Received: 6 December 2021 | Revised: 29 March 2022 | Accepted: 1 June 2022

DOI: 10.1111/1755-0998.13666

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

A target capture approach for phylogenomic analyses at multiple evolutionary timescales in rosewoods (*Dalbergia* spp.) and the legume family (Fabaceae)

Simon Crameri[1] | Simone Fior[1] | Stefan Zoller[1,2] | Alex Widmer[1]

[1]Institute of Integrative Biology, ETH Zurich, Zürich, Switzerland

[2]Genetic Diversity Centre (GDC), ETH Zurich, Zürich, Switzerland

Correspondence
Simon Crameri and Alex Widmer, Institute of Integrative Biology, ETH Zurich, Zürich, Switzerland.
Emails: sfcrameri@gmail.com (S.C.); alex.widmer@usys.ethz.ch (A.W.)

Funding information
This work was supported by ETH Zurich and a grant from the Rübel Foundation to AW. The funders had no role in study design, data collection and analysis, or preparation of the manuscript

Handling Editor: C. Alex Buerkle

Abstract

Understanding the genetic changes associated with the evolution of biological diversity is of fundamental interest to molecular ecologists. The assessment of genetic variation at hundreds or thousands of unlinked genetic loci forms a sound basis to address questions ranging from micro- to macroevolutionary timescales, and is now possible thanks to advances in sequencing technology. Major difficulties are associated with (i) the lack of genomic resources for many taxa, especially from tropical biodiversity hotspots; (ii) scaling the numbers of individuals analysed and loci sequenced; and (iii) building tools for reproducible bioinformatic analyses of such data sets. To address these challenges, we developed target capture probes for genomic studies of the highly diverse, pantropically distributed and economically significant rosewoods (*Dalbergia* spp.), explored the performance of an overlapping probe set for target capture across the legume family (Fabaceae), and built the general purpose bioinformatic pipeline CaptureAl. Phylogenomic analyses of Malagasy *Dalbergia* species yielded highly resolved and well supported hypotheses of evolutionary relationships. Population genomic analyses identified differences between closely related species and revealed the existence of a potentially new species, suggesting that the diversity of Malagasy *Dalbergia* species has been underestimated. Analyses at the family level corroborated previous findings by the recovery of monophyletic subfamilies and many well-known clades, as well as high levels of gene tree discordance, especially near the root of the family. The new genomic and bioinformatic resources, including the Fabaceae1005 and Dalbergia2396 probe sets, will hopefully advance systematics and ecological genetics research in legumes, and promote conservation of the highly diverse and endangered *Dalbergia* rosewoods.

KEYWORDS
Dalbergia, Fabaceae, Leguminosae, phylogenomics, rosewood, target capture

1 | INTRODUCTION

The question how biological diversity evolves is of fundamental interest in ecology and evolution, and addressing it benefits from integrative approaches (Cutter, 2013; Rissler, 2016). Investigating evolutionary processes acting at the level of populations or groups of spatially interconnected populations (metapopulations) within species typically falls within the fields of population genetics and phylogeography. By contrast, analyses of evolutionary relationships among species and patterns of diversification in higher taxonomic groups fall within the realm of phylogenetics. Although it has long been recognized that "the same ecological and evolutionary processes that cause lineage divergence can also drive speciation" (Rissler, 2016), research in these fields has traditionally relied on different conceptual approaches, analytical methods, and molecular markers, generating a false dichotomy between fields aiming to address the same underlying processes. Today, the conceptualization of common theory combined with advances in methodology leveraging on next-generation sequencing (NGS) data offer the opportunity to jointly study the processes that drive the evolution of biological diversity from micro- to macroevolutionary timescales.

Target capture (Mamanova et al., 2010) provides an efficient approach to acquire molecular information across broad evolutionary timescales when genomic regions with varying level of diversity are included in the experimental design (Jones & Good, 2016). It requires the design of capture probes that target unique regions in the genome to prevent conflation of orthologues and paralogues, and are characterized by a conserved core for in-solution hybridization and more variable flanking regions expected to provide parsimony informative sites (Lemmon et al., 2012). Combined with high-throughput sequencing, this approach allows for the analysis of hundreds or thousands of orthologous loci in dozens to hundreds of individuals at moderate per-sample costs, and therefore strikes a good balance between locus information content and scalability to high numbers of individuals, including museum specimens (Brewer et al., 2019; de La Harpe et al., 2017). Hence, target capture holds a great potential to bridge the divide between phylogenetics, phylogeography and population genetics (de La Harpe et al., 2017; Nicholls et al., 2015; Rissler, 2016) and has increasingly been applied at macrozevolutionary, phylogeographic and microevolutionary timescales in a wide range of animals (e.g., Faircloth et al., 2012; Lemmon et al., 2012; Prum et al., 2015) and plants (e.g., de La Harpe et al., 2018; Koenen, Kidner, et al., 2020; Mandel et al., 2014).

A global probe set targeting 353 putatively single-copy protein-coding genes has recently been developed for flowering plants (Angiosperms353; Johnson et al., 2019). Recent studies in various plant families have shown that the Angiosperms353 probe set represents a cost-effective resource to resolve phylogenetic relationships at the level of plant orders (e.g., Thomas et al., 2021), families (e.g., Siniscalchi et al., 2021), or at the infrageneric level (e.g., Ottenlips et al., 2021). However, several comparisons revealed that microevolutionary relationships are often better resolved when targeting more loci using taxon-specific probe sets (e.g., Shah et al., 2021; Siniscalchi et al., 2021; Ufimov et al., 2021). The development of taxon-specific probe sets therefore remains valuable for detailed phylogenetic and population genetic analyses (Yardeni et al., 2022).

In addition to challenges associated with the de novo probe design, processing and analysis of high-throughput sequencing data often involves complex and computationally demanding calculations. Target capture data are often analysed using the PHYLUCE (Faircloth, 2016) or HybPiper (Johnson et al., 2016) bioinformatic pipelines. PHYLUCE was developed for analysis of sequences flanking ultraconserved genomic elements and has mainly been used at macro-evolutionary and phylogeographic timescales in animal systems, whereas HybPiper is optimized for data sets derived from probes designed in exons using Hyb-Seq (Weitemier et al., 2014). There is thus a need for existing tools to be expanded with pipelines that are applicable at deep to shallow evolutionary timescales (de La Harpe et al., 2017), while being independent from high-quality annotated genomes or transcriptomes.

Dalbergia L.f. (Fabaceae) is a pantropical and ecologically diverse plant genus with c. 270 currently accepted species (WCVP, 2021), some of which have been described relatively recently (e.g., Adema et al., 2016; Lachenaud, 2016; Wilding, Phillipson, Andriambololonera, et al., 2021; Wilding, Phillipson, & Crameri, 2021). Numerous arborescent species are a source of rosewood (Bosser & Rabevohitra, 2002; Prain, 1904), a high-quality timber sought-after on the international market and cause of conservation concern (Schuurman & Lowry II, 2009; Waeber et al., 2019). National and international regulations have been established, aiming at sustainable exploitation and revenues (Barrett et al., 2013; CITES, 2020), but illegal logging and trade continues (UNODC, 2016b, 2020; Vardeman & Runk, 2020). The effective implementation of regulations demands that species are reliably recognized and that extant population sizes are estimated to assess the potential threat status. Developing a comprehensive understanding of species diversity in *Dalbergia* and their evolutionary history, as well as a thorough knowledge of the ecology and distribution of many traded species, has been hampered by several factors. There is a shortage of collections and experts focusing on this taxonomically challenging genus, and current treatments heavily rely on leaves and flowers and/or fruits for identification (Bosser & Rabevohitra, 2002; de Carvalho, 1997; Lachenaud, 2016), which are rarely encountered together in the field. As a result, the taxonomy of the genus is in need of extensive revision (Wilding, Phillipson, Andriambololonera, et al., 2021), which could be supported by phylogenomic analyses targeting the nuclear genome (Crameri, 2020).

Motivated by the need for genomic resources to inform a reliable taxonomy and foster conservation practice, we introduce a target capture approach for anchored phylogenomic analyses in *Dalbergia* (Dalbergia2396 probe set). This genus belongs to the third largest angiosperm family (Fabaceae, a.k.a. Leguminosae or legume family), which is subject to extensive research in areas such as systematics (LPWG, 2017), ecology (Sprent et al., 2017), evolution (Koenen et al., 2021), speciation and rapid radiations (Hughes & Eastwood, 2006), and contains many agricultural crops (Mousavi-Derazmahalleh et al., 2018; Zhuang et al., 2019). This motivated us to further explore

the applicability of our approach for analyses across the entire legume family, which resulted in a second probe set (Fabaceae1005 probe set). Both probe sets represent a subset of 6555 conserved target regions distributed across the nuclear genome, derived from a combination of divergent reference capture using five published legume genomes, and a de novo assembly of a *Dalbergia* transcriptome. We also introduce a dedicated bioinformatic pipeline named CaptureAl supporting the analysis of high-throughput target capture sequencing data, with special emphasis on streamlined applicability, parallelization, and graphical output for informed parameter choices. The pipeline is designed for general application to target capture data sets, modular, and therefore easily customizable. We demonstrate the application of our approach to resolve phylogenetic relationships in the economically important and conservation-relevant genus *Dalbergia*. We then explore the utility for phylogenomic analyses at much deeper timescales by analysing target capture data of various legume subfamilies. Finally, we test the utility of this approach at a microevolutionary scale, and assess genetic variation among individuals and populations of two closely related *Dalbergia* species from Madagascar.

2 | MATERIALS AND METHODS

2.1 | Design of target capture probes and reference sequences

We produced a transcriptome assembly of a cultivated individual of *Dalbergia madagascariensis* subsp. *antongilensis* Bosser & R. Rabev., based on 63 million paired-end sequencing reads generated on an Illumina HiSeq 2000 platform. We performed de novo assembly of the transcriptome using Trinity release 2012-01-25 (Grabherr et al., 2011), resulting in 146,484 scaffolds, which were between 201 and 17,129 bp long, with a mean length of 815 bp (see Appendix S1). We then pairwise aligned the *Dalbergia* transcriptome with reference genomes of five legume species available in public databases to generate a set of 12,049 probes extracted from the *Cajanus cajan* (L.) Millsp. v1.0 genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_000340665.1) targeting 6555 conserved genomic regions (see Appendix S1). This probe set was used for synthesis of hybridizing probes at myBaits Custom Target Capture Kits (Arbor Biosciences; https://arborbiosci.com).

2.2 | Taxon sampling for target capture probes validation

We created three taxon sets with contrasting levels of evolutionary divergence, ranging from subfamilies to species to populations. The subfamily set (Table S1) included five of the six legume subfamilies, as recognized in the most recent treatment (LPWG, 2017), and comprised 104 individuals (110 samples, six replicates; 99 species including three outgroups). Three species of *Polygala* Tourn. ex L. (Polygalaceae) were included as the outgroup for the subfamily set. The species set (Table S2) included members of the closely related genera *Dalbergia* (at least 19 species), *Machaerium* Pers. (three species) and *Ctenodon* Baill. sensu Cardoso et al. (2020) (two species) and comprised 60 individuals (63 samples, three replicates; at least 26 species including two outgroups). Two species of *Aeschynomene* L. sensu stricto (s.str.) sensu Cardoso et al. (2020) were included as the outgroup for the species set. The population set (Table S3, Figure S4) included 51 individuals in total, 29 attributed to *D. monticola* Bosser & R. Rabev. from four sampling locations, and 22 attributed to *D. orientalis* Bosser & R. Rabev. from 11 sampling locations.

2.3 | Library preparation, target capture and sequencing

We extracted total genomic DNA from silica gel dried leaf tissue (185 extractions) or museum specimens deposited at the Paris (P) herbarium (11 extractions) using the CTAB protocol (Doyle & Doyle, 1987) or the DNeasy Plant Mini Kit (Qiagen). We quantified DNA using the QuantiFluor dsDNA system for a Quantus fluorometer (Promega) and checked DNA integrity on 1.5% agarose gels for a subset of samples. We prepared genomic DNA libraries for each sample using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs), following the manufacturer's instructions. We individually indexed samples to be pooled within the same sequencing lane during the PCR enrichment step using NEBNext Multiplex Oligos for Illumina (single-indexed with E7335 and E7500 kits, or dual-indexed with E6440 kit, New England Biolabs). We performed in-solution hybridization and target enrichment using our 12,049 tiled RNA probes. We pooled up to six individually indexed libraries during the hybridization step using a stratified random assignment of libraries to hybridization reactions. Stratification aimed at optimizing the sequencing coverage across samples and consisted in avoiding pooling of close relatives of *Cajanus cajan* with more distantly related samples, and of museum specimens with silica gel dried leaf material. We obtained short read data by combining sequencing runs from an Illumina MiSeq (2 × 300 bp paired-end sequencing, 99 libraries) at the Genetic Diversity Centre (GDC) Zurich, an Illumina HiSeq 4000 (2 × 150 bp paired-end sequencing, 88 libraries) at the Functional Genomics Center Zurich (FGCZ) or Fasteris SA (Plan-les-Ouates, Switzerland), and an Illumina NovaSeq 6000 SP flow cell (2 × 150 bp paired-end sequencing, 9 libraries) at the FGCZ. We repeated DNA extraction, hybridization and target enrichment sequencing for nine individuals (replicates) to assess reproducibility. One sample (*Hassold 565*) was represented in each taxon set, nine samples were represented in both the species and population sets, and nineteen samples were represented in both the subfamily and species sets.

2.4 | CaptureAl bioinformatics pipeline

The bioinformatic pipeline CaptureAl was developed for this project and is accessible on GitHub (https://github.com/scrameri/Captu

reAl) as a documented sequence of scripts. These include bash and R scripts (R Core Team, 2022) to manage and visualize data with APE version 5.3 (Paradis & Schliep, 2018), DATA.TABLE version 1.12 (Dowle & Srinivasan, 2019), and TIDYVERSE version 1.3.0 (Wickham et al., 2019). Where appropriate, computations are carried out for multiple samples or target regions in parallel using GNU PARALLEL (Tange, 2011). The CAPTUREAL pipeline streamlines the mapping of quality-trimmed reads to target regions, the exclusion of loci targeting multicopy genes and taxa with insufficient data coverage, the alignment of orthologous loci for downstream phylogenetic analyses, and the generation of longer and taxon-specific reference sequences for population genomic analyses. At various critical steps, the pipeline outputs summary statistics and graphs that inform the user on the effects of specific filtering parameters, allowing for informed parameter choices.

The pipeline is divided into seven steps to process quality-filtered reads (Figure 1). Steps 1 to 5 are always required, and (1) map the sequencing reads to target regions, (2) assemble mapped reads separately for each target region, (3) identify the most-likely orthologous contigs, (4) identify taxa and target regions with high capture sensitivity and specificity, and (5) create trimmed alignments of the kept taxa and target regions. Steps 6 and 7 are optional, and (6) combine physically neighbouring and overlapping alignments to (7) generate longer and more representative reference sequences as starting points for population genomic analyses and reiteration of steps 1 to 5. Such reiteration can improve mapping success, and can mitigate potential biases arising from the initial reference sequences used (Hahn et al., 2013). This can improve the assembly and alignment of target regions for phylogenetic analyses, and remapping Illumina reads to taxon-specific reference sequences can also improve variant calling and subsequent population genomic analyses. CAPTUREAL includes scripts to filter raw variants and to convert them to *genind* and *genlight* objects for various population genetic and genomic analyses in R, and to sample single nucleotide polymorphisms (SNPs) for STRUCTURE analyses (Pritchard et al., 2000) and similar programs.

In our analyses, we executed the pipeline separately and iteratively for different taxon sets. We first applied steps 1 to 5 to 12 representative samples each of the subfamily and species sets, followed by steps 6 and 7 to generate longer and taxon-specific reference sequences for target regions that were efficiently enriched in these taxon sets. We then reiterated steps 1 to 5 for all samples of the subfamily and species sets using the new reference sequences and more stringent target region filtering parameters (see Appendix S1). We also performed steps 6 and 7 after the second iteration of the species set analysis to produce reference sequences for population genomic analyses of the population set. Bioinformatic analyses were carried out on a multicore LINUX server (GDC Zurich) or on the EULER scientific compute cluster (ETH Zurich). The sequence of executed commands and the chosen parameters are provided in Appendix S1.

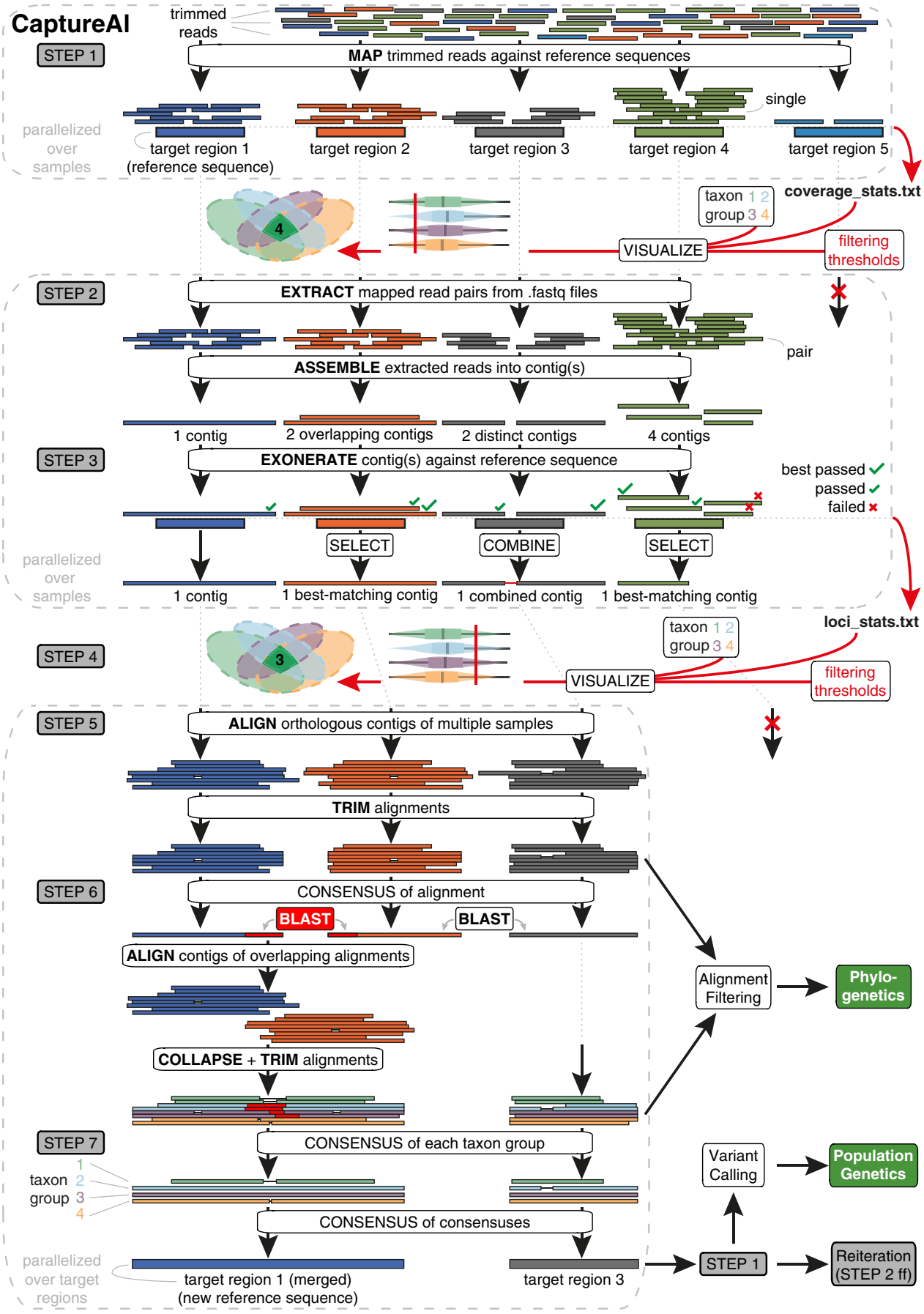
2.4.1 | Step 1: Read mapping

Quality-filtered reads of each sample are mapped against the reference sequences (one sequence per target region) using the BWA-MEM algorithm (Li, 2013). The minimum alignment score and mapping quality can be adjusted as needed. Coverage statistics are computed using SAMTOOLS (Li & Durbin, 2009) and BEDTOOLS (Quinlan & Hall, 2010), and target regions are filtered for adequate coverage across samples using filter.visual.coverages.R, which allows to apply filtering thresholds that are informed by visualizations of coverage statistics (see Appendix S1). The main output of step 1 are BAM files, coverage statistics, and a list of retained target regions.

2.4.2 | Step 2: Sequence assembly

Read pairs are extracted from quality-filtered reads when at least one read mapped to any of the retained target regions with the specified minimum mapping quality. Extracted reads are assembled separately

FIGURE 1 Seven bioinformatic steps of the CAPTUREAL pipeline. Steps 1–3 are shown for a single sample, but are executed for multiple samples in parallel, and steps 5–7 are shown for three target regions, but are executed for many target regions in parallel (indicated by dashed grey lines). In STEP 1, trimmed reads (narrow bars) are mapped to target reference sequences (broad bars; five regions are shown). Coverage statistics are generated and written to coverage_stats.txt, which informs filters applied to poorly sequenced samples and target regions (shown in red; four regions are retained, region 5 does not pass the filters due to low coverage in one or more taxon groups). In STEP 2, read pairs are extracted and assembled separately for each sample and each target region, resulting in zero (not shown) or one (region 1) to multiple contigs (region 4). In STEP 3, contigs are aligned to their respective reference sequence using EXONERATE to select the likely orthologous contig(s). Nonoverlapping contigs with normalized alignment scores passing a user-specified threshold (green ticks) are combined to a single sequence (region 3). If contigs overlap, only the best-matching contig is selected (large green ticks; region 2), and contigs with alignment scores below the threshold (red crosses) are discarded (region 4). Contig statistics are generated and written to loci_stats.txt, which informs the filters applied to poorly assembled samples and target regions in STEP 4 (shown in red; three regions are retained, region 4 does not pass the filters due to high prevalence of multiple contigs in one or more taxon groups). In STEP 5, the contigs of multiple samples (eight are shown for three target regions) are aligned and trimmed, generating a data set potentially suitable for phylogenomic analyses. In the optional STEP 6, a consensus sequence is generated for each target region alignment, and overlaps between neighbouring regions are identified using BLAST+ (shown in red). Individual contigs from such regions are aligned, collapsed to a single sequence per sample, and trimmed. The merged alignments can be used as replacements for overlapping alignments and filtered for phylogenomic analyses as from Step 5. In the optional STEP 7, consensus sequences for each alignment are generated for each taxon group, as well as overall consensus sequences across all taxon groups. These can serve as longer and taxon-specific reference sequences for STEP 1. Remapped reads can then be used for variant calling and population genomic analyses, or to refine target region assembly, alignment, and downstream analyses by repeating Steps 1–5 or 1–7


CaptureAI
trimmed reads
STEP 1
MAP trimmed reads against reference sequences
single
parallelized over samples
target region 1 (reference sequence)
target region 2
target region 3
target region 4
target region 5
coverage_stats.txt
taxon 1 2
group 3 4
4
VISUALIZE
filtering thresholds
STEP 2
EXTRACT mapped read pairs from .fastq files
pair
ASSEMBLE extracted reads into contig(s)
1 contig
2 overlapping contigs
2 distinct contigs
4 contigs
STEP 3
EXONERATE contig(s) against reference sequence
best passed
passed
failed
SELECT
COMBINE
SELECT
parallelized over samples
1 contig
1 best-matching contig
1 combined contig
1 best-matching contig
loci_stats.txt
STEP 4
3
taxon 1 2
group 3 4
VISUALIZE
filtering thresholds
STEP 5
ALIGN orthologous contigs of multiple samples
TRIM alignments
STEP 6
CONSENSUS of alignment
BLAST
BLAST
ALIGN contigs of overlapping alignments
COLLAPSE + TRIM alignments
Alignment Filtering
Phylo-genetics
STEP 7
CONSENSUS of each taxon group
taxon group
1
2
3
4
Variant Calling
Population Genetics
CONSENSUS of consensuses
parallelized over target regions
target region 1 (merged) (new reference sequence)
target region 3
STEP 1
Reiteration (STEP 2 ff)

for each sample and target region using dipSPAdes (Safonova et al., 2015). In the process, dipSPAdes uses contiguous sequences generated by SPAdes (Bankevich et al., 2012) and seeks to collapse overlapping contiguous sequences representing alternate alleles of heterozygous individuals (haplocontigs) to a single sequence, and to extend collapsed sequences to fewer and longer consensus contiguous sequences (contigs hereafter; Safonova et al., 2015). The main output of step 2 is a FASTA file with zero to multiple contigs for each sample and each target region.

2.4.3 | Step 3: Orthology assessment

Sequence assembly may yield multiple contigs per sample for some target regions, e.g., due to capture of several fragments of the same genomic region (e.g., in degraded museum specimens), due to unspecific capture of paralogues (Johnson et al., 2016), or haplocontigs that failed to be merged into a consensus contig by dipSPAdes. The most likely orthologous contig(s) of each sample and each target region are determined using the exhaustive Smith-Waterman alignment (Smith & Waterman, 1981) implemented in exonerate (Slater & Birney, 2005). For each sample and target region, the contig showing the highest exonerate alignment score to the target region reference sequence is assumed to be the most likely orthologous contig of that target region, if the normalized alignment score (defined as the raw exonerate alignment score divided by the target alignment length) meets a user-specified threshold. Any other contig that overlaps with the best-matching contig but shows a lower alignment score is assumed to be paralogous or an alternative haplocontig, and is discarded from downstream analyses. Further contigs that do not overlap with one another or the best-matching contig, but align to other parts of the target region with the required normalized alignment score, are retained. Such contigs were often observed in assemblies of degraded museum specimens and probably represent fragments of the same region. These fragments can therefore be prepended or appended to the best-matching contig using the directionality and number of gap-filling characters (hyphens) as indicated by the exonerate alignment statistics, to form a single contiguous sequence per sample and target region (orthologous contig hereafter, see Appendix S1). The main output of step 3 is a text file containing the exonerate alignment statistics, and a single-sequence FASTA file with the putative orthologous contig for each sample and each target region.

2.4.4 | Step 4: Sample and region filtering

Successful target capture depends on whether sequence data can be collected for a high proportion of target regions (capture sensitivity; Jones & Good, 2016) in a high proportion of focal taxa, and whether the captured sequences are orthologues of the target regions (capture specificity). Target regions are filtered for high capture sensitivity and specificity across focal taxa using filter.visual.assemblies.R, which applies filtering thresholds that are set by the user as informed by visualizations of exonerate alignment statistics generated in step 3, before any contigs are selected or combined. Capture sensitivity thresholds can be set globally to remove generally poorly sequenced samples (i.e., minimum fraction of target regions with at least one contig) or target regions (i.e., minimum fraction of samples with at least one contig). Capture specificity thresholds can be set as the required fraction of samples belonging to a specified taxon group and passing a certain filtering threshold in order for a target region to be retained. If taxon groups are defined, the indicated capture sensitivity and specificity thresholds need to be met in each considered taxon group, thus preventing target regions from being systematically missing in taxon groups with few available samples (see Appendix S1). The main output of step 4 is a list of samples and a list of target regions to keep.

In our analyses, we defined the four subfamilies represented by multiple taxa as taxon groups in the subfamily set. In the species set we defined four taxon groups based on our preliminary phylogenomic results and phylogenetic relationships inferred by Hassold et al. (2016). These were subgroup (SG) 1 (species with large flowers and paniculate inflorescences), SG2 (species with large flowers and racemose inflorescences), SG3 (species with small flowers from East Madagascar), and SG4 (species with small flowers from West and North Madagascar). The second and more stringent iteration required at least 75% (subfamily set) or 85% (species set) of samples to have one or more contigs, as indicated by the exonerate alignment statistics generated in step 3, for a target region to be retained. Furthermore, we only retained target regions for which no more than two contigs had been assembled in at least 50% (subfamily set) or 70% (species set) of samples of each taxon group (see Appendix S1).

2.4.5 | Step 5: Target region alignment and alignment trimming

A multi-sequence FASTA file is generated for all retained target regions, containing the respective orthologous contigs of all retained samples. Sequences are then aligned using mafft (Katoh & Standley, 2013), allowing for different alignment options. Alignments are trimmed at both ends until an alignment site shows nucleotides across a specified minimum fraction of aligned sequences, along with a specified maximum nucleotide diversity (i.e., the mean number of base differences between all sequence pairs). In addition, internal trimming is performed by only keeping sites with nucleotides in a specified minimum fraction of aligned sequences. Potential mis-assemblies or misalignments at contig ends are further resolved using a sliding window approach that identifies and masks sequences with large deviations from the alignment consensus (see Appendix S1). The main output of step 5 are trimmed alignments for each kept target region.

2.4.6 | Step 6: Merging of overlapping alignments

This optional step aims at resolving potential overlaps between physically close alignments generated in step 5, which may

negatively affect downstream analyses. Overlaps can be identified by aligning consensus sequences of target region alignments. Specifically, consensus sequences are generated by calling IUPAC ambiguity codes if a given minor allele frequency threshold across the alignment is reached, or a gap if a given base frequency threshold is not reached. Local alignments between different consensus sequences are identified using BLAST+ version 2.7.1 (Camacho et al., 2009), and filtered for nonreciprocal hits between alignment ends of target regions located on the same genomic scaffold or chromosome (identified from the name of the target region, if available). All orthologous contigs that are part of different, overlapping alignments are then written to a single FASTA file and aligned using MAFFT. The resulting alignments consist of more sequences than samples and need to be collapsed to represent neighbouring target regions of the same individual as a single sequence (supercontig), an automatic process that can be visually inspected. Trimming is then applied as in step 5, and sets of two to several consecutively overlapping alignments are then each replaced by a single merged alignment if merging was successful (see Appendix S1). The main output of step 6 are nonoverlapping trimmed alignments for each kept target region.

2.4.7 | Step 7: Generation of representative reference sequences

In this optional step, the target region alignments generated in the two previous steps are used to produce longer and more representative target region reference sequences. This can mitigate potential shortcomings or biases arising from the reference sequences used in step 1. For this purpose, a consensus sequence is generated for each alignment as in step 6, but separate consensus sequences can be generated for different specified taxon groups (see step 4). These sets of taxon group specific consensus sequences are then aligned, and representative consensus sequences are generated as in step 6 (see Appendix S1). These taxon-specific reference sequences are the main output of step 7 and can be used to refine mapping, assembly and alignment by reiterating steps 1 to 5.

2.4.8 | Alignment assessment and filtering

We characterized all nonoverlapping trimmed alignments for the number of gaps, gap ratio (i.e, the fraction of non-nucleotides in the alignment), total nucleotide diversity, average nucleotide diversity per site, and alignment length, as well as the number and proportion of segregating and parsimony informative sites. We then filtered alignments using filter.visual.alignments.R, which allows filtering thresholds that are informed by visualizations of alignment statistics (see Appendix S1) to be applied. We used the filtered alignments after the second iteration of step 6 for phylogenomic analyses.

2.5 | Phylogenomic analyses

We performed phylogenomic analyses with both the subfamily and species sets, using a supermatrix (concatenation) approach and a gene tree summary approach. For the supermatrix approach, we ran maximum likelihood searches on the concatenated alignments using RAxML version 8.2.11 (Stamatakis, 2014) with rapid bootstrap analysis and search for the best-scoring tree in the same run (–f a option), 100 bootstrap replicates, and the GTRCAT approximation of rate heterogeneity (see Appendix S1). For the gene tree summary approach, we ran RAxML jobs separately for each alignment using the same settings as for the supermatrix approach to generate gene trees. Following Zhang et al. (2018), we collapsed branches in gene trees if they had bootstrap support values below 10 using NEWICK utilities (Junier & Zdobnov, 2010), and we performed species tree analyses with ASTRAL-III version 5.6.3 (Mirarab et al., 2014; Zhang et al., 2018) and standard parameters, except for full branch annotation (see Appendix S1). For the subfamily set, we additionally evaluated the quartet support for 15 different subfamily topologies (i.e., all possible topologies with Caesalpinioideae, Dialioideae, Papilionoideae and [Cercidoideae, Detarioideae] as ingroups; Figure S2), using the tree scoring option in ASTRAL-III and a file with the assignment of taxa to subfamilies or the outgroup. All phylogenetic trees were displayed using GGTREE version 2.0.2 (Yu et al., 2016).

2.6 | Population genomic analyses

We carried out population genomic analyses for the population set by relying on SNPs inferred from variant calling, which was based on mapped reads and revealed individual-level allelic variation. We mapped quality-filtered reads against the target region reference sequences that were representative of the species set after the second iteration using BWA-MEM. We verified efficient recovery of target regions by plotting heatmaps of coverage statistics, removed PCR duplicates using PICARD TOOLS version 2.21.3 (Broad Institute, 2019), and capped excessive coverage to 500 using biostar154220.jar (Lindenbaum, 2015). We then called variants using FREEBAYES version 1.1.0-3-g961e5f3 (Garrison & Marth, 2012) and standard parameters, except for a minimum alternate fraction of 0.05, a minimum repeat entropy of 1, and evaluation of only the four best alleles. Variants were filtered using VCFTOOLS version 0.1.15 (Danecek et al., 2011) and VCFLIB version 1.0.1 (Garrison, 2012), which was also used to decompose complex variants (see Appendix S1). We then used VCFR version 1.10.0 (Knaus & Grünwald, 2017) and ADEGENET version 2.1.1 (Jombart, 2008; Jombart & Ahmed, 2011) to generate *genind* and *genlight* objects that represented the SNP allele table with associated metadata such as individual missingness, species identification, and sampling location. We used the SNP subset with zero missingness to conduct principal component analysis (PCA) based on the centred covariance matrix, as well as to calculate a neighbour-joining

(NJ) tree (Saitou & Nei, 1987) on Nei's genetic distances, as implemented in POPPR version 2.8.1 (Kamvar et al., 2014). We also used the allele table to create a SNP subset for population clustering analysis using STRUCTURE version 2.3.4 (Pritchard et al., 2000). Specifically, we kept SNPs with genotype data in at least 95% of individuals, and we randomly sampled three SNPs per target region (or all SNPs if less than three SNPs remained) for computational ease. STRUCTURE analyses were performed for one to ten demes (*K*), using 110,000 iterations, including a burnin period of 10,000 iterations, with ten replicates per simulation (see Appendix S1). Replicate STRUCTURE results were aligned and visualized using CLUMPAK (Kopelman et al., 2015) and default settings.

2.7 | Comparison to the Angiosperms353 probe set

We used BLAST+ to search for correspondence between our two final sets of reference sequences (i.e., the consensus sequences of genomic regions targeted by the Fabaceae1005 and Dalbergia2396 probe sets, respectively) and 11,781 supercontigs representing 42 angiosperm taxa and 353 target regions of the Angiosperms353 probe set (Johnson et al., 2019). We performed reciprocal BLAST+ searches applying a maximum expect value (evalue) of 1E-04, a minimum of 100 bp alignment length, and retaining only the best hits between the same pair of target regions (see Appendix S1).

3 | RESULTS

3.1 | Two probe sets for target capture across legumes and *Dalbergia*

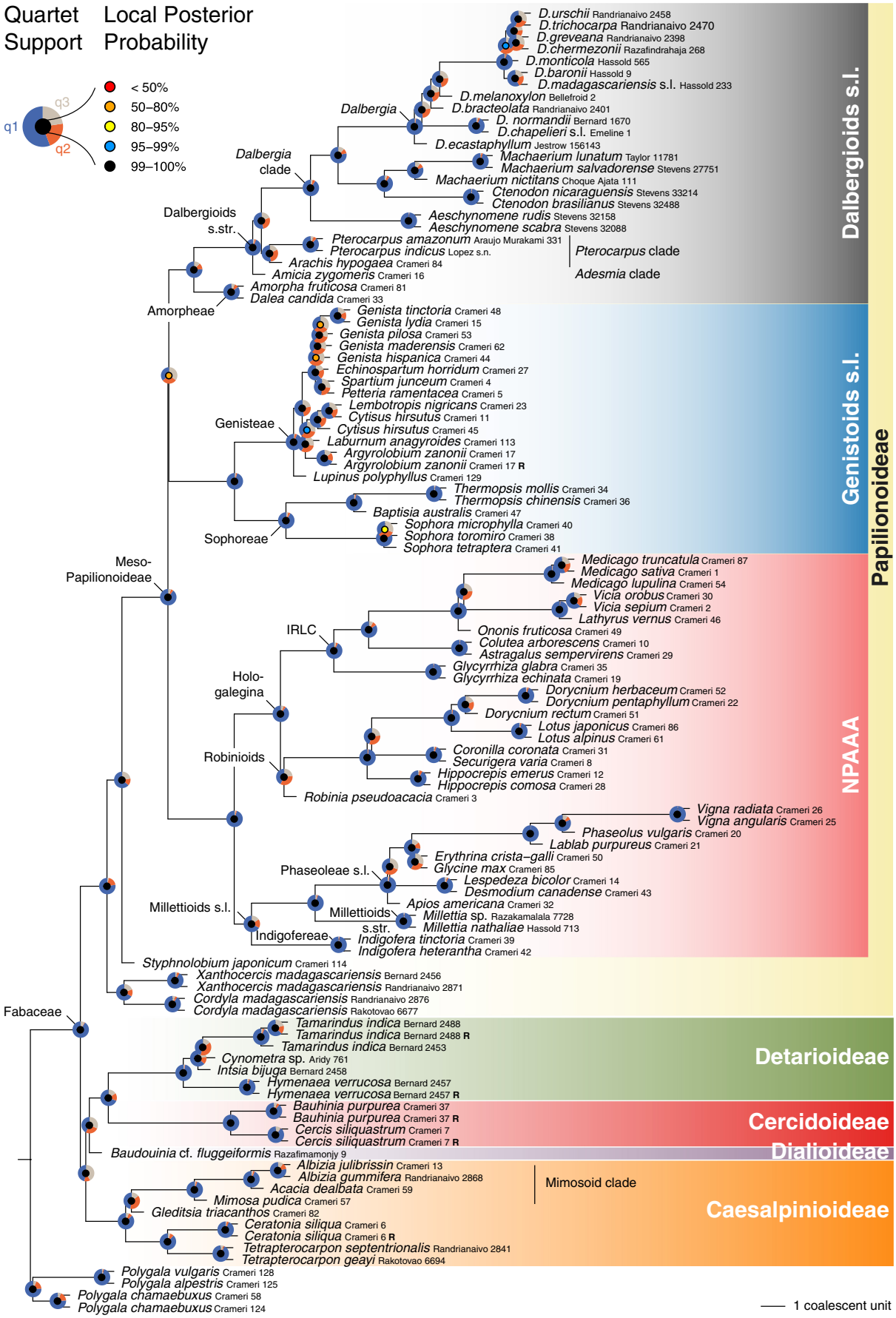
We obtained 0.13 to 13.76 (median: 1.56) million raw read pairs per sample, of which we retained 86.55% to 99.34% (median: 93.82%) after quality trimming (Tables S1–S3). In the first iteration applied to 12 representative samples, reads mapped to 6519 or 6287 of the 6555 initial target regions in the subfamily or species set, respectively (step 1). Of these we retained 3436 or 4908 target regions, which showed adequate coverage across taxon groups. After assembly (step 2) and orthology assessment (step 3), 2710 or 4181 target regions passed the region specificity and sensitivity filters of lower stringency (step 4). Following alignment and trimming (step 5), overlapping portions in 207 or 377 regions were successfully merged, resulting in 2468 or 3736 nonoverlapping trimmed alignments (step 6). Longer and more representative consensus sequences were generated from these target regions (step 7) and used as references for mapping quality-trimmed reads of the complete taxon sets (step 1, see Tables S1 and S2). In the second iteration, we retained 1917 or 3418 target regions with adequate coverage (Figures S5 and S6), of which 1020 or 2407 passed the specificity and sensitivity filters of higher stringency (step 4) after assembly (see Figures S7 and S8 for visualizations). Merging of overlapping alignments in 15 or 11 regions yielded 1005 (subfamily set) or 2396 (species set) distinct alignments (step 5), of which 726 represented the same regions in both sets. The corresponding tiled probe sequences (3273 for the Fabaceae1005 probe set and 6190 for the Dalbergia2396 probe set) extracted from the initial 12,049 probes are deposited on Dryad (https://doi.org/10.5061/dryad.73n5tb2z7) along with refined taxon-specific reference sequences. Corresponding gene annotations in the *Cajanus cajan* genome are given in Tables S6 and S7. For phylogenomic analyses, we excluded 19 or 7 alignments with a gap ratio above 0.35 or 0.3 or a nucleotide diversity above 0.35 or 0.15, leaving 986 (subfamily set) or 2389 (species set) alignments.

Quality-trimmed reads mapped to all 2396 target regions in the population set (step 1) using reference sequences that were representative of the species set after the second iteration for mapping (Figure S9). Variant calling resulted in 203,916 raw variants and 116,500 filtered SNPs after decomposing complex variants, of which 60,204 (51.68%) were bi-allelic with no missing data and were used for PCA and NJ tree reconstruction. Random sampling of three SNPs per target region (or all SNPs if less than three SNPs remained) resulted in a subset of 7156 SNPs for STRUCTURE analyses.

3.2 | Phylogenomic analyses across legumes

Phylogenomic analysis of 986 alignments recovered each of the five sampled subfamilies as monophyletic, and many well-established clades and relationships received ≥95% support (blue and black node support dots) using both the gene tree summary method ASTRAL-III (Figure 2) and the supermatrix method (Figure S1). These included the subfamilies Cercidoideae and Detarioideae found to be sister taxa, the mimosoid clade within the recently re-circumscribed subfamily Caesalpinioideae (LPWG, 2017), as well as the Angylocalyceae-Dipterygeae-Amburaneae (ADA, Cardoso et al., 2012), Cladrastis (Wojciechowski, 2013) and Meso-Papilionoideae (Wojciechowski, 2013) clades within Papilionoideae. We also recovered the Sophoreae and Genisteae clades (Cardoso et al., 2013) within Genistoids sensu lato (s.l.) (Cardoso et al., 2012; Wojciechowski et al., 2004). Within the Dalbergioids s.l. (Wojciechowski et al., 2004), we recovered the Amorpheae clade (McMahon & Hufford, 2004) as sister to the

FIGURE 2 Coalescent-based phylogeny of the Fabaceae subfamily set ($n = 110$) inferred using ASTRAL-III on 986 gene trees. Pie charts on each node denote the fraction of gene trees that are consistent with the shown topology (q1; blue), and with the first (q2; orange) and second (q3; grey) alternative topologies. Local posterior probabilities are shown as small colour-coded dots in the center of each pie chart, black dots indicate clades with 99%–100% local posterior probability (see inset legend). Replicate specimens are labelled with a bold "R". 860 gene trees (87.22%) had missing taxa. The overall normalized quartet score was 88.82%


Quartet Support
Local Posterior Probability
q1
q2
q3
< 50%
50–80%
80–95%
95–99%
99–100%
D.urschii Randrianaivo 2458
D.trichocarpa Randrianaivo 2470
D.greveana Randrianaivo 2398
D.chermezonii Razafindrahaja 268
D.monticola Hassold 565
D.baronii Hassold 9
D.madagascariensis s.l. Hassold 233
D.melanoxylon Bellefroid 2
D.bracteolata Randrianaivo 2401
D. normandii Bernard 1670
D.chapelieri s.l. Emeline 1
D.ecastaphyllum Jestrow 156143
Dalbergia
Machaerium lunatum Taylor 11781
Machaerium salvadorense Stevens 27751
Machaerium nictitans Choque Ajata 111
Ctenodon nicaraguensis Stevens 33214
Ctenodon brasilianus Stevens 32488
Dalbergia clade
Aeschynomene rudis Stevens 32158
Aeschynomene scabra Stevens 32088
Dalbergioids s.str.
Pterocarpus amazonum Araujo Murakami 331
Pterocarpus indicus Lopez s.n.
Pterocarpus clade
Arachis hypogaea Crameri 84
Amicia zygomeris Crameri 16
Adesmia clade
Amorpha fruticosa Crameri 81
Dalea candida Crameri 33
Amorpheae
Dalbergioids s.l.
Genista tinctoria Crameri 48
Genista lydia Crameri 15
Genista pilosa Crameri 53
Genista maderensis Crameri 62
Genista hispanica Crameri 44
Echinospartum horridum Crameri 27
Spartium junceum Crameri 4
Petteria ramentacea Crameri 5
Lembotropis nigricans Crameri 23
Cytisus hirsutus Crameri 11
Genisteae
Cytisus hirsutus Crameri 45
Laburnum anagyroides Crameri 113
Argyrolobium zanonii Crameri 17
Argyrolobium zanonii Crameri 17 R
Lupinus polyphyllus Crameri 129
Thermopsis mollis Crameri 34
Thermopsis chinensis Crameri 36
Baptisia australis Crameri 47
Sophora microphylla Crameri 40
Sophoreae
Sophora toromiro Crameri 38
Sophora tetraptera Crameri 41
Genistoids s.l.
Papilionoideae
Meso-Papilionoideae
Medicago truncatula Crameri 87
Medicago sativa Crameri 1
Medicago lupulina Crameri 54
Vicia orobus Crameri 30
Vicia sepium Crameri 2
Lathyrus vernus Crameri 46
IRLC
Ononis fruticosa Crameri 49
Colutea arborescens Crameri 10
Astragalus sempervirens Crameri 29
Glycyrrhiza glabra Crameri 35
Glycyrrhiza echinata Crameri 19
Holo-galegina
Dorycnium herbaceum Crameri 52
Dorycnium pentaphyllum Crameri 22
Dorycnium rectum Crameri 51
Lotus japonicus Crameri 86
Lotus alpinus Crameri 61
Coronilla coronata Crameri 31
Robinioids
Securigera varia Crameri 8
Hippocrepis emerus Crameri 12
Hippocrepis comosa Crameri 28
Robinia pseudoacacia Crameri 3
NPAAA
Vigna radiata Crameri 26
Vigna angularis Crameri 25
Phaseolus vulgaris Crameri 20
Lablab purpureus Crameri 21
Erythrina crista–galli Crameri 50
Phaseoleae s.l.
Glycine max Crameri 85
Lespedeza bicolor Crameri 14
Desmodium canadense Crameri 43
Apios americana Crameri 32
Millettioids s.l.
Millettioids s.str.
Millettia sp. Razakamalala 7728
Millettia nathaliae Hassold 713
Indigofereae
Indigofera tinctoria Crameri 39
Indigofera heterantha Crameri 42
Styphnolobium japonicum Crameri 114
Xanthocercis madagascariensis Bernard 2456
Xanthocercis madagascariensis Randrianaivo 2871
Cordyla madagascariensis Randrianaivo 2876
Cordyla madagascariensis Rakotovao 6677
Fabaceae
Tamarindus indica Bernard 2488
Tamarindus indica Bernard 2488 R
Tamarindus indica Bernard 2453
Cynometra sp. Aridy 761
Intsia bijuga Bernard 2458
Hymenaea verrucosa Bernard 2457
Hymenaea verrucosa Bernard 2457 R
Detarioideae
Bauhinia purpurea Crameri 37
Bauhinia purpurea Crameri 37 R
Cercis siliquastrum Crameri 7
Cercis siliquastrum Crameri 7 R
Cercidoideae
Dialioideae
Baudouinia cf. fluggeiformis Razafimamonjy 9
Albizia julibrissin Crameri 13
Albizia gummifera Randrianaivo 2868
Acacia dealbata Crameri 59
Mimosoid clade
Mimosa pudica Crameri 57
Gleditsia triacanthos Crameri 82
Ceratonia siliqua Crameri 6
Ceratonia siliqua Crameri 6 R
Tetrapterocarpon septentrionalis Randrianaivo 2841
Tetrapterocarpon geayi Rakotovao 6694
Caesalpinioideae
Polygala vulgaris Crameri 128
Polygala alpestris Crameri 125
Polygala chamaebuxus Crameri 58
Polygala chamaebuxus Crameri 124
1 coalescent unit

rest of the group, which includes the Dalbergioids s.str. Clade (Lavin et al., 2001), containing the *Adesmia* (represented by a single accession of *Amicia zygomeris* DC.), *Pterocarpus* and *Dalbergia* subclades (Lavin et al., 2001), respectively. *Ctenodon brasilianus* (Poir.) D.B.O.S.Cardoso, P.L.R.Moraes & H.C.Lima and *C. nicaraguensis* (Oerst.) A.Delgado were found to be more closely related to *Machaerium* than to *Aeschynomene*. Within the NPAAA we recovered the Millettioid s.l. clade (Wojciechowski et al., 2004), containing the genera *Indigofera* and *Millettia*, and the Phaseoleae s.l. (Vatanparast et al., 2018), as well as the Hologalegina (Wojciechowski, 2013) clade, including the Robinioids and the inverted-repeat-lacking clade (IRLC, Wojciechowski et al., 2004).

Other relationships among subfamilies remained unresolved using both phylogenetic methods (Figures 2, S1). In particular, a clade comprising Caesalpinioideae, Cercidoideae, Detarioideae and Dialioideae as sister group to Papilionoideae was not supported in the supermatrix tree, and was recovered in only 47% of quartet trees. We evaluated quartet scores (i.e., the fraction of induced quartet trees) of 14 further topologies for relationships among sampled subfamilies (Figure S2) using the tree scoring option in ASTRAL-III in combination with a file that mapped taxa to subfamilies or to the outgroup. The subfamily topology presented in Figure 2 showed the highest normalized quartet score (38.40%). Two alternative topologies received a similar normalized quartet score of 38.36% (Figure S2) and involved a clade composed of Caesalpinioideae and Papilionoideae. Further contentious relationships between major groups concerned the three clades within Meso-Papilionoideae, where the clade formed by Dalbergioids s.l. and Genistoids s.l. was recovered only in 36% of quartet trees, and in relationships within Caesalpinioideae, Detarioideae, and Genisteae. All except one genus with multiple sampled accessions were recovered as monophyletic, the exception being *Cytisus*, which was paraphyletic with respect to *Lembotropis nigricans*. Pairs of replicates each grouped together (Figures 2, S1).

3.3 | Phylogenomic analyses in *Dalbergia*

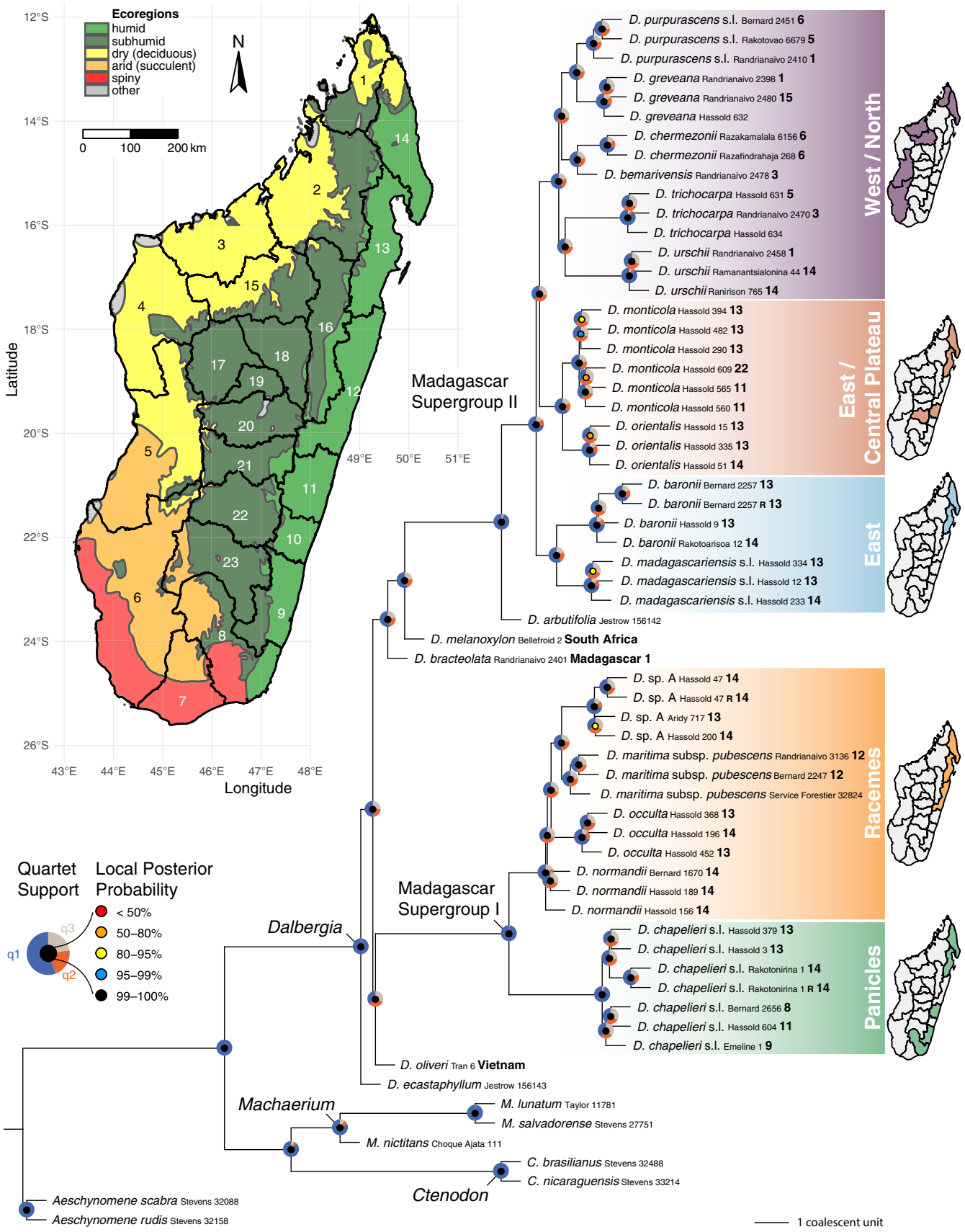
Phylogenomic analysis of 2389 alignments recovered samples of *Dalbergia* as monophyletic with ≥95% support (blue and black node support dots) using both ASTRAL-III (Figure 3) and the supermatrix method (Figure S3). Within *Dalbergia*, we recovered two large and exclusively Malagasy clades, which we name Madagascar Supergroup I and II. All Malagasy species represented by multiple accessions were recovered as highly supported clades, with the exception of *D. normandii*. Four non-Malagasy *Dalbergia* specimens and *D. bracteolata* Baker were each found to represent separate lineages.

Within Supergroup I, one clade comprised samples of *Dalbergia chapelieri* s.l., while the remaining samples belonged to a morphologically divergent sister group containing three monophyletic species and a basal and paraphyletic *D. normandii*. Within Supergroup II, two clades contained species distributed in the humid east of Madagascar, while the third contained species distributed in the seasonally dry west and north of the island. Within *D. chapelieri* s.l. and *D. monticola*, which were each represented by six individuals, we observed geographic structure, with specimens from northeast and southeast Madagascar forming sister groups. Pairs of replicates each grouped together (Figures 3, S3).

3.4 | Population genomic analyses

Principal component analysis revealed three distinct clusters of individuals along principal component (PC) 1 (explaining 27.58% of the total variation) and PC 2 (11.26%; Figure 4a). Individuals of *Dalbergia orientalis* separated along PC1, while individuals originally attributed to *D. monticola* formed two distinct groups mainly along PC2. The unexpected smaller cluster (*D.* sp. B, which corresponds to *D.* sp. 17 in Crameri (2020); in purple in Figure 4) comprised samples from a single broad sampling location in north-eastern Madagascar (location 5, see Figure S4 and Table S3) where both *D. monticola* and *D. orientalis* were also collected. The same three clusters were also recovered in STRUCTURE analyses (Figures S3b–c and S11), where biologically meaningful clustering solutions were found for $K = 2$ (separating *D. orientalis* from the rest) and $K = 3$ (further separating the unexpected smaller cluster). Within *D. orientalis* and the larger cluster of true *D. monticola*, the NJ tree reflects isolation by distance at a broad geographical scale, separating specimens from north-eastern (locations 1 to 6), central-eastern (locations 7 and 8) and south-eastern Madagascar (locations 9 to 13; Figures 3a and S4). A similar geographic pattern was recovered by STRUCTURE assuming $K = 5$ and $K = 7$ (Figure 4c), although these clustering solutions received much lower support (Figure 4b). Clustering solutions assuming higher K did not recover significant additional structure (Figure S11).

FIGURE 3 Coalescent-based phylogeny of the Malagasy *Dalbergia* species set ($n = 63$) inferred using ASTRAL-III on 2389 gene trees. Pie charts on each node denote the fraction of gene trees that are consistent with the shown topology (q1; blue), and with the first (q2; orange) and second (q3; grey) alternative topologies. Local posterior probabilities are shown as small colour-coded dots in the center of each pie chart, black dots indicate clades with 99%–100% local posterior probability (see inset legend). The geographic origins of accessions from Madagascar are indicated as bold numbers in the tree, which correspond to political regions of Madagascar, as well as to ecological regions following Dinerstein et al. (2017), see large inset map. Known countries of origin of non-Malagasy accessions are indicated in bold. Five major clades within Madagascar Supergroups I and II showing ecogeographic or morphological coherence are named and their distribution is indicated (see small maps to the right). Replicate specimens are labelled with a bold "R". 1014 gene trees (42.44%) had missing taxa. The overall normalized quartet score was 85.42%


Ecoregions
humid
subhumid
dry (deciduous)
arid (succulent)
spiny
other
N
0 100 200 km
12°S
14°S
16°S
18°S
20°S
22°S
24°S
26°S
43°E 44°E 45°E 46°E 47°E 48°E 49°E 50°E 51°E
Latitude
Longitude
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Madagascar Supergroup II
Madagascar Supergroup I
Dalbergia
Machaerium
Ctenodon
D. purpurascens s.l. Bernard 2451 6
D. purpurascens s.l. Rakotovao 6679 5
D. purpurascens s.l. Randrianaivo 2410 1
D. greveana Randrianaivo 2398 1
D. greveana Randrianaivo 2480 15
D. greveana Hassold 632
D. chermezonii Razakamalala 6156 6
D. chermezonii Razafindrahaja 268 6
D. bemarivensis Randrianaivo 2478 3
D. trichocarpa Hassold 631 5
D. trichocarpa Randrianaivo 2470 3
D. trichocarpa Hassold 634
D. urschii Randrianaivo 2458 1
D. urschii Ramanantsialonina 44 14
D. urschii Ranirison 765 14
West / North
D. monticola Hassold 394 13
D. monticola Hassold 482 13
D. monticola Hassold 290 13
D. monticola Hassold 609 22
D. monticola Hassold 565 11
D. monticola Hassold 560 11
D. orientalis Hassold 15 13
D. orientalis Hassold 335 13
D. orientalis Hassold 51 14
East / Central Plateau
D. baronii Bernard 2257 13
D. baronii Bernard 2257 R 13
D. baronii Hassold 9 13
D. baronii Rakotoarisoa 12 14
D. madagascariensis s.l. Hassold 334 13
D. madagascariensis s.l. Hassold 12 13
D. madagascariensis s.l. Hassold 233 14
East
D. arbutifolia Jestrow 156142
D. melanoxylon Bellefroid 2 South Africa
D. bracteolata Randrianaivo 2401 Madagascar 1
D. sp. A Hassold 47 14
D. sp. A Hassold 47 R 14
D. sp. A Aridy 717 13
D. sp. A Hassold 200 14
D. maritima subsp. pubescens Randrianaivo 3136 12
D. maritima subsp. pubescens Bernard 2247 12
D. maritima subsp. pubescens Service Forestier 32824
D. occulta Hassold 368 13
D. occulta Hassold 196 14
D. occulta Hassold 452 13
D. normandii Bernard 1670 14
D. normandii Hassold 189 14
D. normandii Hassold 156 14
Racemes
D. chapelieri s.l. Hassold 379 13
D. chapelieri s.l. Hassold 3 13
D. chapelieri s.l. Rakotonirina 1 14
D. chapelieri s.l. Rakotonirina 1 R 14
D. chapelieri s.l. Bernard 2656 8
D. chapelieri s.l. Hassold 604 11
D. chapelieri s.l. Emeline 1 9
Panicles
D. oliveri Tran 6 Vietnam
D. ecastaphyllum Jestrow 156143
M. lunatum Taylor 11781
M. salvadorense Stevens 27751
M. nictitans Choque Ajata 111
C. brasilianus Stevens 32488
C. nicaraguensis Stevens 33214
Aeschynomene scabra Stevens 32088
Aeschynomene rudis Stevens 32158
Quartet Support
Local Posterior Probability
q1
q2
q3
< 50%
50–80%
80–95%
95–99%
99–100%
1 coalescent unit

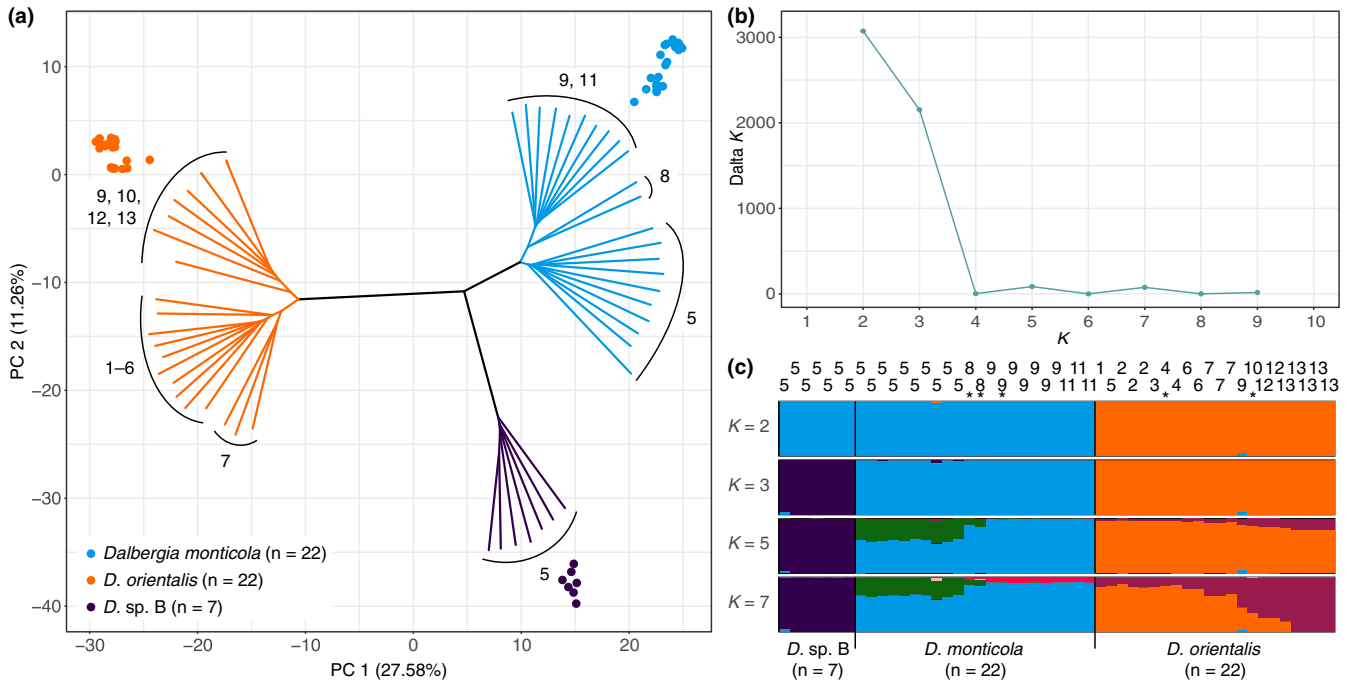


FIGURE 4 Population genomic analyses in *Dalbergia monticola* and *D. orientalis*. (a) Principal component analysis (PCA) and superimposed neighbour-joining (NJ) tree of the population set ($n = 51$) inferred from 60,204 biallelic SNPs with no missing data. Dots in PCA space and NJ tips represent individuals colour-coded according to taxa. Numbers adjacent to NJ tree branches denote sampling locations as shown in Figure S4. See Figure S10 for NJ tip labels. (b) STRUCTURE probability at different values of *K*, as indicated by the delta *K* statistic (Evanno et al., 2005). (c) STRUCTURE results for the 51 individuals and 7156 single nucleotide polymorphisms (SNPs). The four clustering solutions with elevated delta *K* values are shown (see Figure S11 for all results assuming two to 10 clusters) and represent major clusters averaged across 10 replicate runs using CLUMPAK (Kopelman et al., 2015). Individuals (columns) are colour-coded and sorted by taxa and then by increasing degrees south latitude. Numbers at the top indicate broad sampling locations as in Figure S4 and Table S3. Columns marked with an asterisk (*) denote individuals obtained from museum specimens

3.5 | Comparison to the Angiosperms353 probe set

Results of the BLAST+ analyses are shown in Tables S8 and S9. In summary, 34 (3.38%) of the Fabaceae1005 target regions and 75 (3.13%) of the Dalbergia2396 target regions were reciprocal best hits to 32 (9.07%) and 65 (18.41%) target regions of the Angiosperms353 set, respectively. In twelve cases, there were reciprocal hits between a single Angiosperms353 region and two of our regions. All these cases involved contiguous subregions in longer supercontigs of the Angiosperms353 set corresponding to two contiguous regions in the *Cajanus cajan* genome.

4 | DISCUSSION

Understanding the diversity and diversification of species and evolutionary lineages requires an integrative approach that links studies of micro-evolutionary processes to analyses of macroevolutionary relationships (de La Harpe et al., 2017). Genetic data form a preferable source of information for investigations across broad evolutionary scales, as a large number of loci distributed across the nuclear genome can represent the spectrum of evolutionary rates at different scales of sample divergence. The present study introduces two overlapping sets of target capture probes for phylogenomic studies at micro- to macroevolutionary timescales in rosewoods (Dalbergia2396 probe set) and across the legume family (Fabaceae1005 probe set), together with the flexible and modular bioinformatic pipeline CAPTUREAL, which streamlines the processing of sequencing reads for phylogenomic and population genomic analyses while visually informing users on the effect of critical parameter choices. We demonstrated the utility of individual assemblies per target region to produce alignments of hundreds of loci suitable for concatenation and multispecies coalescent approaches, which confirmed phylogenomic conflicts at the root of the legume family, and provided an unprecedented resolution of evolutionary relationships among lineages and species of the taxonomically complex genus *Dalbergia*. Remapping of sequencing reads onto refined reference sequences of the target regions further made it possible to identify thousands of informative sites amenable to population genomic analyses, which revealed the existence of a potentially new cryptic *Dalbergia* species. Together, these results illustrate that our newly developed probe sets are efficient tools for studies of species diversity and diversification in rosewoods (*Dalbergia* spp.) and more broadly in the economically important and highly diverse legume family.

4.1 | Target capture probes

The target capture probes presented here are part of a growing collection of genomic resources for legume phylogenomics. Other

probe sets for target capture in legumes have been developed, focusing on different groups within the family, and designed or validated at the level of legume species (Peng et al., 2017), genera (e.g., de Sousa et al., 2014; Nicholls et al., 2015; Shavvon et al., 2017), or above (Koenen, Kidner, et al., 2020; Vatanparast et al., 2018), and across angiosperms (Johnson et al., 2019). Our reciprocal BLAST+ analyses revealed that at least 65 (18.41%) genomic regions targeted by the Angiosperms353 probe set are likely also targeted by the Dalbergia2396 probe set. The fraction of target regions overlapping with the Angiosperms353 set was nearly identical in the Fabaceae1005 set (3.38%) compared to the Dalbergia2396 set (3.13%). This suggests that a possibly higher degree of conservation in genomic regions targeted by the Fabaceae1005 set compared to the Dalbergia2396 set may refer to divergence of the *Cajanus cajan* genome on which our probe sequences were based, rather than a higher degree of conservation across angiosperms. In light of the considerable overlap between the Dalbergia2396 and Angiosperms353 probe sets, it would be interesting to extend such comparisons to identify overlaps and complementarity in probe sets specifically designed for legume phylogenomics. Capture of additional, less conserved target regions across the legume family could be achieved by designing multiple probes for hybridization in the same target region in different legume groups, as applied for studies across angiosperms (Johnson et al., 2019). Such a probe design could profit from existing legume probe sets but should rely on a stringent selection of targets that accounts for paralogues (Vatanparast et al., 2018), which originated as a consequence of multiple whole-genome duplication events in legumes (Egan & Vatanparast, 2019; Koenen et al., 2021).

In this study, we enriched DNA libraries from three taxon sets spanning microevolutionary (populations) to macroevolutionary (family) timescales, using a single set of 12,049 RNA probes targeting 6555 genomic regions conserved across five Meso-Papilionoideae genomes and a *Dalbergia* transcriptome. We then identified 2396 and 1005 target regions with high capture specificity and sensitivity within the species-rich genus *Dalbergia* (Dalbergia2396 probe set) and more broadly across legumes (Fabaceae1005 probe set). We used our CAPTUREAL pipeline to refine phylogenomic and population genomic analyses using taxon-specific and longer reference sequences. This procedure has both benefits and drawbacks. An advantage is that different but overlapping probe sets amenable for efficient target capture in different focal groups can be identified, and that a single enriched DNA library can be included in multiple data sets spanning different evolutionary timescales. On the other hand, bioinformatic analyses took longer due to the iterative refinement, and only a portion of captured sequence data was ultimately used for phylogenomic or population genomic analyses in each focal group (see Tables S1–S3). Higher costs per used sequence could be compensated by enriching DNA of up to six individuals in a single hybridization reaction, a strategy that has been used successfully in other studies (e.g., de La Harpe et al., 2018; Yardeni et al., 2022). To further reduce costs, future target enrichment experiments in *Dalbergia* could be focused on genomic regions targeted by the 6190 sequences of the Dalbergia2396 probe set.

4.2 | CAPTUREAL bioinformatic pipeline

The CAPTUREAL pipeline starts with the mapping of quality-trimmed reads to conserved target regions identified during probe design, followed by assembly on a per-region basis, orthology assessment, and filtering for target regions with high capture sensitivity and specificity for downstream analyses. This approach differs from the PHYLUCE pipeline (Faircloth, 2016), where quality-trimmed reads are first assembled, and then matched to target regions. CAPTUREAL simplifies the assembly of reads specific to each locus, circumventing the challenging task of de novo assembly of contigs from the large pool of sequencing reads representative of thousands of loci (reviewed by Chaisson et al., 2015). Likewise, alignments are conducted in clearly defined target regions in which overlap among individual contigs is higher. However, assembly per region is more time-consuming and requires reference sequences for the initial mapping step. This might introduce a reference bias when divergent sequences are not mapped (Lunter & Goodson, 2011). We addressed this problem by generating consensus sequences that are representative of a given taxon set and by limiting analyses to target regions that can be efficiently recovered in all groups of that taxon set. These set-specific reference sequences can then be used to iteratively refine mapping, assembly, and target region filtering for any set of taxa. Our approach is conceptually similar to the HYBPIPER pipeline (Johnson et al., 2016), which also employs a mapping-assembly strategy, and uses depth of coverage and percent identity to the target region to choose between multiple contigs, before it identifies intron/exon boundaries using target peptide sequences and extracts coding sequences for alignment. While the HYBPIPER pipeline is designed specifically for the HYB-SEQ approach (Johnson et al., 2016), in which exons are the primary targets and flanking noncoding regions are used as supplementary data for analyses at shallow evolutionary scales (Weitemier et al., 2014), CAPTUREAI is more general in scope and neither requires nor leverages knowledge about intron/exon boundaries in the targeted regions. It is therefore suitable for application in systems lacking high-quality annotated reference genomes or transcriptomes. The main strengths of this pipeline are its modularity, which allows for an iterative refinement of read mapping, assembly and alignment, its flexibility given by user-defined parameters, the merging of alignments representing physically overlapping target regions, and the visualization of key summary statistics and alignments along the workflow to inform the user on critical analysis parameters.

4.3 | Macro- and microevolutionary patters in *Dalbergia*

Dalbergia species endemic to Madagascar were recovered as two large, well-supported and fully resolved clades, each exclusively comprising Malagasy species. These two clades were previously identified on the basis of three chloroplast markers, but phylogenetic

relationships within clades were not resolved, which exposed traditional DNA barcoding as insufficient for genetic discrimination between closely related *Dalbergia* species (Hassold et al., 2016; see Tables S2 and S3). Supergroups I and II are morphologically divergent and largely correspond to Group 1 and 2 reported by Bosser and Rabevohitra (2002). Supergroup I is characterized by a glabrous gynoecium with a long and slender style and relatively large flowers, while Supergroup II is characterized by a pubescent gynoecium with a short and squat style and relatively small flowers. The two supergroups are both more closely related to non-Malagasy taxa than to each other, suggesting at least two independent colonizations of Madagascar followed by species diversification. The only sampled Malagasy species not belonging to either of the two supergroups is *D. bracteolata*, which occurs on Madagascar as well as in mainland East Africa. A further species, which is endemic to Madagascar and morphologically divergent from Supergroups I and II (*D. xerophila* Bosser & R. Rabev.) was not included in this study.

Within Supergroup I, two well-supported subclades were resolved, which differ in their inflorescence structure. Within *Dalbergia chapelieri* s.l., a widely distributed species complex with paniculate inflorescences, northeastern and southeastern populations can be distinguished using the present data as well as chloroplast variation (Hassold et al., 2016). The other subclade within Supergroup I contains species from eastern Madagascar that have a strong tendency to produce racemose inflorescences. It includes a potentially new species, *Dalbergia* sp. A, which corresponds to *D.* sp. 24 in Crameri (2020) and to which no provisional name has yet been assigned. Collections belonging to this entity were previously believed to be conspecific with *D. maritima* subsp. *pubescens* (see Hassold et al., 2016) but show geographic (i.e., north-east vs. central-east), morphological (i.e., more numerous leaflets that are smaller, more oblong and less coriaceous) and genetic (Figure 3, Figure S3) differences compared to the type material (*Service Forestier 32,824*). The type (collected in 1985) showed a slightly longer terminal branch compared to other samples in the concatenation tree (Figure S3) but clearly grouped with two recently collected conspecific samples from central-east Madagascar. The same subclade also contains material of two highly valued rosewood species, *D. occulta* and *D. normandii*; note that in Hassold et al. (2016), sterile material of *D. normandii* was erroneously identified as *D. madagascariensis*.

Supergroup II includes two clades distributed in the humid and subhumid east and northwest of Madagascar, and a large third clade centred in the drier west and north of the island. The identification of morphological synapomorphies characterizing these clades require further genetic and morphological analyses. The geographic separation in major ecogeographic regions of Madagascar suggests that climate regimes may have played a significant role in shaping the evolution of these groups, which thus constitute promising model systems to study processes of ecological divergence, along the same lines of studies that have investigated elements of the Malagasy fauna (Vences et al., 2009).

Our results revealed relationships among Supergroups I and II and non-Malagasy taxa that are incompatible with the plastid phylogeny of Hassold et al. (2016), in particular with regard to *Dalbergia melanoxylon* (Africa), *D. ecastaphyllum* (America and Africa), and *D.* cf. *oliveri* (Asia). Incongruence between nuclear and plastid phylogenies is common at various evolutionary timescales in many plant groups (e.g., Lee et al., 2021; Pelser et al., 2010), and while the multispecies approach applied in this study is expected to return a phylogeny that reflects nuclear evolution accounting for incomplete lineage sorting, conflicts in gene tree topologies due to hybridization and chloroplast capture can further underlie the observed differences.

Our target capture approach demonstrated great potential to facilitate the resolution of several taxonomic conundrums within the genus, which likely resulted from limited observable and diagnostic morphological characters, insufficient collection effort, and the difficulty of distinguishing between heritable and plastic trait variation within and among *Dalbergia* species (Lachenaud, 2016). The integration of highly informative museum specimens, including a nomenclatural type collected in 1985, enabled the accurate identification of recently collected but often sterile specimens, and was crucial in detecting misidentifications or potential taxonomic inadequacies (Buerki & Baker, 2016), as shown for *D. maritima* subsp. *pubescens* or *D. monticola*.

Population genomic analyses of 51 individuals readily separated the two closely related species *Dalbergia monticola* and *D. orientalis*, as well as a sympatric and syntopic but genetically differentiated entity, which could previously not be differentiated from the other two species based on three chloroplast markers (Hassold et al., 2016). The lack of admixture between *D. monticola* and this third cluster, the similarity in leaf characters, and the absence of known morphologically similar species occurring in the region, prompts us to hypothesize the latter to reflect a separate, yet undescribed cryptic species. Both *D. monticola* and *D. orientalis* are distributed from northeastern to south-eastern Madagascar, co-occur in various localities, but differ in their predominant altitudinal distribution (Madagascar Catalogue, 2022). Population structure within both species was uncovered using our target capture approach and appears to be sufficient to distinguish specimens from the northeast (locations 1 to 6), central-east (locations 7 and 8), and southeast of the island (locations 9 to 13). These results indicate that genetic species identification and provenancing, at least to this broad geographic scale, may be feasible, which would have important implications for forensic timber identification and for tracing geographic hotspots of the illegal trade in these valuable timber species (UNODC, 2016a).

4.4 | Phylogenomic analyses across legumes

At the family level, 1005 merged regions of the 6555 initially targeted regions passed our stringent sensitivity and specificity filters, suggesting that many target regions were not efficiently captured across taxa. However, phylogenomic analysis of 986 nuclear target regions recovered multiple known clades within monophyletic subfamilies with strong bootstrap and quartet support,

providing excellent resolution comparable to that obtained in the recent nuclear phylogenomic analysis of transcriptome and genome-wide data across legumes (Koenen, Ojeda, et al., 2020). As in that study, we found high support for Cercidoideae and Detarioideae as sister taxa, a relationship that was never inferred in analyses based on chloroplast genes (LPWG, 2017) or plastomes (Koenen, Ojeda, et al., 2020; Zhang et al., 2020). As shown in both studies, the other relationships among subfamilies are difficult to resolve. Our most supported subfamily topology (38.4% quartet support, Figure S2A) recovered the Papilionoideae as sister to a clade comprised of Caesalpinioideae, Dialioideae, and the Cercidoideae/Detarioideae clade, while Koenen, Ojeda, et al. (2020) demonstrated a successive divergence of the Cercidoideae/Detarioideae clade, Dialioideae, Caesalpinioideae and Papilionoideae in all nuclear analyses. This alternative topology received almost equivalent overall quartet support (38.36%) in our analyses (Figure S2C), as did a third hypothesis in which Caesalpinioideae and Papilionoideae are sister to Dialioideae and the Cercidoideae/Detarioideae clade (Figure S2B). These nearly equally supported subfamily topologies can be explained by short deep internodes associated with conflicting bipartitions and are consistent with the idea of a nearly simultaneous evolutionary origin of all six legume subfamilies, causing incomplete lineage sorting (Koenen, Ojeda, et al., 2020). Taxon sampling may additionally contribute to the contentious deep-branching relationships. The monotypic Duparquetioideae subfamily could not be analysed, and a portion of gene trees may suffer from long branch attraction between *Polygala* and Papilionoideae, which both exhibit markedly higher substitution rates compared to the other legume subfamilies (Koenen, Ojeda, et al., 2020). Additional outgroup taxa such as members of the Quillajaceae family could alleviate this problem, and permit a more accurate inference of subfamily relationships.

Substantial gene tree incongruence was also found with respect to the relationships among the three large clades within Meso-Papilionoidae. The sister relationship between Dalbergioids s.l. and Genistoids s.l. received only slightly higher quartet support than the two alternative hypotheses, which is consistent with previous results (Koenen, Ojeda, et al., 2020). Similarly, conflicting topologies affected most branches within Genisteae. By contrast, our analyses confirm that the genus *Aeschynomene* sensu Rudd (1955), which consisted of the former *A*. sect. *Aeschynomene* and *A*. sect. *Ochopodium* Vogel, is nonmonophyletic (Ribeiro et al., 2007). The recently re-established *Ctenodon* (= *A*. sect. *Ochopodium*, Cardoso et al., 2020) is sister to *Machaerium*, and these two genera form the sister group to *Dalbergia*.

4.5 | Conclusions and perspectives

The resources developed here for Fabaceae and in particular the genus *Dalbergia* bridge micro- and macroevolutionary timescales and will hopefully facilitate community-driven efforts to advance legume genomics. Comprehensive sampling and sequencing by target capture of *Dalbergia* across its distribution range, and in particular from the hotspot of diversity in Madagascar, can yield valuable insights into the origin and diversification of the genus, thereby informing conservation policies and the taxonomic revision of Malagasy *Dalbergia*. The obtained sequence data will further serve to build a reference library for molecular identification of CITES-listed *Dalbergia* species, which would make a significant contribution toward the conservation of the valuable and endangered rosewoods.

AUTHOR CONTRIBUTIONS

Simon Crameri and Alex Widmer designed the study and collected samples. Stefan Zoller assembled the draft *Dalbergia* transcriptome and designed target capture probes. Simon Crameri and Stefan Zoller analysed data and wrote CaptureAl. Simon Crameri, Simone Fior and Alex Widmer wrote the manuscript with contributions from Stefan Zoller.

ACKNOWLEDGEMENTS

We thank the Missouri Botanical Garden (MBG) Madagascar team (especially Sylvie Andriambololonera, Roger Bernard, Nivo Rakotonirina, Charles Rakotovao, Richard Randrianaivo, and Richardson Razakamalala), the University of Antananarivo (especially Harisoa Ravaomanalina), Sonja Hassold and Hoa Thi Tran for organizing and conducting fieldwork, the Botanic Garden of the University of Zurich and the Fairchild Tropical Botanic Garden for the opportunity to sample legume plants from their living collection, and the herbaria of Zurich (ZT), Paris (P), Geneva (G) and MBG (MO) for access to their DNA bank and voucher specimens, respectively. We also thank Peter Phillipson and Nicholas Wilding for discussing sample determinations, Dalbergia taxonomy and Malagasy plant diversity. We are very grateful to Claudia Michel for laboratory work and the Genetic Diversity Centre (GDC) at ETH Zurich for helpful support, in particular to Silvia Kobel for sequencing and Niklaus Zemp for bioinformatics. Finally, we thank Erik Koenen, Colin Hughes, Pete Lowry, Martin Fischer and Nicholas Wilding for their valuable inputs and comments on the manuscript. Open access funding provided by Eidgenossische Technische Hochschule Zurich.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The raw sequencing data is available in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB41848 (https://www.ebi.ac.uk/ena/browser/view/PRJEB41848). The Dalbergia madagascariensis s.l. transcriptome, probe sequences for target capture, taxon-specific reference sequences, final alignments, and SNP data are available in Dryad (https://doi.org/10.5061/dryad.73n5tb2z7).

This article has earned an Open Materials badge for making publicly available the components of the research methodology needed to reproduce the reported procedure and analysis. The bioinformatic pipeline CaptureAl and R scripts for downstream population genomic analyses and visualizations are available and further documented on GitHub (https://github.com/scrameri/CaptureAl).

DATA AVAILABILITY STATEMENT

Raw target capture sequencing reads generated for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB41848 (https://www.ebi.ac.uk/ena/browser/view/PRJEB41848; Crameri et al., 2022a). Transcriptome sequencing reads as well as the draft *Dalbergia* transcriptome, sequences representing the initial 12,049 RNA probes and 6555 target regions, the Fabaceae1005 and Dalbergia2396 probe sets, longer and taxon-specific reference sequences used for mapping, final alignments for the subfamily and species sets (all in FASTA format), and SNP data from the population set (VCF format) are available on Dryad (https://doi.org/10.5061/dryad.73n5tb2z7; Crameri et al., 2022b). The bioinformatic pipeline CaptureAl and R scripts for downstream population genomic analyses and visualizations are available and further documented on GitHub (https://github.com/scrameri/CaptureAl). Because *Dalbergia* species are under threat from illegal exploitation, we have systematically refrained from making detailed distribution maps and precise geo-coordinates publicly available. Specimen records for collections from Madagascar are provided in the Catalogue of the Plants of Madagascar (Madagascar Catalogue, 2022), but with restricted public access to precise geocoordinates (delivered on demand to bona fide researchers).

ORCID

Simon Crameri https://orcid.org/0000-0002-5516-1018
Simone Fior https://orcid.org/0000-0003-1173-1477
Alex Widmer https://orcid.org/0000-0001-8253-5137

REFERENCES

Adema, F., Ohashi, H., & Sunarno, B. (2016). Notes on Malesian Fabaceae (Leguminosae-Papilionoideae) 17 The genus *Dalbergia Blumea*. *Blumea Journal of Plant Taxonomy and Plant Geography*, *61*(3), 186–206. https://doi.org/10.3767/000651916X693905

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Barrett, M. A., Brown, J. L., & Yoder, A. D. (2013). Conservation: Protection for trade of precious rosewood. *Nature*, *499*, 29. https://doi.org/10.1038/499029c

Bosser, J., & Rabevohitra, R. (2002). Tribe Dalbergieae. In D. J. Du Puy, J. N. Labat, R. Rabevohitra, J. F. Villiers, J. Bosser, & J. Moat (Eds.), *The Leguminosae of Madagascar* (pp. 321–361). Royal Botanic Gardens, Kew.

Brewer, G. E., Clarkson, J. J., Maurin, O., Zuntini, A. R., Barber, V., Bellot, S., Biggs, N., Cowan, R. S., Davies, N. M. J., Dodsworth, S., Edwards, S. L., Eiserhardt, W. L., Epitawalage, N., Frisby, S., Grall, A., Kersey, P. J., Pokorny, L., Leitch, I. J., Forest, F., & Baker, W. J. (2019). Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of Angiosperms. *Frontiers in Plant Science*, *10*, 1102. https://doi.org/10.3389/fpls.2019.01102

Broad Institute. (2019). *Picard Toolkit*, *version 2.21.3*. http://broadinstitute.github.io/picard

Buerki, S., & Baker, W. J. (2016). Collections-based research in the genomic era. *Biological Journal of the Linnean Society*, *117*, 5–10. https://doi.org/10.1111/bij.12721

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Cardoso, D. B. O. S., de Queiroz, L. P., Pennington, R. T., de Lima, H. C., Fonty, E., Wojciechowski, M. F., & Lavin, M. (2012). Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. *American Journal of Botany*, *99*(12), 1991–2013. https://doi.org/10.3732/ajb.1200380

Cardoso, D. B. O. S., Mattos, C. M. J., Filardi, F., Delgado-Salinas, A., Lavin, M., de Moraes, P. L. R., Tapia-Pastrana, F., & de Lima, H. C. (2020). A molecular phylogeny of the pantropical papilionoid legume *Aeschynomene* supports reinstating the ecologically and morphologically coherent genus *Ctenodon*. *Neodiversity*, *13*(1), 1–38. https://doi.org/10.13102/neod.131.1

Cardoso, D., Pennington, R. T., de Queiroz, L. P., Boatwright, J. S., Van Wyk, B. E., Wojciechowski, M. F., & Lavin, M. (2013). Reconstructing the deep-branching relationships of the papilionoid legumes. *South African Journal of Botany*, *89*, 58–75. https://doi.org/10.1016/j.sajb.2013.05.001

Chaisson, M. J. P., Wilson, R. K., & Eichler, E. E. (2015). Genetic variation and the *de novo* assembly of human genomes. *Nature Reviews Genetics*, *16*, 627–640. https://doi.org/10.1038/nrg3933

CITES. (2020). *Appendices I, II and III*. Convention on International Trade in Endangered Species of Wild Fauna and Flora. https://cites.org/eng/app/appendices.php

Crameri, S. (2020). *Phylogenomics, species discovery and integrative taxonomy in Dalbergia (Fabaceae) precious woods from Madagascar*. Doctoral Thesis, ETH Zurich. https://doi.org/10.3929/ethz-b-000487274

Crameri, S., Fior, S., Zoller, S., & Widmer, A. (2022a). Data from: A target capture approach for phylogenomic analyses at multiple evolutionary timescales in rosewoods (*Dalbergia* spp.) and the legume family (Fabaceae). European Nucleotide Archive (ENA). https://www.ebi.ac.uk/ena/browser/view/PRJEB41848

Crameri, S., Fior, S., Zoller, S., & Widmer, A. (2022b). Data from: A target capture approach for phylogenomic analyses at multiple evolutionary timescales in rosewoods (*Dalbergia* spp.) and the legume family (Fabaceae). *Dryad Digital Repository*. https://doi.org/10.5061/dryad.73n5tb2z7

Cutter, A. D. (2013). Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. *Molecular Phylogenetics and Evolution*, *69*(3), 1172–1185. https://doi.org/10.1016/j.ympev.2013.06.006

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

de Carvalho, A. M. (1997). A synopsis of the genus *Dalbergia* (Fabaceae: Dalbergieae) in Brazil. *Brittonia*, *49*, 87–109. https://doi.org/10.2307/2807701

de La Harpe, M., Hess, J., Loiseau, O., Salamin, N., Lexer, C., & Paris, M. (2018). A dedicated target capture approach reveals variable genetic markers across micro- and macro-evolutionary time scales in palms. *Molecular Ecology Resources*, *19*(1), 221–234. https://doi.org/10.1111/1755-0998.12945

de La Harpe, M., Paris, M., Karger, D. N., Rolland, J., Kessler, M., Salamin, N., & Lexer, C. (2017). Molecular ecology studies of species radiations: Current research gaps, opportunities and challenges. *Molecular Ecology*, *26*(10), 2608–2622. https://doi.org/10.1111/mec.14110

de Sousa, F., Bertrand, Y. J. K., Nylinder, S., Oxelman, B., Eriksson, J. S., & Pfeil, B. E. (2014). Phylogenetic properties of 50 nuclear loci in *Medicago* (Leguminosae) generated using multiplexed sequence capture and next-generation sequencing. *PLoS ONE*, *9*(10), e109704. https://doi.org/10.1371/journal.pone.0109704

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., … Saleem, M. (2017). An ecoregion-based approach to protecting half the terrestrial realm. *BioScience*, *67*(6), 534–545. https://doi.org/10.1093/biosci/bix014

Dowle, M., & Srinivasan, A. (2019). *data.table: Extension of 'data.frame'*. R package version 1.12.8. https://CRAN.R-project.org/package=data.table

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*(1), 11–15.

Egan, A. N., & Vatanparast, M. (2019). Advances in legume research in the genomics era. *Australian Systematic Botany*, *32*(6), 459–483. https://doi.org/10.1071/SB19019

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, *32*(5), 786–788. https://doi.org/10.1093/bioinformatics/btv646

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, *61*(5), 717–726. https://doi.org/10.1093/sysbio/sys004

Garrison, E. (2012). *Vcflib: A C++ library for parsing and manipulating VCF files*. Github. https://github.com/ekg/vcflib

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv*, *1207*(3907), 1–9. https://arxiv.org/abs/1207.3907

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*, 644–652. https://doi.org/10.1038/nbt.1883

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – A baiting and iterative mapping approach. *Nucleic Acids Research*, *41*(13), e129. https://doi.org/10.1093/nar/gkt371

Hassold, S., Lowry, P. P., II, Bauert, M. R., Razafintsalama, A., Ramamonjisoa, L., & Widmer, A. (2016). DNA barcoding of Malagasy rosewoods: towards a molecular identification of CITES-listed *Dalbergia* species. *PLoS ONE*, *11*(6), e0157881. https://doi.org/10.1371/journal.pone.0157881

Hughes, C. E., & Eastwood, R. (2006). Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(27), 10, 334–10,339. https://doi.org/10.1073/pnas.0601928103

Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., Zerega, N. J. C., & Wickett, N. J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences*, *4*(7), 1600016. https://doi.org/10.3732/apps.1600016

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., Eiserhardt, W. L., Epitawalage, N., Forest, F., Kim, J. T., Leebens-Mack, J. H., Leitch, I. J., Maurin, O., Soltis, D. E., Soltis, P. S., Wong, G. K.-S., Baker, W. J., & Wickett, N. J. (2019). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology*, *68*(4), 594–606. https://doi.org/10.1093/sysbio/syy086

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). adegenet 1.3–1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, *25*(1), 185–202. https://doi.org/10.1111/mec.13304

Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*, *26*(13), 1669–1670.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281

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

Knaus, B. J., & Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53. https://doi.org/10.1111/1755-0998.12549

Koenen, E. J. M., Kidner, C., Souza, E. R., Simon, M. F., Iganci, J. R., Nicholls, J. A., Brown, G. K., de Queiroz, L. P., Luckow, M., Lewis, G. P., Pennington, R. T., & Hughes, C. E. (2020). Hybrid capture of 964 nuclear genes resolves evolutionary relationships in the mimosoid legumes and reveals the polytomous origins of a large pantropical radiation. *American Journal of Botany*, *107*(12), 1710–1735. https://doi.org/10.1002/ajb2.1568

Koenen, E. J. M., Ojeda, D. I., Bakker, F. T., Wieringa, J. J., Kidner, C., Hardy, O. J., Pennington, R. T., Herendeen, P. S., Bruneau, A., & Hughes, C. E. (2021). The origin of the legumes is a complex paleopolyploid phylogenomic tangle closely associated with the Cretaceous-Paleogene (K-Pg) mass extinction event. *Systematic Biology*, *70*(3), 508–526. https://doi.org/10.1093/sysbio/syaa041

Koenen, E. J. M., Ojeda, D. I., Steeves, R., Migliore, J., Bakker, F. T., Wieringa, J. J., Kidner, C., Hardy, O. J., Pennington, R. T., Bruneau, A., & Hughes, C. E. (2020). Large-scale genomic sequence data resolve the deepest divergences in the legume phylogeny and support a near-simultaneous evolutionary origin of all six subfamilies. *New Phytologist*, *225*(3), 1355–1369. https://doi.org/10.1111/nph.16290

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Lachenaud, O. (2016). *Dalbergia* L. f., nom. Cons. In M. S. M. Sosef, J. Florence, L. N. Banak, H. P. B. Bourobou, P. Bissiengou, et al. (Eds.), *Flore du Gabon* (Vol. *49*, pp. 101–153). Margraf Publishers.

Lavin, M., Pennington, R. T., Klitgaard, B. B., Sprent, J. I., de Lima, H. C., & Gasson, P. E. (2001). The dalbergioid legumes (Fabaceae):

delimitation of a pantropical monophyletic clade. *American Journal of Botany*, *88*(3), 503–533. https://doi.org/10.2307/2657116

Lee, A. K., Gilman, I. S., Srivastav, M., Lerner, A. D., Donoghue, M. J., & Clement, W. L. (2021). Reconstructing Dipsacales phylogeny using Angiosperms353: issues and insights. *American Journal of Botany*, *108*(7), 1122–1142. https://doi.org/10.1002/ajb2.1695

Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, *61*(5), 727–744. https://doi.org/10.1093/sysbio/sys049

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. https://arxiv.org/abs/1303.3997

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lindenbaum, P. (2015). *Jvarkit: java utilities for bioinformatics*. Github. https://github.com/lindenb/jvarkit

LPWG (Legume Phylogeny Working Group). (2017). A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon*, *66*(1), 44–77. https://doi.org/10.12705/661.3

Lunter, G., & Goodson, M. (2011). Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research*, *21*, 936–939. https://doi.org/10.1101/gr.111120.110

Madagascar Catalogue. (2022). *Catalogue of the Plants of Madagascar*. Missouri Botanical Garden. http://www.tropicos.org/Project/Madagascar

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods*, *7*, 111–118. https://doi.org/10.1038/nmeth.1419

Mandel, J. R., Dikow, R. B., Funk, V. A., Masalia, R. R., Staton, S. E., Kozik, A., Michelmore, R. W., Rieseberg, L. H., & Burke, J. M. (2014). A target enrichment method for gathering phylogenetic information from hundreds of loci: an example from the *Compositae*. *Applications in Plant Sciences*, *2*(2), 1300085. https://doi.org/10.3732/apps.1300085

McMahon, M., & Hufford, L. (2004). Phylogeny of Amorpheae (Fabaceae: Papilionoideae). *American Journal of Botany*, *91*(8), 1219–1230. https://doi.org/10.3732/ajb.91.8.1219

Mirarab, S., Reaz, R., Bayzid, M. S., Zimmermann, T., Swenson, M. S., & Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, *30*(17), i541–i548. https://doi.org/10.1093/bioinformatics/btu462

Mousavi-Derazmahalleh, M., Bayer, P. E., Hane, J. K., Valliyodan, B., Nguyen, H. T., Nelson, M. N., Erskine, W., Varshney, R. K., Papa, R., & Edwards, D. (2018). Adapting legume crops to climate change using genomic approaches. *Plant, Cell & Environment*, *42*(1), 6–19. https://doi.org/10.1111/pce.13203

Nicholls, J. A., Pennington, R. T., Koenen, E. J. M., Hughes, C. E., Hearn, J., Bunnefeld, L., Dexter, K. G., Stone, G. N., & Kidner, C. A. (2015). Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Science*, *6*, 710. https://doi.org/10.3389/fpls.2015.00710

Ottenlips, M. V., Mansfield, D. H., Buerki, S., Feist, M. A. E., Downie, S. R., Dodsworth, S., Forest, F., Plunkett, G. M., & Smith, J. F. (2021). Resolving species boundaries in a recent radiation with the Angiosperms353 probe set: the *Lomatium packardiae*/*L. anomalum* clade of the *L. triternatum* (Apiaceae) complex. *American Journal of Botany*, *108*(7), 1217–1233. https://doi.org/10.1002/ajb2.1676

Paradis, E., & Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

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

Peng, Z., Fan, W., Wang, L., Paudel, D., Leventini, D., Tillman, B. L., & Wang, J. (2017). Target enrichment sequencing in cultivated peanut (*Arachis hypogaea* L.) using probes designed from transcript sequences. *Molecular Genetics and Genomics*, *292*, 955–965. https://doi.org/10.1007/s00438-017-1327-z

Prain, D. (1904). The Species of *Dalbergia* of South-Eastern Asia. *Annals of the Royal Botanic Garden, Calcutta*, *10*(1), 1–114.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, *526*, 569–573. https://doi.org/10.1038/nature15697

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org

Ribeiro, R. A., Lavin, M., Lemos-Filho, J. P., & Filho, C. (2007). The genus *Machaerium* (Leguminosae) is more closely related to *Aeschynomene* Sect. *Ochopodium* than to *Dalbergia*: Inferences from combined sequence data. *Systematic Botany*, *32*(4), 762–771. https://doi.org/10.1043/06-79.1

Rissler, L. J. (2016). Union of phylogeography and landscape genetics. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 8079–8086. https://doi.org/10.1073/pnas.1601073113

Safonova, Y., Bankevich, A., & Pevzner, P. A. (2015). dipSPAdes: Assembler for highly polymorphic diploid genomes. *Journal of Computational Biology*, *22*(6), 528–545. https://doi.org/10.1089/cmb.2014.0153

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425.

Schuurman, D., & Lowry, P. P., II. (2009). The Madagascar rosewood massacre. *Madagascar Conservation and Development*, *4*, 98–102. https://doi.org/10.4314/mcd.v4i2.48649

Shah, T., Schneider, J. V., Zizka, G., Maurin, O., Baker, W., Forest, F., Brewer, G. E., Savolainen, V., Darbyshire, I., & Larridon, I. (2021). Joining forces in *Ochnaceae* phylogenomics: a tale of two targeted sequencing probe kits. *American Journal of Botany*, *108*(7), 1201–1216. https://doi.org/10.1002/ajb2.1682

Shavvon, R. S., Osaloo, S. K., Maassoumii, A. A., Moharrek, F., Erkul, S. K., Lemmon, A. R., Lemmon, E. M., Michalak, I., Muellner-Riehl, A. N., & Favre, A. (2017). Increasing phylogenetic support for explosively radiating taxa: The promise of high-throughput sequencing for *Oxytropis* (Fabaceae). *Journal of Systematics and Evolution*, *55*(4), 385–404. https://doi.org/10.1111/jse.12269

Siniscalchi, C. M., Hidalgo, O., Palazzesi, L., Pellicer, J., Pokorny, L., Maurin, O., Leitch, I. J., Forest, F., Baker, W. J., & Mandel, J. R. (2021). Lineage-specific vs. universal: A comparison of the Compositae1061 and Angiosperms353 enrichment panels in the sunflower family. *Applications in Plant Sciences*, *9*(7), e11422. https://doi.org/10.1002/aps3.11422

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 31. https://doi.org/10.1186/1471-2105-6-31

Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular sequences. *Journal of Molecular Biology*, *147*, 195–197.

Sprent, J. I., Ardley, J., & James, E. K. (2017). Biogeography of nodulated legumes and their nitrogen-fixing symbionts. *New Phytologist*, *215*(1), 40–56. https://doi.org/10.1111/nph.14474

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Tange, O. (2011). GNU parallel: the command-line power tool. *Login*, *36*(1), 42–47. https://www.usenix.org/system/files/login/articles/105438-Tange.pdf

Thomas, S. K., Liu, X., Du, Z. Y., Dong, Y., Cummings, A., Pokorny, L., Xiang, Q. Y., & Leebens-Mack, J. H. (2021). Comprehending Cornales: phylogenetic reconstruction of the order using the Angiosperms353 probe set. *American Journal of Botany*, *108*(7), 1112–1121. https://doi.org/10.1002/ajb2.1696

Ufimov, R., Zeisek, V., Píšová, S., Baker, W. J., Fér, T., van Loo, M., Dobeš, C., & Schmickl, R. (2021). Relative performance of customized and universal probe sets in target enrichment: A case study in subtribe Malinae. *Applications in Plant Sciences*, *9*(7), e11442. https://doi.org/10.1002/aps3.11442

UNODC. (2016a). *Best practice guide for forensic timber identification*. United Nations Office on Drugs and Crime. United Nations. https://www.unodc.org/documents/Wildlife/Guide_Timber.pdf

UNODC. (2016b). *World wildlife crime report: trafficking in protected species*. United Nations Office on Drugs and Crime. United Nations. https://www.unodc.org/documents/data-and-analysis/wildlife/World_Wildlife_Crime_Report_2016_final.pdf

UNODC. (2020). *World wildlife crime report: trafficking in protected species*. United Nations Office on Drugs and Crime. United Nations. https://www.unodc.org/documents/data-and-analysis/wildlife/2020/World_Wildlife_Report_2020_9July.pdf

Vardeman, E., & Runk, J. V. (2020). Panama's illegal rosewood logging boom from *Dalbergia* retusa. *Global Ecology and Conservation*, *23*, e01098. https://doi.org/10.1016/j.gecco.2020.e01098

Vatanparast, M., Powell, A., Doyle, J. J., & Egan, A. N. (2018). Targeting legume loci: A comparison of three methods for target enrichment bait design in Leguminosae phylogenomics. *Applications in Plant Sciences*, *6*(3), e1036. https://doi.org/10.1002/aps3.1036

Vences, M., Wollenberg, K. C., Vieites, D. R., & Lees, D. C. (2009). Madagascar as a model region of species diversification. *Trends in Ecology and Evolution*, *24*(8), 456–465. https://doi.org/10.1016/j.tree.2009.03.011

Waeber, P. O., Schuurman, D., Ramamonjisoa, B., Langrand, M., Barber, C. V., Innes, J. L., Lowry, P. P., II, & Wilmé, L. (2019). Uplisting of Malagasy precious woods critical for their survival. *Biological Conservation*, *235*, 89–92. https://doi.org/10.1016/j.biocon.2019.04.007

WCVP. (2021). *World Checklist of Vascular Plants, version 2.0*. http://wcvp.science.kew.org

Weitemier, K., Straub, S. C. K., Cronn, R. C., Fishbein, M., Schmickl, R., McDonnell, A., & Liston, A. (2014). Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences*, *2*(9), 1400042.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *The Journal of Open Source Software*, *4*(43), 1686.

Wilding, N., Phillipson, P. B., Crameri, S. Andriambololonera, S., Andriamiarisoa, R. L., Andrianarivelo, S. A. F., Bernard, R., Rakotonirina, N., Rakotovao, C., Randrianaivo, R., Razakamalala, R., & Lowry, P. P. (2021). Taxonomic studies on Malagasy *Dalbergia* (Fabaceae). I. Two new species from northern Madagascar, and an emended description for *D. manongarivensis*. *Candollea*, *76*(2), 237–249. https://doi.org/10.15553/c2021v762a4

Wilding, N., Phillipson, P. B., & Crameri, S. (2021). Taxonomic studies on Malagasy *Dalbergia* (Fabaceae). II. A new name for *D. mollis* and the reinstatement of *D. chermezonii*. *Candollea*, *76*(2), 251–257. https://doi.org/10.15553/c2021v762a5

Wojciechowski, M. F. (2013). Towards a new classification of Leguminosae: Naming clades using non-Linnaean phylogenetic nomenclature. *South African Journal of Botany*, *89*, 85–93. https://doi.org/10.1016/j.sajb.2013.06.017

Wojciechowski, M. F., Lavin, M., & Sanderson, M. J. (2004). A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *American Journal of Botany*, *91*(11), 1846–1862. https://doi.org/10.3732/ajb.91.11.1846

Yardeni, G., Viruel, J., Paris, M., Hess, J., Crego, C. G., de La Harpe, M., Rivera, N., Barfuss, M. H. J., Till, W., Guzmán-Jacob, V., Krömer, T., Lexer, C., Paun, O., & Leroy, T. (2022). Taxon-specific or universal? Using target capture to study the evolutionary history of rapid radiations. *Molecular Ecology Resources*, *22*, 927–945. https://doi.org/10.1111/1755-0998.13523

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2016). GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

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

Zhang, R., Wang, Y.-H., Jin, J.-J., Stull, G. W., Bruneau, A., Cardoso, D., de Queiroz, L. P., Moore, M. J., Zhang, S.-D., Chen, S.-Y., Wang, J., Li, D.-Z., & Yi, T.-S. (2020). Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. *Systematic Biology*, *69*(4), 613–622. https://doi.org/10.1093/sysbio/syaa013

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., Chang, W.-C., Zhang, L., Zhang, X., Tang, R., Garg, V., Wang, X., Tang, H., Chow, C.-N., Wang, J., Deng, Y., Wang, D., Khan, A. W., Yang, Q., … Varshney, R. K. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nature Genetics*, *51*, 865–876. https://doi.org/10.1038/s41588-019-0402-2

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Crameri, S., Fior, S., Zoller, S., & Widmer, A. (2022). A target capture approach for phylogenomic analyses at multiple evolutionary timescales in rosewoods (*Dalbergia* spp.) and the legume family (Fabaceae). *Molecular Ecology Resources*, *22*, 3087–3105. https://doi.org/10.1111/1755-0998.13666