



OPEN ACCESS

EDITED BY

Carl J. Rothfels,
University of California, Berkeley,
United States

REVIEWED BY

Joyce Gloria Onyenedum,
Cornell University, United States
Laura Frost,
University of South Alabama, United States

*CORRESPONDENCE

Elizabeth M. Joyce
✉ E.Joyce@lmu.de

†These authors share senior authorship

SPECIALTY SECTION

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

RECEIVED 06 October 2022

ACCEPTED 31 January 2023

PUBLISHED 07 March 2023

CITATION

Joyce EM, Appelhans MS, Buerki S,
Cheek M, de Vos JM, Pirani JR, Zuntini AR,
Bachelier JB, Bayly MJ, Callmander MW,
Devecchi MF, Pell SK, Groppo M,
Lowry PP II, Mitchell J, Siniscalchi CM,
Munzinger J, Orel HK, Pannell CM,
Nauheimer L, Sauquet H, Weeks A,
Muellner-Riehl AN, Leitch IJ, Maurin O,
Forest F, Nargar K, Thiele KR, Baker WJ and
Crayn DM (2023) Phylogenomic analyses
of Sapindales support new family
relationships, rapid Mid-Cretaceous
Hothouse diversification, and
heterogeneous histories of
gene duplication.
Front. Plant Sci. 14:1063174.
doi: 10.3389/fpls.2023.1063174

COPYRIGHT

© 2023 Joyce, Appelhans, Buerki, Cheek,
de Vos, Pirani, Zuntini, Bachelier, Bayly,
Callmander, Devecchi, Pell, Groppo, Lowry,
Mitchell, Siniscalchi, Munzinger, Orel, Pannell,
Nauheimer, Sauquet, Weeks, Muellner-Riehl,
Leitch, Maurin, Forest, Nargar, Thiele, Baker
and Crayn. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Phylogenomic analyses of Sapindales support new family relationships, rapid Mid-Cretaceous Hothouse diversification, and heterogeneous histories of gene duplication

Elizabeth M. Joyce^{1,2,3*}, Marc S. Appelhans^{4,5}, Sven Buerki⁶,
Martin Cheek⁷, Jurriaan M. de Vos⁸, José R. Pirani⁹,
Alexandre R. Zuntini⁷, Julien B. Bachelier¹⁰, Michael J. Bayly¹¹,
Martin W. Callmander¹², Marcelo F. Devecchi⁹, Susan K. Pell¹³,
Milton Groppo⁹, Porter P. Lowry II^{14,15}, John Mitchell¹⁶,
Carolina M. Siniscalchi¹⁷, Jérôme Munzinger¹⁸, Harvey K. Orel¹¹,
Caroline M. Pannell^{7,19,20}, Lars Nauheimer^{2,3}, Hervé Sauquet²¹,
Andrea Weeks²², Alexandra N. Muellner-Riehl^{23,24}, Ilia J. Leitch⁷,
Olivier Maurin⁷, Félix Forest⁷, Katharina Nargar^{3,25},
Kevin R. Thiele²⁶, William J. Baker^{7†} and Darren M. Crayn^{2,3†}

¹Systematics, Biodiversity and Evolution of Plants, Ludwig-Maximilians-Universität München, Munich, Germany, ²College of Science and Engineering, James Cook University, Cairns, QLD, Australia, ³Australian Tropical Herbarium, James Cook University, Cairns, QLD, Australia, ⁴Department of Systematics, Biodiversity and Evolution of Plants, University of Göttingen, Göttingen, Germany, ⁵Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, DC, United States, ⁶Department of Biological Sciences, Boise State University, Boise, ID, United States, ⁷Royal Botanic Gardens, Kew, Richmond, United Kingdom, ⁸Department of Environmental Sciences, University Basel, Basel, Switzerland, ⁹Departamento de Botânica, Universidade de São Paulo, Herbario SPF, São Paulo, Brazil, ¹⁰Institut für Biologie, Freie Universität Berlin, Berlin, Germany, ¹¹School of BioSciences, The University of Melbourne, Parkville, VIC, Australia, ¹²Conservatoire et Jardin botaniques de la Ville de Genève, Geneva, Switzerland, ¹³United States Botanic Garden, Washington, DC, United States, ¹⁴Missouri Botanical Garden, St. Louis, MO, United States, ¹⁵Institut de Systématique, Évolution, et Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, Centre National de la Recherche Scientifique, Sorbonne Université, École Pratique des Hautes Études, Université des Antilles, Paris, France, ¹⁶New York Botanical Garden, New York, NY, United States, ¹⁷Department of Biological Sciences, Harned Hall, Mississippi State University, Mississippi State, MS, United States, ¹⁸AMAP, Université Montpellier, Institut de Recherche pour le Développement (IRD), Centre de coopération internationale en recherche agronomique pour le développement (CIRAD), Centre National de la Recherche Scientifique (CNRS), Institut national de la recherche agronomique (INRAE), Montpellier, France, ¹⁹Department of Biology, Oxford University, Oxford, United Kingdom, ²⁰Marine Laboratory, Queen's University Belfast, Portaferry, United Kingdom, ²¹National Herbarium of New South Wales (NSW), Royal Botanic Gardens and Domain Trust, Sydney, NSW, Australia, ²²Department of Biology, George Mason University, Fairfax, VA, United States, ²³Department of Molecular Evolution and Plant Systematics & Herbarium, Faculty of Life Sciences, University of Leipzig, Leipzig, Germany, ²⁴German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany, ²⁵National Research Collections Australia, Commonwealth Industrial and Scientific Research Organization (CSIRO), Canberra, ACT, Australia, ²⁶School of Biological Sciences, University of Western Australia, Perth, WA, Australia

Sapindales is an angiosperm order of high economic and ecological value comprising nine families, c. 479 genera, and c. 6570 species. However, family and subfamily relationships in Sapindales remain unclear, making reconstruction of the order's spatio-temporal and morphological evolution difficult. In this study, we used Angiosperms353 target capture data to generate the most densely sampled phylogenetic trees of Sapindales to date, with 448 samples and c. 85% of genera represented. The percentage of paralogous loci and allele divergence was characterized across the phylogeny, which was time-calibrated using 29 rigorously assessed fossil calibrations. All families were supported as monophyletic. Two core family clades subdivide the order, the first comprising Kirkiaceae, Burseraceae, and Anacardiaceae, the second comprising Simaroubaceae, Meliaceae, and Rutaceae. Kirkiaceae is sister to Burseraceae and Anacardiaceae, and, contrary to current understanding, Simaroubaceae is sister to Meliaceae and Rutaceae. Sapindaceae is placed with Nitrariaceae and Biebersteiniaceae as sister to the core Sapindales families, but the relationships between these families remain unclear, likely due to their rapid and ancient diversification. Sapindales families emerged in rapid succession, coincident with the climatic change of the Mid-Cretaceous Hothouse event. Subfamily and tribal relationships within the major families need revision, particularly in Sapindaceae, Rutaceae and Meliaceae. Much of the difficulty in reconstructing relationships at this level may be caused by the prevalence of paralogous loci, particularly in Meliaceae and Rutaceae, that are likely indicative of ancient gene duplication events such as hybridization and polyploidization playing a role in the evolutionary history of these families. This study provides key insights into factors that may affect phylogenetic reconstructions in Sapindales across multiple scales, and provides a state-of-the-art phylogenetic framework for further research.

KEYWORDS

Cenomanian-Turonian Thermal Maximum, phylogenomics, target enrichment, sequence capture, HybSeq, paralogy

Introduction

Sapindales is a flowering plant order of great biological and economic importance; it includes c. 2% of the world's angiosperm diversity, and in 2021, raw products from its taxa were estimated to be worth more than US\$31 billion p.a. (c. 0.2% of the world trade market; [Freiberg et al., 2020](#) onwards; [Stevens, 2001](#); [Simoes and Hidalgo, 2011](#); [Simoes and Hidalgo, 2022](#) onwards). Sapindales currently includes six medium-sized to large families (Anacardiaceae, Burseraceae, Meliaceae, Rutaceae, Sapindaceae, and Simaroubaceae, all with >150 species) and three small families (Nitrariaceae, and the monogeneric Biebersteiniaceae and Kirkiaceae), with c. 479 genera and c. 6750 species. Its species are predominantly tropical woody plants with pinnately compound leaves and small, tetra- or pentamerous flowers with intrastaminal nectar disks. However, remarkable morphological and ecological diversity exists within Sapindales, with species presenting as herbs, lianas, shrubs, trees and mangroves that inhabit tropical, arid, coastal, or montane environments. Taxa such as mangoes, citrus, mahoganies, cashews, maples, pistachio, lychee, frankincense, and myrrh are important to agricultural, pharmaceutical, cosmetic, chemical, and timber industries, and contribute to the high economic value of the order.

Despite its biological and economic significance, Sapindales has a complex taxonomic history, and relationships among families within the order are uncertain. From the nineteenth century, families now placed in Sapindales were variously assigned to 25 different orders. In the 20th Century, two main ordinal concepts persisted, with [Wettstein \(1901\)](#) and [Cronquist \(1968\)](#) both recognizing an expanded order including Rurales + Sapindales, and [Takhtajan \(2009\)](#) assigning families to separate orders (i.e., Sapindales, Rurales and Zygophyllales). More recently, molecular studies supported the expanded ordinal concept, suggesting that Anacardiaceae, Biebersteiniaceae, Burseraceae, Kirkiaceae, Nitrariaceae, Meliaceae, Rutaceae, Sapindaceae, and Simaroubaceae form a monophyletic clade distinct from Zygophyllales ([Chase et al., 1993](#); [Gadek et al., 1996](#); [Muellner et al., 2007](#); [Muellner-Riehl et al., 2016](#); [Li et al., 2019](#); [Li et al., 2021](#)). The inclusion of these nine families in Sapindales is now generally accepted ([Kubitzki, 2011](#); [APG, 2016](#)). Angiosperm-wide molecular studies differ in their placement of Sapindales within the rosids, but most studies suggest Sapindales is most closely related to Malvales, Brassicales, Huerteales, and Picramniales ([Chase et al., 1993](#); [APG, 2016](#); [Li et al., 2019](#); [Ramírez-Barahona et al., 2020](#); [Li et al., 2021](#); [Baker et al., 2022](#)).

Despite extensive systematic research on the order, the relationships of most families within Sapindales remain uncertain (Gadek et al., 1996; Stevens, 2001; Muellner et al., 2007; Wang et al., 2009; Muellner-Riehl et al., 2016; Lin et al., 2018; Li et al., 2019). A close relationship of Burseraceae and Anacardiaceae is well-established, with both families previously included in the family Terebinthaceae and sharing the synapomorphies of vertical intercellular secretory canals in the primary and secondary phloem and the ability to synthesise biflavonyls (Wannan et al., 1985; Wannan, 1986; Wannan and Quinn, 1990; Wannan and Quinn, 1991; Terrazas, 1994). More recent molecular studies support Anacardiaceae and Burseraceae as monophyletic, although infra-familial classifications are in need of revision (Gadek et al., 1996; Pell, 2004; Weeks et al., 2014; Muellner-Riehl et al., 2016; Mitchell et al., 2022). Morphological studies have shown a close affinity in floral structure of members of the monogeneric family Kirkiaceae to Anacardiaceae and Burseraceae (Bachelier and Endress, 2008), and, together, these families form a moderately to well-supported clade in recent molecular studies (Muellner et al., 2007; Muellner-Riehl et al., 2016; Li et al., 2021). However, the relationships of Biebersteiniaceae and Nitrariaceae to the rest of the order, the position of Sapindaceae, and the relationships between Rutaceae, Meliaceae, and Simaroubaceae remain less clear. Nitrariaceae and Biebersteiniaceae are usually retrieved sequentially as sister to the other families in the order (Muellner et al., 2007; Appelhans et al., 2012; Muellner-Riehl et al., 2016; Li et al., 2019; Li et al., 2021), but Ramírez-Barahona et al. (2020) placed them together as sister to a clade comprising Kirkiaceae, Burseraceae, and Anacardiaceae. Regardless of their position within the order, the node between Nitrariaceae and Biebersteiniaceae has remained unsupported in multiple studies (Muellner-Riehl et al., 2016; Li et al., 2019; Ramírez-Barahona et al., 2020; Li et al., 2021). The position of Sapindaceae within the order also remains unresolved, being variously reconstructed as sister to a clade containing Rutaceae, Simaroubaceae, and Meliaceae (Appelhans et al., 2012; Muellner-Riehl et al., 2016; Li et al., 2019; Ramírez-Barahona et al., 2020; Li et al., 2021), as sister to Anacardiaceae and Burseraceae (Gadek et al., 1996; Lin et al., 2018), as sister to Anacardiaceae (Chase et al., 1993), or as a clade in a polytomy (Muellner et al., 2007); in all cases, the family relationships of Sapindaceae are poorly supported. Finally, the consensus of morphological and molecular evidence indicates that Rutaceae, Simaroubaceae, and Meliaceae form a clade within Sapindales; however, the relationships between these three families remain unclear. Many studies have found high support for Meliaceae being sister to Simaroubaceae and Rutaceae (Gadek et al., 1996; Lin et al., 2018; Li et al., 2019; Ramírez-Barahona et al., 2020; Li et al., 2021), but strong contradictory evidence suggests Rutaceae is sister to Simaroubaceae and Meliaceae (Muellner et al., 2007; Appelhans et al., 2012; Muellner-Riehl et al., 2016).

Resolution of family relationships within Sapindales is critical for understanding the evolutionary history of the order. This is particularly pertinent for understanding the development and evolution of unique traits with ecological and commercial significance such as wood characters (e.g. Pace et al., 2022), secondary metabolite synthesis (e.g. Fernandes da Silva et al., 2022),

flower morphology (e.g. Bachelier and Endress, 2008; Bachelier et al., 2011; Alves et al., 2022), pollen morphology (e.g. Gonçalves-Esteves et al., 2022), secretory structures (e.g. Tölke et al., 2022), and cuticular chemical composition (e.g. Roma and Santos, 2022). Furthermore, an understanding of the evolution of the vast variation in nuclear DNA organization and ploidy levels within the order requires a robust, detailed phylogenetic framework (Pennington and Styles, 1975; Guimarães and Forni-Martins, 2022). Likewise, the spatio-temporal origins of the order can only be understood once relationships within it have been resolved. Differences in tree topology have also likely contributed to discrepancies in divergence ages estimated for families in previous phylogenetic work, with some studies reporting a Cretaceous origin for most Sapindales families (stem and crown nodes; Muellner-Riehl et al., 2016; Ramírez-Barahona et al., 2020), but others retrieving a Cenozoic origin for families (Li et al., 2019).

Recent advances in high-throughput sequencing methods provide new opportunities for resolving the familial relationships within Sapindales. Target capture sequencing has become the foremost high-throughput sequencing method for phylogenomics, enabling the reliable retrieval of hundreds or thousands of target loci at an increasingly affordable price (Cronn et al., 2012; Barrett et al., 2016; Bragg et al., 2016). The amount of data generated with target capture sequencing in combination with the development of universal bait kits such as Angiosperms353 (Johnson et al., 2019) has facilitated global efforts to resolve relationships of plants across multiple taxonomic scales (Baker et al., 2021; McDonnell et al., 2021; Baker et al., 2022). In addition, unlike historical Sanger approaches, the sequencing of a high number of reads in target capture approaches allows for the detection and handling of paralogous genes. Paralogous genes are genes with multiple copies that are the product of duplication of an ancestral gene (either by duplication of part of the genome, or of the whole genome) (Fitch, 1970). Duplication of genes can also be produced in the process of allopolyploidization, whereby the hybridization of two species results in the doubling of the genome; as these gene copies do not share a common ancestor they are technically called homeologs, but for the purposes of this paper we do not distinguish between paralogs and homeologs and refer to all loci with multiple copies resulting from duplication as paralogous. Angiosperm genomes often contain a large number of paralogous genes due to the prevalence of polyploidy or whole-genome duplication events in the evolutionary history of plants (Soltis et al., 2009; Jiao et al., 2011; Panchy et al., 2016). Although target capture bait kits such as Angiosperms353 are designed to target low- or single-copy loci, paralogous copies of targeted loci are present in many lineages (Nauheimer et al., 2021; Smith and Hahn, 2021; Ufimov et al., 2022). Paralogous loci can violate the assumption of homology in phylogenetic analysis and confound resulting trees, and so are commonly removed in analyses (e.g. Jones et al., 2019; Larridon et al., 2020). However, the retention and identification of paralogous loci in phylogenetic studies has been shown to be highly valuable for maximizing the amount of informative data, explaining discordance between gene trees, and in pinpointing where genome duplication events such as ancient polyploidization and hybridization have played an important role in the evolution of lineages (e.g. Nauheimer et al., 2021; Morales-Briones et al., 2021; Smith and Hahn, 2021; Ufimov

et al., 2022). Identification, characterization and analysis of paralogy is now possible with target capture sequencing, making it a promising method for improving the resolution of relationships within Sapindales, and for gaining new insight into the role of gene duplication events during the evolution of the order.

In this study, we have achieved the most comprehensive sampling of Sapindales species in a phylogenetic study to date, and use target capture sequencing with the Angiosperms353 bait kit to infer family and subfamily relationships. We characterize patterns of paralogy across the order to investigate whether gene duplication events (whether through hybridization, autopolyploidization, or local gene duplications) have played a role in the evolution of Sapindales lineages and may explain any topological uncertainty. Finally, we go on to infer the temporal evolution of Sapindales, identifying key periods for the evolution of the order and assessing how these change when different crown ages for the angiosperms are assumed. The resulting phylogeny aims to improve our understanding of the order's evolutionary history, and to serve as a robust framework for future phylogenetic, morphological, taxonomic, and systematic studies.

Methods

Sampling

A total of 472 samples were obtained for this analysis, including 448 representatives of Sapindales from all nine families and encompassing c. 85% of genera in the order (Supplementary Material 1). Generally, one sample per genus was included (where possible, the type species for the genus); multiple species were sampled for genera that were suspected to be polyphyletic based on previous studies and expert opinion. The outgroup comprised 24 samples from across the Pentapetales, from the orders Brassicales, Crossosomatales, Ericales, Fabales, Geraniales, Huerteales, Malvales, and Myrtales (Supplementary Material 1). Data for 287 samples were newly generated for this study, sourced from the living collections of the Royal Botanic Gardens, Kew, silica gel-dried field collections, herbarium specimens from multiple institutions, and the DNA banks of the Royal Botanic Gardens, Kew, Australian Tropical Herbarium, United States Botanical Gardens, and Göttingen University (Supplementary Material 1). The dataset was augmented with Angiosperms353 data for 132 species produced for the Sapindaceae phylogeny of Buerki et al. (2021) and with data for 53 species downloaded from the Sequence Read Archive (SRA; NCBI, <https://www.ncbi.nlm.nih.gov/sra>, Supplementary Material 1).

DNA extraction and quality control

For the new data generated in this study, DNA was extracted from silica gel-dried and herbarium samples using the CTAB protocol of Doyle and Doyle (1987). The protocol was modified at the isopropanol precipitation step, with samples left to precipitate at -20°C degrees over 24 hours for silica-dried and fresh samples, and a minimum of 72 hours for herbarium samples. Extractions were cleaned using Agencourt AMPureXP beads (Beckman Coulter, Indianapolis, USA) according

to the manufacturer's protocol and eluted to 50 µL. DNA quality and quantity was ascertained using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Massachusetts, USA) and Quantus Fluorometer (Promega Corporation, Wisconsin, USA), and average fragment size assessed visually after electrophoresis on a 1% agarose gel. For extractions with a concentration of less than 4 ng/µL, yield was increased by combining additional DNA extractions from the same sample and concentrating using a vacuum centrifuge.

Library preparation and sequencing

Library preparation protocols varied depending on DNA quality. For higher-quality extractions (i.e., average fragment size > 350 bp), DNA was sonicated in an M220 Focused-ultrasonicator with microTUBES AFA Fiber Pre-slit Snap-caps (Covaris, Massachusetts, USA) following the manufacturer's protocol. Shearing time varied from 30–90 seconds depending on DNA fragment size profile to obtain an average fragment size of 350 bp. Highly degraded samples with an average fragment size <350 bp were not sonicated. Sonicated samples were diluted to 200 ng DNA in 50 µL Tris, and non-sonicated samples to 100 ng DNA in 25 µL Tris.

Dual-indexed libraries were prepared using the NEBNext Ultra II Library Preparation Kit and the NEBNext Multiplex Oligos for Illumina (New England BioLabs, Massachusetts, USA) using half the manufacturer's recommended volumes. Library size profiles were evaluated on a 4200 TapeStation System using High Sensitivity D1000 ScreenTapes (Agilent Technologies, California, USA), and library concentrations ascertained using a Quantus Fluorometer (Promega Corporation, Wisconsin, USA). All libraries were of an average fragment size of approximately 500 bp (including adapters). For libraries not meeting these standards, PCR, adaptor cleanup and/or size selection steps of the library preparation protocol were repeated. All libraries were normalized to a concentration of 10 nM and combined in 7.5 µL library pools with 20–24 samples per pool of similar fragment lengths.

Pooled libraries were enriched using the myBaits 'Angiosperms 353 v1' Target Sequence Capture Kit (Arbor Bioscience, Michigan, USA) following the manufacturer's protocol. Hybridizations were performed at 60 or 65°C (depending on average fragment length) for 24 hours in a Hybex Microsample Incubator (SciGene, California, USA) using the same volume as the hybridization reaction volume (usually 30 µL) of red Chill-out Liquid Wax (Bio224 Rad, Hercules, CA, USA) to prevent evaporation.

Enriched library pools were amplified with KAPA HiFi 2X HotStart ReadyMix PCR Kit (Roche, Basel, Switzerland) for 14 PCR cycles, and subsequently cleaned using Agencourt AMPure XP Beads. Concentrations of pools were quantified with a Quantus Fluorometer and quality and size profiling were conducted on a 4200 TapeStation System using High Sensitivity D1000 ScreenTapes (Agilent Technologies, California, USA). The hybridised pools were then combined into sequencing runs of approximately 96 libraries in 30 µL at 6 nM concentration. Library pools were multiplexed and sequenced by MacroGen (MacroGen Inc., Seoul, South Korea) on an Illumina HiSeq (Illumina Inc., California, USA) producing 2x 150 bp paired-end reads.

Gene retrieval

Trimmomatic was used to remove adapter sequences, poor-quality base calls and poor-quality reads from sequencing reads with the settings: `illuminaclip 2:30:10`, `leading 30`, `trailing 30`, `sliding window 4:2:30`, `minimum length 36` and `Phred-33` base quality encoding (Bolger et al., 2014). Exon sequences were assembled using the HybPiper v1 pipeline for nucleotide data (Johnson et al., 2016). Trimmed reads were mapped against the Sapindales subset of the mega target file, which led to a substantially higher recovery for Sapindales than with the standard Angiosperms353 target file (McLay et al., 2021). Exons and supercontigs were retrieved using the HybPiper script `retrieve_sequences.py`, and summary gene recovery statistics for each sample were generated for each sample with the HybPiper scripts `get_seq_lengths.py` and `hybpiper_stats.py`.

Data cleaning and paralogy characterization

We cleaned loci or samples where the multiple copies are uninformative (i.e., are likely due to contamination or laboratory error), and retained paralogous loci where the cause is more likely to be gene duplication events. This enabled maximum retention of data, and the identification of lineages in the phylogenetic tree where gene duplication events (such as polyploidization or hybridization) could have played a role. It is uncertain whether Angiosperms353 baits differ in their ability to hybridize with paralogous copies of genes across lineages at the ordinal scale, and theoretically, this could result in an underestimation of paralogy in cases where gene copies are highly divergent or lost. Nevertheless, we suggest that the characterization of paralogy is useful for gaining a general understanding of where genes have been duplicated and retained in Sapindales, how these patterns may differ across the order, and where gene duplication events may affect evolutionary inference from phylogenetic trees and morphology.

To detect, clean and characterize paralogy in our Sapindales target capture data, we used HybPhaser v1 (Nauheimer et al., 2021). HybPhaser v1 uses reference mapping and codes any discrepancies as ambiguity characters. This process enables the identification of single nucleotide polymorphisms (SNPs), which facilitates the characterization of signals of paralogy across the phylogenetic tree (whereby a high number of SNPs compared to related lineages can indicate the presence of multiple gene copies), allows the cleaning of sequences and samples with extremely high signals of paralogy (that are more likely to be due to contamination), and enables reconciliation of polymorphic sites as ambiguities for phylogenetic analysis (rather than consensus bases that may be called from the most common copy of a gene for any given locus). It is possible that the use of ambiguities in sequences could depress branch lengths relative to the use of a tree estimated with consensus sequences; however, we consider the use of ambiguities to be the most conservative and accurate way to code SNPs from paralogous loci, as it avoids the analysis of chimeric contigs assembled from reads of paralogous loci as orthologous loci, and places more weight on the regions of the locus that are conserved across copies within the same sample in the phylogenetic analysis.

Reads were remapped to the contig for each gene generated with the HybPhaser script `Generate_consensus_sequences.sh`. Information on length and coverage of sequences from all samples and loci was collated with the HybPhaser script `Rscript_1a_count_snps_in_consensus_seqs.R`. Following visual inspection of the outputs which summarised heterozygosity in the raw Angiosperms353 data and manual inspection of the remapped.bam files, HybPhaser scripts `R1b_optimize_dataset.R` and `Configure_1_SNPs_assessment.R` were used to clean the data by removing samples with >50% missing loci and loci with <30% locus recovery, >21% missing samples, as well as samples and loci with outlying heterozygosity (> 1.5x the inter-quartile range for heterozygosity, which we considered more likely to be contaminated). Tables of heterozygosity and allele divergence were collated with the script `Rscript_1c_summary_table.R`, and the cleaned consensus sequences with ambiguities exported with `Rscript_1d_generate_sequence_lists.R`. A summary of final sample coverage, sequence length, heterozygosity, and allele divergence after cleaning is given in [Supplementary Material 2](#).

We examined SNP patterns in the cleaned Angiosperms353 dataset and determined the proportion of heterozygous loci and the proportion of loci with more than 0%, 0.5%, 1%, and 2% SNPs, as well as mean allele divergence for each sample. While heterozygosity (as indicated by the presence of SNPs, i.e., any locus with > 0% SNPs) can be expected in any homologous locus due to allelic variation, loci with a high proportion of SNPs (e.g., >1% SNPs per locus) more likely result from multiple gene copies. Therefore, we consider that any locus with >1% SNPs in the cleaned dataset is likely to be paralogous, with paralogy caused by biological processes such as gene duplication, polyploidization, and hybridization, rather than by allelic variation, sequencing error or contamination.

Phylogenomic tree construction

HybPhaser consensus sequences (i.e. sequences including ambiguity codes) were aligned using MAFFT with the `-auto` flag to automatically select alignment strategy (Katoh and Standley, 2013). Sites with >75% missing data were removed from the alignment using the `-clean` option in Phyutility (Smith and Dunn, 2008), and exon alignments concatenated with AMAS following visual inspection (Borowiec, 2016). A maximum likelihood concatenated tree was then estimated from the clean alignment in IQ-TREE (Nguyen et al., 2015), with the appropriate substitution model and partitioning scheme for the alignment chosen using ModelFinder Plus option `-MFP+MERGE` and 1000 ultrafast bootstrap replicates to determine bootstrap support (BS; Lanfear et al., 2012; Nguyen et al., 2015; Kalyaanamoorthy et al., 2017; Hoang et al., 2018). To generate a coalescent species tree, gene trees were estimated from cleaned gene alignments using IQ-TREE with the appropriate substitution model chosen using the ModelFinder and 1000 ultrafast bootstrap replicates (Lanfear et al., 2012; Nguyen et al., 2015; Kalyaanamoorthy et al., 2017; Hoang et al., 2018). Newick Utils v1.6 was used to collapse branches with a BS value of <10, and TreeShrink was used to automatically remove branches with outlying length (Junier and Zdobnov, 2010; Mai and Mirarab, 2018). A species tree was generated from the cleaned gene trees using ASTRAL v5.7.8, and node support was assessed with posterior probability (PP; Mirarab

et al., 2014). Nodes with BS <90 and PP <0.9 were considered to have low support, nodes with BS = 90–97 and PP = 0.9–0.97 were considered moderately supported, while nodes with BS = 97–99 and PP = 0.97–0.99 were considered to have high support. Nodes with BS = 100 and PP = 1.0 received maximum support in our analyses.

Divergence time estimation

To date the Sapindales phylogenetic tree, 29 fossils were selected from the literature as calibrations (Table 1). The reliability of each fossil's identification and age was rigorously assessed and scored following the approach used in a previous angiosperm-wide fossil calibration dataset (Ramírez-Barahona et al., 2020) and using best practices for justifying fossil calibrations (Parham et al., 2012; Supplementary Material 3). A conservative approach to calibration was employed, with fossils assigned to the stem node of the taxon or clade the fossil was assigned to. Full justification of node assignment for each fossil calibration is given in Supplementary Material 3.

Computational efficiency of the dating analysis was optimized through gene-shopping, as implemented in SortaDate (Smith et al., 2018). Gene trees were filtered firstly by their similarity to the species tree, secondly by clock-likeness (as indicated by root-to-tip variance), and thirdly by tree length. The three best loci according to these criteria were selected for downstream dating analyses. Three loci were chosen to facilitate time-efficient completion of dating analyses, and because the inclusion of more data is unlikely to improve results, with recent studies suggesting that age calibration priors are the major influence on dating analysis results rather than the quantity of sequence data included (Dos Reis and Yang, 2013; Foster et al., 2017; Sauquet et al., 2022). These loci were aligned using MAFFT with the -auto flag to automatically select alignment strategy (Kato and Standley, 2013).

Bayesian divergence time estimations were carried out in BEAST v2.6.6 after setting parameters in BEAUti (Bouckaert et al., 2019). Although outgroup topology was not consistent with APG, all outgroup representatives fall within the Pentapetalae, and so the crown age of Pentapetalae was used as the root calibration for the dating analyses. However, the crown age of the angiosperms is uncertain, with molecular and fossil-based studies supporting both a young (Lower Cretaceous) and old angiosperm crown age (Lower Jurassic), resulting in both relatively young and old ages for crown Pentapetalae (Ramírez-Barahona et al., 2020; Silvestro et al. 2021; Sauquet et al., 2022). Given the strong influence of root calibrations on the age of Sapindales families (Muellner-Riehl et al., 2016), the choice of an old or young secondary calibration for the Pentapetalae root prior is likely to affect the results of the current dating analysis. For this reason, three dating analyses were conducted with three alternative ages for the age of Pentapetalae, as estimated by Ramírez-Barahona et al. (2020). Based on dating analyses of the angiosperms with many carefully selected fossil calibrations and three alternative root calibrations, Ramírez-Barahona et al. (2020) reported three possible age ranges for the Pentapetalae, with the crown of Pentapetalae being dated to be between 140.33–144.29 Ma in the 'CC-complete' analysis, 143.91–147.94 Ma in the 'RC-complete' analysis, and 212.25–221.02 Ma in the 'UC-complete' analysis. The 95% HPDs for the age of crown Pentapetalae for the CC-complete,

RC-complete and UC-complete analyses of Ramírez-Barahona et al. (2020) were therefore applied as the bounds of a uniform prior in three separate BEAST analyses, with the CC-complete incorporating ages of Pentapetalae taken from a young-angiosperm scenario, RC-complete analysis representing an scenario where angiosperms are assumed to be older, and the UC-complete analysis incorporates ages for Pentapetalae from an analysis where angiosperms were assumed to be very old. These three analyses were run with a fixed tree topology, a Birth-Death tree prior, an uncorrelated log-normal (relaxed) clock model, and with all primary fossil calibrations as uniform priors, with the maximum age boundary set to the maximum age of crown Pentapetalae. To fix the tree topology and maximise computational efficiency, the starting tree was assigned to the best maximum likelihood concatenated tree and topology exchange operators were disabled (i.e., Wide Exchange, Nanon Exchange, Wilson Balding and Subtree-slide; Bouckaert et al., 2019). Ten runs of each model were conducted, each with a chain length of 50,000,000 and with trees sampled every 1,000 generations, resulting in a combined tree exploration space where most priors and statistics reached an effective sampling size (ESS) >200 and all priors and statistics had an ESS >100. Runs were checked for convergence and stationarity in Tracer v1.7.2, and every 50,000th tree was sampled from each run after a burn-in of 20% and combined using logCombiner, and TreeAnnotator was used to generate the consensus tree (Rambaut et al., 2018; Bouckaert et al., 2019).

To test the effect of the tree prior and distribution of the fossil priors, two additional sensitivity analyses were conducted using the RC-complete root calibration. The tree prior sensitivity analysis was performed as described above but with a Yule tree prior (instead of a Birth-Death tree prior). The fossil prior distribution sensitivity analysis was conducted as described above but with log-normal distributions on the fossil calibrations (instead of uniform distributions), with the minimum age of the fossil set to the offset age of the distribution, the mean rounded up to the nearest 5 Ma, and a sigma value of 1.0.

Results

Locus recovery and paralogy

The final Sapindales dataset comprised 472 species (including 24 outgroup representatives) with an average of 324 loci recovered per sample and 74% target coverage per locus (Supplementary Material 2). On average, 15 (0–44) loci with an outlying proportion of SNPs were removed per sample in the cleaning steps of HybPhaser.

In the cleaned Sapindales dataset, 63% (15–99) of loci contained one or more SNP, and the mean allele divergence was 1.33% (0.89–7.52). While the presence of a low number of SNPs can be expected in any orthologous locus, a high proportion of SNPs in a locus (e.g. >1% SNPs per locus) in the cleaned dataset is more likely to be indicative of multiple copies of that locus in the data (i.e., paralogy). Therefore, to differentiate allelic variation from paralogy, we consider loci with >1% SNPs to be paralogous (i.e., have multiple gene copies). Overall, 28.55% (2.14–96.31) of Angiosperms353 loci for Sapindales contained >1% SNPs (i.e., were paralogous). Variation in the degree of paralogy and allele divergence was unevenly spread across the

TABLE 1 Summary of fossils used to calibrate the genus-level molecular dating analysis of Sapindales.

	Family	Taxon	Organ/s	Country	Oldest stratum	Min. age (Ma)	Node calibrated	Reference
C02	Sapindaceae	† <i>Sapindospermum nitidum</i> Knobloch & Mai	Seed	Czech Republic	Turonian	89.8	crown Sapindales	Knobloch and Mai (1986)
C03	Sapindaceae	† <i>Aesculus hickeyi</i> Manchester	Leaf, fruit, seed	USA	Paleocene	56	crown Sapindaceae	Manchester (2001a)
C04	Sapindaceae	† <i>Dipteronia brownii</i> McClain & Manchester	Fruit	USA	Middle Paleocene	60	stem <i>Dipteronia</i>	McClain and Manchester (2001)
C05	Sapindaceae	† <i>Acer</i> sp.	Leaf		Late Paleocene	56	stem <i>Acer</i>	Crane (1990)
C06	Sapindaceae	† <i>Koelreuteria allenii</i> (Lesq.) W. N. Edwards	Fruit	USA	Eocene	52	stem <i>Koelreuteria</i>	Wang et al. (2013)
C07	Sapindaceae	† <i>Allophylus graciliformis</i> (Berry) Berry	Leaf	Argentina	Early Eocene	46	stem <i>Allophylus</i>	Panti (2020)
C08	Burseraceae	† <i>Bursera inaequalateralis</i> (Lesq.) MacGinitie	Leaf	USA	Eocene	48.5	crown Burserinae	MacGinitie (1969)
C09	Burseraceae	† <i>Bursericarpum aldwickense</i> Chandler	Pyrene, seed	UK	Ypresian	47.8	stem <i>Protium</i> alliance	Chandler (1961)
C10	Anacardiaceae	† <i>Coahuiloxylon terrazasiae</i> Estrada-Ruiz, Martínez-Vabrera & Cevallos-Ferriz	Wood	Mexico	Campanian	72.1	stem Anacardiaceae	Estrada-Ruiz et al. (2010)
C11	Anacardiaceae	† <i>Choerospondias sheppeyensis</i> (Reid & Chandler) Chandler	Fruit, seed	UK	Ypresian	47.8	crown Anacardiaceae	Reid and Chandler (1933); Chandler (1961)
C12	Anacardiaceae	† <i>Dracontomelon macdonaldii</i> (Berry) Herrera, Manchester & Jaramillo	Fruit, seed	Panama	Late Eocene	33.9	crown Spondioideae (- <i>Camnosperma</i>)	Herrera et al. (2012)
C13	Anacardiaceae	† <i>Spondias rothwellii</i> Herrera, Carvalho, Jaramillo & Manchester	Fruit, seed	Panama	Early Miocene	18.5	stem <i>Spondias</i>	Herrera et al. (2019)
C14	Anacardiaceae	† <i>Anacardium germanicum</i> Manchester et al.	Fruit	Germany	Middle Eocene (Lutetian assumed)	41.2	stem <i>Fegimanra</i> + <i>Anacardium</i>	Manchester et al. (2007)
C15	Anacardiaceae	† <i>Mangifera paleoindica</i> Sawangchote, Grote, and Dilcher	Leaf	Thailand	Late Oligocene	23.03	stem <i>Mangifera</i>	Sawangchote et al. (2009)
C16	Anacardiaceae	† <i>Cotinus fraterna</i> (Lesquereux) MacGinitie	Leaf	USA	Late Eocene	33.9	stem <i>Cotinus</i>	MacGinitie (1953); Manchester (2001b)
C17	Anacardiaceae	† <i>Pistacia</i> sp.	Pollen	Austria	Middle Miocene (Middle to Upper Serravallian)	12.7	stem <i>Pistacia</i>	Grimsson et al. (2020)
C18	Anacardiaceae	† <i>Loxopterygium laplayense</i> Burnham and Carranco	Fruit	Ecuador	Middle Miocene	13	stem <i>Loxopterygium</i>	Burnham and Carranco (2004)
C19	Simaroubaceae	† <i>Ailanthus confucii</i> Unger	Fruit	Germany	Middle Eocene (Lutetian assumed)	41.2	stem <i>Ailanthus</i> + <i>Picrasma</i>	Collinson et al. (2012)
C20	Meliaceae	† <i>Cedrela</i> sp.	Leaf	USA	Eocene	51	crown Cedreloideae	Hickey and Hodges (1975)
C21	Meliaceae	† <i>Cedrela merrilli</i> (Chaney) Brown	Leaf, seed	USA	Late Eocene	36.21	stem <i>Cedrela</i> + <i>Toona</i>	Meyer and Manchester (1997); Manchester and McIntosh (2007)
C22	Meliaceae	† <i>Swietenia miocenica</i> Castañeda-Posadas & Cevallos-Ferriz	Flower	Mexico	Late Oligocene	22.5	stem <i>Swietenia</i>	Castañeda-Posadas and Cevallos-Ferriz (2007)

(Continued)

TABLE 1 Continued

	Family	Taxon	Organ/s	Country	Oldest stratum	Min. age (Ma)	Node calibrated	Reference
C23	Meliaceae	† <i>Manchestercarpa vancouverensis</i> Atkinson	Fruit, seed	Canada	Middle Campanian	72.1	crown Meliaceae	Atkinson (2020)
C24	Rutaceae	† <i>Rutaspermum biornatum</i> Knobloch & Mai	Seed	Germany	Maastrichtian	66	crown Rutaceae	Knobloch and Mai (1986)
C25	Rutaceae	† <i>Clausena</i> sp.	Leaf	Ethiopia	Late Oligocene	27.23	stem <i>Clausena</i> + <i>Glycosmis</i>	Pan (2010)
C26	Rutaceae	† <i>Citrus linczangensis</i> Xie et al.	Leaf	China	Late Miocene	11.6	stem <i>Citrus</i>	Xie et al. (2013)
C27	Rutaceae	† <i>Ptelea paliuruoides</i> (Brown) Manchester & O'Leary	Fruit, seed	USA	Middle Eocene	48.5	stem <i>Ptelea</i> + <i>Peltostigma</i> + <i>Plethadenia</i> + <i>Decazyx</i> clade	Manchester and O'Leary (2010)
C28	Rutaceae	† <i>Zanthoxylum</i> sp.	Seed	UK	Ypresian	47.8	stem <i>Zanthoxylum</i>	Chandler (1961)
C29	Rutaceae	† <i>Euodia costata</i> (Chandler) Tiffney	Seed	UK	Latest Paleocene (Thanetian)	56	crown Rutoideae clade	Tiffney (1981)
C30	Rutaceae	† <i>Vepris</i> sp.	Leaf	Ethiopia	Late Oligocene	27.23	stem <i>Vepris</i>	Pan (2010)

Further details of each fossil and justification for their placement are provided in [Supplementary Material 3](#). The dagger symbol (†) indicates fossil taxa.

order. Meliaceae showed substantially higher levels of paralogy and allele divergence relative to other Sapindales families, with an average of $51\% \pm 4.11$ of loci with $>1\%$ SNPs and an average allele divergence of 2.97 ± 0.27 (Figures 1, 2). Kirkiaceae had the lowest level of paralogy and allele divergence, with $11\% \pm 3.89$ of loci with $>1\%$ SNPs and a mean allele divergence of 0.40 ± 0.086 (Figures 1, 2). Similar patterns in paralogy and allele divergence were observed when the threshold for paralogy was raised to $>2\%$ SNPs (shown in [Supplementary Material 4](#)).

Phylogenetic relationships

The concatenated alignment was 194,132 bp long and comprised 330 loci with 135,613 parsimony-informative sites and 14.73% gaps or ambiguities. IQ-TREE identified the best partition scheme and merged the alignment into 45 partitions (lnL = -10642504, df = 1822), all of which were allocated an optimal substitution model with the ModelFinder function of IQ-TREE. After 173 tree search iterations and 1000 bootstrap trees, IQ-TREE produced a consensus tree with a log-likelihood of -10642396 (Figure 2).

Sapindales and all Sapindales families were found to be monophyletic with maximum support in both the concatenated and multispecies coalescent analyses. Nitrariaceae was retrieved as sister to the rest of Sapindales with moderate support (BS = 96, PP = 0.97) in both analyses (Figure 3). In the concatenated analysis, Sapindaceae and Biebersteiniaceae were retrieved as sister families, whereas in the coalescent analysis they were placed as a grade; however, in both cases, the relationships of Sapindaceae and Biebersteiniaceae received poor support (Figure 3). The remaining families grouped into two clades that were consistent and well-supported in both analyses: the 'KAB clade' with Kirkiaceae sister to Anacardiaceae + Burseraceae, and the 'SRM clade' with Simaroubaceae sister to Rutaceae + Meliaceae.

Within Sapindaceae, three major clades were retrieved in both the concatenated and multispecies coalescent analyses (Figure 2A, clades S1, S2, and S3; [Supplementary Material 5](#)). These clades had maximum support, as did most generic relationships within the family. *Xanthoceras* was consistently retrieved as sister to clade S2 with maximum support in both the multispecies coalescent and concatenated analyses (Figure 2A; [Supplementary Material 5](#)). The greatest uncertainty in generic relationships was found in clade S3, particularly among *Eurycorymbus*, *Matayba*, *Mischarytera*, *Mischocarpus*, and *Sarcopteryx*, (BS = 47–57; Figure 2A). The ancestral node of the clade containing *Cupaniopsis* and the clade containing *Rhysotoechia* were also poorly supported (BS = 56; Figure 2A).

In Burseraceae, *Beiselia* was found to be sister to the rest of the family with maximum support in both the concatenated and coalescent analyses (Figure 2B; [Supplementary Material 5](#)). The remaining Burseraceae genera were retrieved in four well-supported main clades (Figure 2B, clades B1–4). The relationships between all Burseraceae genera were well-supported in both the concatenated and coalescent analyses, with the exception of the sister relationship of *Dacryodes* and *Canarium* retrieved in the coalescent analysis ([Supplementary Material 5](#)).

Within Anacardiaceae, two major clades were recovered (Figure 2B, clades A1 & A2). In the concatenated analysis, *Camposperma* was retrieved as a crown member of clade A1 with moderate support (BS = 90), but in the coalescent analysis it was placed as sister to clades A1 and A2 with low support ([Supplementary Material 5](#), PP = 0.69). *Dobinea* and *Campylopetalum* formed a well-supported clade sister to clade A2 in both analyses. *Cyrtocarpa* was not resolved as monophyletic since the two sampled species (*C. procera* and *C. caatingae*) were placed in different subclades within clade A1. Similarly, the genus *Rhus* was not monophyletic, with *R. taitensis* + *R. coriaria* placed separately from *R. thouarsii* + *R. perrieri*

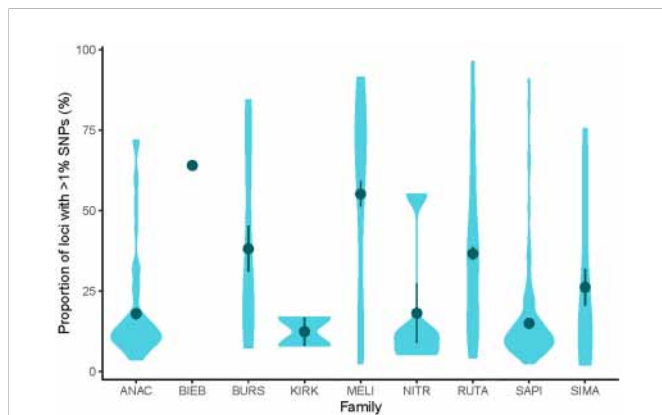


FIGURE 1

Violin plots of mean proportion of paralogous loci (loci with >1% SNPs) of each family in Sapindales as calculated with HybPhaser. ANAC, Anacardiaceae (n = 79), BIEB, Biebersteiniaceae (n = 1), BURS, Burseraceae (n = 16), KIRK, Kirkiaceae (n = 2), MELI, Meliaceae (n = 56), NITR, Nitrariaceae (n = 5), RUTA, Rutaceae (n = 136), SAPI, Sapindaceae (n = 134), SIMA, Simaroubaceae (n = 18).

within clade A2. Other genera for which multiple accessions were included (*Cotinus*, *Mangifera*, *Pistacia*, and *Toxicodendron*) were retrieved as monophyletic.

Three major clades were recovered in Simaroubaceae with maximum support in the concatenated and coalescent analyses (Figure 2B, clades Si1–3; Supplementary Material 5). Clade Si1 comprises the monophyletic genus *Castela* that was placed as sister to the rest of the family. Clade Si2 contains *Ailanthus* and *Picrasma*, with *Ailanthus* retrieved as paraphyletic in relation to *Picrasma crenata* in both the concatenated and coalescent analyses (Figure 2B; Supplementary Material 5). The third clade, Si3, contains the remaining 13 sampled genera, predominantly arranged in a grade with high to maximum support for nodes in the concatenated analysis, and low to maximum support in the coalescent tree (Figure 2B; Supplementary Material 5).

In Meliaceae, three major clades were recovered with high support in both the concatenated and coalescent analyses (Figure 2B, clades M1–3; Supplementary Material 5). Within clade M1, *Chukrasia* and *Schmardaea* form a clade sister to the rest of the genera. *Cedrela* and *Toona* form a clade together, and are mutually paraphyletic. In clade M2, *Pterorhachis* and *Owenia* are successive sisters leading to a clade containing *Azadirachta* and *Melia*. *Melia* was paraphyletic in relation to *Azadirachta* in both analyses. In clade M3, *Munronia* was placed as sister to the rest of the clade. The relationship of the subclade containing *Calodectarya*, *Humbertioturraea*, *Naregamia*, *Nymania*, and *Turraea* was uncertain, being placed as sister to a subclade of *Lepidotrichilia* + *Malleastrum* in the concatenated analysis with low support (BS = 87), and sister to the subclade of *Pseudobersama* + *Trichilia* in the coalescent analysis with no support (PP = 0.38). The placement of *Vavaea* was consistent across analyses with strong support, but the relationships of the subclade comprising *Anthocarapa*, *Chisocheton*, *Heckeldora*, *Leplaea*, *Neoguarea*, *Ruagea*, *Synoum*, and *Turraeanthus* received moderate to weak support (Figure 2B; Supplementary Material 5). Additionally, the placement of *Reinwardtiodendron* differed slightly between coalescent and concatenated analyses, with *Aphanamixis* sister to a clade containing *Aglaia* + *Reinwardtiodendron* in the concatenated

analysis, but *Reinwardtiodendron* retrieved as sister to a clade containing *Aglaia* + *Aphanamixis* in the coalescent analysis (Figure 2B; Supplementary Material 5). *Aglaia* was retrieved as monophyletic with maximum support in both trees.

Rutaceae comprised four main clades, all receiving maximum support in both the concatenated and coalescent analyses (Figure 2C, clades R1–4; Supplementary Material 5). Clades R1 and R2 form a grade (with five and three genera respectively), and the larger clades R3 and R4 were retrieved as sisters. Generic relationships had high to maximum support in clades R1, R2 and R3, with two main exceptions in clade R3: the placement of the clade containing *Aegle*, *Aeglopsis*, *Afraegle*, and *Balsamocitrus* received moderate to low support (BS = 95, PP = 0.51), as did the placement of the clade containing *Burkillanthus*, *Pleiospermium*, *Limnocitrus*, and *Swinglea* (BS = 97, PP = 0.44) (Figure 2C; Supplementary Material 5). Relationships in clade R4 were generally less well supported, with inconsistencies in the topologies of the concatenated and coalescent trees. Major topological differences were in the placement of *Flindersia*, *Nematolepis*, and *Ptelea*, and in the placement of the clade containing *Decazyx*, *Peltostigma*, and *Plethadenia* (Figure 2C; Supplementary Material 5). All genera were monophyletic in both trees, with the exception of *Boronia* and *Melicope* in both the concatenated and coalescent analyses (Figure 2C; Supplementary Material 5).

Divergence time estimation

The three optimal loci selected for dating analyses were loci 5162, 5333 and 6091, resulting in an alignment of 4,738 bp. The ages for major clades in Sapindales were similar across chronograms regardless of whether they were estimated with the CC-complete or RC-complete root priors, with Cretaceous stem and crown ages for the order and all families. However, in the UC-complete analysis (which constrained the analysis to ages from a scenario assuming the angiosperms to be very old) major Sapindales clades were considerably older. In this analysis, the stem and crown nodes of Sapindales were in the Jurassic, as were stem nodes of Nitrariaceae, Biebersteiniaceae, Sapindaceae, Simaroubaceae, and Kirkiaceae. The crown nodes of Sapindaceae and Simaroubaceae were in the Lower Cretaceous. Meliaceae, Burseraceae, Anacardiaceae, and Rutaceae had Lower Cretaceous stem nodes, and crown nodes at the Lower-Upper Cretaceous boundary (Supplementary Materials 6–9). Given the similarity of the CC-complete and RC-complete analyses, we herein focus on the results of the RC-complete and UC-complete analyses, to compare the results of when a younger-Pentapetalae and older-Pentapetalae scenario, respectively, is adopted.

In the analysis using the RC-complete root constraint, the age of the stem node of Sapindales was estimated at c. 131 (124–137) Ma (Figure 4A). Nitrariaceae, Biebersteiniaceae, and Sapindaceae diverged in the Lower Cretaceous with an estimated stem age of 124 (117–130) Ma for Nitrariaceae (making this the crown age of Sapindales), and 122 (114–128) Ma for Biebersteiniaceae and Sapindaceae (Supplementary Material 6; Figure 4A). The major split from the most recent common ancestor of the KAB and SRM clades was estimated to have occurred in the Lower Cretaceous, c. 118 (110–125) Ma. Kirkiaceae emerged approximately 109 (100–119) Ma, and Burseraceae and Anacardiaceae split from their most

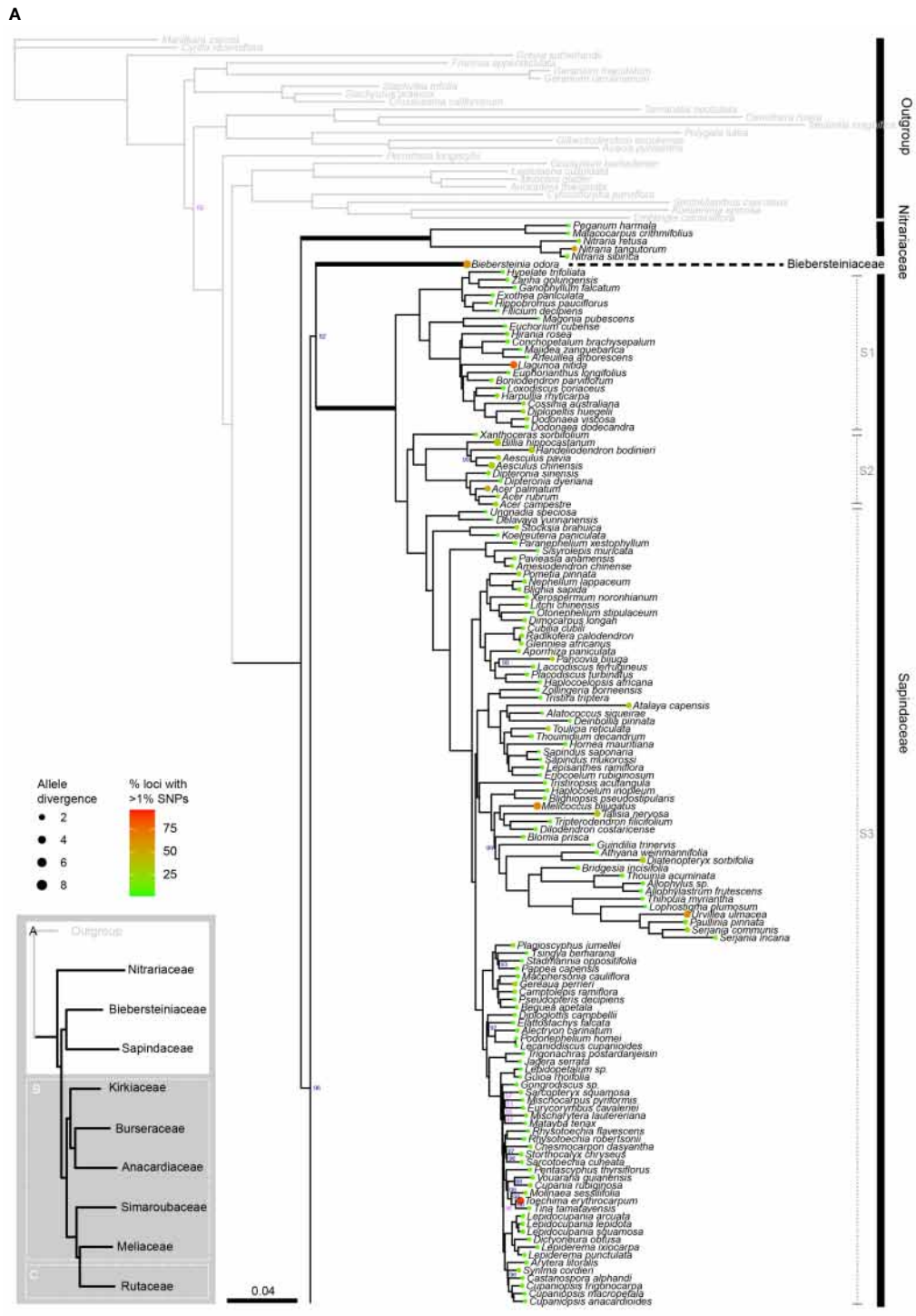


FIGURE 2A

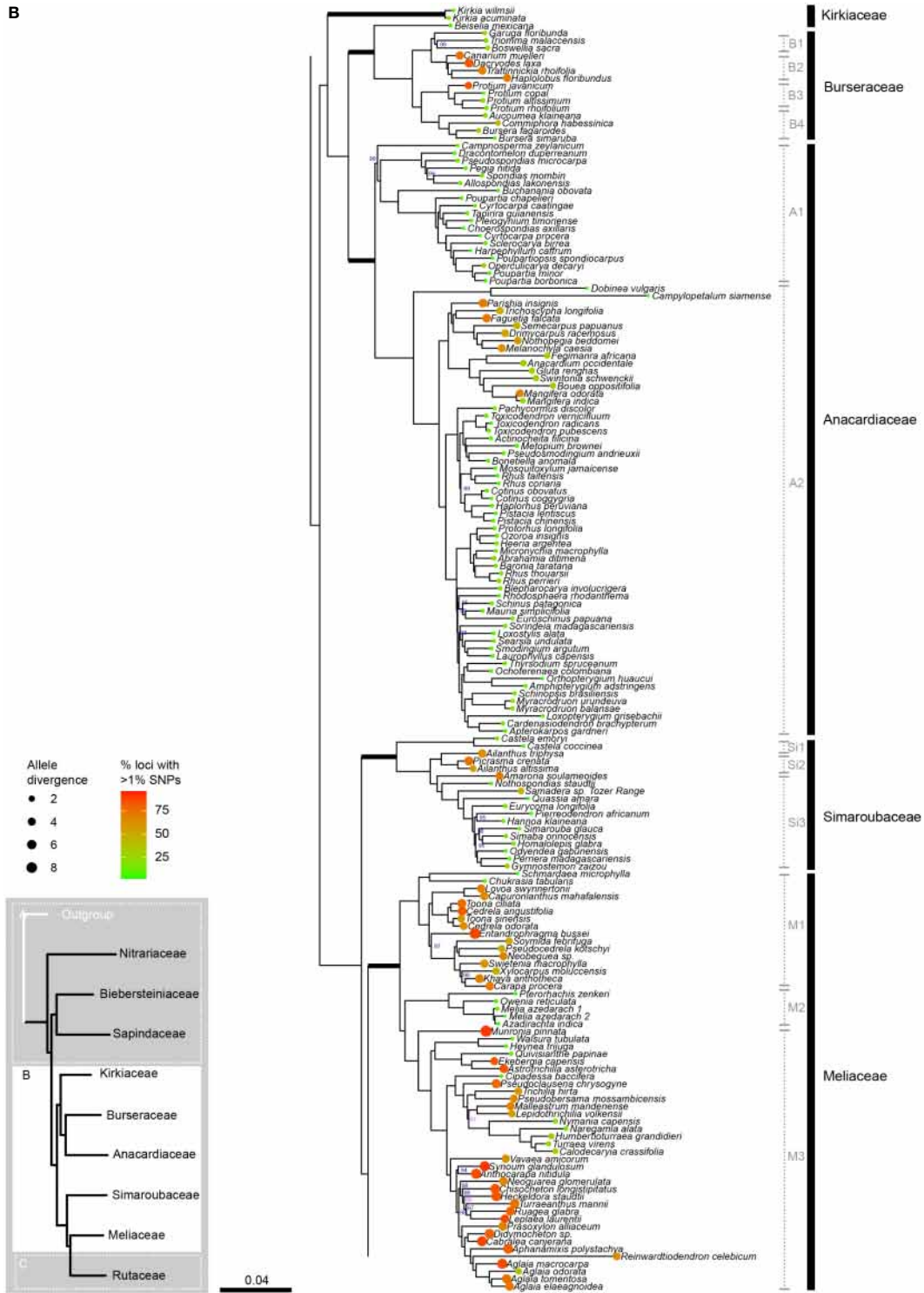


FIGURE 2B

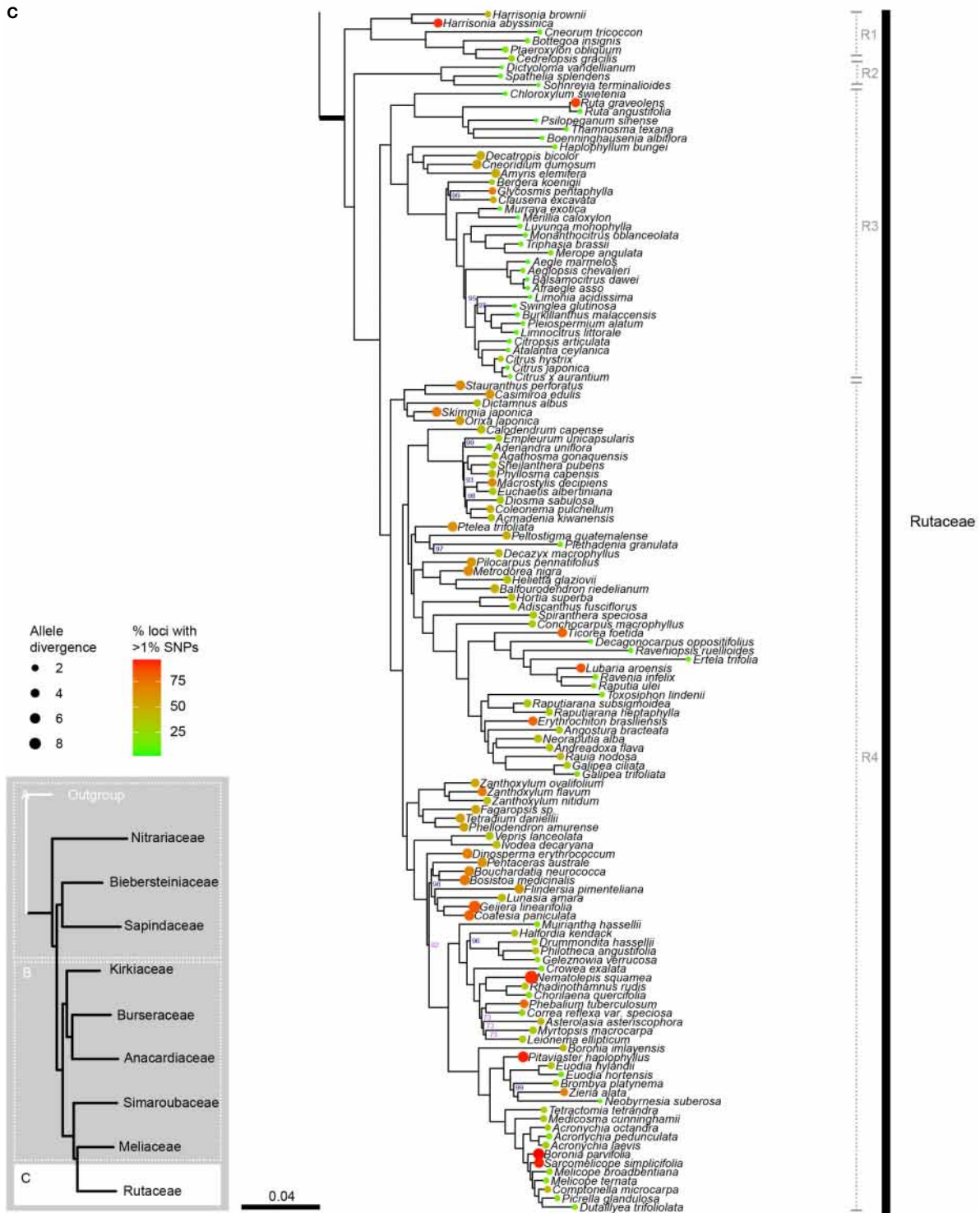
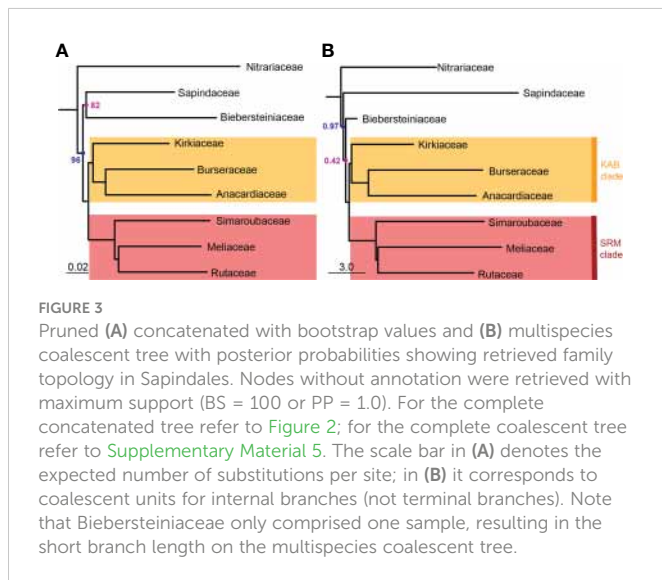


FIGURE 2C
 Phylogenetic relationships in Sapindales based on concatenated analyses of 324 nuclear loci. **(A)**, Nitrariaceae, Biebersteiniaceae, Sapindaceae; **(B)**, Kirkiaceae, Burseraceae, Anacardiaceae, Simaroubaceae, Meliaceae, and **(C)**, Rutaceae. Numbers on branches indicate support for nodes with low support (BS<90%; violet) or moderate to high support (BS>90%; blue); branches without bootstrap values have maximum support (BS = 100%). Thick branches indicate stems of families, with major family clades annotated in grey to the right. Tip circle colour indicates the percentage of paralogous loci (= percentage of loci with >1% SNPs), and tip circle size is proportional to allele divergence for the accession.



recent common ancestor at approximately 100 (89–111) Ma (Figure 4B). Simaroubaceae diverged at c. 108 (100–117) Ma, and Meliaceae and Rutaceae diverged from their most recent common ancestor at c. 104 (96–114) Ma (Figures 4B, C). The estimated crown ages for Nitrariaceae and Sapindaceae were c. 72 (48–102) and 103 (94–113) Ma, respectively. The estimated crown ages for Kirkiaceae, Anacardiaceae, and Burseraceae are 2 (0.2–5), 88 (77–100), and 85 (71–99) Ma, and the crown ages for Simaroubaceae, Rutaceae, and Meliaceae are 83 (66–102), 97 (87–107), and 86 (72–98) Ma, respectively.

In the analysis using the UC-complete root constraint (where an ‘old-Pentapetalae’ scenario is assumed), the age of the stem node of Sapindales was estimated at c. 186 (174–199) Ma (Supplementary Materials 6, 9). Nitrariaceae, Biebersteiniaceae, and Sapindaceae diverged in the Middle Jurassic with an estimated stem age of 172 (159–187) Ma for Nitrariaceae (making this the crown age of Sapindales), and 169 (153–182) Ma for Biebersteiniaceae and Sapindaceae (Supplementary Materials 6, 9). The major split from the most recent common ancestor of the KAB and SRM clades was estimated to have occurred at the Upper/Middle Jurassic boundary, c. 162 (146–176) Ma. Kirkiaceae emerged approximately 148 (131–165) Ma, and Burseraceae and Anacardiaceae split from their most recent common ancestor at approximately 134 (116–151) Ma (Supplementary Materials 6, 9). Simaroubaceae diverged at c. 145 (128–161) Ma, and Meliaceae and Rutaceae diverged from their most recent common ancestor at c. 139 (123–157) Ma (Supplementary Materials 6, 9). The estimated crown ages for Nitrariaceae and Sapindaceae were c. 103 (63–145) and 135 (117–151) Ma, respectively. The estimated crown ages for Kirkiaceae, Anacardiaceae, and Burseraceae are 2 (0.28–5), 114 (96–131), and 108 (84–130) Ma, and the crown ages for Simaroubaceae, Rutaceae, and Meliaceae are 109 (85–134), 128 (111–145), and 109 (89–132) Ma, respectively.

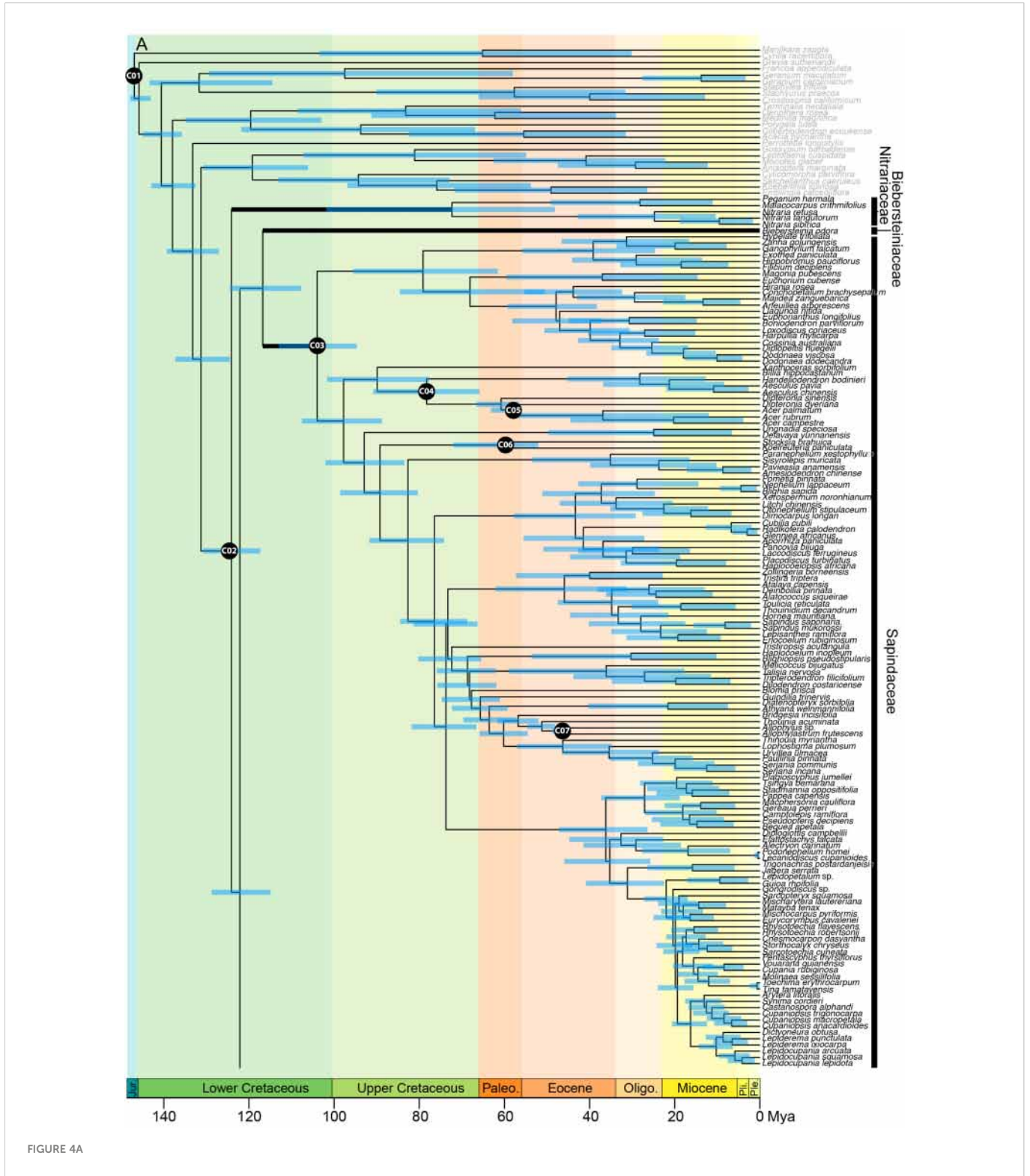
Using a Birth-Death tree prior made little difference to the ages of the chronogram, with the major clades of Sapindales an average of only 0.13 Ma younger when estimated with a Birth-Death tree prior as opposed to a Yule prior (Supplementary Materials 10, 11). Likewise, the application of log-normal prior distributions for fossil calibrations made little difference to the ages of the major nodes of Sapindales, with major nodes only 1–2 Ma younger than the same analysis with uniformly distributed fossil priors, and with overlapping HPD intervals.

Discussion

The rise of Sapindales families in the Mid-Cretaceous Hothouse

Our results strongly support the monophyly of Sapindales, with the stem lineage of the order evolving at c. 131 (124–137) Ma or c. 186 (174–199) Ma, depending on whether angiosperms (and Pentapetalae) are assumed to be younger, or older, respectively. The former estimate is slightly older — and the latter much older — than ages estimated in previous studies, such as the angiosperm-wide study of Magallón et al. (2015) and the Sapindales chronogram of Muellner-Riehl et al. (2016), who dated the stem node of Sapindales to be 104 (98–112) and 111 (106–117) Ma, respectively. Our dating analyses place the emergence of the order c. 41 or 96 Ma before the preservation of the oldest-known fossil of Sapindales (a seed of †*Sapindospermum nitidum* from the Czech Republic; Knobloch and Mai, 1986). In both scenarios, the short stem of the order followed by the rapid succession of family divergences suggests extensive diversification in the Lower-Mid Cretaceous, coincident with a global warming period and the persistence of ancient Sapindales lineages since then (Scotese, 2021). Regardless of the crown age of the angiosperms, the Mid-Cretaceous Hothouse is modelled to have had a substantial effect on the evolution of Sapindales families. When angiosperms are assumed to be relatively young (and thus the age of crown Pentapetalae is modelled to be younger), most families diverged and diversified during the Mid-Cretaceous Hothouse period. If crown angiosperms are assumed to be ancient (and thus the age of crown Pentapetalae is older), Sapindales families emerged just prior to the Mid-Cretaceous Hothouse period and diversification of the families occurred during the Mid-Cretaceous Hothouse. Although both scenarios are possible, there is a chance that ages estimated with the ancient angiosperm scenario are overestimated; in this scenario, the maximum age of crown angiosperms is considered to be coincident with the first appearance of angiosperm-like pollen grains in the fossil record (247 Ma). However, it is possible that these angiosperm-like pollen grain fossils are stem angiosperm relatives rather than crown angiosperm members, and so when applied as a maximum age for the crown of the angiosperms, this calibration could lead to an overestimation of crown-group ages, including the estimated age of the Pentapetalae used in our analysis. This risk is likely compounded by the use of uniform priors in the analysis, whereby older ages close to the appearance of angiosperm-like pollen grains in the fossil record would be just as likely as younger ages.

Novel to this study, our analyses suggest that the well-supported KAB and SRM clades are sisters, and that Sapindaceae does not fall within these clades. This contrasts with previous studies that have tentatively placed Sapindaceae within the KAB clade, or as sister to the SRM clade (Chase et al., 1993; Gadek et al., 1996; Appelhans et al., 2012; Muellner-Riehl et al., 2016; Lin et al., 2018; Li et al., 2019; Ramírez-Barahona et al., 2020; Li et al., 2021). However, despite dense sampling and the inclusion of 330 loci, the relationship between Sapindaceae and Biebersteiniaceae could not be resolved, with these families uncertainly placed as sisters in our concatenated tree and as poorly-supported successive sisters to the KAB + SRM clade in our



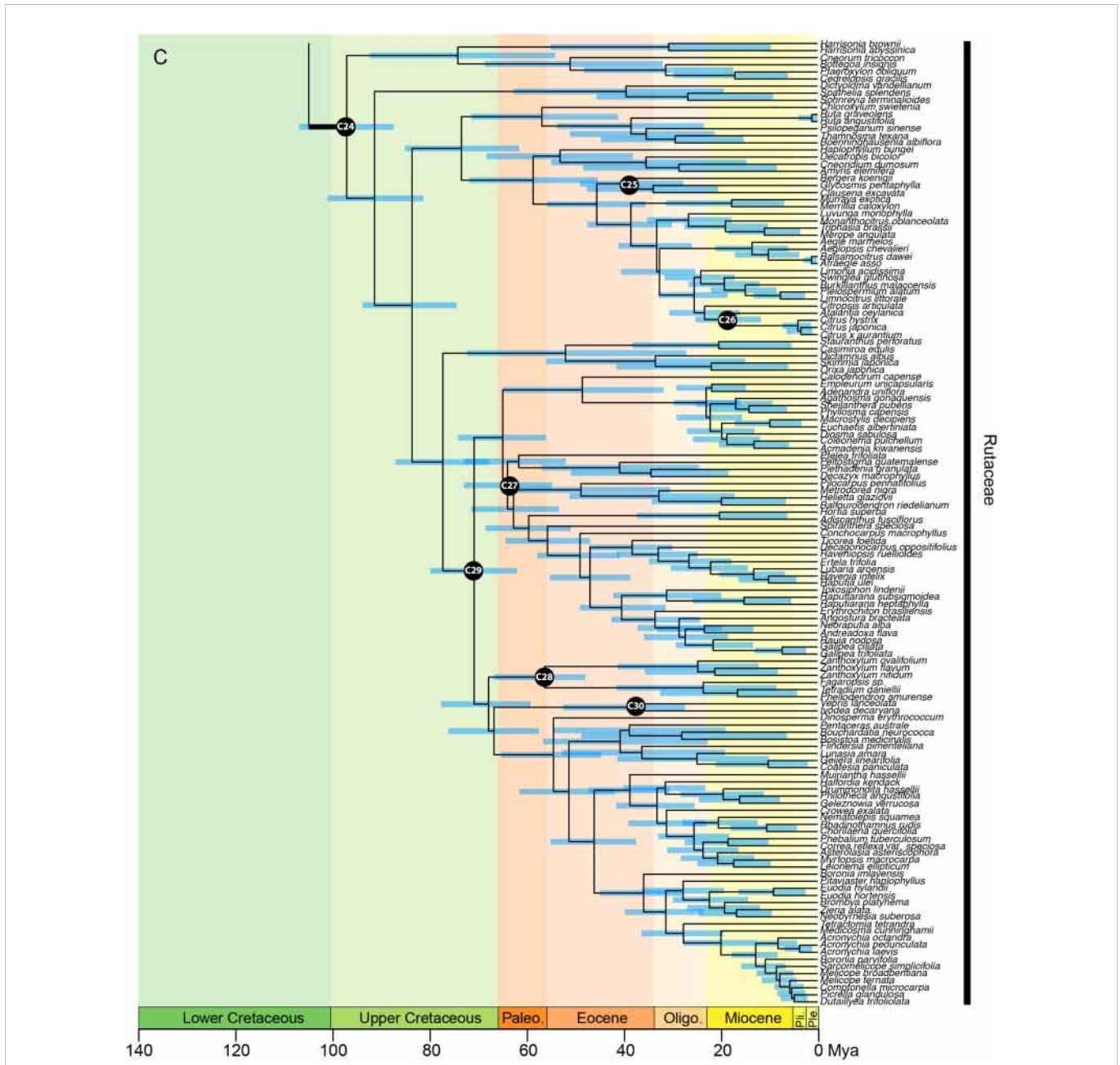


FIGURE 4C
 Chronogram of Sapindales as estimated under a relaxed-clock model with a normal secondary root prior distribution (node C01) and uniform primary calibrations. Tree topology was fixed to the concatenated tree (see Figure 2). (A) shows Nitrariaceae, Biebersteiniaceae, Sapindaceae; (B) shows Kirkiaceae, Burseraceae, Anacardiaceae, Simaroubaceae, Meliaceae, and (C) shows Rutaceae. Black circles with C01-C30 signify calibrated nodes (see Table 1 for fossil information); blue bars represent 95% highest posterior densities (HPDs); thickened branches indicate stem branches of families.

coalescent tree. As a result, the sister family to the KAB and SRM clades cannot be identified. In line with the results of Li et al. (2021), we consistently retrieved Nitrariaceae as sister to all remaining Sapindales with moderate support. However, the poor support for the nodes leading to Sapindaceae and Biebersteiniaceae means that the relationships between Nitrariaceae, Biebersteiniaceae, and Sapindaceae are perhaps best represented as a polytomy. What is clear from all dating analyses, however, is that these families emerged early and rapidly in the evolution of the order, having diverged within an estimated period of approximately 5 Ma. The difficulty of reconstructing relationships between Nitrariaceae, Biebersteiniaceae, and Sapindaceae may be due to this rapid and

ancient divergence of families, in combination with unavoidable sampling heterogeneity caused by a low within-family diversity in Biebersteiniaceae and Nitrariaceae. As previously suggested by Muellner-Riehl et al. (2016), the low extant diversity of Biebersteiniaceae and Nitrariaceae on long stem branches is likely to be indicative of a prevalent history of extinction relative to other Sapindales families (although it could also be due to low speciation rates, or a combination thereof), and such ‘depauperon’ lineages are often difficult to place in phylogenetic analyses (Donoghue and Sanderson, 2015). Further research with custom loci or whole genomes may be required to resolve the relationships of these families, if it is possible at all. Given the modern-day xeric habitat

of Nitrariaceae, if this family is confirmed to be sister to the rest of Sapindales, it could point to an extra-tropical origin for this predominantly tropical order. This hypothesis is supported by a concentration of extant mesic- and xeric-adapted lineages in the Nitrariaceae, Biebersteiniaceae, and Sapindaceae, and the presence of large evaporite and arid belts from the Jurassic-Lower Cretaceous boundary when these families are likely to have evolved (Hay, 2017; Zhang et al., 2018). However, the spatio-temporal diversification dynamics of depauperon lineages such as Nitrariaceae are notoriously complex and difficult to reconstruct (Donoghue and Sanderson, 2015), so this hypothesis needs careful consideration and should be explicitly tested.

Kirkiaceae, Anacardiaceae, and Burseraceae were retrieved as a clade in this study ('the KAB clade'), adding to a substantial body of biochemical, morphological and molecular evidence suggesting that these families are closely related (Chase et al., 1993; Gadek et al., 1996; Muellner-Riehl et al., 2016). When angiosperms are assumed to be relatively young, the KAB clade was reconstructed as diverging from its sister SRM clade in the Lower Cretaceous, with Kirkiaceae, Anacardiaceae, and Burseraceae diverging at the Lower-Upper Cretaceous boundary. This is coincident with the beginning of the Mid-Cretaceous Hothouse, a period characterised by low-lying land masses, rising sea-levels, and high CO₂ concentrations and global temperatures (Barral et al., 2017; Scotese, 2021). Kirkiaceae was found to have diverged from the ancestor of Burseraceae and Anacardiaceae at c. 110 (100–119) Ma, and Burseraceae and Anacardiaceae as diverging at c. 101 (89–111) Ma. These results are slightly older than the analyses of Muellner-Riehl et al. (2016) (who estimated the stem age of Kirkiaceae and the Burseraceae/Anacardiaceae split to be 95 [86–103] and 87 [78–97] Ma, respectively), but in line with Weeks et al. (2014), who estimated the divergence of Anacardiaceae and Burseraceae to be at 116 (105–127) Ma. The crown nodes of Anacardiaceae and Burseraceae at 88 (77–100) and 85 (71–99) Ma, respectively, are just after the Cenomanian–Turonian Thermal Maximum of the Mid-Cretaceous Hothouse period, which is thought to mark the time with the highest temperatures and sea levels in the past 250 million years, suggesting it may have been a key period for the diversification of these families (Scotese, 2021). In an ancient-angiosperm scenario, the KAB and SRM clades are reconstructed as diverging in the Jurassic, with Kirkiaceae, Anacardiaceae, and Burseraceae emerging at the Jurassic-Cretaceous boundary, and crown diversification of Anacardiaceae and Burseraceae coincident with the start of the Mid-Cretaceous Hothouse period (rather than the peak). The young crown node of Kirkiaceae in all analyses could be indicative of extensive extinction since the divergence of the family, or of delayed speciation until the Pleistocene. The contrasting temporal patterns of diversification of Kirkiaceae compared to Anacardiaceae and Burseraceae could provide interesting insights into the evolution of depauperon lineages in future studies (Donoghue and Sanderson, 2015).

Simaroubaceae, Rutaceae, and Meliaceae are retrieved as a clade within Sapindales (the 'SRM clade'), in line with previous studies on the order (Gadek et al., 1996; Muellner et al., 2007; Appelhans et al., 2012; Muellner-Riehl et al., 2016; Lin et al., 2018; Li et al., 2019; Ramírez-Barahona et al., 2020; Li et al., 2021). The SRM clade is united by the presence of nortriterpenoids and the ability to form wood traumatic ducts (Gadek et al., 1996; Kubitzki, 2011; Chuang

et al., 2022; Pace et al., 2022). However, in contrast to previous phylogenetic studies suggesting that Simaroubaceae and Meliaceae are sister families (Muellner et al., 2007; Appelhans et al., 2012; Muellner-Riehl et al., 2016), or that Simaroubaceae and Rutaceae are sister families (Gadek et al., 1996; Lin et al., 2018; Li et al., 2019; Ramírez-Barahona et al., 2020; Li et al., 2021), our analysis unequivocally inferred Meliaceae and Rutaceae as sister families. Regardless of what the crown age of the angiosperms was assumed to be, our dating analyses indicate that the three families diverged in an extremely short time, with Simaroubaceae splitting from the ancestor of Rutaceae and Meliaceae in the Lower Cretaceous, and Rutaceae and Meliaceae diverging from their common ancestor within the next 3–6 million years. As with the families in the KAB clade, the diversification of families in the SRM clade then followed to coincide with the peak or end of the Mid-Cretaceous Hothouse (in the UC-complete and RC-complete analysis, respectively) (Scotese, 2021). The rapid divergence of Simaroubaceae, Rutaceae, and Meliaceae could explain the difficulty of reconstructing familial relationships with the smaller datasets of previous studies. Ancient hybridization in the early history of these families could also explain contrasting family topologies in previous studies based on plastid data, and may be supported by the high degree of paralogy in Rutaceae and Meliaceae detected in this study relative to other Sapindales families. Given that our study includes an order of magnitude more data and denser sampling than previous phylogenetic studies of the order, we suggest that it is the most reliable representation of family relationships published to date. However, future research should be conducted with short- and long-read sequences to investigate whether ancient hybridization could be driving contrasting topologies between nuclear and plastid phylogenetic trees.

Morphological and anatomical data should be reconsidered in light of this new phylogenetic framework. Phytochemically, Meliaceae and Rutaceae are characterised by the presence of limonoid nortriterpenoids, while Simaroubaceae is characterised by quassinoid nortriterpenoids (Gadek et al., 1996; Fernandes da Silva et al., 2022). Recently, Chuang et al. (2022) found that biosynthesis of both quassinoids and limonoids have protolimonoid melianol as an intermediate compound, and that this pathway is controlled by conserved genes. Our new phylogenetic hypothesis for the SRM clade raises the possibility that the downstream change in metabolic pathway from protolimonoid melianol to produce limonoids could be a synapomorphy for sister families Meliaceae + Rutaceae. Further research is needed to test the synapomorphies for these groups and explore the potential of other putative phytochemical synapomorphies.

Divergence of the major Sapindales infra-familial clades

Within Sapindaceae, the three major clades retrieved are congruent with the recent infra-familial classification of Buerki et al. (2021), with S1 corresponding to the subfamily Dodonaeoideae, S2 including monotypic Xanthoceroideae and monophyletic Hippocastanoideae, and S3 equivalent to Sapindoideae (Figure 2A). Our dating analysis suggests that these subfamilies are ancient, with the divergence of Dodonaeoideae, Xanthoceroideae, Hippocastanoideae,

and Sapindoideae aligning with the Cenomanian–Turonian Thermal Maximum of the Upper Cretaceous, regardless of the crown age of the angiosperms (Bentham and Hooker, 1862; Hutchinson, 1926; Gadek et al., 1996; Savolainen et al., 2000). The rich data used in this study has enabled us to reconstruct subfamily topology with high support for the first time, particularly in relation to the position of Xanthoceroideae. This family includes one species (*Xanthoceras sorbifolium*), a deciduous shrub to small tree that inhabits xeric areas of China. *Xanthoceras* was originally included in Dodonaeoideae (Radlkofer, 1931), but due to its distinct morphological features, habitat, and placement as sister to the remainder of Sapindaceae in molecular studies, it was transferred to its own family and eventually subfamily of Sapindaceae (Harrington et al., 2005; Buerki et al., 2010; Buerki et al., 2021). In previous molecular studies with low numbers of loci, Xanthoceroideae was always retrieved as sister to the remainder of Sapindaceae with weak support (e.g., BS = 70 in Harrington et al., 2005; BS = 56 in Buerki et al., 2010; BS = 76 in Muellner-Riehl et al., 2016). The more recent Angiosperms353 phylogeny of Sapindaceae by Buerki et al. (2021), which focused on infra-familial taxonomy and did not test the topology within families, rooted their analysis on Xanthoceroideae, based on the tree of Muellner-Riehl et al. (2016), without outgroup samples from other families. When the tree of Buerki et al. (2021) is re-rooted, their topology agrees with ours, strongly indicating that Xanthoceroideae is in fact nested within Sapindaceae and sister to Hippocastanoideae (clade S2 of Figure 2A). This novel finding has interesting implications for our understanding of the biogeographical and morphological evolution of the family. Based largely on previous inferences of Xanthoceroideae relationships and its modern distribution, the origin of Sapindaceae was thought to be in Eurasia, with expansions into Gondwana during the Late Paleocene (Buerki et al., 2013). The subfamilial relationships presented here for Sapindaceae brings this interpretation of biogeographical history into question, and deserves further attention.

In Burseraceae, the major clades retrieved are largely congruent with the current classification of the family, with clades B1, B2, B3, and B4 corresponding to the *Boswellia*, *Canarium*, *Protium*, and *Bursera* alliances respectively, and *Beiselia* representing the monotypic *Beiselia* alliance (Daly et al., 2011). Most of these alliances are resolved as monophyletic here, the only exception being the *Boswellia* and *Canarium* alliances due to the placement of *Triomma* in the *Boswellia* alliance instead of the *Canarium* alliance (Daly et al., 2011). While the ages estimated in our UC-complete analyses are older than any study previously published, our RC-complete estimates are in agreement with the study of Weeks et al. (2014), where the *Beiselia* alliance (of which *Beiselia* is the sole, monospecific genus) was retrieved as sister to the rest of Burseraceae, having diverged from the ancestor of the remainder of Burseraceae in the Upper Cretaceous (at 85 [71–99] Ma). In contrast to Weeks et al. (2014), however, our RC-complete and CC-complete analyses suggest that the diversification of Burseraceae into the four major alliances was delayed until the Paleogene. More sampling, particularly in the *Protium* alliance and for the genus *Rosselia* (treated by Daly et al., 2011 as an unplaced genus), is needed to further corroborate infra-familial classification and clarify generic relationships.

Anacardiaceae relationships in this study are largely congruent with previous studies of the family, with our analyses supporting the recognition of two subfamilies: Spondioideae (= clade A1, Figure 2B),

and Anacardioidae without *Campnosperma* (= clade A2, Figure 2B) (Pell, 2004; Pell et al., 2011; Weeks et al., 2014). Historically, placement of *Campnosperma* has alternated between Anacardioidae and Spondioideae (Wannan and Quinn, 1990; Wannan and Quinn, 1991; Pell, 2004; Mitchell et al., 2006; Weeks et al., 2014). In our analyses, the position of *Campnosperma* was not fully resolved, being placed with Spondioideae in the concatenated tree, but as sister to Spondioideae + Anacardioidae in the coalescent analysis. *Campnosperma* diverged from its most recent common ancestor either in the Lower Cretaceous (when an ancient-angiosperm scenario is adopted), or the Upper Cretaceous (when angiosperms are considered to be younger), close to the time of the divergence of Spondioideae and Anacardioidae. This long history of independent evolution could explain the difficulty of classifying *Campnosperma* both in both morphological and molecular studies. Additional data may be needed to resolve its subfamilial assignment. As in previous phylogenetic studies of Anacardiaceae, Engler's (1876) tribal classification (*sensu* Mitchell and Mori, 1987) appears artificial, with all but Dobineae retrieved as polyphyletic, suggesting that taxonomic revision at this level is needed (Pell, 2004; Weeks et al., 2014). Revision of the generic limits of *Rhus*, *Poupartia*, and *Cyrtocarpa* is indicated on the basis of their non-monophyly in this analysis, as also suggested by previous studies (Pell et al., 2008; Herrera et al., 2018).

In Simaroubaceae, the three major clades retrieved are broadly congruent with those found in previous studies, with some exceptions. We confirm that *Holacantha* and *Castela* are sister to the rest of the family (clade Si1, Figure 2B); however, *Picrasma* falls within a paraphyletic *Ailanthus*, in clade Si2 (Figure 2B; Clayton et al., 2007; Clayton et al., 2009; Clayton, 2011). This surprising result may be caused in part by the relatively high degree of paralogy in clade Si2 compared to the rest of the family, or the omission of *Leitneria*, and warrants further investigation. The rest of the family (in Si3) forms a monophyletic grade, potentially explaining the labile taxonomic history of Simaroubaceae and the difficulty in identifying morphological synapomorphies in this group (Pirani et al., 2022).

In Meliaceae, the major clades retrieved support a two-subfamily classification system, with clade M1 (Figure 2B) equivalent to Cedreloideae, and clades M2 and M3 (Figure 2B) comprising a monophyletic Melioideae (Muellner et al., 2003; Muellner et al., 2006). However, as in previous molecular analyses, our results suggest that the morphological tribal classification of Pennington and Styles (1975) is in need of revision, with only Aglaieae and Melieae retrieved as monophyletic (Muellner et al., 2003; Muellner et al., 2006; Muellner et al., 2008; Koenen et al., 2015). The resolution of Melioideae has been greatly improved with the Angiosperms353 loci, and the topology (particularly of tribes Trichillieae and Turraeae) differs from previous molecular studies, suggesting that Melioideae comprises two main subclades. Subclade M2 contains *Melia*, *Azadirachta* (from tribe Melieae) and *Owenia* and *Pterorhachis* (from tribe Trichillieae). The relationship of *Owenia* with Melieae was previously shown by Muellner et al. (2008); Koenen et al. (2015) and Muellner-Riehl et al. (2016), but the placement of *Pterorhachis* differs substantially from previous classifications. In Melioideae subclade M3, *Munronia* is sister to the rest of Melioideae, which includes the remainder of Trichillieae split across two clades, Turraeae

(which was paraphyletic in relation to one of the clades of Trichilieae), the monotypic Vavaeae, and a paraphyletic Guareae in relation to Aglaieae. It is notable that in Meliaceae, particularly in Melioideae, there is extreme variation degree of paralogy in Angiosperms353 loci. This extreme variation in paralogy is in line with cytological studies that found Meliaceae has the highest variation in chromosome numbers in Sapindales (with a maximum $2n = 360$ in *Trichilia dregeana*, tribe Trichilieae), likely driven by repeated polyploidization events and occasional dysploidy (Guimarães and Forni-Martins, 2022). Although the influence of paralogy was reduced by the encoding of ambiguity characters in our sequences and our resulting tree is well-supported, we suggest that the extreme variation in paralogy may still affect the reconstructed topology and cause discordance with morphological taxonomic concepts, especially if the cause of paralogy is ancient hybridization (i.e., allopolyploidy). Examples of this could be in the surprising placements of *Pterorhachis* (which has low levels of paralogy relative to related genera) and *Munronia* (which has extremely high levels of paralogy for the order). Therefore, future phylogenetic and systematic studies on Meliaceae should focus on phasing gene copies to infer the type of gene duplication events that occurred in the evolutionary history of the family (i.e. autopolyploidization, allopolyploidization or duplication of certain regions), where they occurred, and how to reconstruct any reticulation events (e.g., Morales-Briones et al., 2021; Nauheimer et al., 2021).

The major clades retrieved in Rutaceae correspond broadly with the most recent subfamily classification (Appelhans et al., 2021): Rutoideae, Aurantoideae and the monotypic Haplophyloideae are monophyletic within clade R3 (Figure 2C). However, *Decatropis* and *Stauranthus* make Amyridoideae (within clade R3) and Zanthoxyloideae (most of R4; Figure 2C) non-monophyletic, with *Decatropis* falling in Amyridoideae instead of Zanthoxyloideae, and *Stauranthus* retrieved with Zanthoxyloideae instead of Amyridoideae. Both *Decatropis* and *Stauranthus* are small genera from Central America that have not been included in previous phylogenetic studies, and their assignment to these subfamilies should be further tested.

Most notably, our results suggest that Rutaceae subfamily Cneoroideae is polyphyletic, with genera split across two clades (R1 and R2; Figure 2C). Genera of clade R1 (Figure 2C) lack the characteristic glandular dots of typical Rutaceae (Appelhans et al., 2011), and until recently were assigned to Simaroubaceae (*Harrisonia*), Cneoraceae (*Cneorum*), and Ptaeroxylaceae (*Bottegoa*, *Cedrelopsis*, *Ptaeroxylon*). Likewise, genera of clade R2 (Figure 2C; *Dictyoloma*, *Sohnreyia*, *Spathelia*) were originally assigned to tribe Spathelieae of Simaroubaceae, based primarily on possession of staminal filament appendages and gynoecium structure, but were placed in Rutaceae by Engler (1931). The aforementioned genera (in clades R1 and R2 of our analysis) were recognised as a subfamily of Rutaceae based on weak to moderate phylogenetic support (Gadek et al., 1996; Chase et al., 1999; Groppo et al., 2008; Appelhans et al., 2011; Appelhans et al., 2021), similar biochemistry, and shared absence or restricted presence of schizogenous oil glands (Waterman, 1993; Waterman, 2007; Appelhans et al., 2021), although Groppo et al. (2008) noted that synapomorphies for the group were lacking. With the dense sampling and large number of loci included in our study, we have shown that Cneoroideae is not

monophyletic and is in need of taxonomic revision. Furthermore, the divergence of the Cneoroideae clades (R1 and R2) is likely to be ancient, occurring in the Upper Cretaceous (or Lower Cretaceous, if an ancient angiosperm scenario is assumed), in parallel with the divergence of Anacardiaceae and Burseraceae. Therefore, in combination with uncertain morphological synapomorphies uniting these clades with core Rutaceae (Appelhans et al., 2011; Groppo et al., 2012), reinstating the families Cneoraceae and/or Ptaeroxylaceae may be warranted.

As in the Melioideae of Meliaceae (M2 & M3; Figure 2B), extreme variation in paralogy was observed across genera in Rutaceae, particularly in clade R4 (mostly comprising subfamily Zanthoxyloideae). This may be driving topological conflict between the phylogenies produced with concatenated and coalescent analyses (particularly in the placement of *Decazyx*, *Flindersia*, *Nematolepis*, *Peltostigma*, *Plethadenia*, and *Ptelea*), and conflicting support for relationships of certain genera that are incongruent with working hypotheses derived mainly from plastome-based molecular phylogenetic trees and morphology (e.g., the non-monophyly of *Boronia* and the placement of *Correa*, *Halfordia*, *Muiriantha*, and *Phebalium*). Chromosome number is known to vary substantially across Rutaceae lineages, with a genome duplication event hypothesised to have occurred early in the evolution of the family followed by multiple polyploidization events (Appelhans et al., 2012; Paetzold et al., 2018; Appelhans et al., 2021; Guimarães and Forni-Martins, 2022). Future studies investigating the nature of gene duplication events in Rutaceae should be undertaken to improve our understanding and reconstruction of the evolution of the family, particularly subfamily Zanthoxyloideae.

Dense sampling of Sapindales genera and sequencing of Angiosperms353 loci has confirmed the monophyly of the nine currently recognised families and improved resolution of their relationships, but also indicated that the recognition of the previously accepted families Cneoraceae and Ptaeroxylaceae, which are currently placed in Rutaceae (clade R1), may be warranted. Regardless of the crown age of the angiosperms, Sapindales is clearly an ancient order, and its families emerged rapidly. Our results support the idea that Mid-Cretaceous climate change drove the diversification of angiosperm families, showing that the Mid-Cretaceous Hothouse likely had a substantial impact on the evolution of Sapindales. If the angiosperms are assumed to be ancient, the rising temperatures and Cenomanian–Turonian Thermal Maximum of the Mid-Cretaceous Hothouse period could have been a key period for the diversification of Sapindales families that were already present; if the angiosperms are assumed to be younger, the rising temperatures of the Mid-Cretaceous Hothouse period may have been coincident with the emergence of Sapindales families, and the cooling temperatures following the Cenomanian–Turonian Thermal Maximum was likely coincident with family diversification. It would be interesting to corroborate these results by investigating how diversification dynamics change with climate change, and if it is more likely that crown nodes of families would coincide with climate minima (such as the start of the Mid-Cretaceous Hothouse) or maxima (such as the Cenomanian–Turonian Thermal Maximum). In most families, infra-familial classifications need some revision, and our analysis may give insight into why infra-familial classification is so difficult: signals of gene duplication are heterogeneously dispersed

throughout the order, and are particularly strong in Rutaceae, Meliaceae and some Anacardiaceae (Anacardiaceae). Taken with evidence from cytological studies and the complex morphological patterns in these clades (Pennington and Styles, 1975; Appelhans et al., 2012; Paetzold et al., 2018; Appelhans et al., 2021; Guimarães and Forni-Martins, 2022), it points to a complex evolutionary history potentially involving local gene duplication, ancient hybridization (allopolyploidization) and autopolyploidization. These processes, especially ancient hybridization, may affect our ability to reconstruct the evolution of these clades as a bifurcating tree and interpret morphology (Morales-Briones et al., 2021; Nauheimer et al., 2021; Heslop-Harrison et al., 2022). Therefore, investigation of the processes responsible for these signals of gene duplication will be critical for furthering our understanding of evolutionary relationships in these clades. Moreover, the heterogeneous signal of gene duplication across Sapindales is interesting in itself: why do some families and clades in Sapindales retain a signal of gene duplication in their genome, while others don't? Thus, further investigation of the processes underlying gene duplication events may give key insight not only into the evolution of this ecologically and commercially important order, but angiosperm evolution more broadly.

Data availability statement

The datasets generated for this study can be found in the NCBI's Sequence Read Archive (SRA): <https://www.ncbi.nlm.nih.gov/sra>. Consensus sequences, alignments and tree files generated in this study are available at: doi: 10.5281/zenodo.7585555.

Author contributions

EJ conceived the study, along with DC, KN, KT and PAFTOL Principal Investigators WB, FF and IL. Sampling was conducted by EJ, MA, SB, JV, JP and MB. OM helped to manage and source samples. EJ conducted laboratory work with assistance from AZ. EJ carried out all analyses with support from AZ and LN. HS provided expertise in dating analysis and access to the PROTEUS database. MA, SB, MC, JV, JP, JB, MB, MC, MD, SP, MG, PL, JMi, CS, JMi, HO, CP, AW and AM-R contributed family expertise. EJ wrote the manuscript, with input from MA, SB, MC, JV, JP, JB, MB, MC, MD, SP, MG, PL, JMi, CS, JMi, HO, CP, LN, HS, AM-R, FF, KN, KT, WB and DC. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the Calvea Foundation to the Plant and Fungal Trees of Life (PAFTOL) project at the Royal Botanic Gardens, Kew. We also acknowledge the contribution of the Genomics for Australian Plants (GAP) Framework Initiative consortium in the generation of data for some samples used in this publication. GAP is supported by funding from Bioplatforms Australia (enabled by NCRIS), the Ian Potter Foundation, Royal Botanic Gardens Foundation (Victoria), Royal Botanic Gardens

Victoria, the Royal Botanic Gardens and Domain Trust, the Council of Heads of Australasian Herbaria, CSIRO, Centre for Australian National Biodiversity Research and the Department of Biodiversity, Conservation and Attractions, Western Australia. This research was also supported by an Australasian Systematic Botany Society Hansjörg Eichler Research Grant and a Wet Tropics Management Authority Research Grant to EJ. EJ was funded by the Australian Federal Government and the Prinzessin Therese von Bayern Foundation throughout the course of this work.

Acknowledgments

We would like to thank G. Brown, L. Simmons, D. Murphy, K. Shepherd, B. Lepschi and R. Fowler for the contribution of data for GAP samples, and the support of M. Lum and L. Simpson in facilitating their acquisition. EJ is grateful to M. Harrison, R. Cowan and L. Frankel for laboratory assistance, B. Wannan for helpful discussions on Anacardiaceae and S. Manchester, P. Wilf and A. Rozefelds for paleobotanical advice. We also thank C.J. Rothfels and three reviewers for their feedback which greatly improved this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Yanã Campos Rizzieri, School of Integrative Plant Sciences, Plant Biology Section, Cornell University, in collaboration with reviewer JO.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1063174/full#supplementary-material>

SUPPLEMENTARY MATERIAL 1

List of ingroup and outgroup samples included in the genus-level phylogenetic analysis of Sapindales and their NCBI SRA accession numbers.

SUPPLEMENTARY MATERIAL 2

Output of HybPhaser showing recovery statistics for loci, allele divergence, and the proportion of loci with more than 0% SNPs (heterozygosity), 0.5% SNPs, 1% SNPs and 2% SNPs.

SUPPLEMENTARY MATERIAL 3

Details of each fossil used for time-calibration of the Sapindales tree.

SUPPLEMENTARY MATERIAL 4

Trees from the concatenated alignment showing the allele divergence of each sample, and their proportion of loci with more than 0%, 0.5%, 1% and 2% SNPs.

SUPPLEMENTARY MATERIAL 5

Coalescent species tree of Sapindales generated from cleaned gene trees using ASTRAL v5.7.8, with node support shown as posterior probability.

SUPPLEMENTARY MATERIAL 6

Table 1 Ages and 95% HPD intervals for major clades in Sapindales under different dating schemes. 'C.' denotes crown node, 'S.' denotes stem node. KAB refers to major Sapindales clade with Kirkiaceae, Anacardiaceae and Burseraceae; SRM refers to major Sapindales clade with Simaroubaceae, Rutaceae and Meliaceae. **Figure 1** Comparison of chronograms estimated under a young-angiosperm (RC-complete; blue) and old-angiosperm (UC-complete; red) scenario. Thick lines mark the stem branch of Sapindales families for each dating analysis. Grey line under chronogram represents global average paleotemperature following Scotese (2021), with orange shading demarcating the Mid-Cretaceous Hothouse and asterisk (*) indicating the Cenomanian–Turonian Thermal Maximum.

References

- Alves, G. G. N., Fonseca, L. H. M., Devecchi, M. F., El Ottra, J. H. L., Demarco, D., and Pirani, J. R. (2022). What reproductive traits tell us about the evolution and diversification of the tree-of-heaven family, simaroubaceae. *Braz. J. Bot.* 45, 367–397. doi: 10.1007/s40415-021-00768-y
- APG (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1–20. doi: 10.1111/boj.12385
- Appelhaus, M. S., Bayly, M. J., Heslewood, M. M., Groppo, M., Verboom, G. A., Forster, P. I., et al. (2021). A new subfamily classification of the citrus family (Rutaceae) based on six nuclear and plastid markers. *Taxon* 70, 1035–1061. doi: 10.1002/tax.12543
- Appelhaus, M. S., Keßler, P. J. A., Smets, E., Razafimandimbison, S. G., and Janssens, S. B. (2012). Age and historical biogeography of the pantropically distributed spathelioideae (Rutaceae, sapindales). *J. Biogeogr.* 39, 1235–1250. doi: 10.1111/j.1365-2699.2012.02686.x
- Appelhaus, M. S., Smets, E., Razafimandimbison, S. G., Haevermans, T., van Marle, E. J., Couloux, A., et al. (2011). Phylogeny, evolutionary trends and classification of the *Spathelia*–*Ptaeroxylon* clade: Morphological and molecular insights. *Ann. Bot.* 107, 1259–1277. doi: 10.1093/aob/mcr076
- Atkinson, B. A. (2020). Fossil evidence for a Cretaceous rise of the mahogany family. *Am. J. Bot.* 107, 139–147. doi: 10.1002/ajb2.1416
- Bachelier, J. B., and Endress, P. (2008). Floral structure of *Kirkia* (Kirkiaceae) and its position in sapindales. *Ann. Bot.* 102, 539–550. doi: 10.1093/aob/mcn139
- Bachelier, J. B., Endress, P. K., and Craene, L. P. R. D. (2011). “Comparative floral structure and development of nitrariaceae (Sapindales) and systematic implications,” in *Flowers on the tree of life*. Eds. L. Wanntorp and L. P. R. D. Craene (Cambridge: Cambridge University Press), 181–217.
- Baker, W. J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., et al. (2022). A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Syst. Biol.* 71, 301–319. doi: 10.1093/sysbio/syab035
- Baker, W. J., Dodsworth, S., Forest, F., Graham, S. W., Johnson, M. G., McDonnell, A., et al. (2021). Exploring Angiosperms353: An open, community toolkit for collaborative phylogenomic research on flowering plants. *Am. J. Bot.* 108, 1059–1065. doi: 10.1002/ajb2.1703
- Barral, A., Gomez, B., Fourel, F., Daviero-Gomez, V., and Lécuyer, C. (2017). CO₂ and temperature decoupling at the million-year scale during the Cretaceous greenhouse. *Sci. Rep.* 7, 8310. doi: 10.1038/s41598-017-08234-0
- Barrett, C. F., Bacon, C. D., Antonelli, A., Cano, Á., and Hofmann, T. (2016). An introduction to plant phylogenomics with a focus on palms. *Bot. J. Linn. Soc.* 182, 234–255. doi: 10.1111/boj.12399
- Bentham, G., and Hooker, J. D. (1862). *Genera plantarum* (London, UK: Lovell Reeve & Co).
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Borowiec, M. L. (2016). AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4, e1660. doi: 10.7717/peerj.1660
- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15, e1006650. doi: 10.1371/journal.pcbi.1006650
- Bragg, J. G., Potter, S., Bi, K., and Moritz, C. (2016). Exon capture phylogenomics: efficacy across scales of divergence. *Mol. Ecol. Resour.* 16, 1059–1068. doi: 10.1111/1755-0998.12449
- Buerki, S., Callmander, M. W., Acevedo-Rodriguez, P., Lowry, P. P., Munzinger, J., Bailey, P., et al. (2021). An updated infra-familial classification of sapindaceae based on targeted enrichment data. *Am. J. Bot.* 108, 1234–1251. doi: 10.1002/ajb2.1693
- Buerki, S., Forest, F., Stadler, T., and Alvarez, N. (2013). The abrupt climate change at the Eocene–Oligocene boundary and the emergence of south-East Asia triggered the spread of sapindaceous lineages. *Ann. Bot.* 112, 151–160. doi: 10.1093/aob/mct106
- Buerki, S., Lowry, P. P., Alvarez, N., Razafimandimbison, S. G., Küpfer, P., and Callmander, M. W. (2010). Phylogeny and circumscription of sapindaceae revisited: Molecular sequence data, morphology and biogeography support recognition of a new family, xanthoceraceae. *Plant Ecol. Evol.* 143, 148–159. doi: 10.5091/plecevo.2010.437
- Burnham, R. J., and Carranco, N. L. (2004). Miocene winged fruits of *Loxopterygium* (Anacardiaceae) from the Ecuadorian Andes. *Am. J. Bot.* 91, 1767–1773. doi: 10.3732/ajb.91.11.1767
- Castañeda-Posadas, C., and Cevallos-Ferriz, S. R. (2007). *Swietenia* (Meliaceae) flower in late oligocene–early Miocene amber from simojovel de allende, chiapas, Mexico. *Am. J. Bot.* 94, 1821–1827. doi: 10.3732/ajb.94.11.1821
- Chandler, M. E. J. (1961). *The lower tertiary floras of southern England: I Palaeocene floras, London clay flora (Supplement)* (London: Order of the Trustees of the British Museum).
- Chase, M. W., Morton, C. M., and Kallunki, J. A. (1999). Phylogenetic relationships of rutaceae: A cladistic analysis of the subfamilies using evidence from rbcL and atpB sequence variation. *Am. J. Bot.* 86, 1191–1199. doi: 10.2307/2656983
- Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., et al. (1993). Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene rbcL. *Ann. Mo. Bot. Gard.* 80, 528–580. doi: 10.2307/2399846
- Chuang, L., Liu, S., Biedermann, D., and Franke, J. (2022). Identification of early quassinoid biosynthesis in the invasive tree of heaven (*Ailanthus altissima*) confirms evolutionary origin from protolimonoids. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.958138
- Clayton, J. W. (2011). “Simaroubaceae,” in *The families and genera of vascular plants - eudicots: Sapindales, cucurbitales, myrtaceae*. Ed. K. Kubitzki (Berlin: Springer-Verlag), 408–423.
- Clayton, J. W., Fernando, E. S., Soltis, P. S., and Soltis, D. E. (2007). Molecular phylogeny of the tree-of-heaven family (Simaroubaceae) based on chloroplast and nuclear markers. *Int. J. Plant Sci.* 168, 1325–1339. doi: 10.1086/521796
- Clayton, J. W., Soltis, P. S., and Soltis, D. E. (2009). Recent long-distance dispersal overshadows ancient biogeographical patterns in a pantropical angiosperm family (Simaroubaceae, sapindales). *Syst. Biol.* 58, 395–410. doi: 10.1093/sysbio/syp041
- Collinson, M. E., Manchester, S. R., and Wilde, V. (2012). *Fossil fruits and seeds of the middle Eocene messel biota, Germany* (Stuttgart: E. Schweizerbart'sche Verlagsbuchhandlung).

- Crane, P. R. (1990). A preliminary survey of fossil leaves and well-preserved reproductive structures from the sentinel butte formation (Paleocene) near almost, north Dakota. *Fieldiana Geol. New Ser.* 20, 1–63. doi: 10.5962/bhl.title.100826
- Cronn, R., Knaus, B. J., Liston, A., Maughan, P. J., Parks, M., Syring, J. V., et al. (2012). Targeted enrichment strategies for next-generation plant biology. *Am. J. Bot.* 99, 291–311. doi: 10.3732/ajb.1100356
- Cronquist, A. (1968). *The evolution and classification of flowering plants* (London: Thomas Nelson & Sons Ltd.).
- Daly, D. C., Harley, M. M., Martínez-Habibe, M. C., and Weeks, A. (2011). “Bursaceae,” in *The families and genera of vascular plants - eudicots: Sapindales, cucurbitales, myrtaceae*. Ed. K. Kubitzki (Berlin: Springer-Verlag), 76–104.
- Donoghue, M. J., and Sanderson, M. J. (2015). Confluence, synnovation, and depauperism in plant diversification. *New Phytol.* 207, 260–274. doi: 10.1111/nph.13367
- Dos Reis, M., and Yang, Z. (2013). The unbearable uncertainty of Bayesian divergence time estimation. *J. Syst. Evol.* 51, 30–43. doi: 10.1111/j.1759-6831.2012.00236.x
- Doyle, J., and Doyle, J. L. (1987). Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem. Bull.* 19, 11–15.
- Engler, A. (1876). “Anacardiaceae,” in *Flora brasiliensis*. Ed. C. F. P. Martius (Munich: F. Fleischer), 367–418.
- Engler, A. (1931). “Rutaceae,” in *Die natürlichen pflanzenfamilien*. Eds. A. Engler and H. Harms (Leipzig: Verlag von Wilhelm Engelmann), 187–359.
- Estrada-Ruiz, E., Martínez-Cabrera, H. I., and Cevallos-Ferriz, S. R. S. (2010). Upper Cretaceous woods from the olmos formation (late campanian–early maastrichtian), coahuila, Mexico. *Am. J. Bot.* 97, 1179–1194. doi: 10.3732/ajb.0900234
- Fernandes da Silva, M. F., das, G., da Silva Pinto, L., Amaral, J. C., Fernandes da Silva, D., Rossi Forim, M., et al. (2022). Nortriterpenes, chromones, anthraquinones, and their chemosystematics significance in meliaceae, rutaceae, and simaroubaceae (Sapindales). *Braz. J. Bot.* 45, 15–40. doi: 10.1007/s40415-021-00733-9
- Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. *Syst. Biol.* 19, 99–113. doi: 10.2307/2412448
- Foster, C. S. P., Sauquet, H., van der Merwe, M., McPherson, H., Rossetto, M., and Ho, S. Y. W. (2017). Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. *Syst. Biol.* 66, 338–351. doi: 10.1093/sysbio/syw086
- Freiberg, M., Winter, M., Gentile, A., Zizka, A., Muellner-Riehl, A. N., Weigelt, A., et al. (2020). LCVP, the Leipzig catalogue of vascular plants, a new taxonomic reference list for all known vascular plants. *Sci. Data* 416, 416. doi: 10.1038/s41597-020-00702-z
- Gadek, P. A., Fernando, E. S., Quinn, C. J., Hoot, S. B., Terrazas, T., Sheahan, M. C., et al. (1996). Sapindales: Molecular delimitation and infraordinal groups. *Am. J. Bot.* 83, 802–811. doi: 10.1002/j.1537-2197.1996.tb12769.x
- Gonçalves-Esteves, V., Cartaxo-Pinto, S., Marinho, E. B., Esteves, R. L., and Mendonça, C. B. F. (2022). Pollen morphology and evolutionary history of sapindales. *Braz. J. Bot.* 45, 341–366. doi: 10.1007/s40415-021-00719-7
- Grimsson, F., Bouchal, J. M., Xafis, A., and Zetter, R. (2020). Combined LM and SEM study of the middle Miocene (Sarmatian) palynoflora from the lavanttal basin, Austria: Part v. magnoliophyta 3 – myrtales to ericales. *Grana* 59, 127–193. doi: 10.1080/00173134.2019.1696400
- Groppo, M., Kallunki, J., Pirani, J., and Antonelli, A. (2012). Chilean *Pitavia* more closely related to Oceania and old world rutaceae than to Neotropical groups: Evidence from two cpDNA non-coding regions, with a new subfamilial classification of the family. *Phytokeys* 19, 9–29. doi: 10.3897/phytokeys.19.3912
- Groppo, M., Pirani, J. R., Salatino, M. L. F., Blanco, S. R., and Kallunki, J. A. (2008). Phylogeny of rutaceae based on two noncoding regions from cpDNA. *Am. J. Bot.* 95, 985–1005. doi: 10.3732/ajb.2007313
- Guimarães, R., and Forni-Martins, E. R. (2022). Chromosome numbers and their evolutionary meaning in the sapindales order: An overview. *Braz. J. Bot.* 45, 77–91. doi: 10.1007/s40415-021-00728-6
- Harrington, M. G., Edwards, K. J., Johnson, S. A., Chase, M. W., and Gadek, P. A. (2005). Phylogenetic inference in sapindaceae *sensu lato* using plastid matK and rbcL DNA sequences. *Syst. Bot.* 30, 366–382. doi: 10.1600/0363644054223549
- Hay, W. (2017). Toward understanding Cretaceous climate—an updated review. *Sci. China Earth Sci.* 60, 5–19. doi: 10.1007/s11430-016-0095-9
- Herrera, F., Carvalho, M. R., Jaramillo, C., and Manchester, S. R. (2019). 19-Million-Year-Old spondioid fruits from Panama reveal a dynamic dispersal history for anacardiaceae. *Int. J. Plant Sci.* 180, 479–492. doi: 10.1086/703551
- Herrera, F., Manchester, S. R., and Jaramillo, C. (2012). Permineralized fruits from the late Eocene of Panama give clues of the composition of forests established early in the uplift of central America. *Rev. Palaeobot. Palynol.* 175, 10–24. doi: 10.1016/j.revpalbo.2012.02.007
- Herrera, F., Mitchell, J. D., Pell, S. K., Collinson, M., Daly, D., and Manchester, S. (2018). Fruit morphology and anatomy of the spondioid anacardiaceae. *Bot. Rev.* 84, 315–393. doi: 10.1007/s12229-018-9201-1
- Heslop-Harrison, J. S., Schwarzacher, T., and Liu, Q. (2022). Polyploidy: its consequences and enabling role in plant diversification and evolution. *Ann. Bot.* mcacl32. doi: 10.1093/aob/mcacl32
- Hickey, L. J., and Hodges, R. W. (1975). Lepidopteran leaf mine from the early Eocene wind river formation of northwestern Wyoming. *Science* 189, 718–720. doi: 10.1126/science.189.4204.718
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281
- Hutchinson, J. (1926). *The families of flowering plants arranged according to a new system based on their probable phylogeny* (London: Macmillan & Co. Ltd.).
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916
- Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., et al. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *App. Plant Sci.* 4. doi: 10.3732/apps.1600016
- Johnson, M. G., Pokorný, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., et al. (2019). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Syst. Biol.* 68, 594–606. doi: 10.1093/sysbio/syy086
- Jones, K. E., Fér, T., Schmickl, R. E., Dikow, R. B., Funk, V. A., Herrando-Moraira, S., et al. (2019). An empirical assessment of a single family-wide hybrid capture locus set at multiple evolutionary timescales in asteraceae. *App. Plant Sci.* 7, e11295. doi: 10.1002/aps3.11295
- Junier, T., and Zdobnov, E. M. (2010). The newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics* 26, 1669–1670. doi: 10.1093/bioinformatics/btq243
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Knobloch, E., and Mai, D. H. (1986). *Monographie der fruchte und samen in der kreide von mitteleuropa* (Prague: Rozpravy Ústředního Ústavu Geologického).
- Koenen, E. J. M., Clarkson, J. J., Pennington, T. D., and Chatrou, L. W. (2015). Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytol.* 207, 327–339. doi: 10.1111/nph.13490
- Kubitzki, K. (2011). *Flowering plants. eudicots: Sapindales, cucurbitales, myrtaceae* (Berlin: Springer-Verlag).
- Lanfear, R., Calcott, B., Ho, S. Y. W., and Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* 29, 1695–1701. doi: 10.1093/molbev/mss020
- Larridon, I., Villaverde, T., Zuntini, A. R., Pokorný, L., Brewer, G. E., Epatwalage, N., et al. (2020). Tackling rapid radiations with targeted sequencing. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01655
- Li, H.-T., Luo, Y., Gan, L., Ma, P.-F., Gao, L.-M., Yang, J.-B., et al. (2021). Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biol.* 19, 1–13. doi: 10.1186/s12915-021-01166-2
- Li, H.-T., Yi, T.-S., Gao, L.-M., Ma, P.-F., Zhang, T., Yang, J.-B., et al. (2019). Origin of angiosperms and the puzzle of the Jurassic gap. *Nat. Plants* 5, 461–470. doi: 10.1038/s41477-019-0421-0
- Lin, N., Moore, M. J., Deng, T., Sun, H., Yang, L., Sun, Y., et al. (2018). Complete plastome sequencing from *Toona* (Meliaceae) and phylogenomic analyses within sapindales. *Appl. Plant Sci.* 6, e1040. doi: 10.1002/aps3.1040
- MacGinitie, H. D. (1953). *Fossil plants of the florissant beds, Colorado* (Washington: Carnegie institution of Washington).
- MacGinitie, H. D. (1969). *The Eocene green river flora of northwestern Colorado and northeastern Utah* (Berkeley: University of California Press).
- Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., and Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* 207, 437–453. doi: 10.1111/nph.13264
- Mai, U., and Mirarab, S. (2018). TreeShrink: Fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genom.* 19, 23–40. doi: 10.1186/s12864-018-4620-2
- Manchester, S. R. (2001a). Leaves and fruits of *Aesculus* (Sapindales) from the Paleocene of north America. *Int. J. Plant Sci.* 162, 985–998. doi: 10.1086/320783
- Manchester, S. R. (2001b). “Update on the megafossil flora of florissant, Colorado,” in *Fossil flora and stratigraphy of the florissant formation, Colorado* (Denver: Denver Museum of Nature and Science), 137–162.
- Manchester, S. R., and McIntosh, W. C. (2007). Late Eocene silicified fruits and seeds from the John day formation near post, Oregon. *PaleoBios* 27, 7–17.
- Manchester, S. R., and O’Leary, E. L. (2010). Phylogenetic distribution and identification of fin-winged fruits. *Bot. Rev.* 76, 1–82. doi: 10.1007/s12229-010-9041-0
- Manchester, S. R., Wilde, V., and Collinson, M. E. (2007). Fossil cashew nuts from the Eocene of Europe: Biogeographic links between Africa and south America. *Int. J. Plant Sci.* 168, 1199–1206. doi: 10.1086/520728
- McClain, A. M., and Manchester, S. R. (2001). *Dipteronia* (Sapindaceae) from the tertiary of north America and implications for the phylogeographic history of the aceroideae. *Am. J. Bot.* 88, 1316–1325. doi: 10.2307/3558343
- McDonnell, A. J., Baker, W. J., Dodsworth, S., Forest, F., Graham, S. W., Johnson, M. G., et al. (2021). Exploring Angiosperms353: Developing and applying a universal toolkit for flowering plant phylogenomics. *Appl. Plant Sci.* 9, e11443. doi: 10.1002/aps3.11443

- McLay, T. G. B., Birch, J. L., Gunn, B. F., Ning, W., Tate, J. A., Nauheimer, L., et al. (2021). New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Appl. Plant Sci.* 9, e11420. doi: 10.1002/aps3.11420
- Meyer, H. W., and Manchester, S. R. (1997). *The oligocene bridge creek flora of the John day formation, Oregon* (Berkeley: University of California Press).
- Mirarab, S., Reaz, R., Bayzid, M., Zimmermann, T., Swenson, M. S., and Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30, i541–i548. doi: 10.1093/bioinformatics/btu462
- Mitchell, J. D., and Mori, S. A. (1987). *The cashew and its relatives (Anacardium: Anacardiaceae)* (New York: Memoirs of the New York Botanical Garden).
- Mitchell, J. D., Daly, D. C., Pell, S. K., and Randrianasolo, A. (2006). *Poupartioopsis* gen. nov. and its context in anacardiaceae classification. *Syst. Bot.* 31, 337–348. doi: 10.1600/036364406777585757
- Mitchell, J. D., Pell, S. K., Bachelier, J. B., Warschewsky, E. J., Joyce, E. M., Canadell, L. C., et al. (2022). Neotropical Anacardiaceae (cashew family). *Braz. J. Bot.* 45, 139–180. doi: 10.1007/s40415-022-00793-5
- Morales-Briones, D. F., Gehrke, B., Huang, C.-H., Liston, A., Ma, H., Marx, H. E., et al. (2021). Analysis of paralogs in target enrichment data pinpoints multiple ancient polyploidy events in *alchemilla* s.l. (Rosaceae). *Syst. Biol.* 71. doi: 10.1093/sysbio/syab032
- Muellner, A. N., Samuel, R., Chase, M. W., Coleman, A., and Stuessy, T. F. (2008). An evaluation of tribes and generic relationships in melioideae (Meliaceae) based on nuclear ITS ribosomal DNA. *Taxon* 57, 98–108. doi: 10.2307/25065951
- Muellner, A. N., Samuel, R., Johnson, S. A., Cheek, M., Pennington, T. D., and Chase, M. W. (2003). Molecular phylogenetics of meliaceae (Sapindales) based on nuclear and plastid DNA sequences. *Am. J. Bot.* 90, 471–480. doi: 10.3732/ajb.90.3.471
- Muellner, A. N., Savolainen, V., Samuel, R., and Chase, M. W. (2006). The mahogany family “out-of-Africa”: Divergence time estimation, global biogeographic patterns inferred from plastid rbcL DNA sequences, extant, and fossil distribution of diversity. *Mol. Phylogenet. Evol.* 40, 236–250. doi: 10.1016/j.ympev.2006.03.001
- Muellner, A. N., Vassiliades, D. D., and Renner, S. S. (2007). Placing biebersteiniaceae, a herbaceous clade of sapindales, in a temporal and geographic context. *Plant Syst. Evol.* 266, 233–252. doi: 10.1007/s00606-007-0546-x
- Muellner-Riehl, A. N., Weeks, A., Clayton, J. W., Buerki, S., Nauheimer, L., Chiang, Y.-C., et al. (2016). Molecular phylogenetics and molecular clock dating of sapindales based on plastid rbcL, atpB and trnL-trnF DNA sequences. *Taxon* 65, 1019–1036. doi: 10.12705/655.5
- Nauheimer, L., Weigner, N., Joyce, E., Crayn, D., Clarke, C., and Nargar, K. (2021). HybPhaser: A workflow for the detection and phasing of hybrids in target capture data sets. *Appl. Plant Sci.* 9, 43–51. doi: 10.1002/aps3.11441
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Pace, M. R., Gerolamo, C. S., Onyenedum, J. G., Terrazas, T., Victorio, M. P., Cunha Neto, I. L., et al. (2022). The wood anatomy of sapindales: Diversity and evolution of wood characters. *Braz. J. Bot.* 45, 283–340. doi: 10.1007/s40415-021-00764-2
- Paetzold, C., Kiehn, M., Wood, K. R., Wagner, W. L., and Appelhans, M. S. (2018). The odd one out or a hidden generalist: Hawaiian *Melicope* (Rutaceae) do not share traits associated with successful island colonization. *J. Syst. Evol.* 56, 621–636. doi: 10.1111/jse.12454
- Pan, A. D. (2010). Rutaceae leaf fossils from the late oligocene (27.23 ma) guang river flora of northwestern Ethiopia. *Rev. Palaeobot. Palynol.* 159, 188–194. doi: 10.1016/j.revpalbo.2009.12.005
- Panchy, N., Lehti-Shiu, M., and Shiu, S.-H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523
- Panti, C. (2020). Fossil leaves of subtropical lineages in the Eocene–Oligocene of southern Patagonia. *Hist. Bot.* 32, 252–266. doi: 10.1080/08912963.2018.1487421
- Parham, J. F., Donoghue, P. C., Bell, C. J., Calway, T. D., Head, J. J., Holroyd, P. A., et al. (2012). Best practices for justifying fossil calibrations. *Syst. Biol.* 61, 346–359. doi: 10.1093/sysbio/syr107
- Pell, S. K. (2004). *Molecular systematics of the cashew family (Anacardiaceae)* (Baton Rouge (LA: Louisiana State University). [PhD dissertation].
- Pell, S. K., Mitchell, J. D., Lowry, P. P., and Randrianasolo, A. (2008). Phylogenetic split of Malagasy and African taxa of *Protorhus* and *Rhus* (Anacardiaceae) based on cpDNA trnL-trnF and nrDNA ETS and ITS sequence data. *Syst. Bot.* 33, 375–383. doi: 10.1600/036364408784571545
- Pell, S. K., Mitchell, J. D., Miller, A. J., and Lobova, T. A. (2011). “Anacardiaceae,” in *The families and genera of vascular plants. eudicots: Sapindales, cucurbitales, myrtaceae*. Ed. K. Kubitzki (Berlin: Springer-Verlag), 7–50.
- Pennington, T. D., and Styles, B. T. (1975). A generic monograph of the meliaceae. *Blumea* 22, 419–540.
- Pirani, J. R., Majure, L. C., and Devecchi, M. F. (2022). An updated account of simaroubaceae with emphasis on American taxa. *Braz. J. Bot.* 45, 201–221. doi: 10.1007/s40415-021-00731-x
- Radlkofer, L. A. T. (1931). “Sapindaceae,” in *Das Pflanzenreich*. Ed. A. Engler (Leipzig: Verlag von Wilhelm Engelmann).
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032
- Ramírez-Barahona, S., Sauquet, H., and Magallón, S. (2020). The delayed and geographically heterogeneous diversification of flowering plant families. *Nat. Ecol. Evol.* 4, 1232–1238. doi: 10.1038/s41559-020-1241-3
- Reid, E. M., and Chandler, M. E. J. (1933). *The London clay flora* (London: British Museum (Natural History)).
- Roma, L. P., and Santos, D. Y. A. C. (2022). A comprehensive review of the chemical composition and epicuticular wax morphology of the cuticle in sapindales. *Braz. J. Bot.* 45, 5–14. doi: 10.1007/s40415-021-00723-x
- Sauquet, H., Ramírez-Barahona, S., and Magallón, S. (2022). What is the age of flowering plants? *J. Exp. Bot.* 73, 3840–3853. doi: 10.1093/jxb/erac130
- Savolainen, V., Chase, M. W., Hoot, S. B., Morton, C. M., Soltis, D. E., Bayer, C., et al. (2000). Phylogenetics of flowering plants based on combined analysis of plastid atpB and rbcL gene sequences. *Syst. Biol.* 49, 306–362. doi: 10.1093/sysbio/49.2.306
- Sawangchote, P., Grote, P. J., and Dilcher, D. L. (2009). Tertiary leaf fossils of *Mangifera* (Anacardiaceae) from Li basin, Thailand as examples of the utility of leaf marginal venation characters. *Am. J. Bot.* 96, 2048–2061. doi: 10.3732/ajb.0900086
- Scotese, C. R. (2021). An atlas of phanerozoic paleogeographic maps: The seas come in and the seas go out. *Annu. Rev. Earth Planet Sci.* 49, 679–728. doi: 10.1146/annurev-earth-081320-064052
- Silvestro, D., Bacon, C. D., Ding, W., Zhang, Q., Donoghue, P. C. J., Antonelli, A., et al. (2021). Fossil data support a pre-Cretaceous origin of flowering plants *Nature Ecology & Evolution* 1–9. doi: 10.1038/s41559-020-01387-8
- Simoes, A. J. G., and Hidalgo, C. A. (2011). “The economic complexity observatory: An analytical tool for understanding the dynamics of economic development,” in *Scalable integration of analytics and visualization: Papers from the 2011 AAAI workshop*. Eds. O. J. Mengshoel, T. Selker and H. Lieberman (Menlo Park, CA: The AAAI Press), 39–42.
- Simoes, A. J. G., and Hidalgo, C. A. (2022). *OEC 5.0, the observatory of economic complexity*. Available at: <https://oec.world/> (Accessed August 2, 2022).
- Smith, M. L., and Hahn, M. W. (2021). New approaches for inferring phylogenies in the presence of paralogs. *Trends Genet.* 37, 174–187. doi: 10.1016/j.tig.2020.08.012
- Smith, S. A., Brown, J. W., and Walker, J. F. (2018). So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *PLoS One* 13, e0197433. doi: 10.1371/journal.pone.0197433
- Smith, S. A., and Dunn, C. W. (2008). Phytutility: A phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24, 715–716. doi: 10.1093/bioinformatics/btm619
- Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., et al. (2009). Polyploidy and angiosperm diversification. *Am. J. Bot.* 96, 336–348. doi: 10.3732/ajb.0800079
- Stevens, P. F. (2001). *Angiosperm phylogeny website version 14*. Available at: <http://www.mobot.org/MOBOT/research/APweb/> (Accessed August 30, 2022).
- Takhtajan, A. (2009). *Flowering plants: Class magnoliopsida (Dicotyledons)* (Dordrecht: Springer), 7–588.
- Terrazas, T. (1994). *Wood anatomy of the anacardiaceae: ecological and phylogenetic interpretation* (Chapel Hill (NC: University of North Carolina). [PhD dissertation].
- Tiffney, B. H. (1981). *Euodia costata* (Chandler) tiffney, (Rutaceae) from the Eocene of southern England. *Paläontol. Z.* 55, 185–190. doi: 10.1007/BF02988138
- Tölke, E. D., Medina, M. C., Souto, A. L., Marques, J. P. R., Alves, G. G. N., Gama, R. L., et al. (2022). Diversity and evolution of secretory structures in sapindales. *Braz. J. Bot.* 45, 251–279. doi: 10.1007/s40415-021-00778-w
- Ufimov, R., Gorospe, J. M., Fer, T., Kandziora, M., Salomon, L., Loo, M., et al. (2022). Utilizing paralogs for phylogenetic reconstruction has the potential to increase species tree support and reduce gene tree discordance in target enrichment data. *Mol. Ecol. Res.* 22, 3018–3034. doi: 10.1111/1755-0998.13684
- Wang, H., Moore, M. J., Soltis, P. S., Bell, C. D., Brockington, S. F., Alexandre, R., et al. (2009). Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl. Acad. Sci. U.S.A.* 106, 3853–3858. doi: 10.1073/pnas.0813376106
- Wang, Q., Manchester, S. R., Gregor, H.-J., Shen, S., and Li, Z. (2013). Fruits of *Koeleruteria* (Sapindaceae) from the Cenozoic throughout the northern hemisphere: their ecological, evolutionary, and biogeographic implications. *Am. J. Bot.* 100, 422–449. doi: 10.3732/ajb.1200415
- Wannan, B. S. (1986). *Systematics of the anacardiaceae and its allies* (Sydney (NSW: University of New South Wales). [PhD dissertation].
- Wannan, B. S., and Quinn, C. J. (1990). Pericarp structure and generic affinities in the anacardiaceae. *Bot. J. Linn. Soc.* 102, 225–252. doi: 10.1111/j.1095-8339.1990.tb01878.x
- Wannan, B. S., and Quinn, C. J. (1991). Floral structure and evolution in the anacardiaceae. *Bot. J. Linn. Soc.* 107, 349–385. doi: 10.1111/j.1095-8339.1991.tb00228.x
- Wannan, B. S., Waterhouse, J. T., Gadek, P. A., and Quinn, C. J. (1985). Biflavonyls and the affinities of *Blepharocarya*. *Biochem. Syst. Ecol.* 13, 105–108. doi: 10.1016/0305-1978(85)90066-3
- Waterman, P. G. (1993). “Phytochemical diversity in the order rutales,” in *Phytochemical potential of tropical plants*. Eds. K. R. Downum, J. T. Romeo and H. A. Stafford (New York: Springer), 203–233.
- Waterman, P. G. (2007). The current status of chemical systematics. *Phytochemistry* 68, 2896–2903. doi: 10.1016/j.phytochem.2007.06.029

Weeks, A., Zapata, F., Pell, S. K., Daly, D. C., Mitchell, J. D., and Fine, P. V. A. (2014). To move or to evolve: Contrasting patterns of intercontinental connectivity and climatic niche evolution in “Terebinthaceae” (Anacardiaceae and burseraceae). *Front. Genet.* 5. doi: 10.3389/fgene.2014.00409

Wettstein, R. (1901). *Handbuch der systematischen botanik* (Leipzig: F. Deuticke).

Xie, S., Manchester, S. R., Liu, K., Wang, Y., and Sun, B. (2013). *Citrus linczangensis* sp. n., a leaf fossil of rutaceae from the late Miocene of yunnan, China. *Int. J. Plant Sci.* 174, 1201–1207. doi: 10.1086/671796

Zhang, M., Dai, S., Du, B., Ji, L., and Hu, S. (2018). Mid-Cretaceous hothouse climate and the expansion of early angiosperms. *Acta Geol. Sin.* 92, 2004–2025. doi: 10.1111/1755-6724.13692