bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Formatted for: *Journal of Biogeography*

Running title: Dispersal and biome shifts in *Pterocarpus*

The diversification of *Pterocarpus* (Leguminosae: Papilionoideae) was influenced by biome-switching and infrequent long-distance dispersal

Rowan J. Schley[1*], Ming Qin[2], Mohammad Vatanparast[3,4], Panagiota Malakasi[1], Manuel de la Estrella[5], Gwilym Lewis[1], Bente Klitgård[1]

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

[2]*South China Botanical Garden, Chinese Academy of Sciences, 1190 Tianyuan Rd, Tianhe, Guangzhou, Guangdong, China*

[3]*Department of Geosciences and Natural Resource Management, University of Copenhagen, 1958, Denmark*

[4]*Department of Botany, National Museum of Natural History, Smithsonian Institution, MRC 166, Washington, DC 20560, USA*

[5]*Departamento de Botánica, Ecología y Fisiología Vegetal, Facultad de Ciencias, Universidad de Córdoba, 14014, Córdoba, Spain*

*Corresponding author. *E-mail address:* rowan.schley@gmail.com (R. J. Schley).

Acknowledgements

All the authors would like to thank the PAFTOL program at RBG, Kew for funding the sequencing of Pterocarpus clade accessions. MQ's lab visit was funded by the National Natural Science Foundation of China (grant no. 31670193) and the Science and Technology Planning Project of Guangdong Province (grant no. 2019B030316020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Aim

Phenotypes which evolved for dispersal over ecological timescales may lead to significant macroevolutionary consequences, such as infrequent long-distance dispersal and diversification in novel environments. We aimed to reconstruct the phylogenetic history of *Pterocarpus* (Leguminosae/ Fabaceae) to assess whether seed dispersal phenotypes help to explain the current biogeographical patterns of this group.

Location

Pantropical

Taxon

The Pterocarpus clade, particularly *Pterocarpus* (Leguminosae/Fabaceae)

Methods

We sequenced ~300 nuclear loci captured using *Angiosperms*-353, a genomic 'bait set' for flowering plants, from which we generated a time-calibrated phylogenomic tree. To corroborate this, we also generated a time-calibrated phylogenetic tree from data-mined Sanger-sequencing data. We then collated distribution data and fruit dispersal morphology traits to compare trait-dependent and trait-independent biogeographical models, allowing us to assess whether dispersal traits influenced the spatio-temporal evolution of *Pterocarpus*. Finally, using the results of these model tests, we estimated the ancestral ranges and biomes of *Pterocarpus* species to better understand their biogeographical history.

Results

We recovered well-supported phylogenetic relationships within *Pterocarpus*, within which there were two subclades – one Neotropical and the other Palaeotropical. Our divergence date estimates suggested that *Pterocarpus* largely diversified from around 12 Ma, during the Miocene. Trait-dependent biogeographical models were rejected for both range and biome evolution within *Pterocarpus*, but models parameterising dispersal were supported. *Pterocarpus* largely diversified in the Neotropics, followed by dispersal and diversification into Africa and Asia, with later dispersal into Australasia/Oceania. The neotropical subclade of *Pterocarpus* underwent multiple biome switches between moist forest and dry forest, while in palaeotropical *Pterocarpus* we reconstructed multiple switches between moist forest and grassland.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Main conclusions

Overall, our analyses suggest that *Pterocarpus* underwent infrequent cross-continental dispersal and adaptation to novel biomes. While this was minimally impacted by fruit dispersal syndromes, shifts between moist and arid environments precipitated by long-distance dispersal and environmental change have played an important role in diversification within *Pterocarpus* since the Miocene.

Keywords: Angiosperms, Biomes, Dispersal, Diversification, Legumes, Phylogenomics, Traits, Tropics

Introduction

Dispersal is one of the major processes structuring the geographic ranges of organisms, both at ecological and evolutionary timescales (Brown et al., 1996, Sanmartín & Ronquist, 2004). Indeed, adaptations which evolved to promote dispersal locally may end up being significant evolutionary innovations influencing the biogeographical fate of lineages by allowing dispersal across continents and subsequent diversification over millions of years (Klaus & Matzke, 2020).

While vicariance has a major role in explaining biogeographical patterns, the role of dispersal was seen as being somewhat less significant (Sanmartín, 2012). However, there is now ample evidence that dispersal, and especially trans-oceanic dispersal, plays a major role in the biogeographical dynamics of a great number of terrestrial lineages (De Queiroz, 2005, Gillespie et al., 2012, Harris et al., 2018). This is especially evident in plants due to the wide range of traits that they have evolved to promote dispersal of their fruits, seeds and spores (Cousens et al., 2008). Seeds in particular have developed a wide range of morphological traits which allow dispersal over long distances, including wings and other aerodynamic structures (Cain et al., 2000), long dormancy periods (van der Pijl, 1982), thick seed coats and air pockets which render the seeds buoyant, thus facilitating dispersal across oceans (Bellot et al., 2020, Renner, 2004). Multiple recent studies have highlighted that the biogeographical distributions of modern plant lineages are largely explained by traits which they evolved for effective dispersal in their ecological context (e.g. Podocarpaceae (Klaus & Matzke, 2020); Annonaceae (Onstein et al., 2019)) using trait-dependent biogeographical modes. Indeed, it has been well established that a wide range of phenotypic traits have a significant impact on macroevolutionary dynamics (e.g. Rabosky et al., 2014).

Given their superlative biological diversity, the tropics are one of the most important places to examine the impact of dispersal on biogeography and diversification. Dispersal is particularly important for terrestrial organisms in the tropics because they are subject to both geographical and ecological constraints to their distribution. The tropics are largely separated into three zones (Neotropical, African and Asian/Australasian) by large areas of ocean, bounded to the north and south by increasingly unsuitable climates with increasing distance from the equator (Corlett & Primack, 2011, Morley, 2000). There is ample evidence of disjunct ranges across continents within tropical organisms, and particularly in plants (e.g., Arecaceae (Eiserhardt et al., 2011), Podocarpaceae (Klaus & Matzke, 2020), Leguminosae/Fabaceae (Schrire et al., 2005, Vatanparast et al., 2013), Vitaceae (Nie et al., 2012)). There are many different biomes within the tropics, spanning climatic extremes from deserts, seasonally dry forests, to mangroves and tropical rainforest. These different biomes are

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

largely defined by their ecological community composition (Pennington & Dick, 2004, Sanmartín, 2014) as well as by the physiognomy of their dominant plant species (Pennington & Ratter, 2006, Woodward et al., 2004), and are seen as independent evolutionary arenas (Hughes et al., 2013, e.g. Pennington et al., 2009) with broad differences in evolutionary dynamics. For example, rainforest tree species are known to disperse incredibly well, and outcross very effectively (Ward et al., 2005), even forming syngameons of interfertile species which hybridise (Schley et al., 2020) resulting in very large effective population sizes with species spread over thousands of hectares. In contrast, species from the much more restricted tropical dry forest biome are characterised by comparatively poor long-distance dispersal and much lower effective population sizes, resulting in high levels of endemism and well-resolved phylogenetic lineages (Pennington & Lavin, 2016). It is precisely these dynamics that make the tropics, and by extension the pantropical clades which occur in many different biomes, an ideal system to study biogeographical processes.

The advent of next-generation sequencing (NGS) techniques has opened a vast landscape of inquiry for investigating biogeographical processes such as those described above, especially at the species level and below. In addition, NGS techniques have enabled use of herbarium specimens, many of which were previously not viable to sample for molecular work, and unlocked a wealth of data held within the world's botanical collections (Staats et al., 2013). This is particularly important for work on tropical floras, as in many cases specimen collection entails extensive fieldwork in remote locations, while using novel genomic techniques we are now able to utilise large collections from herbaria and museums which can be centuries old (Hart et al., 2016). One such method is target-capture sequencing (Mamanova et al., 2010), which utilises RNA probes to extract highly fragmented DNA from plant material and allows high-throughput sequencing of this material. Since these RNA probes are designed to target phylogenetically informative regions of the genome (targeted genes), data may be collected and compared over many highly divergent taxa, such as with the *Angiosperms-353* probe (Johnson et al., 2019) utilised by the plant and fungal tree of life (PAFTOL, https://treeoflife.kew.org/) program at RBG, Kew. *Angiosperms-353* was originally developed as an angiosperm-wide target capture probe, but an increasing number of studies are using this bait for studies at the species level (e.g., *Nepenthes* (Murphy et al., 2020)).

One plant family which is being examined closely using this bait kit is the legumes (Leguminosae/Fabaceae). The legume family is an excellent study system to examine evolutionary processes in tropical environments because they are present in every terrestrial biome in the tropics (Lewis et al., 2005) and are dominant in many of them (de la Estrella et al., 2017). Members of the legume family are also among the 'hyperdominant' species of Neotropical rainforests (ter Steege et al., 2013, ter Steege et al., 2020), are the most prominent woody plant family in Africa and the second most dominant family in Southeast Asia after the Dipterocarpaceae (Gentry, 1988). Moreover, the circumtropical distribution and the breadth of ecological niches occupied by legume species renders them an ideal group with which to explore the eco-evolutionary history of the tropics. The genus *Pterocarpus* Jacq. belongs to the largest subfamily of legumes (Papilionoideae) and consists of around 33 species (Klitgård & Lavin, 2005, Lewis, 1987, Rojo, 1972), many of which are important timber species or are used in traditional medicine (Saslis-Lagoudakis et al., 2011). This genus is diagnosed by either winged or corky (coriaceous) fruits adapted for dispersal by wind or water, respectively, and

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

this is likely one reason why *Pterocarpus* is found pantropically (Klitgård & Lavin, 2005). *Pterocarpus* forms a part of the broader, pantropical 'Pterocarpus clade' which contains 22 genera and ~200 species, of which most are trees (Klitgård et al., 2013). However, the Pterocarpus clade also contains some herbaceous species, including economically important crop species such as the peanut (*Arachis hypogaea* L.).

Pterocarpus is a particularly good group of model species for eco-evolutionary studies within the tropics because of their pantropical distribution, as well as the fact that they occupy many different biomes, including rainforest, seasonally dry tropical forest and savannah. Moreover, the genus *Pterocarpus* is particularly suited to studies of dispersal because its species possess a range of dispersal mechanisms and associated morphologies, such as winged fruit dispersed by wind (anemochory), and buoyant fruits which allow dispersal by water (hydrochory) (Klitgård et al., 2013). Accordingly, we aim to examine the impact of dispersal traits on the diversification of *Pterocarpus* to provide insight into how trait evolution can structure the biogeographical distribution of species, and so influence the evolution of tropical diversity. Particularly, we will:

1. Reconstruct the relationships between species within the Pterocarpus clade using novel next-generation sequencing techniques.
2. Infer a time-calibrated phylogeny for *Pterocarpus* using multiple approaches.
3. Test whether trait-dependent or trait-independent biogeographical models best explain extant biogeographical patterns, both in terms of continental realm and biome.

Materials and Methods

Taxon sampling

Twenty-six of 33 accepted species (75%) were sampled within the genus *Pterocarpus* from across its distributional range for the generation of novel phylogenomic data using target capture sequencing. In total, 27 accessions were sampled within *Pterocarpus* for target capture sequencing, which are detailed in Supporting Information, Appendix S1, Table S1.1. An additional 22 outgroup taxa from the 21 other genera belonging to the Pterocarpus clade within which *Pterocarpus* is nested were sampled for phylogenetic analysis.

In addition to generation of novel target capture data for 26 of 33 species, Sanger sequence data from one nuclear locus (nrITS) and four plastid loci (*mat*K, *ndh*F-*rpl*32, *rbcl*, *trn*L-*trn*F) were downloaded from NCBI GenBank https://www.ncbi.nlm.nih.gov/genbank/ (Appendix S1, Table S1.2) for comparison between phylogenetic dating methods. These data represented 28 out of 33 *Pterocarpus* species, along with 8 outgroup species from the Pterocarpus clade.

A species list of all accepted taxa belonging to *Pterocarpus* was compiled using IPNI (IPNI, 2020) and a generic monograph of *Pterocarpus* (Rojo, 1972). This list was used to ensure that all taxonomically accepted species were included in analyses, as well as to account for synonyms and infraspecific taxa. Additionally, the voucher specimens were examined, and their determination updated as appropriate. Newly described and recently reinstated species from within the *P. rohrii*

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

complex were also sampled (Klitgård et al., *in prep*). Leaf material for DNA extraction was acquired from both silica material (*N*= 1) and herbarium specimens (*N*= 48) collected from the BM, K and MO herbaria.

Library preparation and sequencing

DNA extractions were carried out using 20 mg of dried leaf material with the CTAB method (Doyle & Doyle, 1987), following which DNA concentrations were measured using a Quantus fluorometer (Promega, Wisconsin, USA). Libraries were prepared using the NEBNext® Ultra™ II DNA Kit (New England Biolabs, Massachusetts, USA). Samples with high molecular weight DNA (>1000 bp, measured with a TapeStation 4200 (Agilent Technologies, California, USA)) were sheared with a Covaris focussed ultrasonicator M220 (Covaris, Massachusetts, USA). Following this, end-preparation and Illumina adaptor ligation were undertaken according to the NEBNext protocol, including a 400bp size-selection step with Agencourt Ampure XP magnetic beads (Beckman Coulter, California, USA). Libraries containing a range of insert sizes were then amplified using PCR, following the protocol outlined in Appendix S1, Table S1.2. Targeted bait capture was performed using the MyBaits protocol (Arbor Biosciences, Michigan, USA) to target 353 phylogenetically informative nuclear genes with *Angiosperms-353*, a bait kit designed for all Angiosperms. The final library pools were sequenced using a paired-end 150bp run on the Illumina HiSeq platform by Macrogen 154 Inc. (Seoul, South Korea).

Quality filtering, read assembly and alignment

DNA sequencing reads were quality-checked with the program FASTQC v0.11.3 (Andrews, 2010) and were subsequently trimmed using TRIMMOMATIC v.0.3.6 (Bolger et al., 2014) to remove adapter sequences and quality-filter reads. TRIMMOMATIC settings permitted <4 mismatches, a palindrome clip threshold of 30 and a simple clip threshold of 6. Bases with a quality score <28 and reads shorter than 36 bases long were removed from the dataset. Following quality-filtering, loci were assembled using SPADES v3.11.1 (Bankevich et al., 2012) by the HybPiper pipeline v1.2 (Johnson et al., 2016) with a coverage cut-off of 8x. All loci with <50% recovery among taxa were removed, and potentially paralogous loci were removed using the Python (Python Software Foundation, 2010) script '*paralog_investigator.py*', distributed with the HybPiper pipeline. Recovery of different loci was visualised using the '*gene_recovery_heatmap.R*' script distributed with HybPiper.

Following this, targeted loci were aligned by gene region (excluding those with potential paralogs) using 1,000 iterations in MAFFT (Katoh & Standley, 2013) with the '*—adjustdirectionaccurately*' option to incorporate reversed sequences. These alignments were then cleaned using the *automated1* flag in TRIMAL (Capella-Gutiérrez et al., 2009), following which they were visually inspected for poor sequence quality using *Geneious* v. 8.1.9 (https://www.Geneious.com). This was done in order to

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

remove taxa with mostly missing data and to prevent the inclusion of poorly recovered loci into the dataset, resulting in 303 refined alignments.

The nrITS and four plastid Sanger sequencing loci downloaded from NCBI GenBank were refined to include 68 sequences belonging to 36 species from nine genera (GenBank numbers for these sequences are listed in Appendix S1, Table S1.3). These data were aligned using MAFFT with the same parameters as above. These sequences were also subject to visual data inspection and trimming in Geneious to minimise missing sequence data where multiple accessions of a species were present.

Phylogenomic inference and divergence dating

Three hundred and three gene trees were inferred using RAxML v.8.0.26 (Stamatakis, 2014) with 1,000 rapid bootstrap replicates and the GTRCAT model of nucleotide substitution, which is a parameter-rich model and hence is most suitable for large datasets. Only the best-scoring maximum-likelihood trees were retained for further analysis by using the '*-f a*' option in RAxML. A species tree was generated under the multi-species coalescent model in ASTRAL v.5.6.1 (Zhang, C. et al., 2018). The heuristic version of ASTRAL was used because gene trees were estimated for >30 taxa, and the default parameters for the program were used for species tree inference. Monophyly was not enforced for individuals belonging to the same species (the '*-a*' option in ASTRAL).

Moreover, in order to compare between phylogenetic methods, we concatenated all 303 target capture locus alignments into a single alignment using AMAS (Borowiec, 2016), after which a phylogenetic tree was inferred using RAxML HPC2 (Stamatakis, 2014) on the CIPRES web portal ((Miller et al., 2010), https://www.phylo.org/). Inference was performed using 1,000 rapid bootstrap replicates and the GTRCAT model of nucleotide substitution (which is a parameter rich model and hence is most suitable for large data sets).

MCMCTREE, distributed with the PAML package (Yang, 2007), was used to date our phylogenomic tree. This program utilises approximate likelihood calculation and allows Bayesian divergence time estimation from large phylogenomic datasets (Dos Reis & Yang, 2013). To do this MCMCTREE uses a fixed phylogenetic tree topology without branch lengths to perform molecular clock analysis on the sequence data used to generate the tree. As such, the ASTRAL tree was rooted with an appropriate outgroup (*Cascaronia astragalina*) and the tree's branch lengths were removed with the *R* (R Development Core Team, 2013) package 'PhyTools' (Revell, 2012). This rooted, branchless tree and the concatenated 303-locus target capture alignment were then used as inputs for MCMCTREE. Fossil calibrations (listed in Table 1) were added using skewnormal priors as uncertainty is distributed asymmetrically across the mean age in fossil calibrations (Ho & Phillips, 2009) using the *R* package 'MCMCtreeR' (Puttick, 2019).

The MCMCTREE analysis was run by first calculating the gradient (b) and Hessian (h) parameters using the HKY substitution model, from which the posteriors of divergence times and rates were

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

estimated using two independent runs of 20 million generations to ensure MCMC convergence. Moreover, a third, identical run with only priors was performed and compared to the runs with both priors and data to assess whether prior settings were adequate. MCMC convergence for posterior estimation was checked using TRACER (Rambaut et al., 2015), as well as in the R packages 'ape' and 'bppr' (Angelis & Dos Reis, 2015, Paradis & Schliep, 2019). The final time-calibrated phylogenetic tree estimated with MCMCTREE was then plotted with 'MCMCtreeR'.

Table 1: Fossils used to generate a time-calibrated phylogenomic tree of the Pterocarpus clade in MCMCTREE and a dated phylogenetic tree in BEAST v.18.0. Skewnormal calibration priors were used for fossil ages in MCMCTREE to provide a minimum age estimate, and the corresponding Lognormal priors were used for the BEAST analysis. Mean ages are shown for calibrations, with 95% confidence intervals in parentheses. All ages are in millions of years (Ma).

Calibration point	**Age (Ma)**	**Prior distribution (MCMCTREE)**	**Prior distribution (BEAST)**	**Fossil**
Pterocarpus clade	37.5 (34-41)	Skewnormal	Lognormal	*Luckowcarpa gunnii* (Martínez, 2018)
Pterocarpus-Riedeliella crown group	21.8 (15.9-27.8)	Skewnormal	Lognormal	*Pterocarpus tertiarius* (Buzek, 1992)
Tipuana crown group	8.2 (7.9-8.5)	Skewnormal	Not used (no suitable taxa)	*Tipuana ecuatoriana* (Burnham, 1995)

For comparison to the time-calibrated phylogenomic tree inferred with MCMCTREE, Sanger sequencing data downloaded from GenBank were used to perform divergence time estimation in BEAST v.1.8.0 (Drummond & Bouckaert, 2015) on the CIPRES web portal. Multiple runs of BEAST were undertaken using several different tree models (Yule, birth-death and birth-death (incomplete sampling)) to compare model fit. Substitution rate models were chosen for each of the five loci using AICc comparisons implemented in JMODELTEST (Darriba et al., 2012). Loci and their corresponding substitution rate models are outlined in Appendix S1, Table S1.4. The Birth-Death tree model was chosen for inference based on likelihood comparisons. Dated trees were generated using a relaxed lognormal clock model, informed by coefficient of variation values >0.1, taken from the BEAST log file. Two of the three fossil calibration points in the MCMCTREE phylogenomic dating analysis were used to provide date estimates in BEAST (Table 2).

Phylogenetic inference in BEAST comprised three independent runs of 50 million generations (giving a total of 150 million generations), which were combined using LOGCOMBINER v.1.8.2 with a burn-in level of 10%. Convergence of runs was assessed using effective sample size (ESS) of all parameters in TRACER v. 1.6 with a threshold value of 200. Tree model choice was based on path sampling analysis output from BEAST, using 10 million generations with 30 path steps. This resulted in the use

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of the Birth-Death model for inference**.** The final ultrametric trees were generated in TREEANNOTATOR v1.8.2 and visualised using FIGTREE v1.4.4 (Rambaut, 2014).

Biogeographical analyses

Trait-dependent biogeographical analyses were performed using the R package 'BioGeoBEARS' (Matzke, 2013) to understand the relationships between dispersal trait evolution and the spatio-temporal patterns of diversification within *Pterocarpus*, both in terms of biogeographical realm and biome. To do this, fruit morphology traits were collected from multiple sources, including the first monograph of *Pterocarpus* (Rojo, 1972), a taxonomic revision circumscribing new species within the *P. rohrii* complex (Klitgård *et al. in prep*), and several taxonomic treatments of various *Pterocarpus* species (de Candolle, 1825, Klitgård et al., 2000, Zamora, 2000). Since no fruiting specimen of *P. rohrii* var. *rubiginosus* exists, it was inferred to be winged since it belongs to the *P. rohrii* species complex. Fruit dispersal traits were then plotted on the *Pterocarpus* ASTRAL tree with biome and realm for each species using the *dotplot()* function in the R package 'PhyTools' to visualise the data.

Geographical distribution data were collected from GBIF (www.GBIF.org (06 November 2019), GBIF Occurrence Download https://doi.org/10.15468/dl.tr3h2y) based on the curated taxonomic name list described above. Only records corresponding to 'preserved specimens' and 'material samples' were included in this dataset, and all 'cultivated' specimens were removed. Moreover, cultivars of the peanut (*Arachis hypogaea*), which is part of the Pterocarpus clade, were removed to prevent the inclusion of any cultivated specimens in our geographical dataset. This dataset was then further refined using the R package 'CoordinateCleaner' ((Zizka et al., 2019); https://github.com/ropensci/CoordinateCleaner) and the input for our *BioGeoBears* analysis was produced from these data with Alex Zizka's 'biogeography in R' scripts (https://github.com/azizka/Using_biodiversity_data_for_biogeography).

From these refined data, species were assigned to biogeographical regions using the *wwfLoad()* function in 'speciesgeocodeR'. Coordinates for each species were plotted on two maps: the first was based on the WWF terrestrial realms, representing five major biogeographical realms within the tropics (Australasia, Afrotropics, IndoMalaya, Neotropics and Oceania (Olson et al., 2001)). The second was based on the WWF terrestrial biomes map, representing six of the world's tropical terrestrial biomes (tropical moist forest, montane grassland, desert, mangrove, tropical dry forest and tropical grassland (Olson et al., 2001)). These regions are displayed in Appendix S1, Fig. S1.1a. This allowed us to extract both the biogeographical area and the biome in which each sample was found, which then allowed us to assign each species to one or more biogeographical regions and biomes. Finally, biogeographical assignments for each *Pterocarpus* species were cross-referenced with monographs and previous studies of *Pterocarpus* (Rojo, 1972, Saslis-Lagoudakis et al., 2011, Klitgård et al., in prep), and with the Plants Of the World Online database (http://powo.science.kew.org/). All traits (fruit morphology, realm assignment and biome assignment) are visualised onto the *Pterocarpus* phylogenetic tree in Appendix S1, Fig. S1.1b.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Trait-dependent historical biogeography analyses were performed using 'BioGeoBEARS' to assess the effect of dispersal traits on the biogeographical history of *Pterocarpus.* The fit of trait-dependent biogeographical models outlined by Klaus and Matzke (2020) and non-trait-dependent models were assessed for both biogeographical realms and biomes independently. Six models were tested, all of which incorporated dispersal, extinction and range switching. First, we tested between the standard models implemented in 'BioGeoBEARS', which are DEC (Ree & Smith, 2008), which additionally accounts for vicariance and 'subset' sympatric speciation, DIVALIKE (Ronquist, 1997), which also accounts for vicariance, and BAYAREALIKE (Landis et al., 2013), which additionally accounts for widespread sympatric speciation but not vicariance.

In addition to these standard models, we also included three trait-dependent models in our comparisons. The first of these models parameterised only traits with one rate of switching ('*Trait_1Rate'*, which describes switching from trait 1 to trait 2 (t_{12})). The second was '*Trait_2Rates'*, which describes two rates of trait switching (t_{12} and t_{21}). Finally, the third was *DEC_t12_t21_m2*, which is a trait-dependent extension of the DEC model that describes two rates of trait switching as well as a multiplier on the base anagenetic dispersal rate *d* (known as '*m2'*).

The model which was best described by the data was chosen using AICc and Akaike weights, from which ancestral ranges and range shifts were estimated across *Pterocarpus*. Ancestral ranges were plotted onto the time-calibrated phylogenomic tree inferred with MCMCTREE using the function '*plot_BioGeoBears_results()*' for realm and biome independently. For realm, the maximum number of range areas was set to five, and for biome it was set to six, as these were the total number of biogeographical areas used for each analysis. Fixed parameters in each model were given 'initial' and 'estimate' values from '0.0' to '0.00001' as per the recommendations of the 'BioGeoBEARS' authors.

Results

Phylogeny and divergence dating

A mean of 17.7% of reads were 'on target' (i.e., were successfully assembled) onto the *Angiosperms*-353 bait sequences, with a range from 6.9% to 30.5%. Gene recovery can be seen in the heatmap displayed in Appendix S1, Fig. S1.2. Of the 353 loci targeted by *Angiosperms*-353, we discarded 35 loci which were potentially paralogous and 15 which were assembled poorly or had over 50% missing data across taxa.

Our ASTRAL analysis indicated that most bipartitions were well supported with local posterior probabilities (LPP) >0.8, with 1 being full support (Fig. 1) and a quartet score of 0.68. Within this tree, most generic splits are well supported (LPP >0.9) and largely support the monophyly of *Pterocarpus*. However, one species, *Pterocarpus dubius* (formerly *Etaballia dubia*) was resolved in sister position to the rest of *Pterocarpus.* Within *Pterocarpus*, there were two major subclades, that containing species mostly from the Neotropics (the neotropical clade) and that containing species from the Palaeotropics (the palaeotropical subclade). However, this split received a relatively low level of support (LPP=0.43, quartet score = 34.48%). The RAxML analysis of the same 303 loci

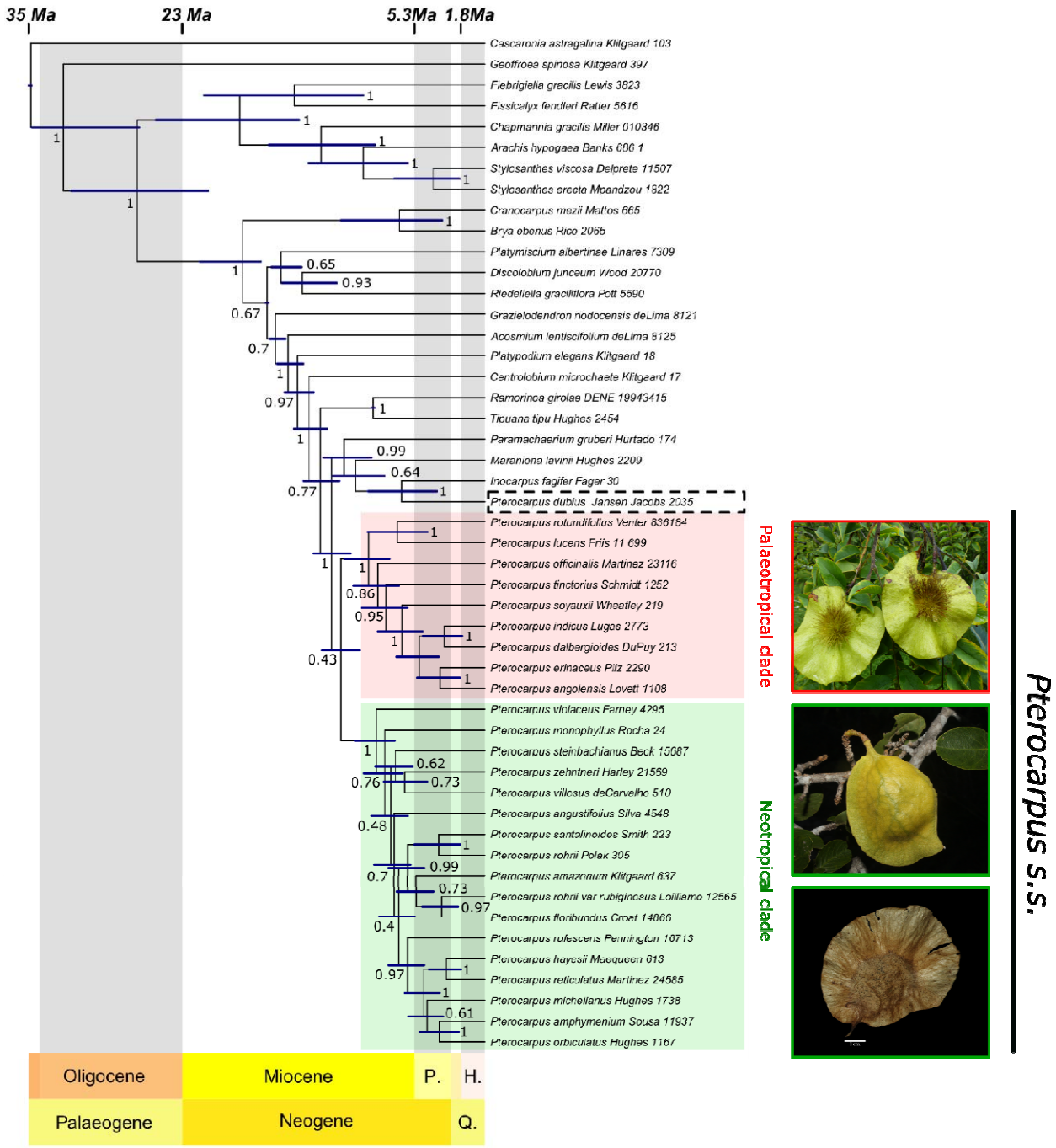
bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

concatenated into a single dataset (Appendix S1, Fig. S1.3) recovered a nearly identical topology to the ASTRAL species tree (Fig. 1), and was much better resolved, with nearly all nodes having bootstrap values >95. Moreover, the split between Neotropical and Palaeotropical *Pterocarpus* species received full support (BS=100) in the inference based on the concatenated dataset.

Divergence time estimation carried out in MCMCTREE based on the ASTRAL species tree of 303 nuclear loci provided a robust, time-calibrated tree of the Pterocarpus clade. The two MCMC runs were convergent, with every node showing a suitable effective sample size (> 200), and posterior estimates fit the priors well as indicated in Appendix S1, Fig. S1.4a, b and c. The time-calibrated phylogenomic tree in Fig. 1 suggests that the Pterocarpus clade arose around 35 Ma during the late Eocene, whereas most diversification of present-day genera occurred during the Miocene, from 23-5.3 Ma. Of these, *Pterocarpus* (*sensu stricto*) diverged from its sister genera around 12 Ma, and underwent most of the diversification which gave rise to current *Pterocarpus* species in the late Miocene. It is worth noting that divergence date estimates are described by larger HPD intervals in the earlier divergence events within the Pterocarpus clade towards the backbone of the tree, and this increased uncertainty is mirrored in Appendix S1, Fig. S1.5, which indicates that uncertainty (and hence the distribution of inferred tree topologies) is broader in the earlier divergence events inferred, as well as being much reduced around fossil calibration points, as would be expected.

Figure 1: Fossil-calibrated MCMCTREE of the Pterocarpus clade. Node bars show 95% highest posterior density (HPD) for node ages. Local posterior probabilities (LPP), a measure of topological support, are shown for each node, and were taken from the ASTRAL analysis which provided the tree topology for the MCMCTREE analysis. Shading on the background of the phylogram represents geological epochs, which are labelled beneath the tree. The red and green shading indicates species belonging to the palaeotropical and neotropical subclades of *Pterocarpus*, respectively, and the extent of the genus *sensu stricto* is shown by the black bar. The dashed box around *Pterocarpus dubius* highlights its paraphyly with the rest of the genus. Photographs show winged fruits of *Pterocarpus erinaceus* (© Gwilym Lewis), coriaceous fruits of *Pterocarpus monophyllus* (© Domingos Cardoso) and winged fruits of *Pterocarpus violaceus* (© Reinaldo Aguilar, CC BY-NC-SA 2.0).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



The topology of the BEAST trees (Appendix S1, Fig. S1.6) inferred using Sanger sequencing data from GenBank recovered a broadly similar topology to that inferred in the MCMCTREE analysis (Fig. 1). Moreover, the divergence time estimates of this phylogeny were largely congruent with those from Fig. 1, including an origin of the Pterocarpus clade within the late Eocene (~38 Ma), followed by most divergence between genera within the Miocene (~23-5.3 Ma), as well as most divergence events between the sampled species within *Pterocarpus*. However, the age estimate for the initial divergence of *Pterocarpus* is later than that inferred with MCMCTREE, at around 25 Ma, again during the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Miocene. In addition, the BEAST analysis displayed a much broader degree of uncertainty on age estimates, as indicated by the breadth of the 95% HPDs, and recovered *Pterocarpus dubius* as nested within *Pterocarpus*, whereas in the MCMCTREE analysis *P. dubius* is in sister clade position to the rest of *Pterocarpus*.

Biogeography

Pterocarpus largely diversified in the Neotropics, with subsequent dispersal events to the Afrotropics, Indo-Malaya, Australasia and Oceania, and the current distribution of *Pterocarpus* species (both in terms of biome and realm), as well as the distribution of fruit dispersal traits across the ASTRAL tree, are visualised in Appendix S1 (Fig. S1.1b). For the analysis of biogeographical realm range evolution using 'BioGeoBEARS', AICc comparison indicated that the best-fitting model was DIVALIKE (AICc= 52.785) (Appendix S1, Table S1.5). This model parameterises dispersal, extinction and vicariance, and was a better fit to the *Pterocarpus* data than the trait-dependent models. This also suggests that narrow sympatric speciation was an important mechanism of range evolution. Using DIVALIKE, we estimated a shared ancestral range of both the Neotropics and Afrotropics for the common ancestor of *Pterocarpus*, with subsequent splitting into neotropical and palaeotropical clades (Fig. 2a). Most of the palaeotropical species occur in Africa, a range shared with most nodes within the palaeotropical clade. Within this palaeotropical clade the common ancestor of *P. indicus* and *P. dalbergioides* appears to have dispersed into Asia, and subsequently into Australasia and Oceania, having a shared Afro-Asian range. It is interesting to note that within both neotropical and palaeotropical clades, there are species which share ranges, being found in both eastern South America and west Africa (*P. santalinoides* and *P. officinalis*). A shared characteristic of these two amphi-Atlantic species are thick-walled, wingless, buoyant fruits which allow water-borne dispersal.

For biome evolution, the model which best fit our data across trait-dependent and trait-independent models was DEC (AICc= -64.882, Appendix S1, Table S1.5). This model parameterises dispersal, extinction and cladogenesis, and suggests an important role for narrow vicariance and sympatric speciation as well as dispersal and extinction in biome change across *Pterocarpus*. Using DEC, we inferred an ambiguous ancestral biome for the common ancestor of *Pterocarpus*, but at this node a shared moist forest+dry forest+grassland biome was the most probable (Fig. 2b). Within the neotropical clade of *Pterocarpus*, there appears to have been multiple events of biome-switching between moist forest and seasonally dry tropical forest (SDTF), with multiple nodes reconstructing a shared moist forest+dry forest biome. Within the palaeotropical *Pterocarpus* clade, the most probable state for the majority of ancestral nodes was a shared moist forest and grassland. In a reflection of the neotropical *Pterocarpus* species, there appears to have been multiple biome-switching events between grassland and moist forest in the African *Pterocarpus*. In addition, it is of interest that one of the two *Pterocarpus* species which inhabit both South America and Africa (*P. officinalis*) occurs both in tropical rainforest as well as mangroves.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

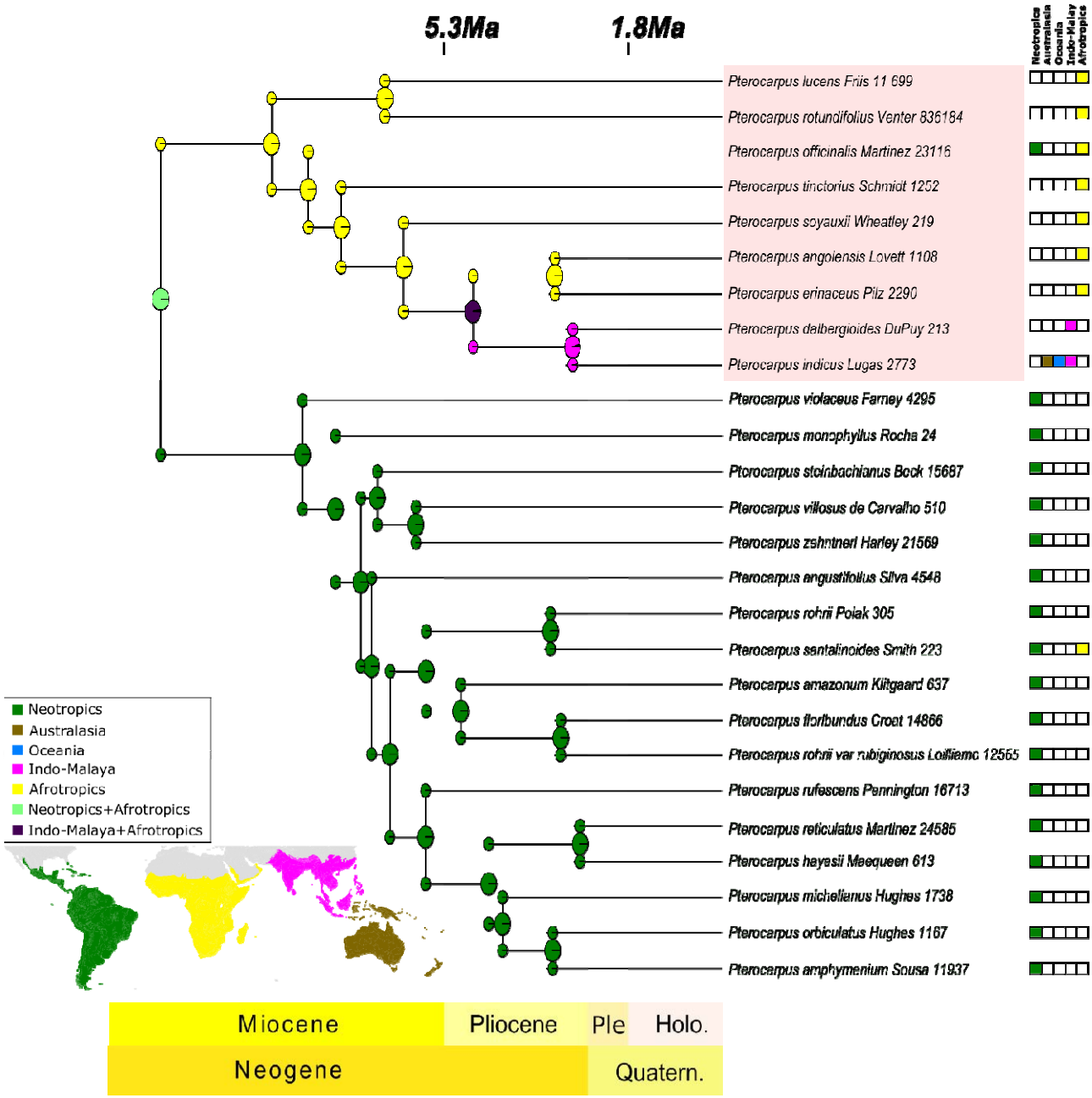


Figure 2a) Biogeographical range estimation of *Pterocarpus*. Pie charts represent likelihoods of ranges derived from ancestral range estimations run on the MCMCTREE topology, using the *R* package 'BioGeoBEARS'. The legend represents five biogeographical realms defined by Olson *et al.* (2001) and shown in Appendix S1, Fig. S1.1a. Shading on the background of the phylogenetic tree represents geological epochs, and within the bottom bar 'Ple' represents 'Pleistocene', 'Holo.' represents 'Holocene' and 'Quatern.' represents 'Quaternary'. Inset is a map displaying biogeographical realms as defined by Olson *et al*. (2001).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

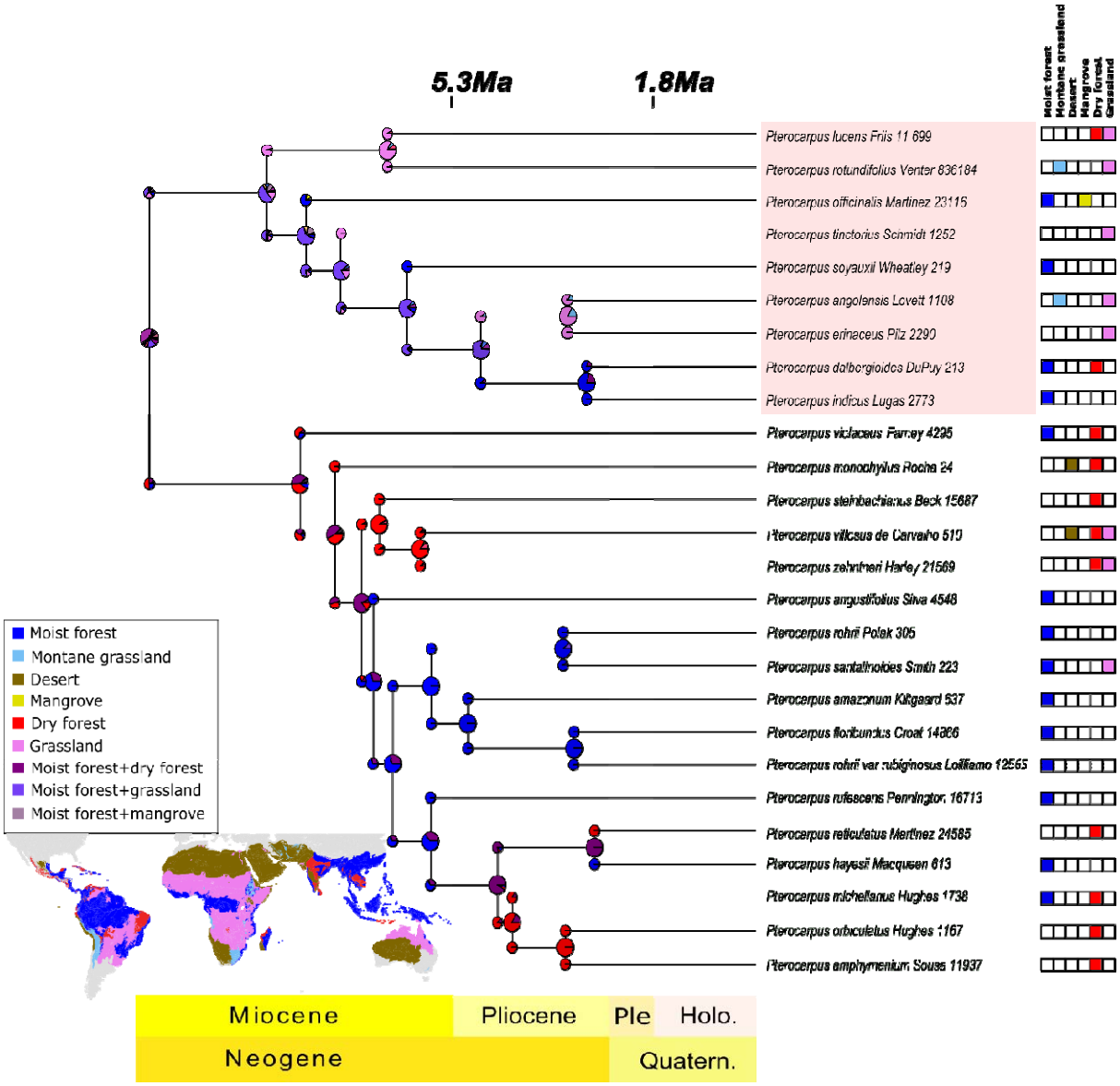


Figure 2b) Historical biome estimation of *Pterocarpus*. Pie charts represent likelihoods of ranges derived from ancestral range estimations run on the MCMCTREE topology, using the *R* package 'BioGeoBEARS'. The legend represents six biomes defined by Olson *et al.* (2001) and shown in Appendix S1, Fig. S1.1a. Shading on the background of the phylogenetic tree represents geological epochs, and within the bottom bar 'Ple' represents 'Pleistocene', 'Holo.' represents 'Holocene' and 'Quatern.' represents 'Quaternary'. Inset is a map displaying biomes as defined by Olson *et al*. (2001).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Divergence dating indicated that the Pterocarpus clade began to diversify in the late Eocene (~35 Ma), followed by extensive diversification during the Miocene (23-5.3 Ma). During the Miocene, the genus *Pterocarpus* largely diversified within the Neotropics, with subsequent dispersal and diversification into the Afrotropics, Indo-Malaya, Australasia and Oceania. The diversification of *Pterocarpus* appears to have involved multiple biome shifts, and while dispersal across large geographical areas played a role in range expansion, the fruit dispersal traits we examined were not significantly associated with macroevolutionary dynamics according to our results.

It is apparent that water-dispersed (hydrochoric) species such as *Pterocarpus santalinoides* and *P. officinalis* are the only species with amphi-atlantic ranges, being found both in South America and Africa (Fig. 2a; Fig S3a), suggesting a role for dispersal in range expansion. This differs with the ranges of wind-dispersed species, which tend to be restricted to a single continent (apart from *P. indicus*). Overall, it appears that traits for hydrochoric dispersal evolved multiple times independently in *Pterocarpus*, likely in concert with adaptation to waterlogged habitats such as mangroves and riparian zones, as evident in *P. officinalis* (Fig. 2b; Appendix S1, Fig. S1.1b). These traits which evolved to promote dispersal in the ecological theatre may have then led to rare long-distance dispersal across oceans (as we see in extant species such as *P. santalinoides* and *P. officinalis*). This was likely then followed by diversification into novel biomes on new continents in the ancestors of multiple *Pterocarpus* species during the Miocene and Pliocene (Fig. 2a; Fig. 2b). A similar pattern has been shown in the Bombacoideae (Malvaceae) (Zizka et al., 2020).

Diversification history of Pterocarpus *and the Pterocarpus clade*

The bulk of the extant diversity of the Pterocarpus clade arose during the Miocene, between 23-5.3 Ma. Our results (Fig. 1; Appendix S1, Fig. S1.6) suggest that the genus *Pterocarpus* diversified in the late Miocene, which was congruent across methods (MCMCTREE, BEAST) and across data types (NGS nuclear data for 303 loci, Sanger sequencing data for one nuclear and four plastid loci). This pattern of Miocene diversification has been shown in multiple other tropical tree genera (e.g., the mahoganies (Meliaceae) (Koenen et al., 2015), the Brownea clade (Leguminosae) (Schley et al., 2018) and the Daniellia clade (Leguminosae) (Choo et al., 2020) as well as herbaceous legumes such as *Eriosema* (Cândido et al., 2020) and was likely concurrent with major climatic, geological and ecological changes which occurred in the tropics during the Miocene (Keeley & Rundel, 2005, Morley, 2000).

One such change which could have had a major effect on populations of *Pterocarpus* ancestors was the global cooling which occurred during the Miocene (Zachos et al., 2001, Zachos et al., 2008). This cooling was precipitated by the closure of the Tethys seaway and a subsequent change in ocean currents (Rommerskirchen et al., 2011, Zhang, Z. et al., 2014), leading to a decrease in global temperature, desiccation of the continents and a retreat of tropical vegetation, especially on the African continent (Couvreur et al., 2021, Morley & Richards, 1993, Senut et al., 2009). Moreover, within the Neotropics, orogenic events within the Andes were beginning in earnest from the early to

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mid-Miocene (23-12Ma), driving further environmental changes within the Neotropics (Hoorn et al., 2010). The biogeographical and ecological impacts of this climatic change are discussed further below.

While our divergence dating analyses were broadly congruent with each other across methods and data type, there were some discrepancies between the two methods. The most obvious difference between date estimates for these two methods was the origin of *Pterocarpus*, for which MCMCTREE inferred a date of 12Ma (Fig. 1), whereas the BEAST analysis carried out on the Sanger sequence dataset suggested that *Pterocarpus* arose around 25Ma (Appendix S1, Fig. S1.6). The most likely reason for this discrepancy is the size of the dataset, since MCMCTREE was run on a nuclear NGS dataset consisting of 303 loci, while BEAST was run to corroborate the MCMCTREE results using previously collected Sanger sequence data for 5 loci, resulting in a much less informative sequence matrix.

It is also likely that the inclusion of plastid loci within the BEAST analysis extended the date estimate for the origin of *Pterocarpus*, due to deeper divergences within the plastid genome (Wolfe et al., 1987). This conflicting signal would have also led to the much higher degree of uncertainty on age estimates in the BEAST analysis, as indicated by the breadth of the 95% HPDs. In addition, the nested position of *Pterocarpus dubius* within *Pterocarpus*, as shown in the BEAST analysis (Appendix S1, Fig. S1.6) contrasts with the sister clade position of *P. dubius* to *Pterocarpus* in the MCMCTREE analysis (Fig. 1). This species was previously circumscribed as *Etaballia dubia* (Klitgård et al., 2013) and is highly divergent in terms of floral morphology. It is possible that the inclusion of chloroplast loci in the BEAST tree caused this incongruence, due to the lower mutation rate of chloroplast regions (Wolfe et al., 1987), or through plastid capture resulting from hybridisation (Naciri & Linder, 2015). While there is undoubtedly useful signal within these plastid loci, further work using plastomes would help to accurately infer discordances between nuclear and organellar loci in tree topology and divergence dates.

Spatio-ecological diversification history of Pterocarpus

Ancestral range reconstruction (Fig. 2a and 2b) indicated that dispersal was likely important for range expansion of *Pterocarpus* species. This is because the best-supported models in ancestral range reconstructions based on biogeographical realm and biome were DIVALIKE and DEC, respectively. Since these models parameterise dispersal (as well as vicariance and sympatric speciation), this suggests a biogeographical role for dispersal (as in Garcia-R & Matzke, 2021, Tsang et al., 2020). Moreover, a high degree of biome-switching was evident among *Pterocarpus* lineages, suggesting a significant role of ecology and dispersal in driving speciation within the group. However, the lack of support for trait-dependent models suggested that fruit dispersal traits had an indirect effect on broad-scale macroevolution of the group.

Using DIVALIKE we inferred that *Pterocarpus* largely diversified in the Neotropics, with subsequent dispersal and diversification in the Afrotropics, Indo-Malaya, Australasia and Oceania, giving rise to separate neotropical and palaeotropical clades within *Pterocarpus* (Fig. 2a). While the inferred

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ancestral range of *Pterocarpus* was shared between the Afrotropics and Neotropics, nearly all other genera within the Pterocarpus clade are confined to tropical America, and so it seems likely that *Pterocarpus* first evolved in the Neotropics, followed by dispersal into Africa and Asia. Ancestral biome reconstruction recovered a highly ambiguous biome for the common ancestor of *Pterocarpus*. This initial ambiguity is likely artefactual, due to the presence of two distinct descendent clades at this node– one in the Neotropics and one in the Palaeotropics, with species within these clades occurring in multiple biomes.

i. *The palaeotropical clade*

In the palaeotropical clade of *Pterocarpus*, most extant species occur in Africa, with only one subclade (that containing *P. indicus* and *P. dalbergioides*) having dispersed and speciated into Asia, Australasia and Oceania. The palaeotropical clade appears to have diversified largely within Africa over the last ~10Ma, in the late Miocene and into the early Pliocene (Fig. 2a). As mentioned above, this was concurrent with the cooling and desiccation of the African continent, and likely spurred on by the expansion of grasslands during this period (Couvreur et al., 2021). This is corroborated by the fact that most of the extant species within the palaeotropical clade are now found in grassland and are wind-dispersed (Fig. 2b).

Within the Palaeotropical clade, the two Asian sister species (*P. dalbergioides* and *P. indicus*) only dispersed to the Asian tropics and beyond into Australasia and Oceania around 4Ma, during the Pliocene (Fig. 2a). It is likely that the divergence of these species was caused by infrequent eastward seed dispersal, promoted by the east African coastal current (EACC) between Africa and Asia which formed with the closing of the Tethys seaway in the late Miocene (Scotese, 2004). It is interesting to note that these species are found in rainforests and mirror the highly dispersible nature of the other rainforest-dwelling member of the Palaeotropical clade (*P. officinalis*), which is found in both Africa and South America. This is congruent with the observation that rainforest tree species tend to be excellent dispersers (Dexter et al., 2017, Pennington & Lavin, 2016), and are often water dispersed. *Pterocarpus officinalis* is also found in mangroves, further suggesting that the corky exocarp of the fruit which it evolved for dispersal at an ecological scale (Appendix S1, Fig. S1.1b) is responsible for its ancestor's dispersal across the Atlantic and speciation in the Neotropics from African ancestors.

The most recent common ancestor of *P. officinalis* appears to have been Afrotropical, and *P. officinalis* is nested within the palaeotropical clade dating back to the late Miocene (ca. 10 Ma; Fig. 1 & 2a). According to the fossil record of the Caribbean region, *P. officinalis* occurred during the Quaternary in the Neotropics (Graham, 1995) and considering the presence of *P. officinalis* subsp. *gilletii* (De Wild.) Rojo in West Africa along with the water dispersal capacity of this species' fruits, long-distance dispersal by ocean currents likely facilitated the geographic expansion of the species from Africa to the Neotropics (Rivera□Ocasio et al., 2002). Indeed, more generally, the observed biogeographical shifts in *Pterocarpus* are likely due to dispersal because by the mid Miocene the continents were largely in their current day positions (Scotese, 2004). Moreover, model comparison performed in our 'BioGeoBEARS' analyses showed that models parameterising dispersal were the best supported for both realm and biome reconstructions (Appendix S1, Table S1.5).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ii. *The neotropical subclade*

Within neotropical *Pterocarpus* we inferred multiple events of biome-switching between rainforest and dry forest, along with dispersal into other arid biomes (e.g., grassland) (Fig. 2b). Our analyses indicated that the most likely biome for the ancestral node of the neotropical *Pterocarpus* was a shared rainforest+dry forest habitat. Moreover, within the neotropical *Pterocarpus*, we inferred multiple events of biome switching within and between clades, largely between rainforest and dry forest habitats, which mainly occurred between the late Miocene and the Pliocene (~8 - 2 Ma). This suggests a significant role of dispersal and ecologically-driven speciation within *Pterocarpus* species, and provides another example of biome switching between these habitats, which has been extensively studied in the Neotropics (Ireland et al., 2010, Klitgård, 2005, Pennington & Dick, 2010).

The biogeographical events we inferred within the Neotropics may be explained by the acceleration of Andean orogeny during the Miocene (Hoorn et al., 2010), and provide a contrast to the major drivers of diversification in African *Pterocarpus*. Rather than being driven by extensive desiccation and cooling as in Africa (Couvreur et al., 2021), mountain building within the Andes would have resulted in the formation of the Pebas and Acre wetland systems during the Miocene (Antonelli & Sanmartín, 2011, Hoorn et al., 2010). The resulting habitat heterogeneity, coupled with hyperdiverse soil assemblages formed by erosion of the newly formed Andes, would have resulted in adaptation to varying edaphic conditions in rainforest-dwelling species of *Pterocarpus*, leading to ecological differentiation and speciation (e.g. Fine et al., 2004, Fine et al., 2005). Moreover, since rainforest communities are relatively more invasible due to variation in mortality and recruitment caused by drought (da Costa et al., 2010), this may have promoted colonisation and speciation of ancestrally dry-forest dwelling *Pterocarpus* species into rainforest habitats. This appears to be the case, given that neotropical *Pterocarpus* species from rainforests are nested within species found in dry forests (Fig. 2b).

Most seasonally dry tropical forests within the Neotropics are found either within the inter-Andean valleys or Central America and largely began to form during the mid-Eocene (Pennington & Ratter, 2006, Pennington & Lavin, 2016). As such, it is likely that dispersal-mediated invasion of these novel habitats (newly accessible due to the formation of the Andes and gradual rise of the Panama Isthmus (Bacon et al., 2012)) from the Miocene onwards could have resulted in the repeated biome switching we observed through dispersal and ecological speciation. Other studies have indicated that the spread of tropical dry forests, coupled with dispersal and speciation within them has been a significant driver of evolution for many other genera within the legume family (Herendeen, 1992, Schrire et al., 2005).

Conclusion

Overall, our analyses suggest that while a dispersal trait-dependent model of range macroevolution was not supported, the biogeographical history of *Pterocarpus* was likely mediated by cross-continental dispersal, associated range change and diversification driven by biome shifts. Indeed, biome switching between moist and arid environments, likely precipitated by infrequent long-distance

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dispersal events and environmental change, appear to have promoted ecological differentiation and diversification within *Pterocarpus* since the early Miocene.

This investigation is the first such generic-level work on *Pterocarpus* using next-generation sequencing methods and highlights the utility of target bait capture techniques for leveraging collections of plant material held within herbaria and other institutions. Our study is one of an increasing number of genus-level studies to utilise *Angiosperms*-353 and highlights how a large volume of data can be obtained at relatively fine taxonomic scale using RNA probes designed to work at genus and family level (Johnson et al., 2019, Larridon et al., 2020, Ogutcen et al., 2021).

Data availability statement

The target capture sequencing data generated for this project are publicly available in the NCBI sequence read archive under the BioProject number PRJNA728569, with the accession numbers SAMN19092119 - SAMN19092167.

References

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

Angelis, K. & Dos Reis, M. (2015) The impact of ancestral population size and incomplete lineage sorting on Bayesian estimation of species divergence times. *Current Zoology* **61,** 874-885.

Antonelli, A. & Sanmartín, I. (2011) Why are there so many plant species in the Neotropics? *Taxon* **60,** 403-414.

Bacon, C.D., Mora, A., Wagner, W.L. & Jaramillo, C.A. (2012) Testing geological models of evolution of the Isthmus of Panama in a phylogenetic framework. *Botanical Journal of the Linnean Society* **171,** 287-300.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S. & Prjibelski, A.D. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19,** 455-477.

Bellot, S., Bayton, R.P., Couvreur, T.L., Dodsworth, S., Eiserhardt, W.L., Guignard, M.S., Pritchard, H.W., Roberts, L., Toorop, P.E. & Baker, W.J. (2020) On the origin of giant seeds: the macroevolution of the double coconut (Lodoicea maldivica) and its relatives (Borasseae, Arecaceae). *The New Phytologist* **228,** 1134.

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114-2120.

Borowiec, M.L. (2016) AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **4,** e1660.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Brown, J.H., Stevens, G.C. & Kaufman, D.M. (1996) The geographic range: size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics* **27,** 597-623.

Burnham, R.J. (1995) A new species of winged fruit from the Miocene of Ecuador: Tipuana ecuatoriana (Leguminosae). *American Journal of Botany* **82,** 1599-1607.

Buzek, C. (1992) Fruits of Pterocarpus tertiarius Weyland from the North-Bohemian Basin, Czechoslovakia. *Advances in legume systematics: Part 4* 19-31.

Cain, M.L., Milligan, B.G. & Strand, A.E. (2000) Long□distance seed dispersal in plant populations. *American Journal of Botany* **87,** 1217-1227.

Cândido, E.S., Vatanparast, M., de Vargas, W., Bezerra, L.M., Lewis, G.P., Mansano, V.F., Simões, A.O., Silva, M.J., Stirton, C. & Tozzi, A.M.G. (2020) Molecular phylogenetic insights into the evolution of *Eriosema* (Fabaceae): a recent tropical savanna-adapted genus. *Botanical Journal of the Linnean Society* **194,** 439-459.

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

Choo, L.M., Forest, F., Wieringa, J.J., Bruneau, A. & de la Estrella, M. (2020) Phylogeny and biogeography of the Daniellia clade (Leguminosae: Detarioideae), a tropical tree lineage largely threatened in Africa and Madagascar. *Molecular phylogenetics and evolution* **146,** 106752.

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

Cousens, R., Dytham, C. & Law, R. (2008) *Dispersal in plants: a population perspective*Oxford University Press.

Couvreur, T.L., Dauby, G., Blach□Overgaard, A., Deblauwe, V., Dessein, S., Droissart, V., Hardy, O.J., Harris, D.J., Janssens, S.B. & Ley, A.C. (2021) Tectonics, climate and the diversification of the tropical African terrestrial flora and fauna. *Biological Reviews* **96,** 16-51.

da Costa, A.C., Galbraith, D., Almeida, S., Portela, B.T.T., da Costa, M., Junior, João de Athaydes Silva, Braga, A.P., de Gonçalves, P.H., de Oliveira, A.A. & Fisher, R. (2010) Effect of 7 yr of experimental drought on vegetation dynamics and biomass storage of an eastern Amazonian rainforest. *New Phytologist* **187,** 579-591.

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9,** 772.

de Candolle, A.P. (1825) *Prodromus Systematis Naturalis Regni Vegetabilis 2*.

de la Estrella, M., Forest, F., Wieringa, J.J., Fougere-Danezan, M. & Bruneau, A. (2017) Insights on the evolutionary origin of Detaroideae, a clade of ecologically dominant tropical African trees. *New Phytologist* **214,** 1722-1735.

De Queiroz, A. (2005) The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology & Evolution* **20,** 68-73.

Dexter, K.G., Lavin, M., Torke, B.M., Twyford, A.D., Kursar, T.A., Coley, P.D., Drake, C., Hollands, R. & Pennington, R.T. (2017) Dispersal assembly of rain forest tree communities across the Amazon

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

basin. *Proceedings of the National Academy of Sciences of the United States of America* **114,** 2645-2650.

Dos Reis, M. & Yang, Z. (2013) MCMCTree tutorials.

Doyle, J.J. & Doyle, J.L. (1987) Genomic plant DNA preparation from fresh tissue- CTAB method. *Phytochemical Bulletin* **19,** 11-15.

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

Eiserhardt, W.L., Svenning, J., Kissling, W.D. & Balslev, H. (2011) Geographical ecology of the palms (Arecaceae): determinants of diversity and distributions across spatial scales. *Annals of Botany* **108,** 1391-1416.

Fine, P.V., Mesones, I. & Coley, P.D. (2004) Herbivores promote habitat specialization by trees in Amazonian forests. *Science* **305,** 663-665.

Fine, P.V., Daly, D.C., Muñoz, G.V., Mesones, I. & Cameron, K.M. (2005) The contribution of edaphic heterogeneity to the evolution and diversity of Burseraceae trees in the western Amazon. *Evolution* **59,** 1464-1478.

Garcia-R, J.C. & Matzke, N.J. (2021) Trait-dependent dispersal in rails (Aves: Rallidae): Historical biogeography of a cosmopolitan bird clade. *Molecular phylogenetics and evolution* **159,** 107106.

Gentry, A.H. (1988) Tree species richness of upper Amazonian forests. *Proceedings of the National Academy of Sciences* **85,** 156-159.

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,** 47-56.

Graham, A. (1995) Diversification of Gulf/Caribbean mangrove communities through Cenozoic time. *Biotropica* 20-27.

Harris, A.J., Ickert□Bond, S. & Rodríguez, A. (2018) Long distance dispersal in the assembly of floras: A review of progress and prospects in North America. *Journal of Systematics and Evolution* **56,** 430-448.

Hart, M.L., Forrest, L.L., Nicholls, J.A. & Kidner, C.A. (2016) Retrieval of hundreds of nuclear loci from herbarium specimens. *Taxon* **65,** 1081-1092.

Herendeen, P.S. (1992) The fossil history of the Leguminosae: phylogenetic and biogeographic implications. *Advances in Legume Systematics, part 4. The fossil record*(ed. by Anonymous), pp. 303-316. Royal Botanic Gardens.

Ho, S.Y. & Phillips, M.J. (2009) Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology* **58(3),** 367-380.

Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., Sanmartín, I., Sanchez-Meseguer, A., Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., Negri, F.R., Hooghiemstra, H., Lundberg, J., Stadler, T., Sarkinen, T. & Antonelli, A. (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330,** 927-931.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hughes, C., Pennington, R.T. & Antonelli, A. (2013) Neotropical plant evolution: assembling the big picture. *Botanical Journal of the Linnean Society* **171,** 1-18.

IPNI. (2020) The international plant names index. **2019,** .

Ireland, H.E., Kite, G.C., Veitch, N.C., Chase, M.W., Schrire, B., Lavin, M., Linares, J. & Pennington, R.T. (2010) Biogeographical, ecological and morphological structure in a phylogenetic analysis of Ateleia (Swartzieae, Fabaceae) derived from combined molecular, morphological and chemical data. *Botanical Journal of the Linnean Society* **162,** 39-53.

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J. & Wickett, N.J. (2016) HybPiper: Extracting coding sequence and introns for phylogenetics from high□throughput sequencing reads using target enrichment. *Applications in Plant Sciences* **4,** 1600016.

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. (2019) A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* **68,** 594-606.

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

Keeley, J.E. & Rundel, P.W. (2005) Fire and the Miocene expansion of C4 grasslands. *Ecology Letters* **8,** 683-690.

Klaus, K.V. & Matzke, N.J. (2020) Statistical Comparison of Trait-Dependent Biogeographical Models Indicates That Podocarpaceae Dispersal Is Influenced by Both Seed Cone Traits and Geographical Distance. *Systematic Biology* **69,** 61-75.

Klitgård, B.B. & Lavin, M. (2005) Tribe Dalbergieae sens. lat. *Legumes of the world*(ed. by G.P. Lewis, B. Schrire, B. Mackinder & M. Lock), pp. 307-335. Kew Publishing, Richmond, Surrey.

Klitgård, B.B., de Queiroz, L.P. & Lewis, G.P. (2000) A remarkable new species of Pterocarpus (Leguminosae: Papilionoideae: Dalbergieae) from Bahia, Brazil. *Kew Bulletin* 989-992.

Klitgård, B.B. (2005) Platymiscium (Leguminosae: Dalbergieae): Biogeography Systematics, Morphology, Taxonomy and Uses. *Kew Bulletin* 321-400.

Klitgård, B.B., Forest, F., Booth, T.J. & Saslis-Lagoudakis, C.H. (2013) A detailed investigation of the Pterocarpus clade (Leguminosae: Dalbergieae): Etaballia with radially symmetrical flowers is nested within the papilionoid-flowered Pterocarpus. *South African Journal of Botany* **89,** 128-142.

Koenen, E.J., Clarkson, J.J., Pennington, T.D. & Chatrou, L.W. (2015) Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytologist* **207,** 327-339.

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,** 789-804.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Larridon, I., Villaverde, T., Zuntini, A.R., Pokorny, L., Brewer, G.E., Epitawalage, N., Fairlie, I., Hahn, M., Kim, J. & Maguilla, E. (2020) Tackling rapid radiations with targeted sequencing. *Frontiers in plant science* **10,** 1655.

Lewis, G.P. (1987) *Legumes of Bahia*Kew Publishing, Richmond, Surrey.

Lewis, G.P., Schrire, B., Mackinder, B. & Lock, M. (2005) *Legumes of the World*Kew Publishing, Richmond, Surrey.

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J. & Turner, D.J. (2010) Target-enrichment strategies for next-generation sequencing. *Nature methods* **7,** 111-118.

Martínez, C. (2018) Dalbergieae (Fabaceae) Samara Fruits from the Late Eocene of Colombia. *International journal of plant sciences* **179,** 541-553.

Matzke, N.J. (2013) BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts.

Miller, M.A., Pfeiffer, W. & Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Environments Workshop (GCE)* 1-8.

Morley, R. (2000) *Origin and evolution of tropical rain forests.*John Wiley & Sons.

Morley, R. & Richards, K. (1993) Gramineae cuticle: a key indicator of Late Cenozoic climatic change in the Niger Delta. *Review of palaeobotany and palynology* **77,** 119-127.

Murphy, B., Forest, F., Barraclough, T., Rosindell, J., Bellot, S., Cowan, R., Golos, M., Jebb, M. & Cheek, M. (2020) A phylogenomic analysis of *Nepenthes* (Nepenthaceae). *Molecular phylogenetics and evolution* **144,** 106668.

Naciri, Y. & Linder, H.P. (2015) Species delimitation and relationships: the dance of the seven veils. *Taxon* **64,** 3-16.

Nie, Z., Sun, H., Manchester, S.R., Meng, Y., Luke, Q. & Wen, J. (2012) Evolution of the intercontinental disjunctions in six continents in the Ampelopsis clade of the grape family (Vitaceae). *BMC evolutionary biology* **12,** 1-13.

Ogutcen, E., Christe, C., Nishii, K., Salamin, N., Möller, M. & Perret, M. (2021) Phylogenomics of Gesneriaceae using targeted capture of nuclear genes. *Molecular phylogenetics and evolution* **157,** 107068.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E. & Morrison, J.C. (2001) Terrestrial Ecoregions of the World: A New Map of Life on EarthA new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *Bioscience* **51,** 933-938.

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,** 1874-1888.

Paradis, E. & Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35,** 526-528.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Pennington, R.T. & Lavin, M. (2016) The contrasting nature of woody plant species in different neotropical forest biomes reflects differences in ecological stability. *New Phytologist* **210,** 25-37.

Pennington, R.T. & Dick, C.W. (2010) Diversification of the Amazonian flora and its relation to key geological and environmental events: a molecular perspective. *Amazonia, Landscape and Species evolution*(ed. by C. Hoorn & F.P. Wesselingh)Blackwell Publishing, Oxford, UK.

Pennington, R.T. & Ratter, J.A. (2006) *Neotropical savannas and seasonally dry forests: plant diversity, biogeography, and conservation*CRC press.

Pennington, R.T. & Dick, C.W. (2004) The role of immigrants in the assembly of the South American rainforest tree flora. *Philosophical Transactions of the Royal Society of London Series B: Biological sciences* **359,** 1611-1622.

Pennington, R.T., Lavin, M. & Oliveira-Filho, A. (2009) Woody plant diversity, evolution, and ecology in the tropics: perspectives from seasonally dry tropical forests. *Annual Review of Ecology, Evolution, and Systematics* **40,** 437-457.

Puttick, M.N. (2019) MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. *Bioinformatics* **35,** 5321-5322.

Python Software Foundation. (2010) Python Language Reference, version 2.7. *http://www.python.org* .

R Development Core Team. (2013) R: A language and environment for statistical computing. Available at: http://www.R-project.org/.

Rabosky, D.L., Donnellan, S.C., Grundler, M. & Lovette, I.J. (2014) Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Systematic Biology* **63,** 610-627.

Rambaut, A. (2014) FigTree, a graphical viewer of phylogenetic trees (tree.bio.ed.ac.uk/software/figtree). **1.4.2,** .

Rambaut, A., Suchard, M.A., Xie, D. & Drummond, A.J. (2015) Tracer. *http://tree.bio.ed.ac.uk/software/tracer/* **1.6,** .

Ree, R.H. & Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* **57,** 4-14.

Renner, S. (2004) Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Sciences* **165,** S23-S33.

Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3,** 217-223.

Rivera□Ocasio, E., Aide, T.M. & McMillan, W.O. (2002) Patterns of genetic diversity and biogeographical history of the tropical wetland tree, Pterocarpus officinalis (Jacq.), in the Caribbean basin. *Molecular ecology* **11,** 675-683.

Rojo, J.P. (1972) *Pterocarpus (Leguminosae - Papilionaceae): Revised for the world.*Cramer, Lehre, Germany.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Rommerskirchen, F., Condon, T., Mollenhauer, G., Dupont, L. & Schefuss, E. (2011) Miocene to Pliocene development of surface and subsurface temperatures in the Benguela Current system. *Paleoceanography* **26,** .

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

Sanmartín, I. (2014) Biogeography. *The Tree of Life, eds P.Vargas and R.Zardoya (Sunderland, MA: Sinauer Associates, Inc.)* 555-576.

Sanmartín, I. (2012) Historical Biogeography: Evolution in Time and Space. *Evolution education & outreach* **5,** 555-568.

Sanmartín, I. & Ronquist, F. (2004) Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic Biology* 216-243.

Saslis-Lagoudakis, C.H., Klitgaard, B.B., Forest, F., Francis, L., Savolainen, V., Williamson, E.M. & Hawkins, J.A. (2011) The use of phylogeny to interpret cross-cultural patterns in plant use and guide medicinal plant discovery: an example from Pterocarpus (Leguminosae). *PloS one* **6,** e22275.

Schley, R.J., de la Estrella, M., Pérez-Escobar, O.A., Bruneau, A., Barraclough, T., Forest, F. & Klitgård, B. (2018) Is Amazonia a 'museum' for Neotropical trees? The evolution of the Brownea clade (Detarioideae, Leguminosae). *Molecular Phylogenetics and Evolution* **126,** 279-292.

Schley, R.J., Pennington, R.T., Pérez□Escobar, O.A., Helmstetter, A.J., de la Estrella, M., Larridon, I., Sabino Kikuchi, Izai Alberto Bruno, Barraclough, T.G., Forest, F. & Klitgård, B. (2020) Introgression across evolutionary scales suggests reticulation contributes to Amazonian tree diversity. *Molecular ecology* **29,** 4170-4185.

Schrire, B.D., Lavin, M. & Lewis, G.P. (2005) Global distribution patterns of the Leguminosae: insights from recent phylogenies. 375-422.

Scotese, C.R. (2004) A continental drift flipbook. *The Journal of geology* **112,** 729-741.

Senut, B., Pickford, M. & Ségalen, L. (2009) Neogene desertification of Africa. *Comptes Rendus Geoscience* **341,** 591-602.

Staats, M., Erkens, R.H., van de Vossenberg, B., Wieringa, J.J., Kraaijeveld, K., Stielow, B., Geml, J., Richardson, J.E. & Bakker, F.T. (2013) Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PloS one* **8,** e69189.

Stamatakis, A. (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **Open access.,** Published online: 10.1093/bioinformatics/btu033.

ter Steege, H., Prado, P.I., de Lima, R.A., Pos, E., de Souza Coelho, L., de Andrade Lima Filho, Diogenes, Salomão, R.P., Amaral, I.L., de Almeida Matos, Francisca Dionízia & Castilho, C.V. (2020) Biased-corrected richness estimates for the Amazonian tree flora. *Scientific reports* **10,** 1-13.

ter Steege, H., Pitman, N.C., Sabatier, D., Baraloto, C., Salomao, R.P., Guevara, J.E., Phillips, O.L., Castilho, C.V., Magnusson, W.E., Molino, J.F., Monteagudo, A., Nunez Vargas, P., Montero, J.C., Feldpausch, T.R., Coronado, E.N., Killeen, T.J., Mostacedo, B., Vasquez, R., Assis, R.L., Terborgh, J., Wittmann, F., Andrade, A., Laurance, W.F., Laurance, S.G., Marimon, B.S., Marimon, B.H.,Jr, Guimaraes Vieira, I.C., Amaral, I.L., Brienen, R., Castellanos, H., Cardenas Lopez, D.,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Duivenvoorden, J.F., Mogollon, H.F., Matos, F.D., Davila, N., Garcia-Villacorta, R., Stevenson Diaz, P.R., Costa, F., Emilio, T., Levis, C., Schietti, J., Souza, P., Alonso, A., Dallmeier, F., Montoya, A.J., Fernandez Piedade, M.T., Araujo-Murakami, A., Arroyo, L., Gribel, R., Fine, P.V., Peres, C.A., Toledo, M., Aymard, C.G.A., Baker, T.R., Ceron, C., Engel, J., Henkel, T.W., Maas, P., Petronelli, P., Stropp, J., Zartman, C.E., Daly, D., Neill, D., Silveira, M., Paredes, M.R., Chave, J., Lima Filho Dde, A., Jorgensen, P.M., Fuentes, A., Schongart, J., Cornejo Valverde, F., Di Fiore, A., Jimenez, E.M., Penuela Mora, M.C., Phillips, J.F., Rivas, G., van Andel, T.R., von Hildebrand, P., Hoffman, B., Zent, E.L., Malhi, Y., Prieto, A., Rudas, A., Ruschell, A.R., Silva, N., Vos, V., Zent, S., Oliveira, A.A., Schutz, A.C., Gonzales, T., Trindade Nascimento, M., Ramirez-Angulo, H., Sierra, R., Tirado, M., Umana Medina, M.N., van der Heijden, G., Vela, C.I., Vilanova Torre, E., Vriesendorp, C., Wang, O., Young, K.R., Baider, C., Balslev, H., Ferreira, C., Mesones, I., Torres-Lezama, A., Urrego Giraldo, L.E., Zagt, R., Alexiades, M.N., Hernandez, L., Huamantupa-Chuquimaco, I., Milliken, W., Palacios Cuenca, W., Pauletto, D., Valderrama Sandoval, E., Valenzuela Gamarra, L., Dexter, K.G., Feeley, K., Lopez-Gonzalez, G. & Silman, M.R. (2013) Hyperdominance in the Amazonian tree flora. *Science* **342,** 1243092.

Tsang, S.M., Wiantoro, S., Veluz, M.J., Sugita, N., Nguyen, Y., Simmons, N.B. & Lohman, D.J. (2020) Dispersal out of Wallacea spurs diversification of Pteropus flying foxes, the world's largest bats (Mammalia: Chiroptera). *Journal of Biogeography* **47,** 527-537.

van der Pijl, L. (1982) Ecological Developments in Leguminous Fruits. *Principles of Dispersal in Higher Plants*(ed. by Anonymous), pp. 150-159. Springer.

Vatanparast, M., Klitgård, B.B., Adema, F.A., Pennington, R.T., Yahara, T. & Kajita, T. (2013) First molecular phylogeny of the pantropical genus Dalbergia: implications for infrageneric circumscription and biogeography. *South African Journal of Botany* **89,** 143-149.

Ward, M., Dick, C.W., Gribel, R. & Lowe, A.J. (2005) To self, or not to self… A review of outcrossing and pollen-mediated gene flow in neotropical trees. *Heredity* **95,** 246.

Wolfe, K.H., Li, W.H. & Sharp, P.M. (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America* **84,** 9054-9058.

Woodward, F.I., Lomas, M.R. & Kelly, C.K. (2004) Global climate and the distribution of plant biomes. *Philosophical Transactions of the Royal Society of London.Series B: Biological Sciences* **359,** 1465-1476.

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24,** 1586-1591.

Zachos, J., Dickens, G.R. & Zeebe, R.E. (2008) An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451,** 279-283.

Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292,** 686-693.

Zamora, N. (2000) Nuevas especies y combinaciones en Leguminosas de Mesoamérica. *Novon* 175-180.

Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19,** 153.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.444160; this version posted May 16, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Zhang, Z., Ramstein, G., Schuster, M., Li, C., Contoux, C. & Yan, Q. (2014) Aridification of the Sahara desert caused by Tethys Sea shrinkage during the Late Miocene. *Nature* **513,** 401-404.

Zizka, A., Carvalho Sobrinho, J.G., Pennington, R.T., Queiroz, L.P., Alcantara, S., Baum, D.A., Bacon, C.D. & Antonelli, A. (2020) Transitions between biomes are common and directional in Bombacoideae (Malvaceae). *Journal of Biogeography* **47,** 1310-1321.

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M. & Scharn, R. (2019) CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* **10,** 744-751.

Significance statement

Dispersal involves organisms spreading from one location to another, and is one of the major processes defining the distribution of organisms. Sometimes, structures which evolved for local dispersal may lead to rare, long-distance dispersal, and this may allow species to colonise new landmasses and diversify in a new environment. As such, here we build an evolutionary tree to understand relationships between rosewood species (*Pterocarpus*), and test whether dispersal traits have impacted their evolution. We found that dispersal and climate change in the Miocene likely impacted the modern distribution of *Pterocarpus,* but that this was not significantly explained by dispersal structures.

Biosketch

Rowan Schley is a postdoctoral researcher interested in the evolution, biogeography and conservation of tropical biodiversity. His research is broadly centred around speciation with a particular focus on the effects of introgression, using phylogenomic and population genomic approaches to address these questions in tropical trees.

R.J.S. conceived the hypotheses, led data analysis and wrote the manuscript. M.Q. and P.M. led the phylogenomic data generation. M.V., M.d.l.E. and G.L. provided many comments and background information on the Legume family for the manuscript. B.K. helped conceive the study and led sample collection. All authors contributed to editing the manuscript.