, J. L., Wikström, N., & Bremer, K. (2004). Molecular evidence on plant divergence times. *American Journal of Botany*, *91*(10), 1656-1665. https://doi.org/10.3732/ajb.91.10.1656

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, *74*(12), 5463-5467. https://doi.org/https://doi.org/10.1073%2Fpnas.74.12.5463

Sanmartín, I. (2012). Historical Biogeography: Evolution in Time and Space. *Evolution: Education and Outreach*, *5*(4), 555-568. https://doi.org/10.1007/s12052-012-0421-2

Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*, *12*(6), 355-367. https://doi.org/10.1016/j.crpv.2013.07.003

Sauquet, H., Ho, S. Y., Gandolfo, M. A., Jordan, G. J., Wilf, P., Cantrill, D. J., Bayly, M. J., Bromham, L., Brown, G. K., Carpenter, R. J., Lee, D. M., Murphy, D. J., Sniderman, J. M., & Udovicic, F. (2012). Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). *Systematic Biology*, *61*(2), 289-313. https://doi.org/10.1093/sysbio/syr116

Schenk, J. J. (2016). Consequences of Secondary Calibrations on Divergence Time Estimates. *PLOS ONE*, *11*(1), e0148228. https://doi.org/10.1371/journal.pone.0148228

Schneider, J. V., Jungcurt, T., Cardoso, D., Amorim, A. M., Paule, J., & Zizka, G. (2022). Predominantly Eastward Long-Distance Dispersal in Pantropical Ochnaceae Inferred From Ancestral Range Estimation and Phylogenomics. *Frontiers in Ecology and Evolution*, *10*. https://doi.org/10.3389/fevo.2022.813336

Schroeder, C. A. (1976). Some useful plants of the botanical family Lauraceae. *Yearbook of the California Avocado Society for the year*, 9 30-34. https://eurekamag.com/research/017/089/017089031.php

Shaul, S., & Graur, D. (2002). Playing chicken (*Gallus gallus*): methodological inconsistencies of molecular divergence date estimates due to secondary calibration points. *Gene*, *300*(1-2), 59-61. https://doi.org/10.1016/s0378-1119(02)00851-x

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. https://doi.org/10.1371/journal.pone.0197433

Smitinand, T. (1958). The genus *Dipterocarpus* Gaertn.f. in Thailand. *Thai Forest Bulletin (Botany)*, *4*, 1-50.

Song, Y., Yu, W. B., Tan, Y. H., Jin, J. J., Wang, B., Yang, J. B., Liu, B., & Corlett, R. T. (2019). Plastid phylogenomics improve phylogenetic resolution in the Lauraceae. *Journal of Systematics and Evolution*, *58*(4), 423-439. https://doi.org/10.1111/jse.12536

Sri-Ngernyuang, K., Kanzaki, M., Mizuno, T., Noguchi, H., Teejuntuk, S., Sungpalee, C., Hara, M., Yamakura, T., Sahunalu, P., Dhanmanonda, P., & Bunyavejchewin, S. (2003). Habitat differentiation of Lauraceae species in a tropical lower montane forest in northern Thailand. *Ecological Research*, *18*(1), 1-14. https://doi.org/10.1046/j.1440-1703.2003.00539.x

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, *17*, 57-86.

Tiffney, B. H. (1985). The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the northern hemisphere. *Journal of the Arnold Arboretum*, *66*(2), 243-273. http://www.jstor.org/stable/43782164

van der Merwe, M., Crayn, D. M., Ford, A. J., Weston, P. H., & Rossetto, M. (2016). Evolution of Australian *Cryptocarya* (Lauraceae) based on nuclear and plastid phylogenetic trees: evidence of recent landscape-level disjunctions. *Australian Systematic Botany*, *29*(2). https://doi.org/10.1071/sb16023

van der Werff, H., & Richter, H. G. (1996). Toward an Improved Classification of Lauraceae. *Annals of the Missouri Botanical Garden*, *83*(3). https://doi.org/10.2307/2399870

Wang, W., Ortiz, R. D. C., Jacques, F. M. B., Xiang, X. G., Li, H. L., Lin, L., Li, R. Q., Liu, Y., Soltis, P. S., Soltis, D. E., & Chen, Z. D. (2012). Menispermaceae and the diversification of tropical rainforests near the Cretaceous-Paleogene boundary. *New Phytologist*, *195*(2), 470-478. https://doi.org/10.1111/j.1469-8137.2012.04158.x

Warnock, R. C., Parham, J. F., Joyce, W. G., Lyson, T. R., & Donoghue, P. C. (2015). Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proceedings: Biological Sciences*, *282*(1798), 20141013. https://doi.org/10.1098/rspb.2014.1013

Warnock, R. C., Yang, Z., & Donoghue, P. C. (2012). Exploring uncertainty in the calibration of the molecular clock. *Biology Letters*, *8*(1), 156-159. https://doi.org/10.1098/rsbl.2011.0710

Yang, T., Lu, L. M., Wang, W., Li, J. H., Manchester, S. R., Wen, J., & Chen, Z. D. (2018). Boreotropical range expansion and long-distance dispersal explain two amphi-Pacific tropical disjunctions in Sabiaceae. *Mol Phylogenet Evol*, *124*, 181-191. https://doi.org/10.1016/j.ympev.2018.03.005

Young, A. D., & Gillung, J. P. (2019). Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, *45*(2), 225-247. https://doi.org/10.1111/syen.12406

Yule, G. U. (1925). A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *213*, 21-87. http://www.jstor.org/stable/92117

Zerega, N. J., Clement, W. L., Datwyler, S. L., & Weiblen, G. D. (2005). Biogeography and divergence times in the mulberry family (Moraceae). *Molecular phylogenetics and evolution*, *37*(2), 402-416. https://doi.org/10.1016/j.ympev.2005.07.004

Zhao, J. L., Yu, X. Q., Kress, W. J., Wang, Y. L., Xia, Y. M., & Li, Q. J. (2022). Historical biogeography of the gingers and its implications for shifts in tropical rain forest habitats. *Journal of Biogeography*, *49*(7), 1339-1351. https://doi.org/10.1111/jbi.14386

Zuckerkandl, E., & Pauling, L. (1965). Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, *8*(2), 357-366. https://doi.org/10.1016/0022-5193(65)90083-4

# Part B: Article manuscript

# Long-distance dispersal shaped the pantropical distribution of the geographically structured Cryptocarya group

Laura Kragh Frederiksen[1*], Jens G. Rohwer[2] and Wolf L. Eiserhardt[1]

[1] *Ecoinformatics and Biodiversity Group, Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark*

[2] *Universität Hamburg, Institut für Pflanzenwissenschaften und Mikrobiologie, Ohnhorststr. 18, 22609 Hamburg, Germany*

*Correspondence: laura.kragh.frederiksen@gmail.com

## ABSTRACT

**Aim** Pantropical disjunctions constitute intriguing patterns in biogeography. Yet, the relative importance of the underlying evolutionary processes (vicariance and dispersal) remains incompletely understood. We aimed to generate the first comprehensive nuclear gene phylogeny of the Cryptocarya group (Lauraceae), a pantropical taxon, to assess congruence with previous plastid phylogenies and explore its biogeographical history. This includes determining the timing of the Cryptocarya group's origination, as well as understanding mechanisms and migration routes underlying the group's pantropical distribution.

**Location** Pantropical

**Taxon** Cryptocarya group (Lauraceae)

**Methods** We obtained sequences based on 342 low-copy nuclear genes for 106 Cryptocarya group species (c. 13 % species-level sampling) and 26 outgroup species to explore phylogenetic relationships, divergence times and conduct ancestral range estimation and biogeographic stochastic modelling (BSM). The phylogeny was inferred using a maximum-likelihood tree inference approach and we estimated divergence times using a Bayesian relaxed clock method.

**Results** Nuclear data provided well-supported phylogenetic relationships within the Cryptocarya group that were highly congruent to comparable plastid phylogenies. Our best model indicated an origin of the Cryptocarya group in a range encompassing Madagascar and Malesia during the Late Cretaceous (78 Ma; 95% highest posterior density [HPD]: 65.6-90.3 Ma). Long-distance dispersal was the principal process by which members of the Cryptocarya group range expanded.

**Main conclusions** We presented the first nuclear phylogeny of the Cryptocarya group, which largely showed congruence to comparable plastid phylogenies. Our analysis suggests that the group likely originated in a range comprising Madagascar and Malesia during the Late Cretaceous and achieved a pantropical distribution by the early Neogene. Long-distance dispersal was the main process for range expanding in this group. Furthermore, our phylogeny revealed strong geographical structure within well-supported clades. This pattern is consistent with several long-distance dispersals followed by persistent periods of sympatric speciation.

**Keywords** ancestral range estimation, Cryptocarya group, divergence time estimation, historical biogeography, Lauraceae, pantropical distribution, phylogeny, tropical rain forest

## INTRODUCTION

Tropical rainforest (TRF) is the most biodiverse terrestrial biome on Earth, harbouring more than 40% of plant species within c. 7% of all continental land (Eiserhardt et al., 2017; Hill & Hill, 2016; Lomolino et al., 2010; Morley, 2007), but their evolutionary origins remain incompletely understood. Currently, TRF is confined to the equatorial zone and has a disjunct distribution spanning Africa, the Americas, southeastern Asia and Australia (Couvreur, Pirie, et al., 2011). Intriguingly, pantropical intercontinental disjunction are common, as seen in several tropical and subtropical families and genera (Givnish & Renner, 2004; Meng et al., 2014). Understanding the origins of TRFs and their disjunct distributions is a crucial challenge in biogeography. In this context, various hypotheses, including vicariance (Raven & Axelrod, 1974) and long-distance dispersal (LDD) events (Renner et al., 2001), have been proposed. Better understanding of plant phylogenetic relationships is fundamental to develop our understanding of biogeographic patterns, such as geographical structure, in pantropical lineages (Eiserhardt et al., 2018; Pérez-Escobar et al., 2021). An examination of geographic structure can as example help elucidate whether lineages have primarily evolved within major blocks of TRF or across regions of tropical rainforests, including the underlying evolutionary mechanisms (Conceição Oliveira et al., 2021). Moreover, considering the role of past connections, such as widespread megathermal climates (e.g., boreotropics) and land-bridges (e.g., the Northern Atlantic Land Bridge; NALB), is essential to comprehend the current distribution of TRF (Lee et al., 2020; Morley, 2007; Tiffney, 1985). So far, several studies have been undertaken including in Annonaceae Juss. (Couvreur, Pirie, et al., 2011), Arecaceae Bercht. & J.Presl (Couvreur, Forest, et al., 2011), Chrysobalanaceae R.Br (Bardon et al., 2016), Fabaceae Lindl. (Lavin et al., 2004), Malvaceae Juss. (Richardson et al., 2015), Melastomataceae Juss. (Renner et al., 2001), Menispermaceae Juss. (Wang et al., 2012), *Paederia* L. (Rubiaceae Juss.; Nie et al., 2013) and Sapotaceae Juss. (Richardson et al., 2014). Analysing more TRF restricted plant families or genera will undoubtedly add to this knowledge (Couvreur & Baker, 2013). Nevertheless, within Laurales, an ancient (mean: 117-131 Ma) angiosperm lineage (Massoni et al., 2015), few comparable studies have been undertaken (e.g., Huang et al., 2016; Michalak et al., 2010), but given the lineage's antiquity, it has the potential to significantly contribute to our understanding of when, where, and how TRFs emerged.

Within Laurales, Lauraceae Juss. is a large pantropical family of mainly woody plants with 56 accepted genera and c. 3500 species (POWO, 2023) and is among the most encountered tree families in TRF (Balslev, 1998; Gentry, 1988). The Cryptocarya group with 826 species (POWO) includes *Eusideroxylon* Teijsm. & Binn., *Potoxylon* Kosterm., *Aspidostemon* Rohwer & H.G.Richt., *Dahlgrenodendron* van der Merwe & van Wyk, *Cryptocarya* R.Br. and the Beilschmiedia complex, and is recognised as sister to all other genera within Lauraceae except *Hypodaphnis* Stapf (Rohwer et al., 2014). As of Li et al. (2020) the Beilschmiedia complex includes *Beilschmiedia* Nees, *Endiandra* R.Br. (including *Brassiodendron* C.K.Allen and *Triadodaphne* Kosterm.), *Hexapora* Hook.f., *Potameia* Thouars, *Sinopora* J.Li, N.H.Xia & H.W.Li, *Syndiclis* Hook.f. and *Yasunia* van der Werff. Of these genera *Beilschmiedia* and *Cryptocarya* are species rich pantropical genera including c. 262 and 362 species, respectively (POWO, 2023), mostly restricted to TRF (81%; Govaerts, 2022). Because of the pantropical distribution and the age of the Cryptocarya group, which must be significantly older than previous age estimates of the split between the Beilschmiedia complex and *Cryptocarya* (~90 Ma) in Chanderbali et al., 2001 and ~50 Ma in Li et al. (2020), this group constitutes a suitable model for investigating the development of TRF (Chanderbali et al., 2001; Couvreur & Baker, 2013). However, in order to make the most of the group's potential as a TRF model group it is necessary to study the phylogenetic relationships as they are incompletely understood. In recent years several studies have worked on resolving these relationships based on plastid regions.

Among others, Rohwer et al. (2014) carried out a phylogenetic analysis of the entire Cryptocarya group, Li et al. (2020) of the Beilschmiedia complex and C. Nehrke et al. (personal communication; June, 2023) of *Cryptocarya*. Yet, further studies including more markers and collections, especially from poorly represented regions (e.g., tropical Africa, South Africa and Madagascar) are needed to increase resolution and understand relationships among species from different biogeographical regions.

Previous studies on Cryptocarya group phylogenetics have, except for the nuclear internal transcribed spacer (nrITS), relied on plastid phylogenies, which often but not always reflect the evolutionary history of species, and data from the nucleus are needed to test and extend the existing phylogenetic hypotheses. Plastid regions have historically provided valuable insights into plant phylogeny, particularly at higher taxonomic levels (Davis et al., 2014). The widespread use of plastid regions can be attributed, in part, to their ease of sequencing (Pérez-Escobar et al., 2021). However, the emergence of high-throughput sequencing methods has made the use of low-copy nuclear genes more accessible and cost-effective, enabling the inclusion of multiple genes in large-scale phylogenetic analyses (Mardis, 2011; Zeng et al., 2014). Several studies, among other in the Lauraceae family (Song et al., 2019), have shown that employing the plastome, as opposed to only a few plastid markers, yields phylogenetic trees with robust support, primarily due to the incorporation of a larger dataset that minimises stochastic and sampling errors (Young & Gillung, 2019). To approximate the 'true' species tree, it is crucial to include numerous genes, to account for processes such as incomplete lineage sorting (ILS) and hybridization that contribute to varying genealogical histories. Consequently, a gene tree may not accurately reflect the species tree (Doyle, 1992). Incongruence between plastid and nuclear phylogenies has been reported in several studies (e.g., Barrett et al., 2014; Perez-Escobar et al., 2021), attributable to the divergent nature of these genomes. Plastids are uniparentally inherited with limited recombination and evolve as a single unit, thus resembling one gene tree (Doyle, 1992; Gonçalves et al., 2019), while nuclear genomes exhibit biparental inheritance, higher levels of recombination and each gene evolves independently (Zeng et al., 2014). Therefore, the choice between nuclear genes and the plastome as the basis for constructing a species tree holds significant importance. Comparative analyses have demonstrated that the implementation of low-copy nuclear genes, specifically those found in the Angiosperms353 bait kit (Johnson et al., 2019), outperforms plastid regions in both shallow and deep-time phylogenetic reconstructions (Pérez-Escobar et al., 2021; Sang, 2002). Nevertheless, nuclear and plastid phylogenies can be largely congruent (Li et al., 2020; Pérez-Escobar et al., 2021; Zeng et al., 2014), but this relationship has yet to be explored within the Cryptocarya group. Therefore, the question arises: Does target sequence capture using the nuclear Angiosperms353 probe set confirm relationships within the Cryptocarya group, based on plastid regions?

Traditionally Gondwanan vicariance (Raven & Axelrod, 1974) has been the preferred hypothesis explaining how numerous TRF lineages have developed a transcontinental disjunct distribution. However, as an increasing number of biogeographic studies incorporating fossil-calibrated divergence time estimation have been undertaken, many of these lineages appear to be younger than the break-up of Gondwana (Baker et al., 2013; Couvreur, Forest, et al., 2011). Long-distance dispersal is thus steadily becoming the more plausible *a priori* assumption of these pantropical disjunctions (Schneider et al., 2022). Therefore, it is now important to look for generalities in dispersal patterns, routes, and vectors to identify important sources and sinks. Earlier attempts to touch upon historical biogeography of the Cryptocarya group provided the first insight into divergence events and the possible roles of vicariance vs. LDD. According to Li et al. (2020) based on 8% species-level sampling, strongly biased towards Asian/Australian

taxa, the split between the Beilschmiedia complex and *Cryptocarya* occurred in the Early Eocene ~50 Ma, supporting LDD as the most plausible explanation for this split. In comparison Chanderbali et al. (2001) estimated this split to the Late Cretaceous ~90 Ma, suggesting that the group could have a Gondwanan history. Given the incongruence between these two estimates, this remains an open question. Although divergence times have been estimated within Lauraceae (Chanderbali et al., 2001) and the Beilschmiedia complex (Li et al., 2020), this remains to be achieved for *Cryptocarya* and allies (Rohwer et al., 2014). Consequently, no study has yet presented divergence time estimates for all genera within this group. Also, Li et al. (2020) reported that sampling from a few key geographical regions, in particular Africa and Madagascar, was not achieved in their study. Hence, at this point we can only speculate whether *Beilschmiedia* colonised Africa and Madagascar by similar routes as seen in previous studies, perhaps by LDD (Federman et al., 2015; Yoder & Nowak, 2006). Moreover, given that previous studies have relied on divergence time estimates and the fossil record alone, they were limited in their capacity to infer where clades originated and how they dispersed. We were, however, able to accomplish this by further including ancestral range estimation as implemented in e.g., BioGeoBEARS (Matzke, 2013). Taken together, a significant biogeographical knowledge gap remains to be filled within the Cryptocarya group. Thus, we ask when and where the Cryptocarya group originated and how it achieved its pan-tropical distribution? Additionally, did the Beilschmiedia complex's divergence from *Cryptocarya* occur during the early Eocene or late Cretaceous? Also, was *Beilschmiedia*'s colonisation of Madagascar the result of LDD?

Geographical structuring is recognised as an important predictor of phylogenetic relationships (Conceição Oliveira et al., 2021). Thus, an assessment of geographic structure in the phylogeny can contribute to understanding the evolution of clades. Specifically, in our study it can help pinpoint whether species within a clade primarily have evolved within individual blocks of TRF, indicating geographical homogeneity. This pattern can be attributed to phylogenetic niche conservatism (PNC; Donoghue, 2008), dispersal limitation, and within-area speciation (Pennington et al., 2006). Consequently, groups with high geographical structure are expected to have evolved in relative isolation over extended periods, with limited dispersal or vicariance events to connect disjunct areas of rainforest (Simon et al., 2011). This pattern of high geographical structure has been observed within well-supported clades of the Cryptocarya group, as reported by Rohwer et al. (2014). Thus, we await to see whether this pattern also endures in our study when including denser taxon sampling.

To examine congruence between plastid and nuclear phylogenies and to investigate the biogeographical history of the TRF-restricted Cryptocarya group, we applied the universal Angiosperms353 probe set (Johnson et al., 2019) to generate phylogenomic nuclear data, hitherto not applied in the Cryptocarya group, including 13% species-level sampling. We inferred their phylogeny using likelihood and ASTRAL analyses and calculated molecular age estimates using a Bayesian approach. Based on this we carried out ancestral range estimation using a maximum likelihood approach and stochastic mapping, to estimate the origin of the Cryptocarya group in addition to the evolutionary processes that have driven the extant distribution.

## MATERIALS AND METHODS

### Taxon sampling

119 samples were obtained from a combination of silica-dried and herbarium specimens, constituting a sampling effort of 13% across the Cryptocarya group (Govaerts et al., 2021).

Hereof 2 *Aspidostemo*n (7%), 36 *Beilschmiedia* (14%), 43 *Cryptocarya* (12%), 13 *Endiandra* (10%), 1 *Eusideroxylon* (100%), 3 *Potameia* (13%), 1 *Potoxylon* (100%), 1 *Sinopora* (100%) and 2 *Syndiclis* (20%) were included to understand intrageneric relationships within this group. This constitutes the yet broadest and densest sampling of the Cryptocarya group (Rohwer, 2014; Li et al., 2020). Most species were represented by a single sample (Table 1).

As outgroup, sequences of 26 species, hereof 24 species from nine lauralean families (Lauraceae, Monimiaceae Juss., Hernandiaceae Blume, Athersopermataceae R.Br., Gomortegaceae Reiche, Siparunaceae Schodde and Calycanthaceae Lindl.; see Table 1) and two magnolialean families (Magnoliaceae Juss. and Myristicaceae R.Br., see Table 1) were retrieved from Baker et al. (2022) to obtain nodes required for fossil calibration.

**DNA extraction and sequencing**

DNA from silica-gel dried material and herbarium specimens was isolated with the innuPrep Plant DNA Kit (Analytik Jena, Germany) according to the manufacturer's protocol with modifications given by Rohwer and Rudolph (2005); Trofimov et al. (2016). Leaf fragments of approximately 5 x 5 mm were ground in a ball mill (MM 400, Retsch GmbH, Haan), hereafter incubated with lysis buffer in a shaker-incubator at 65 °C for 30 min. To increase DNA yield, the incubation time of the elution step was elongated to 15 min. Furthermore, a higher temperature (50 °C) was applied to improve the solution of the residual DNA from the column.

DNA was prepared for target sequence capture using a NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs Ltd, Hitchin, UK) but with half volumes of buffers and adapters (Hale et al., 2020). Where necessary DNA fragments were sheared to c. 300 – 400 bp of size, as targeted during library preparation using a Covaris ME220 Focused-ultrasonicator (Covaris Ltd, Brighton, UK).

DNA libraries were prepared with 50 – 200 ng input DNA. Libraries were except for highly fragmented input DNA size-selected to 300–400 bp insert size using NEBNext Multiplex Oligos dual indexing for Illumina (New England BioLabs Ltd, Hitchin, UK), and 10 – 12 PCR cycles. Quality of the prepared DNA libraries was assessed using an Agilent 4200 TapeStation (Agilent Technologies LDA UK Limited, Stockport, UK). Total genomic DNA libraries were enriched for selected target loci using RNA baits (MyBaits, Arbor Biosciences, Ann Arbor, Michigan, U.S.A.) in six pools of 120 libraries in total, following manufacturer's protocols. All samples were enriched for the nuclear Angiosperm353 target loci (Johnson et al., 2019). Sequencing of enriched libraries was conducted at Novogene (Oxford, UK) on an Illumina NovaSeq 6000 sequencer, generating 2 x 150 bp paired end reads. These sequence data have been submitted to the NCBI Sequence Read Archive under accession number PRJNA939499, https://www.ncbi.nlm.nih.gov/sra/PRJNA939499.

*Outgroup samples*

Data generation of sequences retrieved from https://treeoflife.kew.org (Table 1) is described in Carpenter et al. (2019) and Baker et al. (2022) respectively, whereas SRA data (Sayers et al., 2022) was generated according to the individual study's protocol.

**Table 1** Samples of Cryptocarya group and outgroup with their NCBI accession numbers, specimen preservation, sampling location (Country) and collector. Herbarium acronyms following Thiers (continuously updated). SANBI = South African National Biodiversity. BG = Botanical Garden.

| Taxon name | Accessions | Country | Collector (Herbarium) | Specimen preservation |
|---|---|---|---|---|
| **Ingroup:** | | | | |
| *Aspidostemon fungiformis* | SRR23971709 | Madagascar | van der Werff 12765 (MO) | Silica |
| *Aspidostemon parvifolium* | SRR23971708 | Madagascar | Lowry 5024 (MO) | Silica |
| *Beilschmiedia appendiculata* | SRR23971633 | China | Li Jie 2006077 (KUN) | Silica |
| *Beilschmiedia berteroana* | SRR23971622 | Chile | Zöllner 21411 (HBG) | Herbarium |
| *Beilschmiedia brachythyrsa* | SRR23971611 | China | Huang, J. H-BN026 (KUN) | Silica |
| *Beilschmiedia brenesii* | SRR23971696 | Costa Rica | Yasuda 1314 (MO) | Silica |
| *Beilschmiedia dictyoneura* | SRR23971685 | Indonesia | Ambri W698 (L) | Herbarium |
| *Beilschmiedia emarginata* | SRR23971650 | Brazil | Moraes 5412 (HCRB) | Silica |
| *Beilschmiedia fasciata* | SRR23971639 | China | DNA Barcoding Group CBOWS385 (KUN) | Silica |
| *Beilschmiedia fulva* | SRR23971604 | Gabon | Breteler 15088 (WAG) | Herbarium |
| *Beilschmiedia furfuracea* | SRR23971707 | China | Li Jie 2002178 (KUN) | Silica |
| *Beilschmiedia henghsienensis* | SRR23971672 | China | Li Jie 2002135 (KUN) | Silica |
| *Beilschmiedia latifolia* | SRR23971665 | Peru | Monteagudo 3700 (HBG) | Herbarium |
| *Beilschmiedia latifolia* | SRR23971664 | Peru | van der Werff 18547 (HBG) | Herbarium |
| *Beilschmiedia linharensis* | SRR23971663 | Brazil | Moraes 2550 | Silica |
| *Beilschmiedia linocieroides* | SRR23971662 | China | Huang, J. H-BNO066 (KUN) | Silica |
| *Beilschmiedia macropoda* | SRR23971661 | China | Li Jie 2006084 (KUN) | Silica |
| *Beilschmiedia madagascariensis* | SRR23971660 | Madagascar | Lowry 5015 (MO) | Silica |
| *Beilschmiedia mannii* | SRR23971659 | Ghana | Jongkind 1649 (WAG) | Herbarium |
| *Beilschmiedia manii* | SRR23971634 | Liberia | Jongkind 6110 (WAG) | Herbarium |
| *Beilschmiedia manii* | SRR23971632 | Liberia | Jongkind 12931 (WAG) | Herbarium |
| *Beilschmiedia miersii* | SRR23971631 | Chile | Greissl 657-99 (MJG) | Herbarium |
| *Beilschmiedia moratii* | SRR23971630 | Madagascar | Martial 323 (MO) | Silica |
| *Beilschmiedia pauciflora* | SRR23971629 | China | Huang, X. JYL-3901 (KUN) | Silica |
| *Beilschmiedia pedicellata* | SRR23971628 | Madagascar | Antilahimena 3187 (WAG) | Herbarium |
| *Beilschmiedia percoriacea* | SRR23971626 | China | Zhihua Wang WO8062 (KUN) | Silica |
| *Beilschmiedia percoriacea var. cilliata* | SRR23971627 | China | Huang, X. JYL-3281 (KUN) | Silica |
| *Beilschmiedia purpurascens* | SRR23971625 | China | Huang, X. LSL-0812 (KUN) | Silica |
| *Beilschmiedia robusta* | SRR23971624 | China | Huang, X. JYL-5507 (KUN) | Silica |
| *Beilschmiedia roxburgiana* | SRR23971623 | Germany (HBG) | Hoffmann s.n. (HBG) | Silica |

| | | | | |
|---|---|---|---|---|
| *Beilschmiedia roxburgiana* | SRR23971621 | China | Huang, X. JYL-5215 (KUN) | Silica |
| *Beilschmiedia rufohirtella* | SRR23971620 | China | Lang Li 20070019A (KUN) | Silica |
| *Beilschmiedia rugosa* | SRR23971619 | Madagascar | Turk 205 (WAG) | Herbarium |
| *Beilschmiedia sary* | SRR23971618 | Madagascar Toamasina | van der Werff 12800 (WAG) | Herbarium |
| *Beilschmiedia sericans* | SRR23971617 | Madagascar | Martial 81 (MO) | Silica |
| *Beilschmiedia tarairi* | SRR23971616 | New Zealand | Sykes 529799 A (CHR) | NA |
| *Beilschmiedia tarairi* | SRR23971615 | New Zealand | collector not indicated 616303 (CHR) | Silica |
| *Beilschmiedia tawa* | SRR23971614 | New Zealand | Dawson 565486 A (CHR) | Silica |
| *Beilschmiedia tawa* | SRR23971613 | New Zealand | PB Heenan 616304 (CHR) | NA |
| *Beilschmiedia tawaroa* | SRR23971612 | New Zealand | Wright 451260 (CHR) | NA |
| *Beilschmiedia tilaranensis* | SRR23971706 | Costa Rica | Yasuda 1313 (MO) | Silica |
| *Beilschmiedia tungfangensis* | SRR23971705 | China | Li Jie 2006072 (KUN) | Silica |
| *Beilschmiedia ugandensis* | SRR23971704 | Congo | Malaisse 13528 (WAG) | Herbarium |
| *Beilschmiedia velutina* | SRR23971703 | Madagascar | Antilahimena 8010 (MO) | Silica |
| *Beilschmiedia volckii* | SRR23971702 | Australia | van der Werff 17052 (MO) | Silica |
| *Beilschmiedia yunnanensis* | SRR23971701 | China | Huang, X. JYL-3138 (KUN) | Silica |
| *Cryptocarya acutifolia* | SRR23971700 | China | Huang, X. JYL-4153 (KUN) | Silica |
| *Cryptocarya alba* | SRR23971699 | Chile, cultiv. Edinburgh | Browp 1237 (cult. E) | Silica |
| *Cryptocarya* cf. *albida* | SRR23971698 | Indonesia | Brambach 682 (HBG) | Silica |
| *Cryptocarya* cf. *ampla* | SRR23971697 | Indonesia | Brambach 646 (HBG) | Silica |
| *Cryptocarya aschersoniana* | SRR23971695 | Brazil | Moraes 5405 (HRCB) | Silica |
| *Cryptocarya botelhensis* | SRR23971694 | Brazil | Moraes 2311 (HRCB) | Silica |
| *Cryptocarya calcicola* | SRR23971693 | China | Huang, X. JYL-4611 (KUN) | Silica |
| *Cryptocarya chinensis* | SRR23971692 | China | Li Jie 2006056 (KUN) | Silica |
| *Cryptocarya citriformis* | SRR23971691 | Brazil | Moraes 3199 (HRCB) | Silica |
| *Cryptocarya concinna* | SRR23971690 | China | Huang, J. H-BN054 | Silica |
| *Cryptocarya conduplicata* | SRR23971648 | New Caledonia | McPherson 19131 (MO) | Silica |
| *Cryptocarya depauperata* | SRR23971646 | China | Wang W08064 (KUN) | Silica |
| *Cryptocarya ferrea* | SRR23971688 | Indonesia | Brambach 625 (HBG) | Silica |
| *Cryptocarya* cf. *fusca* | SRR23971687 | Polynesia | Whistler 6685 (HBG) | Herbarium |
| *Cryptocarya hainanensis* | SRR23971686 | China | Huang, J. H-BN049 (KUN) | Silica |
| *Cryptocarya hornei* | SRR23971684 | Polynesia | Whistler 6685 (HBG) | Herbarium |
| *Cryptocarya krameri* | SRR23971683 | Madagascar | Gautier 3158 (HBG) | Herbarium |
| *Cryptocarya leptospermoides* | SRR23971658 | New Caledonia | McPherson 6192 (L) | Herbarium |
| *Cryptocarya liebertiana* | SRR23971657 | South Africa | van Wyk; 2902 (SANBI) | Herbarium |
| *Cryptocarya litoralis* | SRR23971656 | Madagascar | Ratovoson 1745 (MO) | Silica |
| *Cryptocarya litoralis* | SRR23971655 | Madagascar | Ravelonarivo 2983 (MO) | Silica |

| | | | | |
|---|---|---|---|---|
| *Cryptocarya mandioccana* | SRR23971654 | Brazil | Moraes 1245 (ESA) | Silica |
| *Cryptocarya medicinalis* | SRR23971653 | Papua New Guinea | Damas SAJ1123 (BISH) | Silica |
| *Cryptocarya micrantha* | SRR23971652 | Brazil | Moraes 2426 (ESA) | Herbarium |
| *Cryptocarya moschata* | SRR23971651 | Brazil | Moraes 2551 (HRCB) | Silica |
| *Cryptocarya nitens* | SRR23971649 | Indonesia | Hyland 14218 (MO) | Silica |
| *Cryptocarya ovalifolia* | SRR23971647 | Madagascar | Miller, J.S. 8758 (MO) | Silica |
| *Cryptocarya pervillei* | SRR23971645 | Madagascar | Rakotonirina 427 (MO) | Silica |
| *Cryptocarya pervillei* | SRR23971644 | Madagascar | Rasoazanany 230 (MO) | Silica |
| *Cryptocarya polyneura* | SRR23971643 | Madagascar | Antilahimena 8458 (MO) | Silica |
| *Cryptocarya polyneura* | SRR23971642 | Madagascar | Rasoazanany 365 (MO) | Silica |
| *Cryptocarya rhodosperma* | SRR23971641 | Australia | Gray 7556 (MO) | Silica |
| *Cryptocarya riedeliana* | SRR23971640 | Brazil | Moraes 2465 (ESA) | NA |
| *Cryptocarya rigidifolia* | SRR23971638 | Madagascar | Ravelonarivo 3242 (MO) | Silica |
| *Cryptocarya rolletii* | SRR23971637 | China | Huang, J. H-BN050 (KUN) | Silica |
| *Cryptocarya saligna* | SRR23971636 | Brazil | Moraes 3226 (HRCB) | Silica |
| *Cryptocarya sellowiana* | SRR23971635 | Brazil | Oliveira Filho s.n. (HRCB) | Herbarium |
| *Cryptocarya spathulata* | SRR23971610 | Madagascar | Rakotonirina 158 (MO) | Silica |
| *Cryptocarya spathulata* | SRR23971609 | Madagascar | Ravelonarivo 3849 (MO) | Silica |
| *Cryptocarya subtriplinervia* | SRR23971608 | Madagascar | van der Werff 12775 (MO) | Silica |
| *Cryptocarya thouvenotii* | SRR23971607 | Madagascar | van der Werff 12723 (MO) | Silica |
| *Cryptocarya transvaalensis* | SRR23971606 | South Africa | Jacobsen, W 5303 (SANBI) | Herbarium |
| *Cryptocarya velloziana* | SRR23971605 | Brazil | Moraes 3203 (HRCB) | Silica |
| *Cryptocarya woodii* | SRR23971603 | South Africa | Burrows 8940 (SANBI) | Herbarium |
| *Cryptocarya wyliei* | SRR23971602 | BG Edinburgh, South Africa | Browp 261 (cult. E) | Silica |
| *Cryptocarya yunnanensis* | SRR23971601 | China | Huang, J. H-BN092 (KUN) | Silica |
| *Endiandra phaeocarpa* | SRR23971595 | Australia | Gray 8805 (MO) | Silica |
| *Endiandra impressicosta* | SRR23971600 | Australia | Gray 7539 (MO) | Silica |
| *Endiandra jonesii* | SRR23971599 | Australia | van der Werff 17025 (MO) | Silica |
| *Endiandra latifolia* | SRR23971598 | Papua New Guinea | James SAJ1209 (BISH) | Silica |
| *Endiandra lecardii* | SRR23971597 | New Caledonia | McPherson 19147 (MO) | Silica |
| *Endiandra palmerstonii* | SRR23971596 | Australia | van der Werff 17050 (MO) | Silica |
| *Endiandra* cf. *pilosa.* | SRR23971594 | Papua New Guinea | Damas SAJ1183 (BISH) | Silica |
| *Endiandra pouboensis* | SRR23971682 | New Caledonia | McPherson 18988 (MO) | Silica |
| *Endiandra pubens* | SRR23971681 | Australia | Constable 4876 (HBG) | Herbarium |
| *Endiandra sankeyana* | SRR23971680 | Australia | van der Werff 17054 (MO) | Silica |
| *Endiandra scrobiculata* | SRR23971679 | Indonesia | Ambri 502 (HBG) | Silica |
| *Endiandra sulavesiana* | SRR23971678 | Indonesia | Ambri 1467 (HBG) | Silica |
| *Endiandra xanthocarpa* | SRR23971677 | Australia | van der Werff 17046 (MO) | Silica |
| *Eusideroxylon zwageri* | SRR23971676 | BG Singapore | 2004-42 (SING) | Herbarium |
| *Potameia confluens* | SRR23971675 | Madagascar | Rakotonirina 200 (MO) | Silica |

| | | | | |
|---|---|---|---|---|
| *Potameia microphylla* | SRR23971674 | Madagascar | van der Werff 12655 (MO) | Silica |
| *Potameia obtusifolia* | SRR23971673 | Madagascar | Birkinshaw 2007 (MO) | Silica |
| *Potameia obtusifolia* | SRR23971671 | Madagascar | Martial 2 (MO) | Silica |
| *Potoxylon melagangai* | SRR23971670 | Brunei | Wong 325 (AAU) | Herbarium |
| *Sinopora hongkongensis* | SRR23971669 | China | Xia s.n. (MO) | Silica |
| *Syndiclis kwangsiensis* | SRR23971668 | China | Liu 1552 (PE) | Silica |
| *Syndiclis marlipoensis* | SRR23971667 | China | Liu 1282 (PE) | Silica |
| *Syndiclis marlipoensis* | SRR23971666 | China | Liu 1420 (PE) | Silica |
| *Syndiclis chinensis*‡ | SRR1012868 | NA | NA | NA |
| *Endiandra macrophylla*† | ERR7620351 | NA | Bohk 417 (K) | NA |
| *Beilschmiedia tsangii*† | ERR7618678 | NA | Kok, R.P.J. de 1585 (K) | Herbarium |
| **Outgroup:** | | | | |
| *Myristica fragrans** | ERR2040140 | NA | NA | NA |
| *Magnolia grandiflora** | ERR2040139 | NA | NA | NA |
| *Calycanthus floridus** | ERR2040125 | NA | NA | NA |
| *Idiospermum australiense** | ERR2040126 | NA | NA | NA |
| *Siparuna guianensis*† | ERR4180066 | NA | Clarke, HD C9035 (K) | Herbarium |
| *Gomortega keule** | ERR2040127 | NA | NA | NA |
| *Atherosperma moschatum*† | ERR7620993 | NA | De Lirio E.J. s.n. (K) | Herbarium |
| *Hernandia nymphaeifolia*† | ERR4180128 | NA | Chase, M.W. 3837 (K) | Live plant |
| *Sparattanthelium botocudorum*† | ERR7620332 | NA | Kollmann, L. 13254 (MBML) | Herbarium |
| *Peumus boldus*† | ERR2040134 | NA | NA | NA |
| *Monimia rotundifolia*† | ERR7620324 | NA | M.Fay & M.W.Chase 10030 (REU) | Silica |
| *Tambourissa ficus*† | ERR7620307 | NA | Lorence, D. 8847 (PTBG) | Silica |
| *Mollinedia gilgiana*† | ERR7620315 | NA | Lirio, E.J. 131 (RB) | Silica |
| *Neocinnamomum caudatum*† | ERR7621654 | NA | De Kok, R. 1556 (K) | Herbarium |
| *Laurus nobilis*‡ | SRR10064026 | China, Chenshan Botanical Garden, Shanghai | NA | NA |
| *Cassytha filiformis** | ERR2040133 | NA | NA | NA |
| *Caryodaphnopsis tonkinensis*‡ | SRR10064043 | NA | NA | NA |
| *Alseodaphnopsis petiolaris*‡ | SRR10064005 | NA | NA | NA |
| *Dehaasia hainanensis*‡ | SRR10064006 | NA | NA | NA |
| *Machilus salicinus*‡ | SRR10064004 | NA | NA | NA |
| *Nothaphoebe umbelliflora*† | ERR7620353 | NA | Arbainsyah, AA 3141 (K) | Herbarium |
| *Phoebe lanceolata*† | ERR7618680 | NA | De Kok 1518 (K) | DNA bank |
| *Mezilaurus ita-uba*† | ERR7620345 | NA | Perdiz, R.O. 1511 (K) | Herbarium |
| *Persea borbonia** | ERR2040132 | NA | NA | NA |
| *Cinnamomum camphora** | ERR2040129 | NA | NA | NA |

| | | | | |
|---|---|---|---|---|
| *Chimonanthus salicifolius*† | SRR11109040 | NA | NA | NA |

*: Data from One Thousand Plant Transcriptomes Initiative (Carpenter et al., 2019). †: Data from Baker et al. (2022) ‡: Data from Sequence Read Archive (Sayers et al., 2022)

## Pipelines for species tree inference

*Sequence assembly*

FastQC 0.11.9 (Andrews, 2010) was applied to assess sequence read quality before and after trimming applying the MultiQC 1.13 wrapper (Ewels et al., 2016). Trimmomatic 0.39 (Bolger et al., 2014) was applied to trim reads from adapters and low-quality reads. We found that SLIDINGWINDOW with a 4 base pair frame, 33 thred minimum sequence length, and quality threshold of 30 thred was superior to MAXINFO in our study and improved sequence quality satisfactorily.

Trimmed paired and unpaired reads were extracted using the HybPiper 2.0.1 pipeline (Johnson et al., 2016) using BWA v.0.7.17 (Li & Durbin, 2009) to map reads to the 353 nuclear loci in the 'mega353' target file (McLay et al., 2021), an expanded 'Angiosperms353' target file (Johnson et al., 2019) including more species. Reads matching each target locus were subsequently assembled independently with SPAdes 3.15.4 (Bankevich et al., 2012) using default cut-off coverage. To generate supercontigs (exons including the flanking intron and spacer-regions) the 'run_intronerate' flag was selected for ingroup and 10% outgroup species (see 'Contig selection'). Moreover, we assessed for potential paralogs by running the --paralog_retriever pipeline. Duplicate contigs covering $\geq$ 75% of the reference sequence length were flagged as paralogs. We identified ten genes that included up to five potential paralogs per gene, for most samples (see Appendix S1 in Supporting Information). We choose to keep these based on the study by Yan et al. (2022).

Next, paired, and unpaired reads were mapped to raw reads using BWA, duplicated reads were identified and de-dupicated using MarkDuplicates implemented in Picard v.2.22.3 (Broad Institute, 2020), depth was calculated using samtools 1.15.1 (H. Li et al., 2009), and bases only covered by a single read were masked with a read coverage < 2, as done in Gardner et al. (2020), and a trimmed sample-level fasta file was generated (Eiserhardt et al., 2022).

Sequences of < 150 bp or sequence length < 20 % of the median gene length were filtered out since short sequences may be difficult to align (Eiserhardt et al., 2022). For three samples (*Chimonanthus salicifolius* S.Y.Hu, *Peumus boldus* Molina and *Syndiclis chinensis* C.K.Allen), contigs from Baker et al. (2022) were used instead of newly assembled contigs, as the sequence files of the abovementioned files were corrupted.

## Alignment and cleaning

Targeted loci were aligned individually with MAFFT 7.508 (Katoh & Standley, 2013) using global pairwise alignment with the Needleman-Wunsch algorithm, applying 1000 cycles of iterative refinement and accurately adjusted direction. Subsequent, alignment trimming was conducted with trimAL 1.4.1 (Capella-Gutiérrez et al., 2009) to remove fragmentary sites from the alignment for each of the specified (0.1-0.9 by 0.05) gt (gap threshold) values. We used the 'optrimAL' R script of Shee et al. (2020) to set the optimal trimming threshold for each gene, based on summary statistics collected for each alignment using AMAS 1.0 (Borowiec, 2016). Subsequently we used CIalign 1.0.18 (Tumescheit et al., 2022) applying default settings to clean the multiple sequence alignments by removing sources of noise, such as cropping poorly

aligned ends. Final alignment cleaning was conducted with Julia 1.8.2 in Taper (Zhang et al., 2021) applying default settings, to mask errors in small species-specific areas in the alignment. Alignments were visually inspected after trimming to check alignment quality using Geneious Prime 2022.1.1 (https://www.geneious.com). No manual trimming or masking of sequences was conducted to preserve reproducibility and because manual editing is a time-consuming process.

**Phylogenetic analysis**

We used IQ-TREE 2.2.0.3 (Minh et al., 2020) to infer gene trees based on maximum likelihood and 1000 iterations of ultrafast bootstrap (Hoang et al., 2018). Genes were partitioned into intron and exon regions using partitioning (Chernomor et al., 2016) and substitution models were selected individually for each gene, hereunder partitioned exon and intron regions using ModelFinder (Kalyaanamoorthy et al., 2017) in IQ-TREE.

To infer a species tree from the gene trees we used the quartet-based method ASTRAL-III 5.7.8 (Zhang et al., 2017). Prior to species tree assembly, internal branches in gene trees with bootstrap support < 10% were collapsed, using Newick Utilities 1.6 (Junier & Zdobnov, 2010) to avoid unsupported topologies influencing subsequent analyses in ASTRAL-III (Zhang et al., 2018). Finally, the species tree was rooted on *Myristica fragrans* Houtt. and *Magnolia grandiflora* L.

**Contig selection**

In our pipeline to generate a species tree of the Cryptocarya group including outgroup we first generated a tree, using supercontigs for everything (ingroup and outgroup). However, we got an unrealistic topology in the outgroup contradicting Baker et al. (2022) and additional studies (Chanderbali et al., 2001; Song et al., 2019), probably attributable to alignment issues in the non-exonic regions. We used Baker et al. (2022) as benchmark to remove non-exonic regions. Rerunning the analysis with exons we found that the non-exonic regions caused this issue. As compromise we generated a species tree from supercontigs but removing non-exonic regions from the outgroup as necessary. We did this by using SortaDate (Smith et al., 2018) to first identify which of the supercontig gene trees that were most like the outgroup topology of a previously generated exon species tree (only exons in ingroup and outgroup). Hereafter, we generated new gene trees (supercontig + exon) for the outgroup by running IQ-TREE (Minh et al., 2020) and combining different combinations of the previously generated supercontigs and exons until the anticipated outgroup topology was achieved. The correct outgroup included the 10% genes with the highest bipartition scores as supercontig genes and the remaining 90% as exons (Appendix S2). Finally, we reran the analyses using supercontigs for the ingroup and the abovementioned combination of supercontigs and exons for the outgroup.

**Fossil calibration**

In this study we selected eight primary calibration points (Table 2) based on Ramirez-Barahona et al. (2020) and Li et al. (2020; see Appendix S3 for details). Following a conservative calibration approach (Sauquet et al., 2012), uniform priors were chosen, setting minimum ages to estimated fossil ages and maximum ages to infinity. For example, to estimate the minimum age of crown Laurales the fossil *Virginianthus calycanthoides* (Friis et al., 1994) with an estimated age of 107.7 Ma (Massoni et al., 2015) was assigned, applying a uniform prior with minimum age = 107.7 Ma and maximum age = ∞.

The root was calibrated by secondary calibration based on the age from Massoni et al. (2015) using a normal distribution prior, to reflect an equal distribution of uncertainty around the

mean. The mean was set to 124.27 Ma to match the estimated minimum age for stem Laurales in line with the study.

**Table 2** Calibration points used for molecular dating analyses to generate the MCC tree (Figure 1) including information about age, priors, and source.

| Node | Calibration fossil | Minimum age (Ma) | Prior dist. | Prior parameters | Source |
|---|---|---|---|---|---|
| C1: stem Laurales | Secondary | 113 | normal | M: 124.27; SD: 1.43 | Massoni et al. (2015) |
| C2: crown Laurales | *Virginianthus calycanthoides* | 107.7 | uniform | L: 107.7; U: ∞ | Friis et al. (1994) |
| C3: crown Calycanthoideae | *Jerseyanthus calycanthoides* | 85.8 | uniform | L: 85.8; U: ∞ | Crepet et al. (2005) |
| C4: stem Neocinnamomum | *Neusenia tetrasporangiata* | 83.6 | uniform | L: 83.6; U: ∞ | Eklund (2000) |
| C5: crown Monimiaceae | *Monimiophyllum callidentatum* | 52.22 | uniform | L: 52.22; U: ∞ | Knight and Wilf (2013) |
| C6: crown Hernandiaceae | *Illigera eocenica* | 41.2 | uniform | L: 41.2; U: ∞ | Manchester and O'Leary (2010) |
| C7: crown clade *C. aristata*, *C. conduplicata*, *C. pluricostata* and *C. rhodosperma* | *Cryptocaryoxylon gippslandicum* | 39 | uniform | L: 39; U: ∞ | Leisman (1986) |
| C8: crown Perseae | *Alseodaphne changchangensis* | 37-48 | uniform | L: 37; U: ∞ | J. Li et al. (2009) |
| C9: stem Machilus Nees | *Machilus maomingensis* | 33.7-33.9 | uniform | L: 33.7; U: ∞ | Tang et al. (2016) |

M: mean, SD: standard deviation, L: lower, U: upper, dist. = distribution.

## Molecular dating

For molecular dating and biogeographical analysis, we selected one individual per species to comply with ancestral range estimation (Matzke, 2013) model assumptions, selecting the sample with fewest gaps in the alignment.

Genes were selected by 'gene shopping' with SortaDate to assess topological species tree conflict. Only the 27 genes for which non-exonic regions were included in both ingroup and outgroup were considered to ensure a similar evolutionary rate across the tree. The three genes (*5594*, *5620* and *6139*) with least topological species tree conflict were selected for subsequent divergence-time inference to minimise gene-tree conflict (Carruthers et al., 2022).

Molecular dating was performed using BEAST 1.10.4 (Suchard et al., 2018). The best-performing evolutionary model for each gene was determined using the Akaike Information Criterion (AIC; Akaike, 1974) as implemented in IQ-TREE ModelFinder (Kalyaanamoorthy et al., 2017). The GTR+Γ+I with 4 Γ categories was selected for gene *5620* and *5594*, whereas HKY+ Γ with 4 Γ categories was selected for gene *6139*. An uncorrelated relaxed clock model assuming an uncorrelated lognormal rate prior and a birth-death speciation prior was selected. The ASTRAL tree was selected as user-specified starting tree and used as constraint by deleting the commands: subtreeSlide; narrowExchange; wideExchange; wilsonBalding in

BEAUti 1.10.4, because a topology inferred from 300+ genes is more reliable than one inferred from three.

Tracer 1.7.2 was used to check for convergence of the runs (Rambaut et al., 2018) after discarding the initial 10% as burn-in. The analysis was run twice for up to 500 million generations until the ESS of all parameters was >200, logging every 10.000 generations after which log and tree files were combined in LogCombiner 1.10.4 (Drummond et al., 2012). The median node ages with 95% highest posterior density (HPD) intervals were summarised and a maximum clade credibility tree was annotated with TreeAnnotator 1.10.4 (Drummond et al., 2012) and visualised using FigTree 1.4.4 (Rambaut, 2009).

**Ancestral range estimation**

We estimated ancestral ranges of the Cryptocarya group using 'BioGeoBEARS' 1.1.2 (Matzke, 2013). The analysis was performed on a pruned version of the MCC tree including the ingroup only. Ten biogeographic regions were delimited according to the phytogeographic regions developed by Carta et al. (2022): A, Andean-Argentinian; B, Neotropical; C, Southern African; D, African; E, Madagascan; F, Northern Australia; G, Malesian; H, Indian-Indochinese; I, Neozealandic-Patagonian and J, Eurasiatic (see Figure 1) and species were assigned to regions based on distribution data from the World Checklist of Vascular Plants (Govaerts, 2022). The maximum number of areas any ancestral species could occupy was set to three, given the maximum number of areas any extant species occupies.

We carried out a total of 12 analyses based on the three implemented Maximum Likelihood (ML) models of range evolution: Dispersal-Extinction-Cladogenesis (DEC; Ree & Smith, 2008), DIVALIKE, a likelihood version of DIVA (Ronquist, 1997) and BAYAREALIKE, a likelihood version of BayArea (Landis et al., 2013). These models were run with default settings and parameters, but each model was additionally run with the parameter *j* to include founder-event speciation and time-stratification (TS) to account for the relative probability of dispersal the past 80 Ma, based on palaeogeographical changes (Seton et al., 2012). When constructing a stratified biogeographical model, it is important not to divide it too finely to get enough phylogenetic events in each time slice (Ree & Sanmartín, 2009). Here we chose to divide the model into three time slices: 80-40 Ma, 40-20 Ma and 20 Ma to present, and assigned different dispersal rate scalers (s) between the biogeographical areas (see Appendix S4), following the rules set up in Toussaint et al. (2017). The best model was selected using AIC (Burnham & Anderson, 2002).

We estimated the number and type of biogeographical events using biogeographical stochastic mapping (BSM; Matzke, 2015) implemented in 'BioGeoBEARS'. The BSM was conducted on the best model (DEC+TS+*j*) that produced the best fit to the data (Table 3). 100 BSMs were carried out and event frequencies were estimated by taking the mean and standard deviation of event counts.

**Table 3** Results of BioGeoBEARS models. The best model according to AIC is highlighted.

| **Models** | **Dispersal multipliers** | **Free parameters** | | | | **LnL** | **AIC** | **ΔAIC** |
|---|---|---|---|---|---|---|---|---|
| **Basic Models** | | **Number** | ***d*** | ***e*** | ***j*** | | | |
| DEC | Non-TS | 2 | 0.0022 | 0.0016 | 0.0000 | -200.587 | 405.2 | 36.3 |
| DIVALIKE | Non-TS | 2 | 0.0028 | 0.0000 | 0.0000 | -212.622 | 429.2 | 60.3 |
| BayAreaLIKE | Non-TS | 2 | 0.0017 | 0.0265 | 0.0000 | -210.136 | 424.3 | 55.4 |
| **Time-Stratified Models** | | | | | | | | |
| DEC + TS | TS | 2 | 0.0061 | 0.0009 | 0.0000 | -199.752 | 403.5 | 34.6 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| DIVALIKE + TS | TS | 2 | 0.0077 | 0.0006 | 0.0000 | -197.687 | 399.4 | 30.5 |
| BayAreaLIKE + TS | TS | 2 | 0.0051 | 0.0257 | 0.0000 | -202.750 | 409.5 | 40.6 |
| ***+j* Models** | | | | | | | | |
| DEC + *j* | Non-TS | 3 | 0.0017 | 0.0000 | 0.0064 | -196.047 | 398.1 | 29.2 |
| DIVALIKE + *j* | Non-TS | 3 | 0.0020 | 0.0000 | 0.0080 | -204.737 | 415.5 | 46.6 |
| BayAreaLIKE + *j* | Non-TS | 3 | 0.0009 | 0.0149 | 0.0070 | -199.620 | 405.2 | 36.3 |
| DEC + TS + *j* | TS | 3 | 0.0046 | 0.0000 | 0.0178 | -181.423 | 368.9 | 0.00 |
| DIVALIKE + TS + *j* | TS | 3 | 0.0054 | 0.0000 | 0.0225 | -189.438 | 384.9 | 16 |
| BayAreaLIKE + TS + *j* | TS | 3 | 0.0030 | 0.0110 | 0.0246 | -202.750 | 391.4 | 22.5 |

AIC, Akaike information criterion; ΔAIC, difference in AIC; d, dispersal; e, extinction; j, weight of jump dispersal; TS, time stratified dispersal multipliers; LnL, log-likelihood of the model.

On https://github.com/LKFrederiksen/Cryptocarya, detailed descriptions of the analyses including scripts can be found. In addition, alignments (Appendix S5) and gene trees (Appendix S6) have been deposited here.

## RESULTS

### Target sequence capture and phylogenetic analyses

Targeted sequencing of the Angiosperms353 markers produced between 375,278 and 6,611,636 trimmed reads per Cryptocarya group sample (Appendix S7). Of these, a median of 25.6% were on target. For the outgroup between 284,142 and 45,226,714 trimmed reads per outgroup sample was produced and a median of 9.3% were on target (Appendix S2). Some potential genes with paralogs were detected, specifically for ten genes with up to five paralogs per sample (Appendix S1). We found that the species tree topology was identical with and without these ten paralogous genes (results not shown).

After alignment and all cleaning steps, for alignments of the combined ingroup and outgroup, 16 genes had been lost (4 %). The alignments included a median of 139 samples (96%), ranging between 10 and 145 samples per alignment. The median alignment length was 1199 bp, with lengths varying between 285 and 3757 bp. Overall, the alignments had 26% missing data, 70% of the alignment columns were variable and 50% were parsimony informative (see Appendix S8 for more alignment statistics).

### Phylogenetic relationships

The species tree generated with ASTRAL-III was overall well supported, with 83% of branches receiving high support (Local Posterior Probability, LPP ≥ 0.95; Sayyari & Mirarab, 2016) and a well-supported backbone, except for six nodes within the Beilschmiedia complex (Figure 1, Appendix S9).

The outgroup topology was in accordance with Baker et al. (2022) and was consistently separated from the monophyletic ingroup i.e., the Cryptocarya group with maximal support (LPP = 1). The following review of the phylogenetic relationships is based on the pruned species tree, including one sample per species (Figure 1).

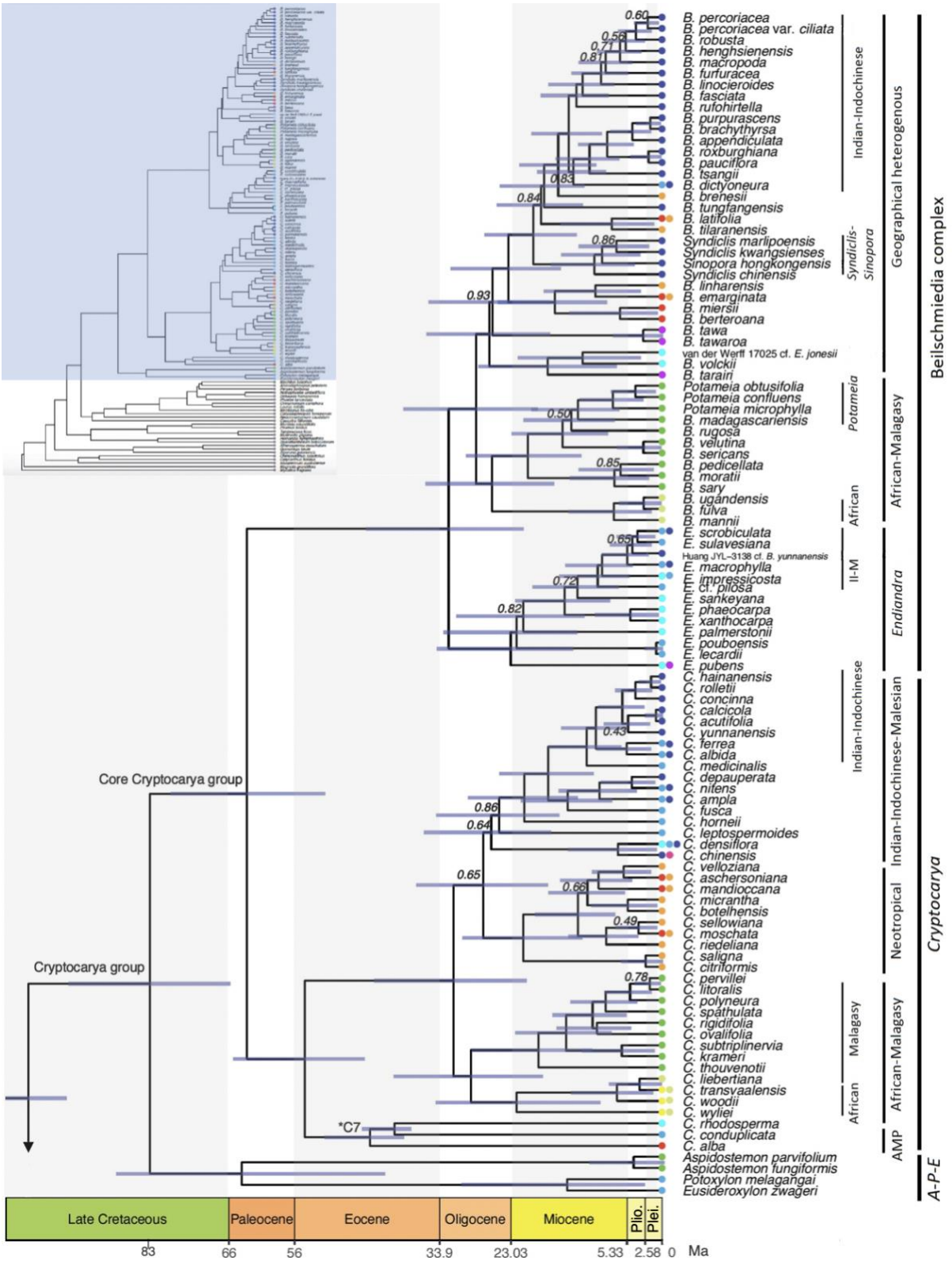
The Cryptocarya group comprises two main clades, a clade including the monophyletic genus *Aspidostemon* and the monotypic genera *Potoxylon* and *Eusideroxylon* (A-P-E). This clade forms the sister group to the remaining species, the core Cryptocarya group, consisting of *Cryptocarya* and the Beilschmeidia complex with maximum support.

*Cryptocarya* formed a monophyletic clade with maximum support. Within, *Cryptocarya*, an amphi-Pacific clade, consisting of *C. alba* from Chile as sister to *C. conduplicata* from New Caledonia and *C. rhodosperma* from Australia, forms the sister group to the remaining species. The latter then split into an African-Malagasy clade, and a rather poorly supported clade including species from other regions. Within the African-Malagasy clade, the four species from the African and South African region (*C. liebertiana*, *C. transvaalensis*, *C. woodii* and *C. wyliei*) form the sister group to the Malagasy species. Among the species from other regions, the Neotropical and Andean species form a maximally supported clade, sister to another weakly supported clade from the Indian-Indochinese-Malesian region, including mainly species from SE Asia, plus a few from New Guinea (*C. medicinalis*), Polynesia (*C.* cf. *fusca* and *C. hornei*) and New Caledonia (*C. leptospermoides*).

The Beilschmeidia complex formed a polyphyletic clade with maximum support. Within the Beilschmiedia complex, the genus *Endiandra* was also resolved as monophyletic with maximum support, apart from an only tentatively determined specimen 'Huang JYL-3138 cf. *B. yunnanensis*'. *Endiandra* includes species with extant distributions linked to the Northern Australian, Indian-Indochinese or Malesian region, which together form the sister group to the remaining species. The latter then split into two maximally supported clades, an African-Malagasy clade, and a clade including species from other regions. Within the African-Malagasy clade, the three species from the African region (*B. mannii*, *B. fulva* and *B. ugandensis*) form the sister group to the Malagasy species. Three of these Malagasy species (*P. micrphylla*, *P. confluens* and *P. obtusifolia*) with maximum support form a monophyletic *Potameia*, embedded within the Malagasy clade. Among the species from other regions (the geographical heterogenous clade), the species from respectively the Neozealandic and Northern Australian region formed two small clades with maximum support, of which the only tentatively determined specimen 'van der Werff 17025 cf. *E. jonesii*' was resolved as sister to *B. volckii* with maximum support. These clades were resolved as sister to a maximally supported clade including species, whose current distribution largely includes the Indian-Indochinese, Andean-Argentinian or Neotropical region, or a combination of these. Within this clade, the genera *Syndiclis* and *Sinopora* together formed a maximally supported clade, sister to a clade among other including three Neotropical species (*B. tilaranensis*, *B. latifolia* and *B. brenesii*) and a large subclade with species from the Indian-Indochinese region, both with maximum support. However, the quartet scores (proportion of gene trees supporting the main topology i.e., red part of pies; see Appendix S9) were generally low for the backbone of the geographical heterogenous clade. Thus, the intercladal relationships are uncertain.

Within the non-pruned species tree including all samples (Appendix S9, S10), some species where more than one sample was included, were not resolved as being most closely related to samples specified as the same species. Within *Cryptocarya* two *C. litoralis* samples were not resolved as sister samples, instead *C. litoralis* (733AL1; Ratovoson 1725) was with maximum support resolved next to C. *pervillei* (780AL1; Rakotonirina 427) and *C. litoralis* (734AL1; Ravelonarivo 2983) next to *C. polyneura*, however with weak support. In continuation the two *C. pervillei* samples were not resolved as sister samples, as evident from the above and that *C. pervillei* (AL781; Rasoazanany 230) was resolved as sister to *C. spathulata* (791AL1; Ravelonarivo 3849) with maximum support. Thus, samples of *C. spathulata* are moreover not resolved as sister samples and *C. spathulata* (790AL1; Rakotonirina 158) is resolved with *C. subtriplinervia* (793AL; van der Werff 12775) with full support. Within the Beilschmiedia complex two *B. roxburghiana* samples were with maximum support not resolved as sister samples, instead *B. roxburghiana* (716AL1; Hoffmann) formed a clade with *B. purpurascens*,

*B. brachythyrsa* and *B. appendiculata*, and *B. roxburghiana* (717AL1; Huang JYL-5215) with *B. pauciflora*.



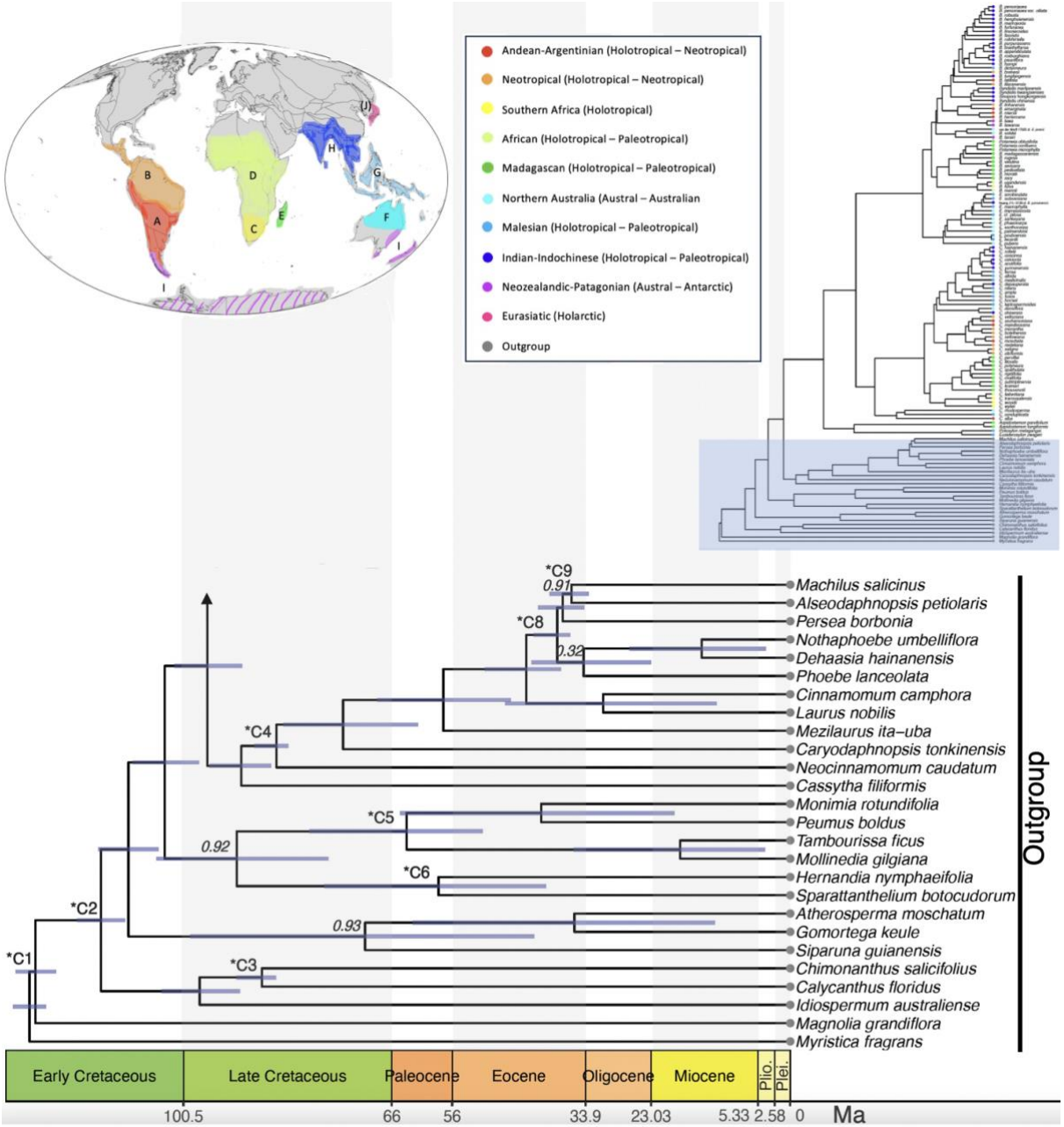



**Figure 1** Maximum Clade Credibility (MCC) tree of the Cryptocarya group and outgroup derived from BEAST with 95% highest probability density (HPD) bars. Branch labels are local posterior probabilities (LPP), not shown if LPP > 0.95. *C represent nodes used for calibrating the tree (see Table S2). Colours next to species names represent coding of regions (see legend with corresponding map). Regions: A = Andean-Argentinian, B = Neotropical, C = Southern Africa, D = African, E = Madagascan, F = Northern Australia, G = Malesian, H = Indian-Indochinese, I = Neozealandic-Patagonian, J = Eurasiatic.

### Divergence time estimates and ancestral range estimation

The DEC+TS+*j* model was based on AIC (Burnham & Anderson, 2002) strongly supported as the best fitting model over the next best model (DIVALIKE+TS+*j*; Table 3). In addition, the inclusion of time-stratification (TS) and founder-event speciation (*j*) consistently improved model fit.

According to our BioGeoBEARS analysis under the DEC+TS+*j* model, the per-area probabilities indicate ambiguous ancestral ranges of the deepest nodes, e.g., the most recent common ancestor (MRCA) of the Cryptocarya group, core Cryptocarya group, *Cryptocarya* and the Beilschmiedia complex (Appendix S11). Based on the best node estimate, the Cryptocarya group most likely originated in a vast range, comprising Madagascar and Malesia about 80 Ma (Late Cretaceous; node I in Figure 2, Table 4), but a range comprising Madagascar, Northern Australia and Malesia is also possible (Appendix S11). For the core Cryptocarya group an origin in Malesia about 65 Ma (Late Cretaceous, node III in Figure 2, Table 4) received strongest support, however a range encompassing Northern Australia, in the constellation of Northern Australia or the combined range Northern Australia and Malesia received equiprobable support (Appendix S11). Thus, depending on the inferred range, i.e., Malesia, Northern Australia or a range comprising both, of the core Cryptocarya group, either within-area speciation, a founder event or narrow vicariance, respectively, could have led to the origin of *Cryptocarya* in Malesia about 55 Ma (Paleocene-Eocene boundary, node IV in Figure 2, Table 4). For *Cryptocarya* an ancestral range comprising the Northern Australian and Malesian region, or in extension to that the Andean-Argentinian region received weak support (Appendix S11). The most probable origin of the Beilschmiedia complex was an origin in Northern Australia about 33 Ma (Eocene-Oligocene boundary, node V in Figure 2, Table 4). The two second most plausible ranges also comprised the Northern Australian region but showed some support for a range also including the Neozealandic-Patagonian region or an area additionally encompassing Madagascar (Appendix S9). For the shallower nodes, the estimated ancestral ranges mostly received stronger support with probabilities mostly > 75%.

Within *Cryptocarya*, given Malesia was the most probable range (present in all most probable range combinations; Appendix S11) within-area speciation led to an amphi-Pacific clade about 45 Ma (mid-Eocene; node XII in Figure 2; Table 4) and a clade including the remaining *Cryptocarya* species. Within the amphi-Pacific clade a founder-event from the Malesian to Neotropical region caused the diversification that has founded the extant species *C. alba*. Following a founder-event in the clade comprising the remaining *Cryptocarya* species, led to the origin of the African-Malagasy clade about 30 Ma (Early Oligocene; node VI in Figure 2; Table 4) after which extant Malagasy species started to diversify about 19 Ma (Early Miocene; node X) and extant species of the African clade started to diversify following a founder event about 22 Ma (Oligocene-Miocene boundary; node IX). On the other hand, a within-area speciation in Malesia resulted in extant species with a range comprising the Indian-Indochinese and Malesian region that started to diversify about 26 Ma (mid-Oligocene; node VII), after which a subsequent founder event led to the colonisation of the Indian-Indochinese region about 17 Ma (mid-Miocene; node XI). A founder event also led to the arrival of *Cryptocarya* in the Neotropical region, where extant species started to diversify about 21 Ma (Early Miocene; node VIII) and this clade showed several range expansions reaching the Andean-Argentinian region three times.

The Beilschmiedia complex originated in Northern Australia after a founder event about 33 Ma (Eocene-Oligocene boundary; node V in Figure 2; Table 4). We infer that within-area speciation led to the origin of *Endiandra* about 23 Ma (Oligocene/Miocene boundary; node XIII). Descendants of the inferred ancestor of this clade subsequently diverged and colonised Malesia through two independent founder events, one of which led to the origin of the Indian-Indochinese-Malesian clade about 10 Ma (mid-Miocene; node XVI). The geographical heterogenous clade most probably originated in the Neozealandic-Patagonian region about 26 Ma (mid-Oligocene; node XIX) following a founder event from Northern Australia and subsequent within-area speciation. Descendants of the geographical

heterogenous clade subsequently diverged and colonised the Northern Australia, Neotropical, Andean-Argentinian and Indian-Indochinese region due to a combination of founder events and within-area subset speciation (Figure 2), of which the *Syndiclis-Sinopora* clade originated about 10 Ma (mid-Miocene; node XX). Instead, within-area speciation followed by a narrow vicariance event, caused the Indian-Indochinese clade to split from the remaining species about 14 Ma (mid-Miocene; node XI; Table 4). The remaining species include the African and Malagasy species that are sister to the geographical heterogenous clade. The MRCA to this clade reached Madagascar following a founder-event from the Neozealandic-Patagonian region about 26 Ma (Late Oligocene, node XIV in Figure 2; Table 4). Subsequent, within-area speciation led to the origin of the Malagasy clade about 20 Ma (Early Miocene; node XVIII), whereas our results indicate that a founder event led to the arrival of *Beilschmiedia* in Africa, whereof extant African *Beilschmiedia* started to diversify about 7 Ma (Late Miocene; node XVII).

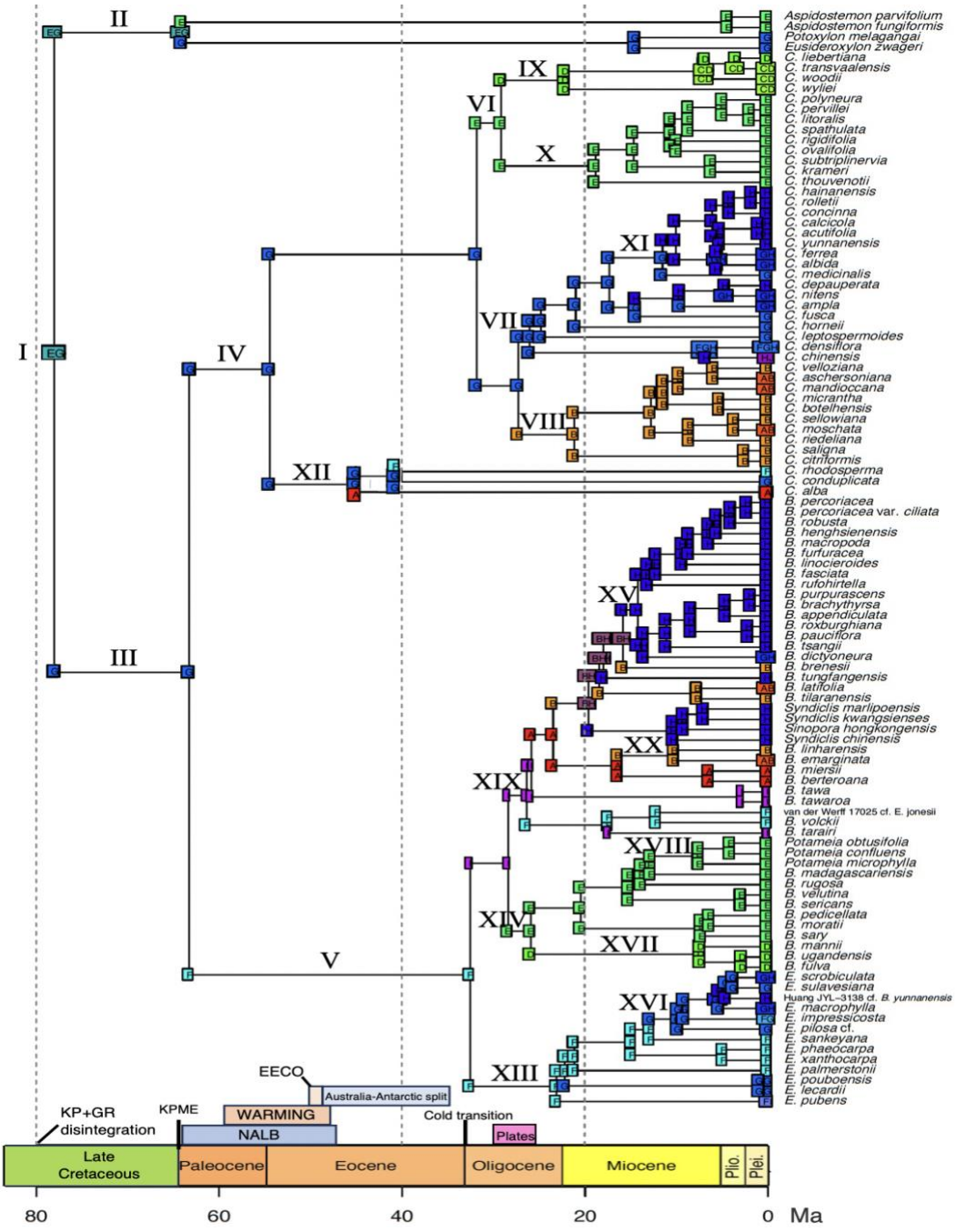


**Figure 2** Ancestral range estimation for the Cryptocarya group using BioGeoBEARS, with ancestral ranges, and dispersal rate probabilities between regions for each time slice (TS1, TS2, TS3; Appendix S4). Roman numerals on the nodes represent crown nodes of important colonisation events, discussed in the text and shown in Table 4. Coloured boxes show the ancestral ranges that received highest likelihood at each node. These ranges can comprise one to three combined regions, as indicated by single letters or a combination of these (e.g., FG = a range comprising Northern Australia and the Malesia region). Regions: A = Andean-Argentinian, B = Neotropical, C = Southern Africa, D = African, E = Madagascan, F = Northern Australia, G = Malesian, H = Indian-Indochinese, I = Neozealandic-Patagonian, J = Eurasiatic. Biogeographic events: KP + GR disintegration = Kerguelen Plateau and Gunnerus Ridge disintegration, KPME = Cretaceous–Palaeogene mass extinction, NALB = North, Atlantic Land Bridge, EECO = Cenozoic thermal maxima, Australian-Antarctic split = No longer a land connection between Australia and Antarctica, Plates = Australian and Pacific plate collides.

**Table 4** Crown group ages of major clades of the Cryptocarya group as recovered in the BEAST analyses. Median ages are displayed together with their 95% highest posterior density (HPD) intervals.

| **Clade name** | **Node** | **Crown group age (Ma)** | |
|---|---|---|---|
| | | **Median** | **(95% HPD)** |
| Cryptocarya group | I | 78.0 | (65.6-90.3) |
| *Aspidostemon-Eusideroxylon-Potoxylon* | II | 64.1 | (41.4-82.4) |
| Core Cryptocarya group | III | 63.3 | (51.7-75.2) |
| *Cryptocarya* | IV | 54.4 | (45.9-65.9) |
| Beilschmiedia complex | V | 32.5 | (21.8-45.8) |
| ***Cryptocarya* clades** | | | |
| African-Malagasy | VI | 29.1 | (18.6-41.3) |
| Indian-Indochinese-Malesian | VII | 26.0 | (17.4-37.9) |
| Neotropical | VIII | 21.2 | (12.6-30.8) |
| African | IX | 22.2 | (9.7-34.7) |
| Malagasy | X | 18.8 | (11.2-29.6) |
| Indian-Indochinese | XI | 17.3 | (6.5-17.7) |
| Amphi-Pacific | XII | 44.5 | (40.0-52.1) |
| **Beilschmiedia complex clades** | | | |
| *Endiandra* | XIII | 23.1 | (14.1-34.9) |
| African-Malagasy | XIV | 25.9 | (17.0-36.6) |
| Indian-Indochinese | XV | 14.2 | (9.4-20.2) |
| Indian-Indochinese-Malesian* | XVI | 9.8 | (5.3-16.0) |
| African | XVII | 7.3 | (2.3-15.0) |
| Malagasy | XVIII | 20.4 | (12.8-30.3) |
| Geographical heterogenous | XIX | 26.3 | (17.6-36.4) |
| *Syndicli*s-*Sinopora* | XX | 10.3 | (4.6-17.3) |

**Endiandra*

The BSM analysis revealed that no single type of biogeographical event dominated across all time slices (TS's; Table 5). Range expansions (31.27 %) and 'narrow vicariance' (30.13%) were dominant in TS1, whereas within-area speciation was most common in TS2 (35.64%) and TS3 (64.84 %). 'Narrow vicariance' refers to range expansion followed by speciation due to a subsequent period of restricted gene flow, resulting in subclade division. Overall, the role of founder events and 'narrow vicariance' became less important over time, while within-area speciation increased in importance. The high number of within-area speciation events was expected due to the extent of the defined biogeographical regions (see Carta et al., 2022).

Focusing on dispersal (founder events and range expansion) we found that the dispersal patterns within the Cryptocarya group varied between areas, genera and over time (Figure 3; Table 5; Appendix S12). All the formerly described major clades, except the Indian-Indochinese *Beilschmiedia* (node XV in Figure 2, 3) and the *Syndiclis-Sinopora* clade (node XX in Figure 2) originated after one to three colonisation events. According to the BSM analyses (Table 5) we had four dispersals in TS1, where the Malesian region was the primary dispersal source (55.5% of all dispersals) and North Australia and the Andean-Argentinian region the primary sinks (45.6% of all dispersals). Apart from the dispersal event between the Malesian and Andean-Argentinian region, representing the founder-event of *C. alba*, dispersals from Malesia to North Australia and vice versa were most frequent (uncertainty regarding standard deviation), could reflect the alternative possible origins and initial dispersal patterns within *Cryptocarya* (node IV in Figure 2, 3; Appendix S12). In TS2 there were nine dispersals, and all major clades originated within this period (Table 5).

In comparison to the beforementioned time slice, there were no clear sources and sinks in TS2, but the predominant dispersal paths also supported by the model (Figure 2) were from the Malagasy to African, the Northern Australian to Malesian, Neotropical to Indian-Indochinese, and Malesian to Neotropical region (Appendix S12). Here, *Cryptocarya* and *Beilschmiedia* colonised Madagascar and the Neotropical region (Figure 2, 3). *Cryptocarya* dispersed to the Neotropical region and Madagascar from Malesia, whereas *Beilschmiedia* dispersed to the Neotropical region and Madagascar via the Neozealandic-Patagonian region.

Within TS3, 22 dispersals took place with subsequent diversification as a result (Table 5). Here the Malesian region was the largest dispersal source (33.1%), particularly due to dispersals within *Endiandra* and the Indian-Indochinese-Malesian Cryptocarya clade (node XIII; VII in Figure 2; Appendix S12), whereas the Indian-Indochinese region was the largest dispersal sink (28.4%; Appendix S12). In Australasia (Indian-Indochinese, Malesian and Northern Australian region) several dispersals took place, especially between the Malesian and Indian-Indochinese region (~9 combined) and the Malesian and Northern Australian region (~2; Figure 2, 3; Appendix S12). Moreover, within the Neotropics (Andean-Argentinian and Neotropical region) with c. four dispersals from the Neotropical to the Andean-Argentinian region and with ~one dispersal from the Andean-Argentinian to the Neotropical region.

**Table 5** Biogeographic stochastic mapping results for the Cryptocarya group using the DEC+TS+*j* across the three time slices (TS's): 80-40 Ma; 40-20 Ma and 20-0 Ma. Mean values, standard deviations (SDs) and the percent of each speciation process are shown. No range contractions were estimated because the relevant model parameter (e) was not relevant in the best fitting mode.

| Speciation process | Type | ~80-40 Ma | | | 40 – 20 Ma | | | 20 – 0 Ma | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | SD | Percent (%) | Mean | SD | Percent (%) | Mean | SD | Percent (%) |
| Dispersal | Founder events | 1.43 | 0.99 | 16.38 | 5.43 | 1.42 | 24.84 | 2.45 | 1.4 | 2.44 |
| | Range expansions | 2.73 | 1.24 | 31.27 | 3.86 | 1.42 | 17.66 | 19.21 | 1.3 | 19.23 |
| Vicariance | Narrow vicariance | 2.63 | 1.13 | 30.13 | 2.79 | 1.33 | 12.76 | 2.07 | 0.88 | 2.07 |
| Within-area | Speciation | 0.53 | 0.66 | 6.07 | 7.79 | 1.06 | 35.64 | 64.98 | 2.15 | 64.84 |
| | Speciation - subset | 1.41 | 0.87 | 16.15 | 1.99 | 1.28 | 9.1 | 11.5 | 2.79 | 11.48 |
| **Total** | | 8.73 | 1.24 | 100 | 21.86 | 1.42 | 100 | 100.2 | 1.3 | 100 |

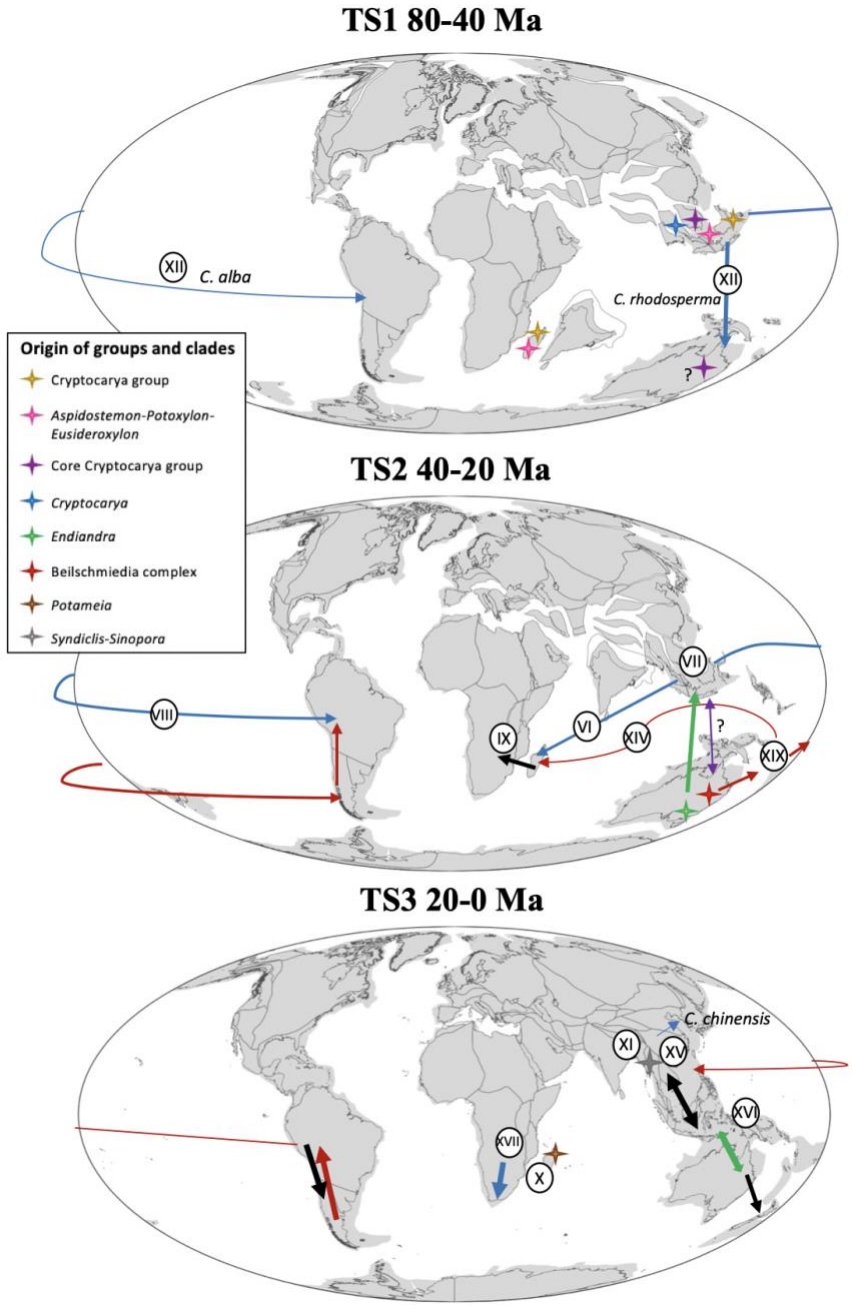


**Figure 3** Summary of major dispersal events within the Cryptocarya group for the three different time slices (TS1, TS2 and TS3). The frequency of dispersals was estimated with biogeographical stochastic mapping (Appendix S12). Coloured stars represent origination events of major groups or genera (see legend; Table 4; Figure 3). Arrows between areas represent direction and the estimated frequency of dispersal events, coloured according to the group dispersing (see legend for colouration). Black arrows indicate dispersal of more than one group or genus in the same direction. Circles with roman numerals refer to the arrival of major groups (see Table 4, Figure 3).

## DISCUSSION

### Are nuclear and plastid evolutionary histories congruent in the Cryptocarya group?

Our phylogenomic analyses provide an unprecedented solid evolutionary framework of the Cryptocarya group inferred from 353 low-copy nuclear genes and increased sampling, together with an assessment of congruence to previously estimated trees based on plastid DNA. Previous Lauraceae studies have with few exceptions (Chanderbali et al., 2001; Rohwer et al., 2014; van der Merwe et al., 2016) relied on plastid markers (Li et al., 2020; Li et al., 2011; Rohwer & Rudolph, 2005; Song et al., 2019). An overview of phylogenetic relationships within Lauraceae including the Cryptocarya group was first produced by Rohwer (2000). We recovered a topology that was mostly congruent with previously conducted plastid (Chanderbali et al., 2001; Rohwer, 2000; Rohwer et al., 2014) and plastome studies (Li et al., 2020; Song et al., 2019) and that altogether showed similar generic and regional relationships (Figure 1).

Our topology identified a monophyletic *Eusideroxylon*, *Potoxylon*, *Aspidostemon*, *Potameia*, *Endiandra* and *Cryptocarya*, but polyphyletic *Beilschmiedia* and *Syndiclis* (Figure 1) in accordance with relationships previously identified by Li et al. (2020), Rohwer et al. (2014) and Song et al. (2019). Our analysis supports the previous perception that *Endiandra* is monophyletic, also with our increased sampling of continental Asia, as requested by Rohwer et al. (2014), despite appearing polyphyletic in the tree. We found that the samples 'Henk van der Werff 17025 cf. *E. jonesii*' and 'Huang JYL-3138 cf. *B. yunnanensis*' were likely misidentified (H. van der Werff; L. Jie, personal communication, January 2023). This conclusion is also supported by the placement of 'Huang JYL-3138 cf. *B. yunnanensis*' as sister to *B. tungfangensis* in Li et al. (2020). Our study was additionally congruent with the recent realisation of Li et al. (2020) that *Sinopora hongkongensis* is most closely related to *Syndiclis* (placed in the same clade), and not *Potameia* as previously thought (Rohwer et al., 2014). However, it should be noted that our study only includes Malagasy species. Therefore, we cannot be certain that species from *Potameia*'s range in Thailand are not more closely related to the *Sinopora-Syndiclis* clade. Additionally, we validated the placement of *Syndiclis* as most closely related to the Indian-Indochinese *Beilschmiedia* clade and the three neotropical species *B. brenesii*, *B. latifolia* and *B. tilaranensis*.

Nevertheless, a few supported relationships were incongruent in comparison to plastid-based studies, e.g., the position of *Endiandra* relative to *Beilschmiedia*. Based on the plastid marker trnk, *Endiandra* was nested within *Beilschmiedia* (Rohwer et al., 2014). With plastome data *Endiandra* formed a well-supported subclade nested within a larger clade also including *Beilschmiedia* (Li et al., 2020). In comparison *Endiandra* was not nested within *Beilschmiedia* in our study (Figure 1) or when using the nuclear ribosomal marker, ITS (Rohwer et al., 2014). Another example includes the position of *B. tungfangensis* that in Li et al. (2020) in contrast to our study was more closely related to *B. fasciata* than *B. brenesii*. A possible explanation for

this incongruence includes ILS and hybridisation (Gonçalves et al., 2019; Pelser et al., 2010; Pérez-Escobar et al., 2021). The quartet support for *B. tungfangensis* indicates that an equal proportion of gene quartets support each of the three possible topologies (Appendix S9). This suggests that nuclear gene tree discordance could be responsible for an equally plausible arrangement, where *B. tungfangensis* is placed within the Indian-Indochinese *Beilschmiedia* clade (Figure 1). In the case of *Endiandra*, ILS does not seem to be an issue (Appendix S9) but instead reflects divergent evolutionary histories between plastid and nuclear regions in this group.

**When, where, and how? Divergence-times and biogeography of the Cryptocarya group**

*Divergence time estimates*

Our results show that the Cryptocarya group originated in the Late Cretaceous, with a crown age of 78 Ma. Our estimate is significantly younger compared to Chanderbali et al. (2001), who found a crown age of ~120 Ma. We further show that the divergence of the Beilschmiedia complex from *Cryptocarya* took place in the Mid-Paleocene, with a crown age of 63.3 Ma, and not the Early-Eocene as expected given Li et al. (2020). Nevertheless, we can confirm that the crown age of the Beilschmiedia complex was at the Oligocene-Miocene boundary, 34 Ma, and we for the first time show that extant *Cryptocarya* started to diversify around the Paleocene-Eocene boundary, 54.4 Ma. Additionally, our analyses showed that most of the generic diversity of the extant Cryptocarya group emerged around the Oligocene-Miocene boundary (Figure 2). The difference in age estimates shown may in part result from our usage of more calibration points which should improve age estimates, when placed with care (Sauquet et al., 2012). Regarding the estimate of the diversification of core *Cryptocarya*, this can be explained by our more extensive sampling of *Cryptocarya* (Linder et al., 2005). For example, our study pushed back the crown age of the core Cryptocarya group relative to Li et al. (2020), likely because we included samples of the amphi-Pacific clade that started to diversify prior to the remaining *Cryptocarya* clades. Finally, the maximum age constraint set for the root is known to have a major impact on the generated divergence time estimates (Sauquet, 2013), which may be part of the explanation for the difference in the age estimates.

*Gondwanan origin?*

Our biogeographical history reconstruction of the pantropical Cryptocarya group revealed some uncertainty regarding the groups ancestral range. However, an eastern Southern Hemisphere origin seems likely given that an area comprising Madagascar and Malesia received strongest support in the DEC+TS+*j* analysis (Figure 2). This agrees with Rohwer et al. (2014) who assumed an Old-World origin of the Cryptocarya group given the Paleotropical distribution of early diversifying lineages. Cryptocarya group species are characteristic elements of TRF vegetation (Gentry, 1988), thus this result indicates that TRF was widespread outside of the Neotropics and Tropical Africa before Late Paleocene (Morley, 2007). Nevertheless, a better estimate of the group's origin would require more extensive taxon sampling and a carefully selected outgroup. The age estimate of Chanderbali et al. (2001) calibrated based on an arbitrarily selected node, however led them to hypothesise that the Cryptocarya group had a Gondwanan origin, which could be explained by Gondwanan vicariance (Raven & Axelrod, 1974). According to our result this scenario seems very unlikely, because the breakup of Gondwana was completed by the Late Cretaceous (Upchurch, 2008). Indeed, we found that the biogeographical evolution of the group, was mainly driven by dispersal, (i.e., range expansion and founder-event speciation; see Figure 2; Table 5). Thus, it is more plausible that the transoceanic disjunctions are attributed to LDD events (e.g., transoceanic dispersal), as has increasingly been proposed for many pantropically disjunct

clades (e.g., Baker et al., 2013; Bartish et al., 2011; Clayton et al., 2009; Couvreur, Pirie, et al., 2011; Renner et al., 2001; Richardson et al., 2014; Schneider et al., 2022; Zerega et al., 2005). These dispersals can be achieved by transport of diaspores by different dispersal vectors in water (Harries & Clement, 2014), by wind (Muñoz et al., 2004), animals (Onstein et al., 2019) or by rafting of seeds or plants on floating land-masses (Conti et al., 2002).

Limited information is available regarding LDD vectors for Lauraceae. However, it is known that lauraceous fruits are generally large with a fleshy mesocarp, containing relatively large seeds. These fruits are described as either one-seeded berries or drupes with weakly developed endocarp (Rohwer, 1993). These traits are associated with specialised frugivores, such as birds, which suggests that strong-flying birds could play a role in LDD of Cryptocarya group species (Carpenter et al., 2010; Onstein et al., 2019; Snow, 1981; Wheelwright, 1986). For instance, a study by Huang et al. (2016) found that fruit pigeons (Columbidae) may have assisted dispersal of *Cinnamomum* Schaeff. seeds from Asia into Australia through seasonal migration. Additionally, bellbirds (Cotingidae) and hornbills (Bucerotidae) have been proposed as potential dispersal vectors (Onstein et al., 2019; Wheelwright, 1986). Additionally, Carpenter et al. (2010) argue that giant tortoises, e.g., in the Indian Ocean, could be possible LDD dispersers of fruit.

LDD via water is another option that cannot be ruled out as a possible dispersal vector of Lauraceae. However, aside from ocean currents, a propagule that can remain viable for a longer period of time in salt water (Huang et al., 2016) and can stay afloat (Carpenter et al., 2010) is required. At least some propagules of Lauraceae species have been reported to meet these criteria. For example, *C. latifolia* Sond., which is an element of the drift component found on southern African beaches (Gunn & Dennis, 1976 as cited in Carpenter et al., 2010). Also, Carpenter et al. (2010) implies that many propagules of Lauraceae could have entered the ocean from wet forests on the south-eastern coast of Australia, during the Palaeocene, as several cuticle fossils, among other things with affinity to the Cryptocarya group, have been reported (Vadala & Greenwood, 2001). Nevertheless, more likely is rafting of seeds or plants, as the same requirements for the propagule are not present. Gillespie et al. (2012) for example approximated that rafting across the Pacific Ocean, from New Zealand to South America via West Wind Drift could be achieved within a year. On the contrary, we infer that wind-dispersal is an unlikely dispersal vector of Cryptocarya group species due to the beforementioned traits of the fruit, that do not coincide with typical traits of wind-dispersed taxa such as Asteraceae Bercht. & J.Presl (Muñoz et al., 2004).

*Cretaceous–Palaeogene mass extinction and the Core Cryptocarya group*

The origin of the core Cryptocarya group and the MRCA of *Aspidostemon*, *Potoxylon* and *Eusideroxylon* coincides with the Cretaceous–Palaeogene mass extinction (KPME), in 66 Ma (Figure 2). Following the KPME a period of slow recovery is presumed to have shaped several of today's characteristic plant lineages (Benton et al., 2022; Schneider et al., 2022). Thus, we wonder whether the KPME triggered Cryptocarya group diversification.

Based on our analysis of ancestral ranges, we suggest that the diversification of the core Cryptocarya group could have begun in the Malesian region (Figure 2, 3). However the origin of the core Cryptocarya group in Malesia does not seem to find support within the Malesian macrofossil record, which has been reported to be largely unexplored (Kooyman et al., 2019). Thus, a Malesian origin of the core Cryptocarya group cannot be confirmed, however also not rejected out on this basis. Based on the palynological record from the Kayan Formation in Western Sarawak, the climate in the Late Cretaceceous to Paleocene was reported to include

pollen of extant taxa, such as *Nypa* Steck (Arecaceae), that inhabit the Malesian rainforest (Kooyman et al., 2019). Thus, a fossil with affinity to the core Cryptocarya group from Malesia may simply not have been discovered yet. Nevertheless, we cannot ignore the possibility that the core Cryptocarya group originated in a range at least encompassing the Northern Australian region, considering the relative strong support in our analyses (Appendix S11). Lauraceous leaves known from the Australian fossil flora from the Early Paleogene, supports this possibility (van der Merwe, 2016).

During the next 30 My, succeeding the KPME, a period of shifting climate and rapid cooling towards the Late Eocene followed (Zachos et al., 2001), with several southern hemisphere megathermal rainforest contractions e.g., in South America. If the Cryptocarya group did originate in Northern Australia, continued diversification within the Northern Australian region, during this time, could have given rise to the Beilschmiedia complex, enabled by the northward drift of the Australian Plate towards the Malesian region (Morley, 2007). Alternatively, in the Malesian origin scenario, the origin of the Beilschmiedia complex, could have been a result of a founder event from Malesia to Northern Australia. This event probably occurred, before the Late Eocene through LDD, when temperatures were still warm at those latitudes, or perhaps by steppingstone dispersal, when Australia was located further north.

Based on the ambiguous ancestral range estimate of the core Cryptocarya group, we conclude that it would be premature to assess, which biogeographic history is the more likely, given the probability of both scenarios. To establish a more definitive estimate of the ancestral range of the core Cryptocarya group, broader and denser taxon sampling is required.

*Floristic exchange between the Sunda and Sahul region*

According to our analysis, given the abovementioned scenarios, *Cryptocarya* may either be the result of dispersal from the Australian region to the Malesian region, as hypothesised by van der Merwe et al. (2016; Figure 3). Else the result of within-area speciation (Figure 2). Floristic exchange between these two regions, also known as the Sahul and Sunda regions is well-known (Holzmeyer et al., 2023), which is supported by our BSM that shows steady levels of dispersal between these regions through time (Appendix S12). According to previous studies south-eastward dispersal between these regions seem to have predominated (e.g., Huang et al., 2016; Richardson et al., 2012 ) and Australian members of *Cryptocarya* also seem to have arisen via an eastward migration pattern in the mid-Miocene, like *Cinnamomum* (Huang et al., 2016). However, we saw dispersal in both directions, where the predominant dispersal direction depended on the time (Figure 2; Appendix S12). As example Malesian members of the otherwise Northern Australian based *Endiandra* probably arose via this north-westward migration pattern in the mid-Miocene according to our model, in connection to the Australian and Pacific plates collision that brought these areas into close proximity (Zerega et al., 2005), possibly facilitating stepping stone dispersal.

*Transoceanic dispersal into the Neotropics*

At the Early Eocene Climatic Optimum (EECO; 52 to 50 Ma), megathermal forests reached their widest distribution, almost reaching the polar regions, in north and south (Morley, 2011), and a general period of global warming persisted during the mid-Paleocene to early Eocene (59-48 Ma; Pross et al., 2012; Zachos et al., 2001). At this point land connections were also available, between Australia and Antarctica until ~50 to 35 Ma, between South America and Antarctica until ~35 Ma and the North American land bridge (NALB), was available between Europe and North America from the early Paleocene to Eocene (~65 to 47 Ma; Morley, 2003; Tiffney, 1985). The Beringian land bridge (BLB) also connecting Europe and North America,

however, is considered unsuitable for tropical species, due to its very northern position (Tiffney & Manchester, 2001). Nevertheless, the warming considerably reduced the distance to be overcome.

The first Cryptocarya group species to colonise the neotropics was *C. alba* within the geographically disjunct amphi-Pacific clade (Figure 2; 3). Rohwer et al. (2014) inferred that LDD across the Pacific was the most likely explanation, but not necessarily for other Neotropical species. Based on just the divergence time estimate, geodispersal across Antarctica was feasible. Nonetheless, consulting the biogeographical model, *C. alba* was inferred to split from the remaining amphi-Pacific clade, due to a founder event from Malesia (Figure 2). We thus confirm that LDD across the Pacific, was the most probable migratory route for *C. alba* into the neotropics, as this was the most direct route (Figure 3).

Likewise, the divergence of the remaining Old and New World lineages (*Cryptocarya*, *Beilschmiedia* and *Syndiclis*; Figure 2) does not seem to be associated with either the NALB (Neotropical *Cryptocarya* and *Syndiclis*) or the Antarctica land connection, as the land connections no longer persisted, and the climate was unfavourable (Tiffney, 1985). Thus, the transoceanic disjunctions of *Cryptocarya* and *Beilschmiedia* must be results of eastward transoceanic LDD, as found for Ochnaceae (Schneider et al., 2022). Interestingly we found that *Beilschmiedia* only colonised the Neotropics once. Hence, the low geographical structure among these clades was according to our model (Figure 2), due to several subsequent range contractions, among other causative of *Syndiclis* origin in the Indian-Indochinese region. Thus, we cautiously reject the hypothesis of Li et al. (2020) and Rohwer et al. (2014) that *Beilschmiedia* colonised the neotropics at least twice, as we remain to see if this holds when additional species are added.

*Transoceanic dispersal into Madagascar*

In this study we were for the first time able to increase sampling of Malagasy *Beilschmiedia* species and increase the continental African sampling. Hitherto, the phylogenetic relationships and thus also the biogeography of this region was known from just two samples, *B. pierreana* Robyns & R.Wilczek distributed in Cameroon and Gabon, included in Li et al. (2020) and Rohwer et al. (2014). In addition 'Breteler 5283 cf. *Beilschmiedia*', included in Rohwer et al. (2014). The MRCA of *Beilschmiedia* and *Potameia* in Madagascar was inferred to have arrived from the Neozealandic-Patagonian region (Figure 2). Intuitively, one could think that the migratory route involved a land bridge via Antarctica and the Gunnerus Ridge (GR) or Kerguelen Plateau (KP) but this could not have been feasible, as the GR and KP had disintegrated by the mid-Late Cretaceous (Yoder & Nowak, 2006) and given the Oligocene climate (Zachos et al., 2001).

Previous studies (Federman et al., 2015; Schneider et al., 2022; Yoder & Nowak, 2006) suggest that LDD across the Indian Ocean was highly plausible during the Eocene to Oligocene, as easterly winds and westward marine surface currents prevailed across the Indian Ocean. Therefore, we infer that *Cryptocarya* reached Madagascar from Malesia in the mid-Oligocene by LDD (Figure 2, 3). In both *Cryptocarya* and *Beilschmiedia* (+ *Potameia*) the MRCA of the African and Malagasy clade split shortly after arrival to Madagascar (5 Ma later). Additionally, both *Beilschmiedia* and *Cryptocarya* started diversifying in the mid-Miocene, during a period where also *Parkia* diversified in Madagascar, which according to Conceição Oliveira et al. (2021) indicates that conditions for rainforest taxa were optimal in Madagascar at this time. Furthermore, *Beilschmiedia* arrived in Africa around the Miocene-Pliocene border (5 Ma), whereas Malagasy *Cryptocarya* arrived much earlier in the early Miocene (20 Ma).

According to Yoder and Nowak (2006), 'Africa appears by far to be the most important source of floral dispersal to Madagascar', and in e.g., Annonaceae (Couvreur, Pirie, et al., 2011), Sapindaceae (Buerki et al., 2011) and Ochnaceae (Schneider et al., 2022), dispersal from Africa has indeed contributed to Malagasy floral diversity. In our study Africa was likewise, according to the BSM (Appendix S12) an important source and sink, though our default model (Figure 2) indicates that dispersal leading to speciation only took place in westward direction, from Madagascar to Africa. In addition, Malesia played some role in the origin of the Malagasy flora, as suggested by Schatz (1996) in Schneider et al. (2022; Appendix S10).

**Geographical structure in major clades of the Cryptocarya group**

In this study, we were able to confirm that the geographical structure within well-supported clades of the Cryptocarya group, except within the amphi-Pacific clade, is overall high, as first pointed out by Rohwer et al. (2014). Strong geographical structure can be attributed to within area speciation and dispersal limitation according to theory (Pennington et al., 2006). In the context of TRF, this implies that interbreeding and the resulting gene flow between major blocks of TRF has ceased (Bechteler et al., 2017). Indeed, we found this to be the common denominator in almost all our major clades (see results, Figure 2), as shown by the outnumber of within-area speciation events within the last 20 Ma (Table 5).

Nevertheless, considering the group as an entity, low geographical structure at the group level is a prerequisite for wide distribution ranges and intercontinental pantropical disjunction. This could occur, for example, through LDD events (range expansion or founder event) or through vicariance that may subsequently lead to strong geographic structure due to isolation (Bechteler et al., 2017). However, a new event would again break down the geographical structure. Based on our BioGeoBEARS model, we were able to determine the evolutionary processes behind the disintegration of geographical structure more confidently. As example, we suggest that two succeeding founder events (LDD) from Malesia, led to low geographical structure within the amphi-Pacific clade (Figure 2). Additionally, that several range expansions (LDD) followed by within-area subset speciation decreased structure in the geographical heterogenous Beilschmiedia complex clade, and that LDD to Madagascar led to transoceanic disjunction in the Indian-Indochinese-Malesian *Cryptocarya* clade. Overall LDD, whether through range expansion or founder-event speciation, constituted the main process driving the origin of major clades with strong geographical structure (Figure 2, Table 5), aligning with previous studies (Bechteler et al., 2017; Conceição Oliveira et al., 2021; Lavin et al., 2004) and suggesting that LDD events are not that rare.

## CONCLUSION

This study has contributed to a more advanced understanding of the pantropical Cryptocarya group, by providing the first nuclear phylogenomic framework for the group. One aspect of our research was the evaluation of congruence between nuclear and plastid evolutionary histories of the Cryptocarya group. We found that our nuclear phylogeny was largely congruent with previously conducted plastid phylogenies but not interchangeable, as our study highlights the utility of nuclear data for assessing the influence of ILS and hybridisation. Furthermore, our study achieved the broadest (13%) sampling to date within the Cryptocarya group, covering previous key sampling gaps in Africa and Madagascar. This allowed us to explore the biogeographical history of most parts of the Cryptocarya group, but we were for example not able to include Asian *Potameia* species. We found a likely origination of the Cryptocarya group somewhere within a range comprising Madagascar, and Malesia during the late Cretaceous. Furthermore, we showed that the group had achieved its pantropical distribution by the early

Miocene, facilitated by several transoceanic long-distance dispersal events that shaped clades with strong geographical structure.

However, to further improve our understanding of this taxonomic group, future studies could focus on increasing sampling efforts because insufficient sampling can lead to the oversight of important biogeographical events or phylogenetic relationships. For example, we received ambiguous support for the ancestral range of the core Cryptocarya group. In addition, we encourage the development of a Lauraceae specific nuclear probe set, like e.g., the PhyloPalm probe kit (Loiseau et al., 2019) used in palms, as this could greatly improve the resolution of phylogenetic relationships and help address the issue of species tree discordance.

**ACKNOWLEDGEMENT**

All the computing for this project was performed on the GenomeDK cluster. We would like to thank GenomeDK and Aarhus University for providing computational resources and support that contributed to these research results.

**REFERENCES**


Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716-723. https://doi.org/10.1109/TAC.1974.1100705

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. In https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Baker, W. J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigue, L. R., Brewer, G., Carruthers, T., Clarkson, J. J., Cook, J., Cowan, R. S., Dodsworth, S., Epitawalage, N., Francoso, E., Gallego, B., Johnson, M. G., Kim, J. T., Leempoel, K., . . . Forest, F. (2022). A Comprehensive Phylogenomic Platform for Exploring the Angiosperm Tree of Life. *Systematic Biology*, *71*(2), 301-319. https://doi.org/10.1093/sysbio/syab035

Baker, W. J., Couvreur, T. L. P., & Kissling, W. D. (2013). 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. https://doi.org/10.1111/j.1365-2699.2012.02794.x

Balslev, H. V., Renato; Paz-y-Mino-C, Guillermo; Christensen, H., Nielsen, I. (1998). Species Count of Vascular Plants in 1-hectare of humid lowland forest in Amazonian Ecuador. In F. C. Dallmeier, J. (Ed.), *Forest Biodiversity in North, Central and South America, and the Caribbean: Research and Monitoring* (Vol. 21, pp. 585-594). CRC Press.

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., Sirotkin, 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. https://doi.org/10.1089/cmb.2012.0021

Bardon, L., Sothers, C., Prance, G. T., Male, P. J., Xi, Z., Davis, C. C., Murienne, J., Garcia-Villacorta, R., Coissac, E., Lavergne, S., & Chave, J. (2016). Unraveling the biogeographical history of Chrysobalanaceae from plastid genomes. *Am J Bot*, *103*(6), 1089-1102. https://doi.org/10.3732/ajb.1500463

Barrett, R. A., Bayly, M. J., Duretto, M. F., Forster, P. I., Ladiges, P. Y., & Cantrill, D. J. (2014). A chloroplast phylogeny of *Zieria* (Rutaceae) in Australia and New Caledonia shows widespread incongruence with species-level taxonomy. *Australian Systematic Botany*, *27*(6). https://doi.org/10.1071/sb14033

Bartish, I. V., Antonelli, A., Richardson, J. E., & Swenson, U. (2011). Vicariance or long-distance dispersal: historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). *Journal of Biogeography*, *38*(1), 177-190. https://doi.org/10.1111/j.1365-2699.2010.02389.x

Bechteler, J., Schafer-Verwimp, A., Lee, G. E., Feldberg, K., Perez-Escobar, O. A., Pocs, T., Peralta, D. F., Renner, M. A., & Heinrichs, J. (2017). Geographical structure, narrow species ranges, and Cenozoic diversification in a pantropical clade of epiphyllous leafy liverworts. *Ecology and Evolution*, *7*(2), 638-653. https://doi.org/10.1002/ece3.2656

Benton, M. J., Wilf, P., & Sauquet, H. (2022). The Angiosperm Terrestrial Revolution and the origins of modern biodiversity. *New Phytologist*, *233*(5), 2017-2035. https://doi.org/10.1111/nph.17822

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

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

Buerki, S., Forest, F., Alvarez, N., Nylander, J. A. A., Arrigo, N., & Sanmartín, I. (2011). An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. *Journal of Biogeography*, *38*(3), 531-550. https://doi.org/10.1111/j.1365-2699.2010.02432.x

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.). Springer New York, NY. https://doi.org/https://doi.org/10.1007/b97636

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972-1973. https://doi.org/10.1093/bioinformatics/btp348

Carpenter, E. J., Matasci, N., Ayyampalayam, S., Wu, S., Sun, J., Yu, J., Jimenez Vieira, F. R., Bowler, C., Dorrell, R. G., Gitzendanner, M. A., Li, L., Du, W., K, K. U., Wickett, N. J., Barkmann, T. J., Barker, M. S., Leebens-Mack, J. H., & Wong, G. K. (2019). Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). *Gigascience*, *8*(10). https://doi.org/10.1093/gigascience/giz126

Carpenter, R. J., Truswell, E. M., & Harris, W. K. (2010). Lauraceae fossils from a volcanic Palaeocene oceanic island, Ninetyeast Ridge, Indian Ocean: ancient long-distance dispersal? *Journal of Biogeography*, *37*(7), 1202-1213. https://doi.org/10.1111/j.1365-2699.2010.02279.x

Carruthers, T., Sun, M., Baker, W. J., Smith, S. A., de Vos, J. M., & Eiserhardt, W. L. (2022). The Implications of Incongruence between Gene Tree and Species Tree Topologies for Divergence Time Estimation. *Systematic Biology*, *71*(5), 1124-1146. https://doi.org/10.1093/sysbio/syac012

Carta, A., Peruzzi, L., & Ramirez-Barahona, S. (2022). A global phylogenetic regionalization of vascular plants reveals a deep split between Gondwanan and Laurasian biotas. *New Phytologist*, *233*(3), 1494-1504. https://doi.org/10.1111/nph.17844

Chanderbali, A. S., van der Werff, H., & Renner, S. S. (2001). Phylogeny and Historical Biogeography of Lauraceae: Evidence from the Chloroplast and Nuclear Genomes. *Annals of the Missouri Botanical Garden*, *88*(1). https://doi.org/10.2307/2666133

Chernomor, O., von Haeseler, A., & Minh, B. Q. (2016). Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Systematic Biology*, *65*(6), 997-1008. https://doi.org/10.1093/sysbio/syw037

Clayton, J. W., Soltis, P. S., & Soltis, D. E. (2009). Recent long-distance dispersal overshadows ancient biogeographical patterns in a pantropical angiosperm family (Simaroubaceae, Sapindales). *Systematic Biology*, *58*(4), 395-410. https://doi.org/10.1093/sysbio/syp041

Conceição Oliveira, L., Picanço Rodrigues, D., Fortune Hopkins, H. C., Peter Lewis, G., & John Gilbert Hopkins, M. (2021). Phylogeny and historical biogeography of the pantropical genus *Parkia* (Leguminosae, Caesalpinioideae, mimosoid clade). *Molecular phylogenetics and evolution*, *163*, 107219. https://doi.org/10.1016/j.ympev.2021.107219

Conti, E., Eriksson, T., Schönenberger, J., Sytsma, K. J., & Baum, D. A. (2002). Early Tertiary out-of-India Dispersal of Crypteroniaceae: Evidence from Phylogeny and Molecular Dating. *Evolution*, *56*(10), 1931-1942. https://doi.org/10.1111/j.0014-3820.2002.tb00119.x

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

Couvreur, T. L. P., Forest, F., & Baker, W. J. (2011). Origin and global diversification patterns of tropical rain forests: inferences from a complete genus-level phylogeny of palms. *BMC Biology*, *9*, 44. https://doi.org/10.1186/1741-7007-9-44

Couvreur, T. L. P., Pirie, M. D., Chatrou, L. W., Saunders, R. M. K., Su, Y. C. F., Richardson, J. E., & Erkens, R. H. J. (2011). Early evolutionary history of the flowering plant family Annonaceae: steady diversification and boreotropical geodispersal. *Journal of Biogeography*, *38*(4), 664-680. https://doi.org/10.1111/j.1365-2699.2010.02434.x

Crepet, W. L., Nixon, K. C., & Gandolfo, M. A. (2005). An extinct calycanthoid taxon, *Jerseyanthus calycanthoides*, from the Late Cretaceous of New Jersey. *American Journal of Botany*, *92*(9), 1475-1485. https://doi.org/10.3732/ajb.92.9.1475

Davis, C. C., Xi, Z., & Mathews, S. (2014). Plastid phylogenomics and green plant phylogeny: almost full circle but not quite there. *BMC Biology*, *12*, 11. https://doi.org/10.1186/1741-7007-12-11

Donoghue, M. J. (2008). Colloquium paper: a phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *105 Suppl 1*(Suppl 1), 11549-11555. https://doi.org/10.1073/pnas.0801962105

Doyle, J. J. (1992). Gene Trees and Species Trees: Molecular Systematics as One-Character Taxonomy. *Systematic Botany*, *17*(1). https://doi.org/10.2307/2419070

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969-1973. https://doi.org/10.1093/molbev/mss075

Eiserhardt, W. L., Antonelli, A., Bennett, D. J., Botigue, L. R., Burleigh, J. G., Dodsworth, S., Enquist, B. J., Forest, F., Kim, J. T., Kozlov, A. M., Leitch, I. J., Maitner, B. S., Mirarab, S., Piel, W. H., Perez-Escobar, O. A., Pokorny, L., Rahbek, C., Sandel, B., Smith, S. A., . . . Baker, W. J. (2018). A roadmap for global synthesis of the plant tree of life. *American Journal of Botany*, *105*(3), 614-622. https://doi.org/10.1002/ajb2.1041

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

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. https://doi.org/10.1111/nph.14516

Eklund, H. (2000). Lauraceous flowers from the Late Cretaceous of North Carolina, U.S.A. *Botanical Journal of the Linnean Society*, *132*(4), 397-428. https://doi.org/10.1111/j.1095-8339.2000.tb01220.x

Ewels, P., Magnusson, M., Lundin, S., & Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047-3048. https://doi.org/10.1093/bioinformatics/btw354

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*, 216. https://doi.org/10.1186/s12862-015-0483-1

Friis, E. M., Eklund, H., Pedersen, K. R., & Crane, P. R. (1994). *Virginianthus calycanthoides* gen. et sp. nov.-A Calycanthaceous Flower from the Potomac Group (Early Cretaceous) of Eastern North America. *International Journal of Plant Sciences*, *155*(6), 772-785. https://doi.org/10.1086/297217

Gardner, E. M., Johnson, M. G., Pereira, J. T., Puad, A. S. A., Arifiani, D., Wickett, N. J., & Zerega, N. J. C. (2020). Paralogs and off-target sequences improve phylogenetic resolution in a densely-sampled study of the breadfruit genus (Artocarpus, Moraceae). *Systematic Biology*, *70*(3), 558-575. https://doi.org/10.1093/sysbio/syaa073

Gentry, A. H. (1988). Changes in Plant Community Diversity and Floristic Composition on Environmental and Geographical Gradients. *Annals of the Missouri Botanical Garden*, *75*(1). https://doi.org/10.2307/2399464

Gillespie, R. G., Baldwin, B. G., Waters, J. M., Fraser, C. I., Nikula, R., & Roderick, G. K. (2012). Long-distance dispersal: a framework for hypothesis testing. *Trends in Ecology & Evolution*, *27*(1), 47-56. https://doi.org/10.1016/j.tree.2011.08.009

Givnish, T. J., & Renner, S. S. (2004). Tropical Intercontinental Disjunctions: Gondwana Breakup, Immigration from the Boreotropics, and Transoceanic Dispersal. *International Journal of Plant Sciences*, *165*(S4), S1-S6. https://doi.org/10.1086/424022

Gonçalves, D. J. P., Simpson, B. B., Ortiz, E. M., Shimizu, G. H., & Jansen, R. K. (2019). Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. *Molecular phylogenetics and evolution*, *138*, 219-232. https://doi.org/10.1016/j.ympev.2019.05.022

Govaerts, R. (2022). *The World Checklist of Vascular Plants (WCVP)* GBIF.org. https://doi.org/https://doi.org/10.15468/6h8ucr

Govaerts, R., Nic Lughadha, E., Black, N., Turner, R., & Paton, A. (2021). The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Scientific Data*, *8*(1), 215. https://doi.org/10.1038/s41597-021-00997-6

Hale, H., Gardner, E. M., Viruel, J., Pokorny, L., & Johnson, M. G. (2020). Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. *Applications in Plant Sciences*, *8*(4), e11337. https://doi.org/10.1002/aps3.11337

Harries, H. C., & Clement, C. R. (2014). Long-distance dispersal of the coconut palm by migration within the coral atoll ecosystem. *Annals of Botany*, *113*(4), 565-570. https://doi.org/10.1093/aob/mct293

Hill, J. L., & Hill, R. A. (2016). Why are tropical rain forests so species rich? Classifying, reviewing and evaluating theories. *Progress in Physical Geography: Earth and Environment*, *25*(3), 326-354. https://doi.org/10.1177/030913330102500302

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518-522. https://doi.org/10.1093/molbev/msx281

Holzmeyer, L., Hauenschild, F., & Muellner-Riehl, A. N. (2023). Sunda–Sahul floristic exchange and pathways into the Southwest Pacific: New insights from wet tropical forest trees. *Journal of Biogeography*. https://doi.org/10.1111/jbi.14606

Huang, J. F., Li, L., van der Werff, H., Li, H. W., Rohwer, J. G., Crayn, D. M., Meng, H. H., van der Merwe, M., Conran, J. G., & Li, J. (2016). Origins and evolution of cinnamon and camphor: A phylogenetic and historical biogeographical analysis of the Cinnamomum group (Lauraceae). *Molecular phylogenetics and evolution*, *96*, 33-44. https://doi.org/10.1016/j.ympev.2015.12.007

Institute, B. (2020). *Picard Tools*. In (Version 2.22.3 ) http://broadinstitute.github.io/picard/

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigue, 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., Baker, W. J., & Wickett, N. J. (2019). A Universal Probe Set for Targeted Sequencing of 353 Nuclear

Genes from Any Flowering Plant Designed Using k-Medoids Clustering. *Systematic Biology*, *68*(4), 594-606. https://doi.org/10.1093/sysbio/syy086

Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*, *26*(13), 1669-1670. https://doi.org/10.1093/bioinformatics/btq243

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587-589. https://doi.org/10.1038/nmeth.4285

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

Knight, C. L., & Wilf, P. D. (2013). Rare leaf fossils of Monimiaceae and Atherospermataceae (Laurales) from Eocene Patagonian rainforests and their biogeographic significance. *Palaeontologia Electronica*. https://doi.org/10.26879/386

Kooyman, R. M., Morley, R. J., Crayn, D. M., Joyce, E. M., Rossetto, M., Slik, J. W. F., Strijk, J. S., Su, T., Yap, J.-Y. S., & Wilf, P. (2019). Origins and Assembly of Malesian Rainforests. *Annual Review of Ecology, Evolution, and Systematics*, *50*(1), 119-143. https://doi.org/10.1146/annurev-ecolsys-110218-024737

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. https://doi.org/10.1093/sysbio/syt040

Lavin, M., Schrire, B. P., Lewis, G., Pennington, R. T., Delgado-Salinas, A., Thulin, M., Hughes, C. E., Matos, A. B., & Wojciechowski, M. F. (2004). Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *359*(1450), 1509-1522. https://doi.org/10.1098/rstb.2004.1536

Lee, G. E., Condamine, F. L., Bechteler, J., Perez-Escobar, O. A., Scheben, A., Schafer-Verwimp, A., Pocs, T., & Heinrichs, J. (2020). An ancient tropical origin, dispersals via land bridges and Miocene diversification explain the subcosmopolitan disjunctions of the liverwort genus *Lejeunea*. *Scientific Reports*, *10*(1), 14123. https://doi.org/10.1038/s41598-020-71039-1

Leisman, G. A. (1986). *Cryptocaryoxylon gippslandicum* gen. et sp. nov., from the Tertiary of eastern Victoria. *Alcheringa: An Australasian Journal of Palaeontology*, *10*(3), 225-234. https://doi.org/https://doi.org/10.1080/03115518608619157

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754-1760. 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., & Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Li, H., Liu, B., Davis, C. C., & Yang, Y. (2020). Plastome phylogenomics, systematics, and divergence time estimation of the Beilschmiedia group (Lauraceae). *Molecular phylogenetics and evolution*, *151*, 106901. https://doi.org/10.1016/j.ympev.2020.106901

Li, J., Qiu, J., Liao, W., & Jin, J. (2009). Eocene fossil *Alseodaphne* from Hainan Island of China and its paleoclimatic implications. *Science in China Series D: Earth Sciences*, *52*(10), 1537-1542. https://doi.org/10.1007/s11430-009-0120-1

Li, L., Li, J., Rohwer, J. G., van der Werff, H., Wang, Z. H., & Li, H. W. (2011). Molecular phylogenetic analysis of the Persea group (Lauraceae) and its biogeographic implications on the evolution of tropical and subtropical Amphi-Pacific disjunctions. *American Journal of Botany*, *98*(9), 1520-1536. https://doi.org/10.3732/ajb.1100006

Loiseau, O., Olivares, I., Paris, M., de La Harpe, M., Weigand, A., Koubinova, D., Rolland, J., Bacon, C. D., Balslev, H., Borchsenius, F., Cano, A., Couvreur, T. L. P., Delnatte, C., Fardin, F., Gayot, M., Mejia, F., Mota-Machado, T., Perret, M., Roncal, J., . . . Salamin, N. (2019). Targeted Capture of Hundreds of Nuclear Genes Unravels Phylogenetic Relationships of the Diverse Neotropical Palm Tribe Geonomateae. *Frontiers in Plant Science*, *10*, 864. https://doi.org/10.3389/fpls.2019.00864

Lomolino, M. V., Riddle, B. R., Whittaker, R. J., & Brown, J. H. (2010). *Biogeography* (4 ed.). Sinauer Associates.

Manchester, S. R., & O'Leary, E. L. (2010). Phylogenetic Distribution and Identification of Fin-winged Fruits. *The Botanical Review*, *76*(1), 1-82. https://doi.org/10.1007/s12229-010-9041-0

Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. *Nature*, *470*(7333), 198-203. https://doi.org/10.1038/nature09796

Massoni, J., Couvreur, T. L., & Sauquet, H. (2015). Five major shifts of diversification through the long evolutionary history of Magnoliidae (angiosperms). *BMC Evolutionary Biology*, *15*, 49. https://doi.org/10.1186/s12862-015-0320-6

Matzke, N. J. (2013). *Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model testing* University of California]. https://escholarship.org/content/qt44j7n141/qt44j7n141_noSplash_a25fcfcd2e4c86c599d6f7da5ee1d7f8.pdf?t=pga0we

Matzke, N. J. (2015). *Stochastic mapping under biogeographical models*. Retrieved 10 April from http://phylo.wikidot.com/biogeobears#stochastic_mapping

McLay, T. G. B., Birch, J. L., Gunn, B. F., Ning, W., Tate, J. A., Nauheimer, L., Joyce, E. M., Simpson, L., Schmidt-Lebuhn, A. N., Baker, W. J., Forest, F., & Jackson, C. J. (2021). New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences*, *9*(7). https://doi.org/10.1002/aps3.11420

Meng, H. H., Jacques, F. M., Su, T., Huang, Y. J., Zhang, S. T., Ma, H. J., & Zhou, Z. K. (2014). New Biogeographic insight into *Bauhinia* s.l. (Leguminosae): integration from fossil records and molecular analyses. *BMC Evolutionary Biology*, *14*, 181. https://doi.org/10.1186/s12862-014-0181-4

Michalak, I., Zhang, L.-B., & Renner, S. S. (2010). Trans-Atlantic, trans-Pacific and trans-Indian Ocean dispersal in the small Gondwanan Laurales family Hernandiaceae. *Journal of Biogeography*, *37*(7), 1214-1226. https://doi.org/10.1111/j.1365-2699.2010.02306.x

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

Morley, R. J. (2003). Interplate dispersal paths for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics*, *6*(1-2), 5-20. https://doi.org/10.1078/1433-8319-00039

Morley, R. J. (2007). Cretaceous and Tertiary climate change and the past distribution of megathermal rainforests. In *Tropical Rainforest Responses to Climatic Change* (pp. 1-31). Springer. https://doi.org/10.1007/978-3-642-05383-2

Morley, R. J. (2011). Cretaceous and Tertiary climate change and the past distribution of megathermal rainforests. In M. Bush, J. Flenley, & W. Gosling (Eds.), *Tropical Rainforest Responses to Climatic Change* (2 ed., pp. 1-34). Springer Berlin. https://doi.org/https://doi.org/10.1007/978-3-642-05383-2

Muñoz, J., Felicísimo, Á. M., Cabezas, F., Burgaz, A. R., & Martínez, I. (2004). Wind as a Long-Distance Dispersal Vehicle in the Southern Hemisphere. *Science*, *304*(5674), 1144-1147. http://www.jstor.org/stable/3836873

Nie, Z. L., Deng, T., Meng, Y., Sun, H., & Wen, J. (2013). Post-Boreotropical dispersals explain the pantropical disjunction in *Paederia* (Rubiaceae). *Annals of Botany*, *111*(5), 873-886. https://doi.org/10.1093/aob/mct053

Onstein, R. E., Kissling, W. D., Chatrou, L. W., Couvreur, T. L. P., Morlon, H., & Sauquet, H. (2019). Which frugivory-related traits facilitated historical long-distance dispersal in the custard apple family (Annonaceae)? *Journal of Biogeography*, *46*(8), 1874-1888. https://doi.org/10.1111/jbi.13552

Pelser, P. B., Kennedy, A. H., Tepe, E. J., Shidler, J. B., Nordenstam, B., Kadereit, J. W., & Watson, L. E. (2010). Patterns and causes of incongruence between plastid and nuclear Senecioneae (Asteraceae) phylogenies. *American Journal of Botany*, *97*(5), 856-873. https://doi.org/10.3732/ajb.0900287

Pennington, R. T., Richardson, J. E., & Lavin, M. (2006). Insights into the historical construction of species-rich biomes from dated plant phylogenies, neutral ecological theory and phylogenetic community structure. *New Phytologist*, *172*(4), 605-616. https://doi.org/10.1111/j.1469-8137.2006.01902.x

Pérez-Escobar, O. A., Dodsworth, S., Bogarin, D., Bellot, S., Balbuena, J. A., Schley, R. J., Kikuchi, I. A., Morris, S. K., Epitawalage, N., Cowan, R., Maurin, O., Zuntini, A., Arias, T., Serna-Sanchez, A., Gravendeel, B., Torres Jimenez, M. F., Nargar, K., Chomicki, G., Chase, M. W., . . . Baker, W. J. (2021). Hundreds of nuclear and plastid loci yield novel insights into orchid relationships. *American Journal of Botany*, *108*(7), 1166-1180. https://doi.org/10.1002/ajb2.1702

POWO. (2023). *Plants of the World Online* http://www.plantsoftheworldonline.org/

Pross, J., Contreras, L., Bijl, P. K., Greenwood, D. R., Bohaty, S. M., Schouten, S., Bendle, J. A., Rohl, U., Tauxe, L., Raine, J. I., Huck, C. E., van de Flierdt, T., Jamieson, S. S., Stickley, C. E., van de Schootbrugge, B., Escutia, C., Brinkhuis, H., & Integrated Ocean Drilling Program Expedition, S. (2012). Persistent near-tropical warmth on the Antarctic continent during the early Eocene epoch. *Nature*, *488*(7409), 73-77. https://doi.org/10.1038/nature11300

Rambaut, A. (2009). *FigTree*. In (Version v1.4.4) 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. https://doi.org/10.1093/sysbio/syy032

Ramirez-Barahona, S., Sauquet, H., & Magallon, S. (2020). The delayed and geographically heterogeneous diversification of flowering plant families. *Nature Ecology and Evolution*, *4*(9), 1232-1238. https://doi.org/10.1038/s41559-020-1241-3

Raven, P. H., & Axelrod, D. I. (1974). Angiosperm Biogeography and Past Continental Movements. *Annals of the Missouri Botanical Garden*, *61*(3). https://doi.org/10.2307/2395021

Ree, R. H., & Sanmartín, I. (2009). Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography*, *36*(7), 1211-1220. https://doi.org/10.1111/j.1365-2699.2008.02068.x

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. https://doi.org/10.1080/10635150701883881

Renner, S. S., Clausing, G., & Meyer, K. (2001). Historical biogeography of Melastomataceae: the roles of Tertiary migration and long-distance dispersal. *American Journal of Botany*, *88*(7), 1290-1300. https://doi.org/10.2307/3558340

Richardson, J. E., Bakar, A. M., Tosh, J., Armstrong, K., Smedmark, J., Anderberg, A. A., Slik, F., & Wilkie, P. (2014). The influence of tectonics, sea-level changes and dispersal on migration and diversification of Isonandreae (Sapotaceae). *Botanical Journal of the Linnean Society*, *174*(1), 130-140. https://doi.org/10.1111/boj.12108

Richardson, J. E., Costion, C. M., & Muellner, A. N. (2012). The Malesian floristic interchange:: plant migration patterns across Wallace's Line. In B. Rosen, D. Gower, J. Richardson, K. Johnson, L. Rüber, & S. Williams (Eds.), *Biotic Evolution and Environmental Change in Southeast Asia* (pp. 138-163). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511735882.008

Richardson, J. E., Whitlock, B. A., Meerow, A. W., & Madriñán, S. (2015). The age of chocolate: a diversification history of *Theobroma* and Malvaceae. *Frontiers in Ecology and Evolution*, *3*. https://doi.org/10.3389/fevo.2015.00120

Rohwer, J. G. (1993). Lauraceae. In K. Kubitzki, J. G. Rohwer, & V. Bittrich (Eds.), *Flowering Plants · Dicotyledons: Magnoliid, Hamamelid and Caryophyllid Families* (pp. 366-391). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-02899-5_46

Rohwer, J. G. (2000). Toward a Phylogenetic Classification of the Lauraceae: Evidence from matK Sequences. *Systematic Botany*, *25*(1). https://doi.org/10.2307/2666673

Rohwer, J. G., De Moraes, P. L. R., Rudolph, B., & Werff, H. V. D. (2014). A phylogenetic analysis of the Cryptocarya group (Lauraceae), and relationships of *Dahlgrenodendron*, *Sinopora*, *Triadodaphne*, and *Yasunia*. *Phytotaxa*, *158*(2). https://doi.org/10.11646/phytotaxa.158.2.1

Rohwer, J. G., & Rudolph, B. (2005). Jumping Genera: The Phylogenetic Positions of *Cassytha*, *Hypodaphnis*, and *Neocinnamomum* (Lauraceae) Based on Different Analyses of trnK Intron Sequences. *Annals of the Missouri Botanical Garden*, *92*(2), 153-178. http://www.jstor.org/stable/3298512

Ronquist, F. (1997). Dispersal-Vicariance Analysis: A New Approach to the Quantification of Historical Biogeography. *Systematic Biology*, *46*(1). https://doi.org/10.2307/2413643

Sang, T. (2002). Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology*, *37*(3), 121-147. https://doi.org/10.1080/10409230290771474

Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus Palevol*, *12*(6), 355-367. https://doi.org/10.1016/j.crpv.2013.07.003

Sauquet, H., Ho, S. Y., Gandolfo, M. A., Jordan, G. J., Wilf, P., Cantrill, D. J., Bayly, M. J., Bromham, L., Brown, G. K., Carpenter, R. J., Lee, D. M., Murphy, D. J., Sniderman, J. M., & Udovicic, F. (2012). Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). *Systematic Biology*, *61*(2), 289-313. https://doi.org/10.1093/sysbio/syr116

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., . . . Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *50*(D1), D20-D26. https://doi.org/10.1093/nar/gkab1112

Sayyari, E., & Mirarab, S. (2016). Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Molecular Biology and Evolution*, *33*(7), 1654-1668. https://doi.org/10.1093/molbev/msw079

Schneider, J. V., Jungcurt, T., Cardoso, D., Amorim, A. M., Paule, J., & Zizka, G. (2022). Predominantly Eastward Long-Distance Dispersal in Pantropical Ochnaceae Inferred From Ancestral Range Estimation and Phylogenomics. *Frontiers in Ecology and Evolution*, *10*. https://doi.org/10.3389/fevo.2022.813336

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

Shee, Z. Q., Frodin, D. G., Camara-Leret, R., & Pokorny, L. (2020). Reconstructing the Complex Evolutionary History of the Papuasian *Schefflera* Radiation Through Herbariomics. *Frontiers in Plant Science*, *11*, 258. https://doi.org/10.3389/fpls.2020.00258

Simon, M. F., Grether, R., de Queiroz, L. P., Sarkinen, T. E., Dutra, V. F., & Hughes, C. E. (2011). The evolutionary history of *Mimosa* (Leguminosae): toward a phylogeny of the sensitive plants. *American Journal of Botany*, *98*(7), 1201-1221. https://doi.org/10.3732/ajb.1000520

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. https://doi.org/10.1371/journal.pone.0197433

Snow, D. W. (1981). Tropical Frugivorous Birds and Their Food Plants: A World Survey. *Biotropica*, *13*(1). https://doi.org/10.2307/2387865

Song, Y., Yu, W. B., Tan, Y. H., Jin, J. J., Wang, B., Yang, J. B., Liu, B., & Corlett, R. T. (2019). Plastid phylogenomics improve phylogenetic resolution in the Lauraceae. *Journal of Systematics and Evolution*, *58*(4), 423-439. https://doi.org/10.1111/jse.12536

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1), vey016. https://doi.org/10.1093/ve/vey016

Tang, B., Han, M., Xu, Q., & Jin, J. (2016). Leaf Cuticle Microstructure of *Machilus maomingensi* ssp. nov. (Lauraceae) from the Eocene of the Maoming Basin, South China. *Acta Geologica Sinica - English Edition*, *90*(5), 1561-1571. https://doi.org/10.1111/1755-6724.12802

Thiers, B. (continuously updated). *Index Herbariorum: A Global Directory of Public Herbaria and Associated Staff*. New York Botanical Garden's Virtual Herbarium. http://sweetgum.nybg.org/science/ih/

Tiffney, B. H. (1985). The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the northern hemisphere. *Journal of the Arnold Arboretum*, *66*(2), 243-273. http://www.jstor.org/stable/43782164

Tiffney, B. H., & Manchester, S. R. (2001). The Use of Geological and Paleontological Evidence in Evaluating Plant Phylogeographic Hypotheses in the Northern Hemisphere Tertiary. *International Journal of Plant Sciences*, *162*(S6), S3-S17. https://doi.org/10.1086/323880

Toussaint, E. F. A., Bloom, D., & Short, A. E. Z. (2017). Cretaceous West Gondwana vicariance shaped giant water scavenger beetle biogeography. *Journal of Biogeography*, *44*(9), 1952-1965. https://doi.org/10.1111/jbi.12977

Trofimov, D., Rudolph, B., & Rohwer, J. G. (2016). Phylogenetic study of the genus *Nectandra* (Lauraceae), and reinstatement of *Damburneya*. *Taxon*, *65*(5), 980-996. https://doi.org/10.12705/655.3

Tumescheit, C., Firth, A. E., & Brown, K. (2022). CIAlign: A highly customisable command line tool to clean, interpret and visualise multiple sequence alignments. *PeerJ*, *10*, e12983. https://doi.org/10.7717/peerj.12983

Upchurch, P. (2008). Gondwanan break-up: legacies of a lost world? *Trends in Ecology & Evolution*, *23*(4), 229-236. https://doi.org/10.1016/j.tree.2007.11.006

Vadala, A., & Greenwood, D. (2001). Australian Paleogene vegetation and environments: evidence for palaeo-Gondwanic elements in the fossil records of Lauraceae and Proteaceae. In I. Melcalfe, J. M. B. Smith, & I. Davidson (Eds.), *Faunal and floral migrations and evolution in SE Asia- Australasia* (pp. 196-221). Swets & Zeitlinger Publishers.

van der Merwe, M., Crayn, D. M., Ford, A. J., Weston, P. H., & Rossetto, M. (2016). Evolution of Australian *Cryptocarya* (Lauraceae) based on nuclear and plastid phylogenetic trees: evidence of recent landscape-level disjunctions. *Australian Systematic Botany*, *29*(2). https://doi.org/10.1071/sb16023

van der Merwe, M. C., Darren M.; Ford, Andrew J.; Weston, Peter H.; Rossetto, Maurizio. (2016). Evolution of Australian Cryptocarya (Lauraceae) based on nuclear and plastid phylogenetic trees: evidence of recent landscape-level disjunctions. *Australian Systematic Botany*, *29*(2), 9. https://doi.org/https://doi.org/10.1071/SB16023

Wang, W., Ortiz, R. D. C., Jacques, F. M. B., Xiang, X. G., Li, H. L., Lin, L., Li, R. Q., Liu, Y., Soltis, P. S., Soltis, D. E., & Chen, Z. D. (2012). Menispermaceae and the diversification of tropical rainforests near the Cretaceous-Paleogene boundary. *New Phytologist*, *195*(2), 470-478. https://doi.org/10.1111/j.1469-8137.2012.04158.x

Wheelwright, N. T. (1986). A seven-year study of individual variation in fruit production in tropical bird-dispersed tree species in the family Lauraceae. In A. Estrada & T. H. Fleming (Eds.), *Frugivores and seed dispersal* (pp. 19-35). Springer Netherlands. https://doi.org/10.1007/978-94-009-4812-9_3

Yan, Z., Smith, M. L., Du, P., Hahn, M. W., & Nakhleh, L. (2022). Species Tree Inference Methods Intended to Deal with Incomplete Lineage Sorting Are Robust to the Presence of Paralogs. *Systematic Biology*, *71*(2), 367-381. https://doi.org/10.1093/sysbio/syab056

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*, 405-431. http://www.jstor.org/stable/30033838

Young, A. D., & Gillung, J. P. (2019). Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, *45*(2), 225-247. https://doi.org/10.1111/syen.12406

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Systematic Biology*, *292*(5517), 686-693. https://doi.org/10.1126/science.1059412

Zeng, L., Zhang, Q., Sun, R., Kong, H., Zhang, N., & Ma, H. (2014). Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature Communications*, *5*, 4956. https://doi.org/10.1038/ncomms5956

Zerega, N. J., Clement, W. L., Datwyler, S. L., & Weiblen, G. D. (2005). Biogeography and divergence times in the mulberry family (Moraceae). *Molecular phylogenetics and evolution*, *37*(2), 402-416. https://doi.org/10.1016/j.ympev.2005.07.004

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(Suppl 6), 153. https://doi.org/10.1186/s12859-018-2129-y

Zhang, C., Sayyari, E., & Mirarab, S. (2017, 2017//). ASTRAL-III: Increased Scalability and Impacts of Contracting Low Support Branches. Comparative Genomics, Cham.

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

## SUPPLEMENTARY INFORMATION

Supplementary information is available from:
https://github.com/LKFrederiksen/Cryptocarya/tree/ca7605d64c03b81a8da3582cc45ee6278cad123d/Supplementary_information

### Appendix S1: Paralog analyses from HybPiper

HybPiper generated output showing potential paralogs for each gene and number of gene duplications for each sample. 0 indicates no genes recovered for the specific sample, 1 indicates gene recovered for the sample, >1 indicated that the gene potentially is paralogous.

### Appendix S2: Gene tree concordance to 'PAFTOL' outgroup topology

Nuclear Angiosperms353 (Johnson et al., 2019) genes used to generate the PAFTOL (Baker et al., 2022) outgroup topology. We used SortaDate (Smith et al., 2018) to identify the supercontig genes that were most concordant to the PAFTOL outgroup topology based on exons (Baker et al., 2022). We then tested how many exons we were able to substitute with supercontigs without losing the desired topology used in Ramirez-Barahona (2020). This was 10%. The genes in the supercontig table are the 10% most concordant genes, whereas the genes in the exon table are the remaining 90 %.

### Appendix S3: Fossil selection

Justification for the eight fossils included as calibration points in our molecular dating with BEAST based on Ramirez-Barahona et al. (2020) and Li et al. (2020).

### Appendix S4: Dispersal rate scalers for BioGeoBEARS

Adapted from Toussaint et al. (2017). Dispersal rate scaling matrices used in the BioGeoBEARS (Matzke, 2013) analyses from the oldest time slice to the newest (TS1 to TS3). A matrix of dispersal rate scalers relative to potential geographic barriers is given for each of the three time slices: 80 to 40 Ma, 40 to 20 Ma and 20 Ma to present. Four symbols are used to indicate the kind of barrier between the selected ranges. Regions are abbreviated as follows: A = Andean-Argentinian, B = Neotropical, C = Southern Africa, D = African, E = Madagascan, F = Northern Australia, G = Malesian, H = Indian-Indochinese, I = Neozealandic-Patagonian and J = Eurasiatic (see Carta et al., 2022).

### Appendix S5: Alignments

342 fasta files generated with MAFFT and trimmed with Trimal, Optrimal, CIalign and Taper. Each file represents one gene retrieved from the 'mega353' target file (McLay et al., 2021). The faster headers correspond to each individual in short notation. See attached text file) for a conversion of the short notation to the species name.

### Appendix S6: Gene trees

342 ML (maximum likelihood) gene trees generated with IQtree in Newick format. Each gene tree corresponds to one gene. The faster headers correspond to each individual in short notation. See attached text file for a conversion of the short notation to the species name.

### Appendix S7: HybPiper recovery stats

HybPiper target enrichment and gene recovery efficiency summary for the 115 ingroup samples (*Syndiclis chinensis* included after the 'coverage' step and 26 outgroup samples (*Peumus boldus* and *Chimonanthus salicifolia* included after the 'coverage' step. The columns in the table correspond to: Sample name , Number of reads, Number of reads on target,

Percentage of 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. See HybPiper wiki for an explanation of what the different outputs mean. See attached text file for a conversion of the short notation to the species name.

**Appendix S8: AMAS alignment statistics** AMAS Borowiec (2016)summary statistics for all the alignments, following all cleaning steps (ready to make gene trees). Median, Mean, Max, Min and standard error (SD) has been calculated for all parameters. See here for a description of the gathereed statistics.

**Appendix S9: Non-pruned species tree with quartet scores**

ASTRAL (Zhang et al., 2018) species tree including ingroup and outgroup for all samples (also species with multiple samples) with quartet scores (QS). The quartet scores are included as pies on each node. Red = main topology, turquoise = alternative topology, grey = second alternative topology.

**Appendix S10: Non-pruned species tree with local posterior probability scores**

ASTRAL (Zhang et al., 2018) species tree including ingroup and outgroup for all samples (also species with multiple samples) with local posterior probability scores (LPP) showing the support for the main branch topology. The LPP scores can span from 0 to 1, where LPP = 1 indicates maximum support for the specific branch.

**Appendix S11: Biogeograchic ancestral range estimation using BioGeoBEARS including range probability scores**

Ancestral range estimation for the Cryptocarya group using BioGeoBEARS based on the DEC+TS+*j* model, with ancestral ranges, and dispersal rate scalers between regions for each time slice (TS1, TS2, TS3). Roman numerals on the nodes represent crown nodes of important colonisation events, discussed in the text and shown in Table 4. Pies on the nodes and branches show the probality of each range, based on 1. These ranges can comprise of one to three combined regions, as indicated by single letters or a combination of these (e.g., FG = Northern Australia and Malesia). Regions are abbreviated as follows: A = Andean-Argentinian, B = Neotropical, C = Southern Africa, D = African, E = Madagascan, F = Northern Australia, G = Malesian, H = Indian-Indochinese, I = Neozealandic-Patagonian and J = Eurasiatic (see Carta et al., 2022. Three modes of range inheritance (range expansion, local extinction, and narrow vicariance) are indicated as symbols on branches in the phylogeny. Biogeographic events: KP + GR disintegration = Kerguelen Plateau and Gunnerus Ridge disintegration, KPME = Cretaceous–Palaeogene mass extinction, NALB = North, Atlantic Land Bridge, EECO = Cenozoic thermal maxima, Australian-Antarctic split = No longer a land connection between Australia and Antarctica, Plates = Australian and Pacific plate collides.

**Appendix S12 BioGeoBEARS dispersal sinks and sources**

Number of dispersal (range expansion, e; jump dispersal, *j*) events for the three time slices (see Figure 3; 80 Ma, 40 Ma, 20 Ma and 0 Ma) from the DEC+TS+j model (see Table 3). Counts of events were averaged across the 100 biogeographic stochastic mappings (BSMs) with standard deviations in parentheses. Rows represent source ranges; columns represent

dispersal sinks. Darker shades indicate a higher frequency of dispersal events. The sum and percent of events in each row and column are given on the margins. Regions are abbreviated as follows: A = Andean-Argentinian, B = Neotropical, C = Southern Africa, D = African, E = Madagascan, F = Northern Australia, G = Malesian, H = Indian-Indochinese, I = Neozealandic-Patagonian and J = Eurasiatic.