Diss. ETH No. 27241

PHYLOGENOMICS, SPECIES DISCOVERY AND INTEGRATIVE TAXONOMY IN *DALBERGIA* (FABACEAE) PRECIOUS WOODS FROM MADAGASCAR

A dissertation submitted to

ETH ZURICH

for the degree of

DOCTOR OF SCIENCES

presented by

SIMON FLAVIO CRAMERI

MSc ETH Biology
born April 4, 1989
citizen of Poschiavo (GR)

accepted on the recommendation of

Prof. Dr. Alex Widmer (examiner)
Prof. Dr. Porter P. Lowry II (co-examiner)
Prof. Dr. Rolf Holderegger (co-examiner)

2020

Table of Contents

Summary

Understanding the evolution of biological diversity is of fundamental interest to molecular ecologists and conservation biologists. In particular in biodiversity hotspots, major knowledge gaps exist regarding the numbers of species, their morphological distinction, ecology, distribution and risk of extinction, as well as their possible commercial value and use. The pantropical genus *Dalbergia* L.f. (Fabaceae) includes precious timber species known as rosewoods or palisanders, many of which are threatened by habitat degradation and often illegal selective logging. The use of existing species circumscriptions and application of available identification keys is particularly challenging in Madagascar, where a large number of closely related species co-occur and where the reliance on flower and fruit characters for identification is high, even though these structures are often absent on living specimens. The resulting taxonomic uncertainties have important implications for national and international trade regulations and impede the setting of conservation priorities.

The aim of this Ph.D. thesis was to increase our understanding of the number, morphological distinction and phylogenetic relationships between the *Dalbergia* species present in Madagascar. Towards this goal, I developed a genome-wide DNA analysis approach that I used in combination with morphological and eco-geographic analyses. In the first chapter, my collaborators and I explored the utility of target enrichment sequencing to assess genetic variation in *Dalbergia* at the level of populations and closely related species, and between divergent species within the genus. Specifically, we assessed the sequence capture sensitivity and specificity of 7,201 candidate genomic regions with a conserved core across available genomes of Papilionoideae, and identified 2,396 regions that can be efficiently recovered for genomic analyses across *Dalbergia*. We also explored the performance of the 7,201 candidate regions across legumes, and identified 1,005 target regions that can be efficiently recovered for analyses at the family level. These confirmed high levels of gene tree discordance, previously reported also by other authors, especially near the root of the family. Furthermore, our analyses resulted in the development of a general-purpose bioinformatics pipeline for scalable, streamlined and reproducible analysis of such genome-wide and complex datasets.

In the second chapter, we employed the developed genomic and bioinformatic resources for species discovery within the genus *Dalbergia* from Madagascar and the Comoros. This led to the confirmation of 46 out of 49 already described and accepted species as separately evolving lineages, while three currently accepted species were identified as potential synonyms. Furthermore, we discovered 49 new candidate species, of which 14 could be delimited as separately evolving from one another and from the confirmed described species. Lineage separation could not be conclusively assessed for the

remaining 35 candidate species because of a currently insufficient availability of high-quality collections. Phylogenomic analysis confirmed the existence of two diverse clades of Malagasy *Dalbergia*, one of which appears to be associated with a clade of species from continental Africa, as well as an isolated lineage, which also occurs in continental East Africa. It further revealed a fourth lineage present in Madagascar, which is connected to another clade of species from continental Africa, and a biogeographic connection between species of West Madagascar and the Comoros.

In the third chapter, we completed the integrative taxonomic workflow for three confirmed evolutionary species of high conservation concern. Specifically, we newly described *D. pseudomaritima* Crameri, Phillipson & N. Wilding and *D. razakamalalae* Crameri, Phillipson & N. Wilding as distinct species occurring in littoral and low-elevation evergreen humid forests in southeast Madagascar, and emended the description of *D. maritima* R. Vig. from central-east Madagascar to exclude the two south-eastern species, with which it had previously been confused. For all three species we also performed conservation assessments according to the IUCN Red List categories and criteria, and assigned the category *Endangered* (EN) to all of them.

A major result from this dissertation is the finding that the diversity of *Dalbergia* species in Madagascar has hitherto been vastly underestimated. Thanks to major recent efforts to sample and document *Dalbergia* species, and the developed sequence capture and high-throughput DNA sequencing approach, which provides unprecedented phylogenomic resolution, part of the species diversity existing in Madagascar could be uncovered. In combination with morphological and eco-geographic data we were able to show that several threatened species are not currently represented in existing protected areas and that some highly exploited species with presumably large distribution ranges are actually composed of several species with much narrower distribution ranges. These findings call for further investigations and conservation action, and provide scientific evidence that will hopefully foster science-based conservation efforts.

Zusammenfassung

Das Verständnis der Evolution der biologischen Vielfalt ist für Molekularökologen und Naturschutzbiologen von grundlegendem Interesse. Insbesondere in den Biodiversitäts-Hotspots bestehen grosse Wissenslücken bezüglich der Anzahl Arten, ihrer morphologischen Unterscheidung, Ökologie, Verbreitung und Gefährdung, sowie bezüglich deren Nutzung und kommerziellen Wertes. Die in den Tropen weit verbreitete Gattung *Dalbergia* L.f. (Hülsenfrüchtler, Fabaceae) umfasst zahlreiche Edelholzarten, die als Rosenholz oder Palisander bekannt sind und von denen viele durch Lebensraumverlust und häufig illegalem selektiven Holzeinschlag bedroht sind. In Madagaskar gibt es eine besonders grosse Anzahl an eng verwandten und morphologisch ähnlichen Arten. Die Anwendung der verfügbaren Bestimmungsschlüssel ist aber schwierig, weil man für eine Bestimmung auf Blüten- und Fruchtmerkmale angewiesen ist, die bei lebenden Exemplaren oder auf den Herbarbelegen von biologischen Sammlungen oft fehlen. Es bestehen auch Unsicherheiten darüber, ob die verschiedenen Arten in ihrer Taxonomie so voneinander abgegrenzt werden, dass die real existierenden evolutionären Arten tatsächlich abgebildet werden. Diese taxonomischen Unsicherheiten haben erhebliche negative Auswirkungen auf die Definition, Anwendung und Wirksamkeit von Handelsbestimmungen und Naturschutzmassnahmen.

Das Ziel dieser Doktorarbeit war es, unser Verständnis über die Anzahl, die morphologische Unterscheidung und die phylogenetischen Beziehungen zwischen madagassischen Arten von *Dalbergia* zu verbessern. Um dieses Ziel zu erreichen, entwickelte ich einen Ansatz zur genomweiten DNA-Analyse, den ich in Kombination mit morphologischen und ökogeographischen Untersuchungen verwendete. Im ersten Kapitel erforschten ich und meine Projektpartner den Nutzen der Sequenzierung durch spezifische Anreicherung bestimmter Genom-Regionen (engl. *target enrichment sequencing*), um die genetische Variation in *Dalbergia* auf der Ebene von Populationen und eng verwandten Arten, sowie zwischen divergierenden Arten innerhalb der Gattung, zu bewerten. Insbesondere beurteilten wir die Sensitivität und Spezifität der Anreicherung von 7201 Genom-Regionen und identifizierten dabei 2396 Regionen, die für genomweite DNA-Analysen in verschiedenen Gruppen (Kladen) der Gattung *Dalbergia* effizient analysiert werden können. Zusätzlich identifizierten wir 1005 Regionen, die für genomweite DNA-Analysen in verschiedenen Gruppen der ganzen Familie der Hülsenfrüchtler (Fabaceae) analysiert werden können. Phylogenetische Untersuchungen auf Ebene der Familie bestätigten ein bekanntes Phänomen, nämlich die unterschiedliche stammesgeschichtliche Entwicklung verschiedener Genom-Regionen, wodurch wir unseren genetischen Ansatz validieren konnten. Darüber hinaus entwickelten wir eine bioinformatische Analyse-

Pipeline für eine vereinfachte, erweiterbare und reproduzierbare Analyse solcher komplexer Datensätze.

Im zweiten Kapitel setzten wir die entwickelten genomischen und bioinformatischen Ressourcen für die Entdeckung von verschiedenen Arten innerhalb der Gattung *Dalbergia* aus Madagaskar und den Komoren ein. Dies führte zur Bestätigung von 46 aus 49 bereits beschriebenen und von der wissenschaftlichen Gemeinschaft akzeptierten Arten, und zur Identifizierung von drei derzeit akzeptieren Arten als potenzielle Synonyme anderer Arten. Darüber hinaus entdeckten wir 49 neue Kandidatenarten, von denen wir 14 als sich getrennt voneinander entwickelnde evolutionäre Arten abgrenzen konnten. Die Abstammungstrennung konnte für die verbleibenden 35 Kandidatenarten nicht abschliessend beurteilt werden, da derzeit nicht genügend qualitativ hochwertige Sammlungen zur Verfügung stehen. Eine genomweite Analyse der stammesgeschichtlichen Beziehungen bestätigte die Existenz von zwei verschiedenen Kladen madagassischer *Dalbergia*, von denen eine mit einer Gruppe von Arten aus dem kontinentalen Afrika assoziiert zu sein scheint, sowie eine stammesgeschichtlich isolierte madagassische Art, die auch im kontinentalen Ostafrika vorkommt. Darüber hinaus entdeckten wir eine stammesgeschichtlich isolierte und endemische madagassische Art, die mit einer anderen Gruppe von Arten aus dem kontinentalen Afrika assoziiert ist, sowie eine biogeographische Verbindung zwischen Arten aus West-Madagaskar und den Komoren.

Im dritten Kapitel komplettierten wir unseren integrativen taxonomischen Ansatz für drei bestätigte evolutionäre Arten, die stark gefährdet sind. Wir beschrieben *D. pseudomaritima* Crameri, Phillipson & N. Wilding und *D. razakamalalae* Crameri, Phillipson & N. Wilding als neue Arten aus Südost-Madagaskar. Wir revidierten zudem die Beschreibung von *D. maritima* R. Vig. aus dem zentralöstlichen Madagaskar, um die beiden südöstlichen Arten auszuschließen, mit denen sie zuvor verwechselt worden war. Für alle drei Arten führten wir auch eine Bewertung des Erhaltungszustands nach den Kategorien und Kriterien der Roten Liste (IUCN) durch und wiesen ihnen die Kategorie *stark gefährdet* (EN) zu.

Diese Dissertation zeigt auf, dass die Vielfalt der ökologisch und ökonomisch bedeutsamen *Dalbergia*-Arten in Madagaskar bisher stark unterschätzt wurde. Dank der jüngsten großen Anstrengungen zur Sammlung und Dokumentation von *Dalbergia*-Arten und dem modernen DNA-Sequenzierungsansatz, der eine grosse stammesgeschichtliche Auflösung liefert, konnte ein Teil der in Madagaskar vorkommenden, aber bisher übersehenen Artenvielfalt, aufgedeckt werden. In Kombination mit morphologischen und ökogeographischen Daten konnten wir zeigen, dass mehrere bedrohte Arten in bestehenden

Schutzgebieten möglicherweise nicht vertreten sind und dass einige stark genutzte Arten mit scheinbar grossen Verbreitungsgebieten tatsächlich aus mehreren Arten mit viel kleineren Verbreitungsgebieten bestehen. Diese Erkenntnisse zeigen den Handlungsbedarf zur weiteren Erforschung und liefern wissenschaftlich fundierte Grundlagen, die hoffentlich ebenso fundierte Naturschutzmassnahmen fördern werden.

General Introduction

Species are the fundamental unit in ecology and evolution and are a core component of biodiversity, together with genetic diversity within species, habitat diversity and the interactions among these components (GASTON & SPICER 2013). Despite this fundamental importance of species, we have only a very limited understanding of the global diversity of species (PIMM *et al.* 2014). Sixteen leading hotspots of biodiversity are found in the tropics and include the Tropical Andes, Sundaland and Madagascar ranking highest in terms of vertebrate and plant endemism (MYERS *et al.* 2000). These three regions alone contain an estimated 11% of the global vertebrate diversity, and 15% of the global plant diversity in terms of species richness (MYERS *et al.* 2000). Tropical ecosystems also rank highest among the most threatened ecosystems (BARLOW *et al.* 2018). In this context, Madagascar stands out as a region with an exceptionally high endemism in relation to its area, multiple endemic genera and families (MYERS *et al.* 2000), as well as high past and future predicted rates of deforestation mediated by rapid population growth (VIEILLEDENT *et al.* 2013, VIEILLEDENT *et al.* 2018). Many studies have revealed an underestimation of the actual biodiversity in Madagascar in multiple groups of organisms assessed in the recent past, e.g. in amphibians (VIEITES *et al.* 2009), birds (YOUNGER *et al.* 2018), lemurs (HOTALING *et al.* 2016), and many plant genera, including the endemic genus *Capurodendron* (GAUTIER & NACIRI 2018) or in potential sources of ebony wood (SCHATZ & LOWRY 2011, 2020).

Study system

The genus *Dalbergia* L.f. (Fabaceae) includes valuable timber species, many of which are threatened by overexploitation. The entire genus has been listed in Appendix II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora, except for *D. nigra*, which is listed in Appendix I and therefore entirely banned from international trade (CITES 2019). The CITES Appendix II lists species that may become critically endangered unless trade is regulated by so-called non-detriment findings (but see MURPHY 2006) and export permits. No export is currently allowed for *Dalbergia* populations of Madagascar (decree no. 2010-141 of the Malagasy government). Because of prolonged political instability, poor governance, weak enforcement and rampant corruption, illegal logging of Malagasy *Dalbergia* and other precious woods continues to this date, despite national and international regulations (SCHUURMAN & LOWRY 2009, WAEBER *et al.* 2019). The combination of significant threats defined by habitat degradation and selective illegal logging, and knowledge gaps concerning the actual biological diversity demand more basic research in this genus, especially in Madagascar.

Main methods

The process of species delimitation, i.e., the assessment of species-level biological diversity, often involves two steps. First, putative species are discovered within a set of related individuals (species discovery, CARSTENS *et al.* 2013). This step does not require any prior grouping or assignment of individuals, which is a prerequisite if prior hypotheses about the number of species in a particular area or group are unavailable or unrealistic (SCHLICK-STEINER *et al.* 2010). In a second step, it is assessed whether the discovered putative species are sufficiently distinct from one another to recognise them as different species (species validation, CARSTENS *et al.* 2013). There are multiple possible and often contradicting criteria to assess distinction (reviewed by ZACHOS 2016), which is why systematists increasingly regard species as hypotheses of separately evolving lineages (DE QUEIROZ 2005). Separate evolution is increasingly being inferred by integration of various sources of information, such as genetic, morphological or ecological distinction, in a process named integrative taxonomy (DAYRAT 2005, SCHLICK-STEINER *et al.* 2010).

Genetic (VIEITES *et al.* 2009, HOTALING *et al.* 2016) and genomic approaches (WACHTER *et al.* 2015, YOUNGER *et al.* 2018, NATUSCH *et al.* 2020) have supported the discovery and delimitation of many new species in biodiverse and understudied regions. Before the significant advances in sequencing technologies, DNA barcoding was the standard tool to perform species-level assignments, mainly relying on mitochondrial genes to distinguish among animals and chloroplast genes to distinguish among plants (HEBERT *et al.* 2003, HOLLINGSWORTH *et al.* 2011). A recent evaluation of the DNA barcoding approach using three chloroplast markers to distinguish among *Dalbergia* species from Madagascar and other regions revealed resolution on a macro-geographic scale, enabling accurate distinction of Malagasy from non-Malagasy taxa, as well as between two major Malagasy subgroups, but proved to be insufficient to distinguish between closely related species (HASSOLD *et al.* 2016). New sequencing technologies and significant reduction of costs for genome-wide analysis of nuclear and plastid variation have facilitated the integration of population genomics and phylogenomics into species discovery and species delimitation (CARSTENS *et al.* 2013).

Target enrichment sequencing has emerged as a powerful approach to reveal genetic structure and phylogenetic signals informative at various evolutionary time scales, including those relevant to assess lineage separation of closely related species (JONES AND GOOD 2016). This approach has been successfully used for species delimitation in animals and plants, often in combination with other methods (e.g. KARBSTEIN *et al.* 2020), and has also been adopted to analyse museum collections (HART *et al.* 2016, MCCORMACK *et al.* 2016).

Morphological characters have long been used to describe species and have traditionally been the predominant source of information to distinguish between species using identification keys. In the current taxonomic treatment of *Dalbergia* species from Madagascar and the Comoros (BOSSER & RABEVOHITRA 1996, 2002, 2005), flower and fruit characters are used for identification of most species. These are, however, often absent on living specimens or museum collections. The imparipinnate leaves of Malagasy *Dalbergia* can be examined in most collections. They are variable between species, but a large variation has also been noted within species, and particularly leaflet shape is difficult to represent by standard terminology owing to its variability. Morphometric analysis (BONHOMME *et al.* 2014) is an approach to quantify variation in leaflet shape in a consistent way. It allows a researcher to identify and isolate uncorrelated principal shape components, which can be analysed regarding their mean and variance within individuals, within species and between species (LEXER *et al.* 2009).

In a complete integrative taxonomic workflow, species-level diversity needs to be discovered, validated and formally described (PANTE *et al.* 2015). For species descriptions to have a conservation impact, the described species also need an assessment of their risk of extinction. The International Union for Conservation of Nature (IUCN) provides guidelines and criteria to support consistent and objective assessments, which result in the assignment of threat categories to described species (IUCN 2012).

Thesis outline and research questions

In the following thesis chapters, I will present the application of target enrichment sequencing, phylogenomics and leaf morphometrics to assess the species-level diversity in *Dalbergia* rosewoods and palisanders from Madagascar. The first chapter focuses on the development of genomic resources and a bioinformatics pipeline for analysis of genome-wide target enrichment sequencing data in the legume family (Fabaceae), and specifically in *Dalbergia*. The second chapter focuses on species discovery, phylogenomics and species delimitation (validation) in Malagasy and Comorian *Dalbergia* species. The following research questions were addressed in the second chapter:

(i) What are the phylogenetic relationships among *Dalbergia* species from Madagascar, and what are the biogeographic connections of Malagasy taxa with those from other regions?

(ii) How many candidate species can be discovered in Madagascar based on multi-locus nuclear genetic data?

(iii) Which candidate species are coherent and distinct from other such species based on integration of leaf morphology and eco-geography?

(iv) Do current taxonomic circumscriptions of Malagasy *Dalbergia* species correspond to the inferred candidate species?

(v) How is the eco-geographic diversity distributed across different phylogenetic groups?, and

(vi) Where are the hotspots of species richness and phylogenetic diversity, and are these included in Madagascar's protected areas network?

Finally, the third chapter completes the integrative taxonomic workflow and formalises three species descriptions as well as conservation assessments for these species.

References

Barlow, J., F. França, T. A. Gardner, C. C. Hicks, G. D. Lennox, E. Berenguer, L. Castello, E. P. Economo, J. Ferreira, B. Guénard, C. G. Leal, V. Isaac, A. C. Lees, C. L. Parr, S. K. Wilson, P. J. Young, and N. A. J. Graham. 2018. The future of hyperdiverse tropical ecosystems. Nature **559**:517-526.

Bonhomme, V., S. Picq, and C. Gaucherel. 2014. Momocs: outline analysis using R. Journal of Statistical Software **56**:1-24.

Bosser, J., and R. Rabevohitra. 1996. Taxa et noms nouveaux dans le genre *Dalbergia* (Papilionaceae) à Madagascar et aux Comores. Bull. Mus. Natl. Hist. Nat. B, Adansonia **18**:171-212.

Bosser, J., and R. Rabevohitra. 2002. Tribe Dalbergieae. Pages 321-361 *in* D. J. Du Puy, J. N. Labat, R. Rabevohitra, J. F. Villiers, J. Bosser, and J. Moat, editors. The Leguminosae of Madagascar. Royal Botanical Gardens, Kew.

Bosser, J., and R. Rabevohitra. 2005. Espèces nouvelles dans le genre *Dalbergia* (Fabaceae, Papilionoideae) à Madagascar. Adansonia **27**:209-216.

Carstens, B. C., T. A. Pelletier, N. M. Reid, and J. D. Satler. 2013. How to fail at species delimitation. Molecular Ecology **22**:4369-4383.

CITES. 2019. Appendices I, II and III, https://cites.org/eng/app/appendices.php accessed on Feb 13 2020.

Dayrat, B. 2005. Towards integrative taxonomy. Biological Journal of the Linnean Society **85**:407-415.

de Queiroz, K. 2005. A Unified Concept of Species and Its Consequences for the Future of Taxonomy. Proceedings of the California Academy of Sciences **56**:196-215.

Gaston, K. J., and J. I. Spicer. 2013. Biodiversity: an introduction. John Wiley & Sons, Chicester, UK.

Gautier, L., and Y. Naciri. 2018. Three Critically Endangered new species of *Capurodendron* (Sapotaceae) from Madagascar. Candollea **73**:121-129.

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

Hassold, S., P. P. Lowry, M. R. Bauert, A. Razafintsalama, L. Ramamonjisoa, and A. Widmer. 2016. DNA barcoding of Malagasy rosewoods: towards a molecular identification of CITES-listed *Dalbergia* species. PLoS ONE: doi: 10.1371-journal.pone.0157881.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. deWaard. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society B Biological Sciences **270**:313-321.

Hollingsworth, P. M., S. W. Graham, and D. P. Little. 2011. Choosing and using a plant DNA barcode. PLoS ONE **6**:e19254.

Hotaling, S., M. E. Foley, N. M. Lawrence, J. Bocanegra, M. B. Blanco, R. Rasoloarison, P. M. Kappeler, M. A. Barrett, A. D. Yoder, and D. W. Weisrock. 2016. Species discovery and validation in a cryptic radiation of endangered primates: coalescent-based species delimitation in Madagascar's mouse lemurs. Molecular Ecology **25**:2029-2045.

IUCN. 2012. IUCN Red List Categories and Criteria. Version 3.1. Second edition. IUCN Species Survival Commission, Gland & Cambridge.

Jones, M. R., and J. M. Good. 2016. Targeted capture in evolutionary and ecological genomics. Molecular Ecology **25**:185-202.

Karbstein, K., S. Tomasello, L. Hodac, F. G. Dunkel, M. Daubert, and E. Hörandl. 2020. Phylogenomics supported by geometric morphometrics reveals delimitation of sexual species within the polyploid apomictic *Ranunculus auricomus* complex (Ranunculaceae). bioRxiv:doi: https://doi.org/10.1101/2020.1101.1107.896902

Lexer, C., J. Joseph, M. van Loo, G. Prenner, B. Heinze, M. W. Chase, and D. Kirkup. 2009. The use of digital image-based morphometrics to study the phenotypic mosaic in taxa with porous genomes Taxon **58**:349-364.

McCormack, J. E., W. L. E. Tsai, and B. C. Faircloth. 2016. Sequence capture of ultraconserved elements from bird museum specimens. Molecular Ecology Resources **16**:1189-1203.

Murphy, J. B. 2006. Alternative approaches to the CITES “non-detriment“ finding for appendix II species. Environmental Law **36**:531-563.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature **403**:853-858.

Natusch, D. J. D., D. Esquerré, J. A. Lyons, A. Hamidy, A. R. Lemmon, E. M. Lemmon, A. Riyanto, J. S. Keogh, and S. Donnellan. 2020. Species delimitation and systematics of the green pythons (*Morelia viridis* complex) of melanesia and Australia. Molecular phylogenetics and evolution **142**:106640.

Pante, E., C. Schoelinck, and N. Puillandre. 2015. From integrative taxonomy to species description: one step beyond. Systematic Biology **64**:152-160.

Pimm, S. L., C. N. Jenkins, R. Abell, T. M. Brooks, J. L. Gittleman, L. N. Joppa, P. H. Raven, C. M. Roberts, and J. O. Sexton. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. Science **344**:doi: 10.1126/science.1246752

Schatz, G. E., and P. P. Lowry. 2011. Nomenclatural notes on Malagasy *Diospyros* L. (Ebenaceae). Adansonia **33**:271-281.

Schatz, G. E., and P. P. Lowry. 2020. Taxonomic studies of *Diospyros* L. (Ebenaceae) from the Malagasy region. IV. Synoptic revision of the Squamosa group in Madagascar and the Comoro Islands. Adansonia **42**:201-218.

Schlick-Steiner, B. C., F. M. Steiner, B. Seifert, C. Stauffer, E. Christian, and R. H. Crozier. 2010. Integrative Taxonomy: A Multisource Approach to Exploring Biodiversity. Annual Review of Entomology **55**:421-438.

Schuurman, D., and P. P. Lowry II. 2009. The Madagascar rosewood massacre. Madagascar Conservation and Development **4**:98-102.

Vieilledent, G., C. Grinand, F. Rakotomalala, R. Ranaivosoa, J.-R. Rakotoarijaona, T. F. Allnutt, and F. Achard. 2018. Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. Biological Conservation **222**:189-197.

Vieilledent, G., C. Grinand, and R. Vaudry. 2013. Forecasting deforestation and carbon emissions in tropical developing countries facing demographic expansion: a case study in Madagascar. Ecology and Evolution **3**:1702-1716.

Vieites, D. R., K. C. Wollenberg, F. Andreone, J. Köhler, F. Glaw, and M. Vences. 2009. Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. Proceedings of the National Academy of Sciences of the United States of America **106**:8267-8272.

Wachter, G. A., C. Muster, W. Arthofer, G. Raspotnig, P. Föttinger, C. Komposch, F. M. Steiner, and B. C. Schlick-Steiner. 2015. Taking the discovery approach in integrative taxonomy: decrypting a complex of narrow-endemic Alpine harvestmen (Opiliones: Phalangiidae: *Megabunus*). Molecular Ecology **24**:863-889.

Waeber, P. O., D. Schuurman, B. Ramamonjisoa, M. Langrand, C. V. Barber, J. L. Innes, P. P. Lowry, and L. Wilmé. 2019. Uplisting of Malagasy precious woods critical for their survival. Biological Conservation **235**:89-92.

Younger, J. L., L. Strozier, J. Dylan Maddox, Á. S. Nyári, M. T. Bonfitto, M. J. Raherilalao, S. M. Goodman, and S. Reddy. 2018. Hidden diversity of forest birds in Madagascar revealed using integrative taxonomy. Molecular Phylogenetics and Evolution **124**:16-26.

Zachos, F. E. 2016. Species concepts in Biology. Springer, Switzerland.

Chapter I

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

Unpublished manuscript co-authored by

Simon Crameri[1], Stefan Zoller[2], Alex Widmer[1]

[1] ETH Zurich, Institute of Integrative Biology, Universitätstrasse 16, 8092 Zürich, Switzerland
[2] Orniplan AG, Wiedingstr. 78, 8045 Zürich, Switzerland

Author contributions

SC and AW designed the study and collected samples. SZ assembled the draft *Dalbergia* transcriptome and designed target enrichment baits. SC and SZ analysed data and wrote the bioinformatics pipeline. SC and AW wrote the manuscript with contributions from SZ.

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.

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 macro-evolutionary timescales, and it is now possible thanks to advances in sequencing technology. Major difficulties are associated with i) the lack of genomic resources for many organismal groups, especially from tropical biodiversity hotspots, ii) scaling the numbers of individuals analysed and loci sequenced, and iii) building tools for reproducible bioinformatic analysis of such large and complex datasets. To address these challenges, we developed a set of target enrichment baits for phylogenomic studies of the highly diverse, pantropically distributed rosewoods and palisanders (*Dalbergia* spp.), tested the performance of these loci across the legume family (Fabaceae), and built a general-purpose bioinformatics pipeline. Phylogenetic analyses of *Dalbergia* species from Madagascar yielded highly resolved and well supported hypotheses of evolutionary relationships. Population genetic 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 bioinformatics resources presented should help advance legume genomics and systematics and promote research and conservation of the valuable and endangered *Dalbergia* rosewoods.

Keywords — *Dalbergia*, Fabaceae, target enrichment, sequence capture, phylogenomics

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 interconnected populations (metapopulations) within species typically falls within the field of population genetics, while phylogeography deals with the analysis of evolutionary processes within species or species groups over larger spatial and temporal scales. By contrast, analyses of evolutionary relationships among species and patterns of diversification in higher taxonomic groups fall within the realm of phylogenetics. Until recently, these fields employed different conceptual approaches, molecular methods and markers to assess relationships among individuals, populations or species. However, it has long been recognized that "the same ecological and evolutionary processes that cause lineage divergence can also drive speciation" (RISSLER 2016). The rapid expansion of next-generation sequencing (NGS) and the development of techniques for target enrichment of selected genomic regions jointly offer the opportunity to study the processes that drive the evolution of biological diversity from micro- to macro-evolutionary timescales.

Target enrichment (MAMANOVA *et al.* 2010) in combination with high-throughput sequencing allows for the repeatable analysis of hundreds or thousands of orthologous loci in dozens to hundreds of individuals at a moderate cost per sample (FAIRCLOTH *et al.* 2012, JONES AND GOOD 2016), and therefore strikes a good balance between locus information content and scalability to high numbers of individuals. Since its introduction in the field of phylogenomics by FAIRCLOTH *et al.* (2012) and LEMMON *et al.* (2012), target enrichment has increasingly been applied in a wide range of animals (e.g. FAIRCLOTH *et al.* 2013, FAIRCLOTH *et al.* 2014, PRUM *et al.* 2015, YUAN *et al.* 2016) and plants (MANDEL *et al.* 2014, WEITEMIER *et al.* 2014, NICHOLLS *et al.* 2015, SASS *et al.* 2016, DE LA HARPE *et al.* 2018, LÉVEILLÉ-BOURRET *et al.* 2018, VATANPARAST *et al.* 2018, OJEDA *et al.* 2019), with applications at macro-evolutionary (FAIRCLOTH *et al.* 2012, LEMMON *et al.* 2012, JOHNSON *et al.* 2019), phylogeographic (LEMMON & LEMMON 2012, SMITH *et al.* 2014) and micro-evolutionary timescales (DE LA HARPE *et al.* 2018, VILLAVERDE *et al.* 2018). Target enrichment holds the potential to bridge the divide between phylogenetics, phylogeography and population genetics (NICHOLLS *et al.* 2015, RISSLER 2016) because both slowly evolving sequences informative at macro-evolutionary timescales and faster evolving sequences informative at micro-evolutionary timescales can be targeted simultaneously (JONES & GOOD 2016).

Focusing on selected genomic regions presents a major challenge in the development of target enrichment approaches because it requires the design of target

capture baits for hybridisation and subsequent enrichment of target sequences. An ideal capture bait set targets unique genomic regions to prevent conflation of orthologs and paralogs, while at the same time having a highly conserved core to allow for in-solution hybridisation in divergent taxa (but see JOHNSON *et al.* 2019) along with more variable flanking regions that are informative at the population level (LEMMON *et al.* 2012). Meeting these two requirements depends on the organismal group and timescales of interest, and researchers are therefore required to develop target capture bait sets for their particular focal group. In the absence of high-quality reference genomes of focal species, target capture baits can be designed based on publicly available genome sequences of related species (divergent reference capture), or they can be designed based on transcriptome data (*de novo* assembly capture, JONES & GOOD 2016). Another challenge is posed by the bioinformatic analysis of target enrichment data. Two bioinformatic pipelines, PHYLUCE (FAIRCLOTH 2016) and HybPiper (JOHNSON *et al.* 2016), are widely used to analyse target enrichment datasets. While both are applicable to datasets derived from different capture bait sets, PHYLUCE was developed for ultraconserved genomic elements (UCEs) and has mainly been used at macro-evolutionary and phylogeographic timescales in animal systems, whereas HybPiper is optimized for datasets derived from capture baits designed in exons using Hyb-Seq (WEITEMIER *et al.* 2014). Depending on the characteristics of the capture bait set used and the research questions being addressed, there is a need for existing tools to be expanded or new pipelines developed (DE LA HARPE *et al.* 2018).

Dalbergia L.f. (Fabaceae) is a pantropical and ecologically diverse plant genus with around 275 currently recognized species (POWO 2020). Since the most recent synopsis of the genus by LEWIS *et al.* (2005) more than 25 new species have been described (BOSSER & RABEVOHITRA 2005, JONGKIND 2007, LINARES & SOUSA 2007, DE MORAES 2012, ADEMA *et al.* 2016, LACHENAUD 2016, LACHENAUD & VAN DER MAESEN 2016, SOOSAIRAJ *et al.* 2018). Many species form medium-sized to large trees and are a source of high-quality timber commonly known as rosewood or palisander (PRAIN 1904, BOSSER & RABEVOHITRA 2002). It has long been recognized that some *Dalbergia* species are highly sought-after on the international market and that some species are heavily overexploited (SCHUURMAN & LOWRY 2009, UNODC 2016b). National and international regulations have been established with the goal of fostering sustainable exploitation and revenue generation (BARRETT *et al.* 2013, CITES 2020). However, illegal logging continues to date (SIRIWAT & NIJMAN , SCHUURMAN & LOWRY 2009, WAEBER *et al.* 2019, VARDEMAN & RUNK 2020) and the effective implementation of such regulations demands that species can reliably be recognized and that extant population sizes are estimated to

assess the potential threat status of each species. Despite their huge economic value and the extensive legal and illegal trade in this genus, developing a comprehensive understanding of its species diversity and 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 expert taxonomists focusing on this species-rich, pantropically distributed genus. Many taxonomies heavily rely on fertile structures to delimit and distinguish between species and to build identification keys (PRAIN 1904, DE CARVALHO 1997, BOSSER & RABEVOHITRA 2002), but inflorescences of many species develop at unpredictable times and may be difficult to observe in mature trees. Vegetative traits with potential taxonomic utility, such as leaflet size, often display extensive phenotypic plasticity in response to environmental conditions (MOREIRA *et al.* 2013 and pers. obs.), which is rarely represented in the often limited number of high-quality herbarium specimens, and which hampers the assessment of heritable versus plastic trait variation for species delimitation (WAEBER *et al.* 2019 and pers. obs.). The taxonomic complexity and multiple recently described species suggest that the systematics of the genus remains in need of extensive revision, which could be supported by modern molecular methods. A reliable taxonomy and the ability to study the diversity of *Dalbergia* species from the level of population to genus is therefore of direct conservation relevance and has motivated our development of a target enrichment approach for the genus.

Dalbergia is a member of the legume family, Fabaceae (or Leguminosae), the third largest angiosperm family (LPWG 2017). This remarkable diversity is studied by a large and active research community that focuses on elucidating phylogenetic relationships and its consequences for plant systematics (KOENEN *et al.* 2020b), speciation and rapid radiations (e.g. HUGHES & EASTWOOD 2006, SHAHI SHAVVON *et al.* 2017), analyses of mutualistic relationships between plants and nitrogen-fixing bacteria (SPRENT *et al.* 2017, GRIESMANN *et al.* 2018), and the domestication and breeding of major agricultural crops (MOUSAVI-DERAZMAHALLEH *et al.* 2018, ZHUANG *et al.* 2019), among other topics. The high diversity of legumes and legume-related research questions that are increasingly being addressed using genomics methods (EGAN & VATANPARAST 2019) has prompted us to explore the applicability of the target enrichment approach developed here for macroevolutionary analyses across Fabaceae.

We introduce a target enrichment approach for anchored phylogenomic analyses in the genus *Dalbergia* (Fabaceae, Papilionoideae) and across the legume family. This approach encompasses a capture bait set that targets conserved regions across the nuclear genome, derived from a combination of divergent reference capture using five published legume genomes, and *de novo* assembly of a *Dalbergia* transcriptome. We also introduce

a dedicated bioinformatics pipeline supporting the analysis of high-throughput target enrichment sequencing data, with special emphasis on streamlined applicability and parallelisation, locus filtering, and graphical output. 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 of the developed bait set for phylogenomic analysis at much deeper (macro-) evolutionary timescales by analysing sequence capture data of various legume subfamilies. Finally, we test the utility of this approach at a micro-evolutionary scale and assess genetic variation among individuals and populations of two closely related *Dalbergia* species from Madagascar.

Materials and Methods

Design of target enrichment baits

We selected high-quality genome assemblies of five legume species available in public databases for divergent reference capture (JONES AND GOOD 2016): *Cajanus cajan* (L.) Millsp. v1.0, *Glycine max* (L.) Merr. v1.0, *Lotus japonicus* L. v2.5, *Medicago truncatula* Gaertn. v3.5 and *Phaseolus vulgaris* L. v2.1 (see Supplementary Methods for details). We pairwise aligned these sequences and our draft *Dalbergia* transcriptome (146,484 scaffolds between 201 and 17,129 bp long, with a mean length of 815 bp, see Supplementary Methods) using lastz (HARRIS 2007) and extracted 169,484 sequences from the *C. cajan* assembly that were at least 60 bp long and fully conserved across species. After removing direct and reverse-complement duplicates including exactly matching sub-sequences, we retained 74,604 sequences. We then mapped the sequences back to each genome using Blast (ALTSCHUL *et al.* 1990) and kept 10,867 sequences that returned only one ungapped hit after filtering for a minimum of 90% overlap and 95% identity in all genomes. We tiled sequences longer than 150 bp into 100 bp baits with a 50 bp overlap, extracted a central bait of 100 bp for sequences measuring 100 to 150 bp in length, extended shorter sequences to 100 bp based on the *C. cajan* genome, and discarded sequences that could not be extended, e.g. because they were at the edge of an assembly scaffold. After removal of duplicates and reverse complement duplicates, we removed baits with low complexity and interspersed repeats using RepeatMasker (SMIT *et al.* 2013) and discarded those with a GC-content higher than 70%. Following a further Blast run and filtering with relaxed settings (one hit, overlap >= 85%, identity >= 92.5%), we retained a final set of 12,049 tiled baits from 7,201 conserved regions that we ordered for synthesis as myBaits Custom Target Capture Kits (Arbor Biosciences, Ann Arbor, MI, USA; https://arborbiosci.com).

Definition of initial target region reference sequences

An inspection of the physical distance between the 7,201 conserved regions showed that 378 regions (5.25%) partially overlapped with one another, and that 648 regions (9%) were located no more than 100 bp from one another, whereas 1,703 regions (23.7%) were separated by less than 1,000 bp. We merged the overlapping sequences and combined sequences that were within 100 bp of one another by filling the gaps based on the *Cajanus cajan* assembly. This resulted in reference sequences for 6,555 regions (target regions hereafter), with a mean length of 188 bp (100 to 1,424 bp), a total length of 1,233,157 bp, and a mean physical distance of 52,802 bp (100 to 1,306,862 bp) between sequences of the same genomic scaffold. The target regions covered all eleven linkage groups of the *C. cajan* assembly (3,931 target regions, 104 to 761 target regions per linkage group) while 2,624 target regions belonged to unanchored genomic scaffolds (1 to 48 target regions per scaffold).

Taxon sampling for target enrichment bait 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 subfamilies of Fabaceae, as recognised in the most recent treatment (LPWG 2017), and comprised 104 individuals (110 samples including 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, Figure S4) included members of the closely related genera *Dalbergia* L.f. (at least 19 species)*, Machaerium* Pers. (three species) and *Ctenodon* Baill. sensu CARDOSO *et al.* (2020) (two species) and comprised 60 individuals (63 samples including three replicates; at least 26 species including two outgroups). Two species of *Aeschynomene* L. s.str. sensu CARDOSO *et al.* (2020) were included as the outgroup for the species set. The population set (Table S3, Figure S5) included 51 individuals in total, 29 attributed to *D. monticola* Bosser & R. Rabev. from four localities, and 22 attributed to *D. orientalis* Bosser & R. Rabev. from eleven localities.

Library preparation, target enrichment and sequencing

Total genomic DNA was extracted from silica gel dried leaf tissue (185 extractions) or herbarium sheets (11 extractions) using the CTAB protocol (DOYLE & DOYLE 1987) or the DNeasy Plant Mini Kit (Qiagen, Santa Clara, CA, USA). DNA was quantified using the QuantiFluor dsDNA system for a QuantusTM fluorometer (Promega, Madison, WI, USA)

and DNA integrity was checked on 1.5% agarose gels for a subset of samples. Genomic DNA libraries were prepared for each sample using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA), following manufacturer's instructions. Samples to be pooled within the same sequencing lane were individually indexed during the PCR enrichment step using NEBNext Multiplex Oligos for Illumina (single-indexed with E7335 and E7500 kits, or dual-indexed with E6440 or E7600 kits, New England Biolabs). In-solution hybridisation and target enrichment were performed using our 12,049 tiled RNA baits. We pooled up to six individually indexed libraries during the hybridisation step using a stratified random assignment of libraries to hybridisation reactions. Stratification aimed to prevent samples that were closely related to *Cajanus cajan* from being pooled with more distantly related samples, and to prevent pooling of lower-quality herbarium samples with DNA isolated from silica gel dried leaf material. Following hybridisation and target enrichment, pooled libraries were sequenced in separate lanes on an Illumina MiSeq (2×300 bp paired-end sequencing, 99 libraries) at the Genetic Diversity Centre (GDC) Zurich, on 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), or on an Illumina NovaSeq 6000 SP flow cell (2×150 bp paired-end sequencing, 9 libraries) at the FGCZ. We included nine extraction replicates sequenced on the same or on different sequencing platforms to assess reproducibility of target enrichment and sequencing. 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.

Bioinformatics pipeline

The bioinformatic pipeline developed for this project is accessible on GitHub (https://github.com/scrameri/CaptureAl) as a documented sequence of scripts. Once configured for a given computing environment, these scripts can be executed using most properly formatted target enrichment datasets and arguments can be adapted to specific needs. The pipeline streamlines the mapping of quality-trimmed reads to target regions for downstream population genetic analyses, sequence assembly, orthology assessment, sample and region filtering, alignment and trimming, and filtering of alignments for phylogenetic analysis. To visually inform the selection of analysis parameters, we included R scripts (R CORE TEAM 2020) to manage and visualize data with ape version 5.3 (PARADIS AND 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 regions in parallel using GNU parallel (TANGE 2011).

The sequence of executed commands and the chosen parameters are provided in Supplementary Methods. Bioinformatic analyses were carried out on a multi-core Linux server (GDC Zurich) or on the Euler scientific compute cluster (ETH Zurich), and are divided into seven steps after quality-trimming of raw reads using trimmomatic version 0.32 (BOLGER *et al.* 2014, see Supplementary Methods for details). Steps 1 to 5 are always required, whereas steps 6 and 7 are optional. Step 6 combines physically neighbouring and overlapping alignments and step 7 generates longer and more representative reference sequences as starting points for re-iteration of steps 1 to 5. These steps were applied separately and iteratively for different taxon sets. We first applied the pipeline to twelve representative samples each from the subfamily and species sets to generate longer and more representative reference sequences for target regions that can each be efficiently recovered in these taxon sets, and then reiterated the analysis for all samples of the subfamily and species sets using the new reference sequences (see Tables S1-S3 and Supplementary Methods for details). We then used the resulting reference sequences of the species set for analysis of the population set.

Step 1: Read mapping

We mapped quality-filtered reads of each sample against the 6,555 target region reference sequences using BWA version 0.7.12-r1039 (LI & DURBIN 2009) and the BWA-MEM algorithm (LI 2013). We then performed coverage analysis using samtools version 1.3.1 (LI AND DURBIN 2009) and bedtools version 2.26.0 (QUINLAN AND HALL 2010), visualized results using *filter.visual.coverages.R*, and saved a list of target regions with adequate average coverage across samples (see Supplementary Methods for details). This prevented a time-consuming sequence assembly of target regions which would probably be filtered out in step 4.

Step 2: Sequence assembly

We extracted read pairs when at least one read mapped to a retained target region with a minimum mapping quality of 10. We then assembled the extracted reads separately for each sample and region using dipSPAdes (SAFONOVA *et al.* 2015), which produced consensus contiguous sequences (contigs hereafter) based on haplocontigs generated by SPAdes version 3.6.0 (BANKEVICH *et al.* 2012, see Supplementary Methods for details).

Step 3: Orthology assessment

Sequence assembly may yield multiple contigs per sample for some target regions, e.g. due to sequencing of several fragments of the same region or due to sequencing of paralogous

regions (JOHNSON *et al.* 2016). We determined the most likely orthologous contig(s) of each sample in each target region using an exhaustive Smith-Waterman alignment (SMITH & WATERMAN 1981) between all contigs and the reference sequences using exonerate version 2.2 (SLATER AND BIRNEY 2005). We defined the best-matching contig based on the exonerate alignment statistics as the most likely orthologous contig for each sample and target region, and retained further contigs that did not overlap with one another or the best-matching contig but aligned with a sufficient alignment score to other parts of the target region. These contigs were interpreted to represent fragments of the same region, and were therefore combined with the best-matching contig to form a contiguous sequence (orthologous contig hereafter, see Supplementary Methods for details). The main output of step 3 is a FASTA file with an orthologous contig for each sample and each target region.

Step 4: Sample and region filtering

Successful target enrichment 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 orthologs of the target regions (capture specificity). To identify target regions with high capture sensitivity and specificity across focal taxa, we defined four taxon groups within both the subfamily and the species set by considering known phylogenetic clades. In the subfamily set we defined the four subfamilies represented by multiple taxa as taxon groups. In the species set we defined four taxon groups based on our preliminary phylogenetic 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). We used *filter.visual.assemblies.R* to visualise the exonerate alignment statistics generated in step 3 and to set capture sensitivity and specificity filtering thresholds informed by these visualisations, which needed to be met in all considered taxon groups (see Supplementary Methods for details). The main output of step 4 is a list of samples and a list of target regions to keep.

Step 5: Target region alignment and alignment trimming

We generated a multi-sequence FASTA file for all retained target regions, containing the respective orthologous contigs of all retained samples. Sequences were then aligned using mafft version 7.123b (KATOH & STANDLEY 2013). Raw alignments in each taxon set were trimmed at both ends until an alignment site showed nucleotides in at least 50% of aligned sequences along with a maximum nucleotide diversity (i.e., the mean number of base

differences between all sequence pairs) of 0.25. In addition, we performed internal trimming by only keeping sites with nucleotides in at least 40% of aligned sequences. We also resolved potential mis-assemblies or mis-alignments at contig ends using a sliding window approach that identifies and masks sequences with large deviations from the alignment consensus (see Supplementary Methods for details). The main output of step 5 are potentially overlapping trimmed alignments for each kept target region.

Step 6: Merging of overlapping alignments

Shorter but physically close target regions facilitated sequence assembly in lower-quality samples but led to overlaps in trimmed alignments of neighbouring target regions. We identified such overlaps by aligning consensus sequences of target region alignments. Specifically, we generated consensus sequences by keeping the most frequent allele at each alignment site with a frequency above 1%. We then identified non-reciprocal hits between different consensus sequences using BLAST+ version 2.7.1 (CAMACHO *et al.* 2009), and we filtered for hits between alignment ends of target regions located on the same linkage group in the *Cajanus cajan* genome. Orthologous contigs that were part of overlapping alignments were then aligned using mafft. The resulting merged alignments were then collapsed to represent different orthologous contigs of the same individual as a single sequence, and the merging and collapsing procedure was visually inspected. Trimming was applied as in step 5, and sets of two to several consecutively overlapping alignments were then each replaced by a single merged alignment if merging was successful (see Supplementary Methods for details). The main output of step 6 are non-overlapping trimmed alignments for each kept target region.

Step 7: Generation of representative reference sequences

To mitigate potential biases arising from the reference sequences used, we generated two new sets of target region reference sequences based on the aligned target regions for the subfamily and species sets, respectively. We first calculated a consensus of each alignment separately for each taxon group in the subfamily and species sets, as per step 6. We then aligned the consensus sequences of each taxon group separately for the subfamily and species sets, and generated representative consensus sequences for each set of target regions using the same parameters as before. The representative reference sequences were then used to repeat steps 1-6 using more stringent target region filtering parameters (see Supplementary Methods for details).

Alignment assessment and filtering

We characterized all non-overlapping 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 visualized the alignment statistics and identified alignments with an excessive gap ratio or average nucleotide diversity per site (see Supplementary Methods for details). The filtered alignments of the second iteration were used for phylogenetic analyses.

Phylogenetic analyses

We performed phylogenetic analyses with both the subfamily and species sets, using a supermatrix (concatenation) approach as well as 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 Supplementary Methods for details). For the gene tree summary approach, we ran RAxML jobs separately for each alignment, as in the supermatrix approach, to generate gene trees. As recommended by ZHANG *et al.* (2018), we collapsed branches in gene trees if they had bootstrap support values below 10 using *nw_ed* newick utilities (JUNIER AND 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. All phylogenetic trees were displayed using ggtree version 2.0.2 (YU *et al.* 2016).

Population genetic analyses

We carried out population genetic analyses for the population dataset only. 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. PCR duplicates were removed from the resulting BAM files using picard tools version 2.21.3 (BROAD INSTITUTE 2019), and regions with excessive coverage were capped to a maximum depth of 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 (https://github.com/ekg/vcflib), which was also used to decompose complex variants (see Supplementary Methods for details). We then used vcfR version 1.10.0 (KNAUS & GRÜNWALD 2017) and adegenet version 2.1.1 (JOMBART 2008, JOMBART & AHMED 2011) libraries in R to generate *genind* and *genlight* objects that represented the matrix of single nucleotide polymorphisms (SNPs) with associated metadata such as individual missingness, species identification, and sampling location. We relied on these objects to generate a centered covariance matrix from the allele table to use for principal component analyses (PCA), 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 only kept SNPs with genotype data in at least 95% of individuals, and we randomly sampled up to three SNPs per target region for computational ease. Structure analyses were performed for one to ten deemes (k), with ten replicates per simulation and using 110,000 iterations, including a burn-in period of 10,000 iterations (see Supplementary Methods for details). Replicate Structure results were aligned and visualized using Clumpak (KOPELMAN *et al.* 2015) and default settings.

Results

Sequencing, mapping and target region filtering

We obtained 0.25 to 27.53 (median: 3.12) million raw paired-end reads per sample, of which we retained 86.55% to 99.34% (median: 93.82%) after quality trimming. Reads mapped to 6,519/6,287 of the 6,555 target regions in the subfamily/species set, respectively (step 1). Of these we kept 3,436/4,908 target regions with average coverage between 6/8 and 1,000 in at least 30%/40% of samples in each taxon group. After assembly (step 2) and orthology assessment (step 3), 2,710/4,181 target regions passed region specificity and sensitivity filters of lower stringency (step 4, see Supplementary Methods). After alignment and trimming (step 5), overlapping alignments in 207/377 regions were successfully merged, resulting in 2,468/3,736 non-overlapping trimmed alignments (step 6). Longer and more representative consensus sequences were generated from these target regions (step 7) and used as reference sequences for a second round of mapping of quality-trimmed reads. We retained 1,917/3,418 target regions with average coverage between 8 and 1,000 in at least 70% of samples in each taxon group (Figures S6 and S7), and 1,020/2,407 target regions passed specificity and sensitivity filters of higher stringency

(step 4, see Supplementary Methods). Merging of overlapping alignments in 15/11 further regions followed by removal of 19/7 alignments with a gap ratio above 0.35/0.3 or a nucleotide diversity above 0.35/0.15 yielded 986/2,396 alignments for phylogenetic analysis of the subfamily and species set, respectively. Details including gene annotations for these final target regions are given in Tables S4 and S5.

Quality-trimmed reads mapped to all 2,396 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 S8). Variant calling resulted in 116,500 filtered SNPs after decomposing complex variants, of which 60,204 were bi-allelic with no missing data and were used for PCA and NJ tree reconstruction. Of these, a subset of 5,042 SNPs were selected for Structure analyses.

Phylogenetic analyses in the subfamily set

Phylogenetic analysis of 986 alignments recovered each of the five sampled subfamilies as monophyletic using both the gene tree summary method ASTRAL-III (Figure 1) and the supermatrix method (Figure S1), with 100% posterior probability and bootstrap support (referred to as maximum support hereafter), respectively. Subfamilies Cercidoideae and Detarioideae were found to be sister taxa with maximum support using both methods and in 69% of ASTRAL-III quartet trees. Our analyses further recovered many well-established clades and relationships with maximum support using both methods. These included 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 s.l. (WOJCIECHOWSKI *et al.* 2004, CARDOSO *et al.* 2012). Within the Dalbergioids s.l. (WOJCIECHOWSKI *et al.* 2004), we recovered the Amorpheae clade (MCMAHON & HUFFORD 2004) as sister to the rest of the group, which includes the Dalbergioids s.str. clade (Dalbergioids sensu LAVIN *et al.* (2001)), containing the *Adesmia*, *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 Non-Protein-Amino-Acid-Accumulating (NPAAA) clade (WOJCIECHOWSKI *et al.* 2004, CARDOSO *et al.* 2012), 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,

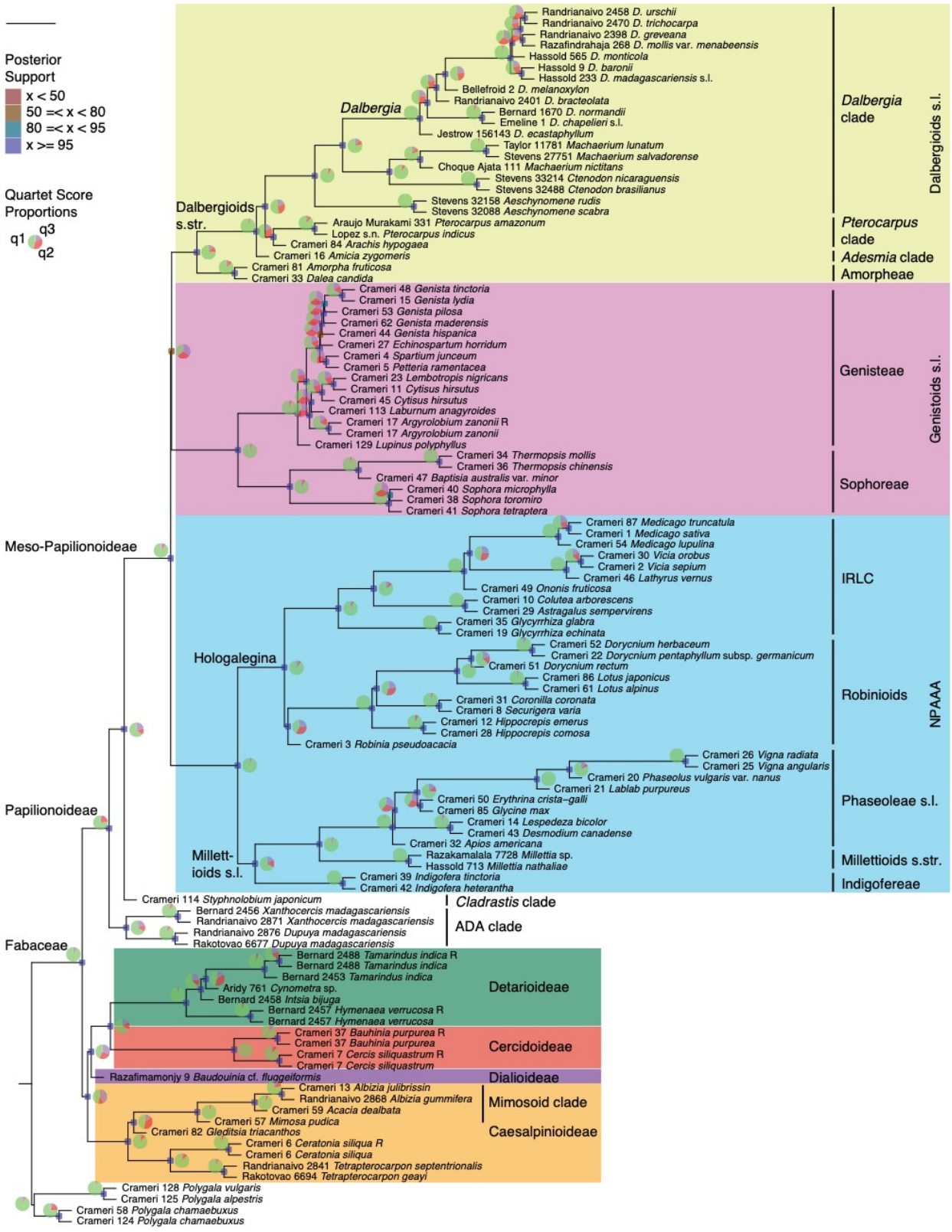

FIGURE 1: Coalescent-based phylogeny of the subfamily set (n = 110) inferred using ASTRAL-III on 986 gene trees with collapsed low-support branches (<10% bootstrap support). 860 gene trees (87.22%) had missing taxa. The overall normalized quartet score was 88.82%. Pie charts denote the fraction of gene trees that are consistent with a given node (green) or with alternative topologies (red, purple). Four subfamilies (Caesalpinioideae, Cercidoideae, Detarioideae and Papilionoideae) represented by multiple taxa are each recovered as monophyletic.

including the Robinioids and the inverted-repeat-lacking clade (IRLC, WOJCIECHOWSKI *et al.* 2004).

Relationships among the other subfamilies remained unresolved using the supermatrix method (Figure S1). ASTRAL-III recovered a clade comprising Caesalpinioideae, Cercidoideae, Detarioideae and Dialioideae as sister group to Papilionoideae with 100% posterior probability. However, this topology was recovered in only 47% of quartet trees, with the remaining trees depicting alternative topologies. We therefore evaluated quartet scores of fifteen different hypothetical relationships among subfamilies (all possible topologies involving Caesalpinioideae, Dialioideae, Papilionoideae and the clade comprising Cercidoideae and Detarioideae; Figure S2) using the tree scoring option in ASTRAL-III in combination with a file that mapped taxa to subfamilies or to the outgroup. In this context, quartet scores denote the fraction of induced quartet trees (i.e, unrooted four-taxon subsets of input gene trees) that were shared by a predefined hypothesis on relationships among subfamilies. The main topology presented in Figure 1 shows the highest normalized quartet score (38.40%). However, gene trees were almost equally congruent with two alternative topologies, each with a normalized quartet score of 38.36% (Figure S2). Both these alternative hypotheses involved a clade composed of Caesalpinioideae and Papilionoideae in addition to the clade composed of Cercidoideae and Detarioideae. Further contentious resolutions of major groups concerned the three clades within Meso-Papilionoideae, with 36% quartet support for a clade formed by Dalbergioids s.l. and Genistoids s.l., as well as 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*.

All extraction replicates (labelled 'R' in Figures 1 and S1) were retrieved as sister to the original sample, independent of whether samples and extraction replicates were sequenced on the same sequencing platform or on a different one.

Phylogenetic analyses in the species set

Phylogenetic analysis of 2,389 alignments recovered samples of *Dalbergia* as monophyletic with maximum support using both ASTRAL-III (Figure 2) and the supermatrix method (Figure S2). Within *Dalbergia*, we recovered two exclusively Malagasy clades using both the supermatrix and the gene tree summary method, which were named Madagascar Supergroup I and Supergroup II. The sole Malagasy specimen not belonging to these supergroups was *D. bracteolata* Baker. The four included non-Malagasy *Dalbergia* specimens were each found to represent a separate lineage.

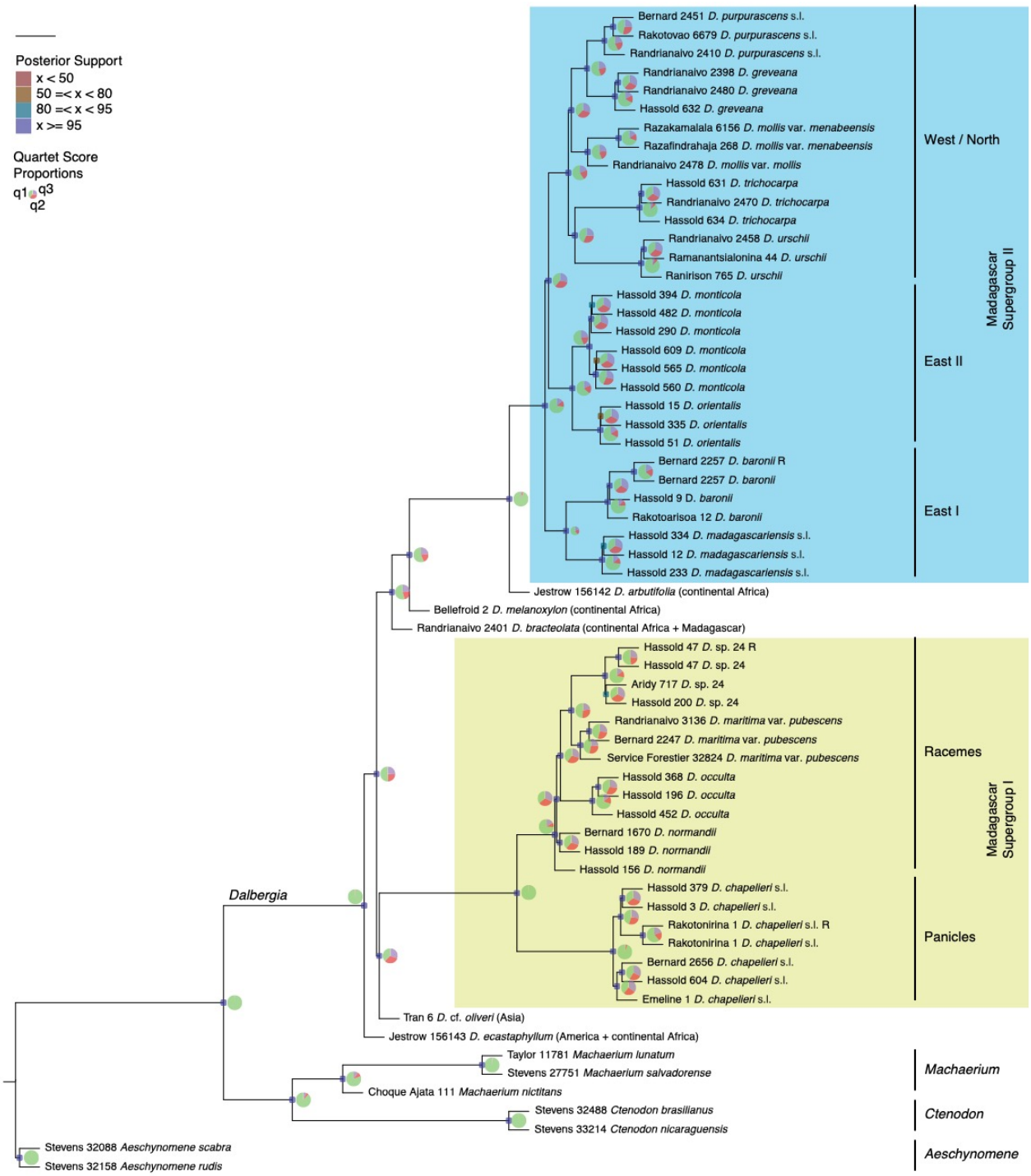

FIGURE 2: Coalescent-based phylogeny of the species set (n = 63) inferred using ASTRAL-III on 2389 gene trees with collapsed low-support branches (<10% bootstrap support). 1,014 gene trees (42.44%) had missing taxa. The overall normalized quartet score was 85.42%. Pie charts denote the fraction of gene trees that are consistent with a given node (green) or with alternative topologies (red, purple). The genus *Dalbergia* is recovered as monophyletic and sister to *Machaerium* and *Ctenodon*. The two highlighted Malagasy *Dalbergia* supergroups are strongly supported. NE = northeast; SE = southeast; s.l. = sensu lato.

Within Supergroup I, two major clades were recovered, each with maximum support using both phylogenetic methods. One clade comprised samples attributed to *Dalbergia chapelieri* sensu lato, whereas the other contained the remaining sampled Supergroup I species. One specimen (*Hassold 156*) was morphologically similar to *D. normandii* but formed a grade, rather than a clade, with other accessions of the same species. Three further specimens (*Dalbergia* sp. 24, naming system following the MADAGASCAR CATALOGUE 2020) resembling *D. maritima* var. *pubescens* formed a well-supported sister clade to the latter and were eco-geographically distinct.

Within Supergroup II, three major clades were recovered, each with maximum support using both phylogenetic methods. Two clades contained species distributed in the humid east of Madagascar, the third contained species distributed in the seasonally dry west and north of the island. All three clades comprised more than one species, all of which were recovered as monophyletic with maximum support using both methods.

We observed geographic structure in both of the species that were represented by more than three collections, regardless of the phylogenetic method. In *D. chapelieri* s.l. and *D. monticola*, specimens from north-eastern Madagascar formed one subclade, while those from the southeast formed another subclade.

The extraction replicates sequenced on the same as well as on different sequencing platforms (specimens labelled with 'R' in Figure 2 and Figure S3) each grouped together. One herbarium sample (*Service Forestier 32824* collected in 1985, the type collection of *Dalbergia maritima* var. *pubescens*) 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 the Betampona reserve.

Population genomic analyses

Principal component analysis based on 60,204 biallelic SNPs with no missing data revealed three distinct clusters of individuals along principal component (PC) 1 (explaining 27.58% of the total variation) and PC 2 (explaining 11.26% of the variation, Figure 3). One cluster was separated from the other two along the PC1 and included all individuals attributed to *Dalbergia orientalis*. Individuals belonging to the other two clusters were initially all attributed to *D. monticola* based on morphology, but were subsequently found to be clearly distinct genetically, mainly along PC2. The unexpected third cluster comprised samples from a single broad sampling location in north-eastern Madagascar (location 5, see Table S3 and Figure S5) where both *D. monticola* and *D. orientalis* had been collected. The same three clusters were also recovered in Structure analyses with three demes (Figure S9). Inspection of the delta *K* statistic (Figure S10) suggested that the biologically meaningful

clustering solutions were those assuming two populations (separating *D. orientalis* from the other samples) or three populations (separating the unexpected third cluster shown in Figure 3). The lack of admixture between *D. monticola* and the unexpected third cluster given their sympatric distribution prompted us to propose that individuals of the third cluster represent a separate species, tentatively referred to as *Dalbergia* sp. 17 (naming system following the MADAGASCAR CATALOGUE 2020).

Genetic structuring was revealed within both *Dalbergia monticola* and *D. orientalis* on the PCA and the NJ tree (Figure 3), indicating 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). A similar pattern was found in Structure analysis with higher numbers of assumed populations (Figure S9).

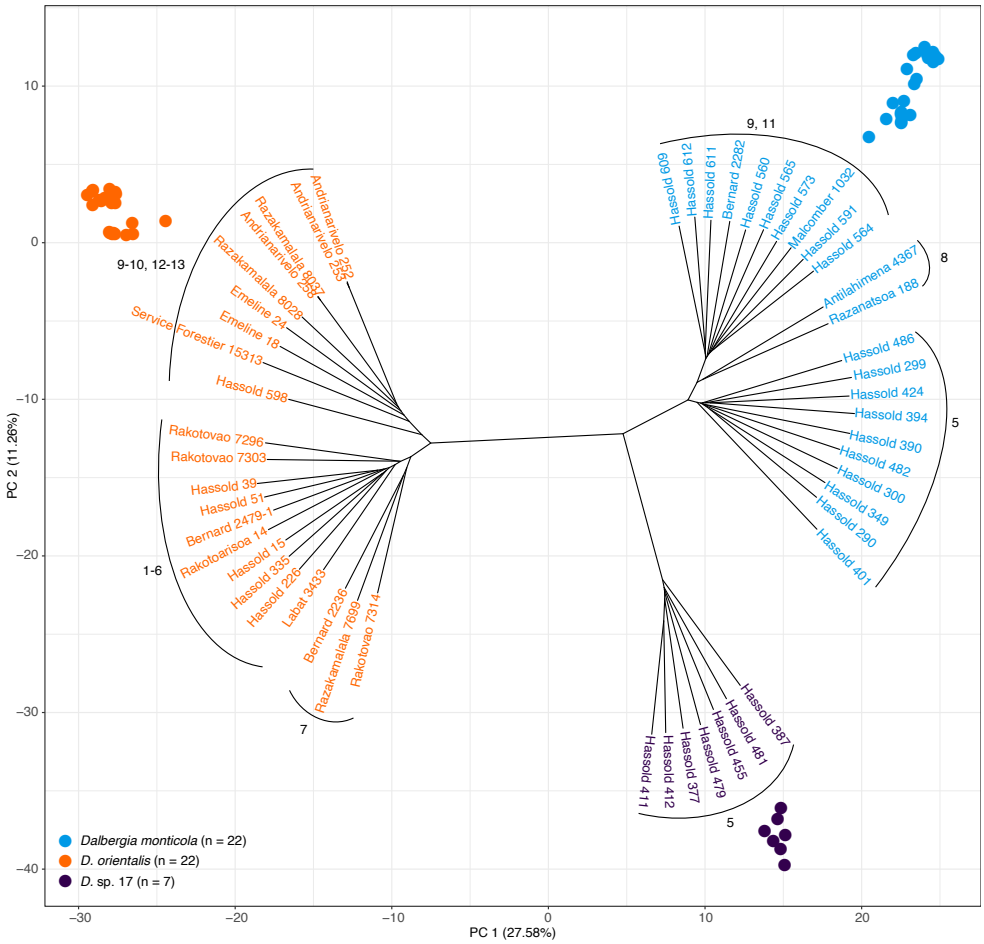

FIGURE 3: First two principal components and NJ tree on Nei's genetic distances of the population dataset (n = 51) inferred from 60,204 biallelic single nucleotide polymorphisms (SNPs) with no missing data. Numbers adjacent to tree branches denote sampling locations as shown in Figure S5. Three distinct species form separate clusters despite their co-occurrence in northeast Madagascar (location 5). *Dalbergia orientalis* is separated from the other two species along the first principal component. Substructure is visible within both *D. monticola* and *D. orientalis* and indicates isolation by distance at a broad geographic scale, dividing specimens from northeast (locations 1 to 6), central-east (locations 7 and 8) and southeast Madagascar (locations 9 to 13).

Discussion

Studies of the diversity and diversification of species and evolutionary lineages require an integrative approach that links studies of micro-evolutionary processes to analyses of macro-evolutionary relationships (DE LA HARPE *et al.* 2017). Genetic data form a preferable source of information for investigations that span across broad evolutionary scales, as common mechanisms underlie the build-up of variation that characterises evolutionary lineages and their divergence processes. Such genetic studies ideally interrogate a large number of loci that are distributed across the nuclear genome and that represent the spectrum of evolutionary rate variation that is relevant for the question at hand. The present study introduces a set of target enrichment baits for phylogenomic studies at micro- to macro-evolutionary timescales in rosewoods (*Dalbergia* spp.) and more generally across the legume family (Fabaceae), together with a flexible bioinformatic pipeline that streamlines the processing of raw reads for both phylogeny reconstruction and SNP-based population genomic analyses. Analyses at the family level recovered well-known subfamilies and clades of legumes as monophyletic but revealed gene tree incongruence and unresolved deep-branching relationships among subfamilies, confirming recent findings based on transcriptome sequencing (KOENEN *et al.* 2020b). Our phylogenomic analyses in the genus *Dalbergia* provide species-level resolution of phylogenetic relationships, significantly improving on an earlier DNA barcoding study (HASSOLD *et al.* 2016), and indicate that species diversity among Malagasy *Dalbergia* has been underestimated. Population genomic analyses revealed the existence of a currently undescribed and apparently cryptic *Dalbergia* species that is genetically differentiated from sympatric individuals of the renowned precious timber species *D. monticola*. Together, our results illustrate the power of the target enrichment approach we have developed for studies of species diversity and diversification in rosewoods (*Dalbergia* spp.) and more broadly in the economically important and highly diverse legume family.

Phylogenetic analyses in Dalbergia

Phylogenetic relationships in the species-rich, pantropically distributed genus *Dalbergia* have been studied with a limited number of loci and often with a focus on DNA barcoding (VATANPARAST *et al.* 2013, BHAGWAT *et al.* 2015, HARTVIG *et al.* 2015, HASSOLD *et al.* 2016). In our study, we sampled sequence variation across 2,389 nuclear regions in fifteen Malagasy and four non-Malagasy species. The resulting phylogeny represents the best resolved tree of Malagasy *Dalbergia* species to date. Our taxon sampling largely overlaps with a previous study (HASSOLD *et al.* 2016) that explored sequence variation in three

standard chloroplast DNA barcode markers (*mat*K, *rbc*L, *trn*L-UAA), enabling a comparison of phylogenetic results. In both studies, *Dalbergia* species endemic to Madagascar were recovered as two large and well-supported clades, each exclusively comprising Malagasy species. These two supergroups are morphologically divergent and largely correspond to the not validly published sections *Grandiflorae* (largely corresponds to Supergroup I) and *Parviflorae* (largely corresponds to Supergroup II) originally classified by VIGUIER (1944) and later reinstated as Group 1 and Group 2 by BOSSER & RABEVOHITRA (2002). Supergroup I is characterised by a glabrous reddish gynoecium with a long and slender style and relatively large flowers, and Supergroup II is characterised 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 a minimum of two independent colonisations 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.

The results from both our study and HASSOLD *et al.* (2016) revealed two subgroups within Supergroup I. One subgroup comprised samples of *Dalbergia chapelieri* s.l., a morphologically highly variable species with paniculate inflorescences that is widely distributed in eastern Madagascar and that could be further divided into material from north-eastern and south-eastern Madagascar based on evidence from the nuclear markers used in the present study (Figure 2) as well as the three chloroplast markers employed by HASSOLD *et al.* (2016). The other subgroup contained various species from eastern Madagascar with mostly racemose inflorescences, including a potentially new species, *Dalbergia* sp. 24. Collections belonging to this operational taxonomic unit (OTU) were previously believed to be conspecific with *D. maritima* var. pubescens (see HASSOLD *et al.* 2016) but show genetic, geographic, and morphological differences (in particular more numerous leaflets that are smaller, more oblong and less coriaceous) compared to the type material, and will be examined taxonomically as soon as fertile material becomes available. The same subgroup also contains material of two highly valued rosewood species, *D. occulta* and *D. normandii* (note that HASSOLD *et al.* (2016) confused *D. normandii* and *D. madagascariensis* due to a lack of flowering material of these species, which are difficult to distinguish when sterile).

Our results also indicated that Supergroup II could be separated into three clades. Two of these clades contain species whose centre of distribution is in the humid and sub-humid east and northwest of Madagascar, while the third clade contains species centred in the drier west and north of the island. This suggests that the geographic separation and

different climate regimes of the major eco-geographic regions of Madagascar might have played a significant role in shaping the patterns of diversity and diversification in the island's *Dalbergia* species, which might constitute a suitable model system to study the underlying diversification mechanisms, along the same lines as studies that have investigated elements of the Malagasy fauna (VENCES *et al.* 2009). Morphological characters that may characterise clades and could represent synapomorphies for them are not yet well understood and elucidating them would require genetic and morphological analysis of more species, which is now feasible using our multi-locus analyses as they provide resolution at the species level and below.

Our multi-locus nuclear topology also 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 has been observed in at various evolutionary timescales in many plant families, such as Asteraceae, and incomplete lineage sorting as well as ancient hybridisation are often invoked as explanations (PELSER *et al.* 2010). Comprehensive, world-wide taxon sampling employed for multi-locus nuclear analyses holds the potential to gain a better understanding of the complex biogeography observed in *Dalbergia* (VATANPARAST *et al.* 2013).

Our study using nine extraction replicates of recent collections also confirmed the reproducibility of sequencing and analysis of thousands of nuclear loci. The integration of highly informative herbarium collections, including a nomenclatural type, greatly facilitated the accurate identification of recently made but often sterile collections, and enabled the detection of misidentifications, as shown in both *Dalbergia maritima* var. *pubescens* and *D. madagascariensis* sensu lato. Our target enrichment approach thus has the potential to greatly facilitate the resolution of several taxonomic conundrums we have identified within the genus, which likely result from the difficulty of distinguishing between heritable and plastic trait variation within and among *Dalbergia* species.

Population genomic analyses

Population genomic analyses of 51 individuals readily separate the two closely related species *Dalbergia monticola* and *D. orientalis*, as well as a sympatric but genetically differentiated population that probably represents another undescribed species (*Dalbergia* sp. 17). By contrast, analyses based on three chloroplast genes (HASSOLD *et al.* 2016) recovered polymorphisms that distinguished samples of *D. monticola* from north-eastern Madagascar from material collected in the southeast, but variation was insufficient to

distinguish *D. monticola* from *D. orientalis* and other species. The NJ tree resulting from our analyses clearly distinguished three closely related species and further showed geographic structure within *D. monticola* and *D. orientalis*, both of which occur from north-eastern to south-eastern Madagascar but differ in their altitudinal distribution. Geographic resolution within species 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 10). These results indicate that genetic species identification and provenancing, at least to this broad geographic scale, could be feasible, which would have important implications for forensic timber identification and for tracing geographic hotspots of the illegal timber trade (UNODC 2016a).

Phylogenetic analyses of Fabaceae

The set of target enrichment baits presented here was mainly developed for *Dalbergia* but has proven to be useful beyond this genus, especially for studies within the Meso-Papilionoideae, from which the five genomes and the *Dalbergia* transcriptome were taken to serve as a basis for bait design. At the family level, less than 20% of targeted regions passed the stringent sensitivity and specificity filters, suggesting that many regions that appeared to be highly conserved with respect to five Papilionoid genomes and the *Dalbergia* transcriptome are in fact not highly conserved and therefore not efficiently recovered at the family level. However, phylogenetic analysis of 986 nuclear target regions was feasible across the family and provided excellent resolution, comparable to that resulting from a recent phylogenomic analysis of 3,473 and a subset of 1,103 nuclear loci derived from transcriptome data across legumes (KOENEN *et al.* 2020b).

We found strong evidence for monophyly of all four subfamilies with multiple species sampled, but ASTRAL-III quartet support values suggested ambiguities regarding the relationships among most subfamilies. KOENEN *et al.* (2020b) postulated a different most-likely tree of legume subfamilies in which Caesalpinioideae, Dialioideae and Papilionoideae formed a clade with 48% quartet support, whereas our analyses suggest that Caesalpinioideae and Dialioideae form a clade with Cercidoideae and Detarioideae with 47% quartet support (note that the monotypic Duparquetioideae was not included). However, the most likely subfamily tree postulated by KOENEN *et al.* (2020b) received almost equivalent overall quartet support in our analyses, as did a third hypothesis in which Caesalpinioideae and Papilionoideae formed one clade, and Dialioideae, Cercidoideae and Detarioideae formed another sister clade (Figure S2). This is consistent with the idea of a nearly simultaneous evolutionary origin of legume subfamilies (LPWG 2017, KOENEN *et al.* 2020b). Our outgroup sampling is sparse and may not permit accurate inference of the

placement of the root of the family. It is likely that Papilionoideae and *Polygala*, which both exhibit markedly higher substitution rates compared to other legume subfamilies (KOENEN *et al.* 2020b), are here inferred to be sister lineages in unrooted gene trees due to a long branch attraction artefact, leading to an erroneous placement of the root of the family. Nevertheless, our analyses confirm the postulated sister relationship between Cercidoideae and Detarioideae with 69% quartet support compared to 37% found by KOENEN *et al.* (2020b), who extensively sampled the nuclear genome. This relationship was not inferred in analyses based on the widely used *matK* chloroplast gene, nor when using 72 chloroplast genes (LPWG 2017, KOENEN *et al.* 2020b). Uncertainty regarding deep-branching relationships could be explained by incomplete lineage sorting, resulting in low phylogenetic signal and conflicting gene genealogies, rather than by missing data, given that we analysed over one million distinct alignment patterns distributed across all linkage groups of the *Cajanus cajan* reference genome, and given that we did not limit the analyses to exons.

Our analyses recovered multiple known clades within several legume subfamilies, a finding that further contributes to the validation of the target enrichment baits and bioinformatic approach presented here. The clades indicated in Figure 1 all showed >99% bootstrap or posterior support, as well as >50% quartet support, with the exception of the Robinioids (79% bootstrap support, >99% posterior support, 42% quartet support), which merit further investigation. 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 slightly higher quartet support than the two alternative hypotheses, and this is consistent with previous analyses of both nuclear and plastid sequences (KOENEN *et al.* 2020b). The genus *Aeschynomene* s.l. sensu RUDD (1955), which included the former *A.* sect. *Aeschynomene* and *A.* sect. *Ochopodium* Vogel, was confirmed to be non-monophyletic (RIBEIRO *et al.* 2007, CARDOSO *et al.* 2020), with *Ctenodon* (= *A.* sect. *Ochopodium*) identified as sister to *Machaerium*, and together as sister to *Dalbergia*, and together as sister to *Aeschynomene* s.str. (= *A.* sect. *Aeschynomene*).

Target enrichment baits

In this study we designed our bait set for target enrichment analysis based on conserved sequences across representatives of the highly diverse plant family Fabaceae using assemblies available in public databases and a draft *Dalbergia* transcriptome. We used the same bait set for target enrichment across Fabaceae, within the species-rich genus *Dalbergia*, and among closely related *Dalbergia* species. This procedure has both benefits and drawbacks. A major advantage is that an individual sample only needs to be sequenced once with a single bait set and can then be included in a diversity of analyses at different evolutionary timescales. On the other hand, capture sensitivity, defined as the "percentage of targets covered by at least one mapped read" (JONES AND GOOD 2016), can be low for species that are divergent from the focal group for which the bait set was optimized. Phylogenetic distance, along with the quality of DNA extractions and dilution of DNA libraries, likely contribute to the variation intrinsic to library preparation in generating the large differences observed in raw read numbers obtained per sample (Tables S1-S3). In the present study, only a fraction of sequence reads could be used in analyses at the macro- and micro-evolutionary scale (Tables S1-S3). Moreover, the identification of target regions that can be efficiently recovered across *Dalbergia* or the entire family required additional bioinformatic steps. Possible improvements could include the use of different bait sequences for target enrichment of the same locus across legumes, an approach whose utility has been demonstrated for bait design across angiosperms (JOHNSON *et al.* 2019). The selection of appropriate targets and bait sequences is particularly important because whole-genome duplication events occurred multiple times in Fabaceae (KOENEN *et al.* 2020a), which complicates orthology assessment and must be taken into consideration in the design of legume-specific bait sets (VATANPARAST *et al.* 2018, EGAN AND VATANPARAST 2019).

Bioinformatics pipeline

The present bioinformatic pipeline starts with the mapping of quality-trimmed reads to target regions, followed by sequence assembly on a per-region basis. This approach differs from the PHYLUCE pipeline (FAIRCLOTH 2016), in which quality-trimmed reads are first assembled to contigs, and these are then matched to target regions. An advantage of our approach is that it divides contig assembly into separate tasks performed on a subset of reads specific to the target region, thus simplifying the challenging *de novo* assembly of contigs from sequencing reads from across the large set of baits (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 AND GOODSON 2011). We addressed this problem by generating consensus sequences that were representative of a given taxon set and by limiting analyses to target regions that could be efficiently recovered in all groups of that taxon set. These set-specific reference sequences were then used in a second round of mapping, assembly and alignment. Our approach is conceptually similar to the HybPiper pipeline (JOHNSON *et al.* 2016), which also employs a mapping-assembly strategy with BWA (for nucleotide targets) and SPAdes, respectively. The HybPiper pipeline also uses exonerate to align contigs to target reference sequences, and combines non-overlapping contigs into supercontigs. Instead of using a minimum alignment length and (normalized) exonerate alignment scores, HybPiper uses depth of coverage to choose between multiple contigs that span more than 85% of the target sequence (full-length contigs), and percent identity to the target in cases where depth of coverage is similar between two full-length contigs. In its third phase, HybPiper identifies intron/exon boundaries and extracts coding sequences from the assembled contigs for alignment. This is possible because HybPiper uses coding sequences (peptide sequences or corresponding coding nucleotide sequences of one or several concatenated exons per gene) as target reference sequences for mapping of sequence reads and alignment of contigs. The HybPiper pipeline has been "designed specifically for the Hyb-Seq approach" (JOHNSON *et al.* 2016), in which exons are the primary targets (i.e., the target enrichment baits) and flanking introns and intergenic regions are used as supplementary regions for analyses at shallow evolutionary scales (WEITEMIER *et al.* 2014). Our approach is more general in scope and neither requires nor leverages knowledge about intron/exon boundaries in the targeted regions. Further strengths of the analysis pipeline we have developed are the flexibility it provides to set various analysis parameters, to merge alignments of physically overlapping target regions, and to readily save and visualize key summary statistics and alignments at different steps along the workflow to inform the selection of appropriate analysis parameters.

Conclusions and perspectives

The resources developed for Fabaceae and in particular for the genus *Dalbergia* bridge micro- and macro-evolutionary timescales and will hopefully contribute to community-driven efforts to advance legume genomics. We plan to expand analyses within *Dalbergia* to include additional species from Madagascar, a hotspot of diversity for the genus, as well as from other parts if its distributional range. Comprehensive sampling of taxa from

Madagascar would yield valuable insights into the complex patterns of diversity observed in the genus, thereby informing the taxonomic revision of Malagasy *Dalbergia* that is currently under way. Given that our efforts to include material from herbarium samples in our analyses were successful, it should be possible to infer relationships among the several species for which no freshly collected leaf material is currently available. The resulting sequence data could 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.

Acknowledgements

We thank Sonja Hassold and the Missouri Botanical Garden (MBG) Madagascar team for organising and conducting field work. We are also grateful to the Botanic Garden of the University of Zurich for the opportunity to sample legume plants from their living collection, as well as to MBG in St. Louis (USA) and the Muséum National d'Histoire Naturelle (MNHN) in Paris for access to their DNA bank and herbarium vouchers, respectively. We also thank Pete Lowry, Peter Phillipson and Nicholas Wilding for discussing sample determinations, *Dalbergia* taxonomy and Malagasy plant diversity. We are grateful to Claudia Michel for conducting laboratory work, and to Silvia Kobel (Genetic Diversity Centre Zurich, GDC) and Maria Domenica Moccia (Functional Genomics Center Zurich, FGCZ) for illumina sequencing. Furthermore, we thank Niklaus Zemp and the GDC for access to computing resources. Finally, we thank Simone Fior, Erik Koenen, Colin Hughes, Pete Lowry, Martin Fischer and Nicholas Wilding for their valuable inputs and comments.

Data accessibility

Raw sequence reads, the draft *Dalbergia* transcriptome, alignments for the subfamily and species sets, and a VCF file for the population set will be added on the European Nucleotide Archive (ENA) under the primary accession number PRJEB41848. Tables S1 – S3 (collections comprising the subfamily, species and population sets, respectively, including sequencing statistics and mapping statistics in both iterations) and Tables S4 – S5 (gene annotations of the 2,396 and 1,005 *C. cajan* reference sequences, respectively) are available as Supporting Online Material at https://github.com/scrameri/Dalbergia_Phylogenomics.

References

Adema, F., H. Ohashi, and B. Sunarno. 2016. Notes on Malesian Fabaceae (Leguminosae-Papilionoideae) 17. The genus Dalbergia. Blumea **61**:186-206.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of Molecular Biology **215**:403-410.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology **19**:455-477.

Barrett, M. A., J. L. Brown, and A. D. Yoder. 2013. Conservation: protection for trade of precious rosewood. Nature **499**:29-29.

Bhagwat, R. M., B. B. Dholakia, and N. Y. Kadoo. 2015. Two new potential barcodes to discriminate Dalbergia species. PLoS One:doi: 10.1371-journal.pone.0142965.s0142005.

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

Bosser, J., and R. Rabevohitra. 2002. Tribe Dalbergieae. Pages 321-361 *in* D. J. Du Puy, J. N. Labat, R. Rabevohitra, J. F. Villiers, J. Bosser, and J. Moat, editors. The Leguminosae of Madagascar. Royal Botanical Gardens, Kew.

Bosser, J., and R. Rabevohitra. 2005. Espèces nouvelles dans le genre Dalbergia (Fabaceae, Papilionoideae) à Madagascar. Adansonia **27**:209-216.

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

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: architecture and applications. BMC Bioinformatics **10**:doi:10.1186-1471-2105-1110-1421.

Cardoso, D., L. P. de Queiroz, R. T. Pennington, H. C. de Lima, E. Fonty, M. F. Wojciechowski, and M. Lavin. 2012. Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. American Journal of Botany **99**:1991-2013.

Cardoso, D., C. Mattos, F. Filardi, A. Delgado-Salinas, M. Lavin, P. Moraes, F. Tapia-Pastrana, and H. Lima. 2020. A molecular phylogeny of the pantropical papilionoid legume Aeschynomene supports reinstating the ecologically and morphologically coherent genus Ctenodon. Neodiversity **13**:1-38.

Cardoso, D., R. T. Pennington, L. P. de Queiroz, J. S. Boatwright, B. E. Van Wyk, M. F. Wojciechowski, and M. Lavin. 2013. Reconstructing the deep-branching relationships of the papilionoid legumes. South African Journal of Botany **89**:58-75.

Chaisson, M. J. P., R. K. Wilson, and E. E. Eichler. 2015. Genetic variation and the de novo assembly of human genomes. Nature Reviews Genetics **16**:1-14.

CITES. 2020. Appendices I, II and III. Retrieved from https://cites.org/eng/app/appendices.php (accessed on 27 January 2021).

Cutter, A. D. 2013. Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. Molecular Phylogenetics and Evolution **69**:1172-1185.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and G. P. A. Group. 2011. The variant call format and VCFtools. Bioinformatics **27**:2156-2158.

de Carvalho, A. M. 1997. A synopsis of the genus Dalbergia (Fabaceae: Dalbergieae) in Brazil. Brittonia **49**:87-109.

de La Harpe, M., J. Hess, O. Loiseau, N. Salamin, C. Lexer, and M. Paris. 2018. A dedicated target capture approach reveals variable genetic markers across micro- and macro-evolutionary time scales in palms. Molecular Ecology Resources **19**:221-234.

de La Harpe, M., M. Paris, D. N. Karger, J. Rolland, M. Kessler, N. Salamin, and C. Lexer. 2017. Molecular ecology studies of species radiations: current research gaps, opportunities and challenges. Molecular Ecology **26**:2608-2622.

De Moraes, P. L. R. 2012. Linnaeus's Plantae Surinamenses revisited. Phytotaxa **41**:1-86.

Dowle, M., and A. Srinivasan. 2019. data.table: Extension of `data.frame`. R package version 1.12.8. https://CRAN.R-project.org/package=data.table.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin **19**:11-15.

Egan, A. N., and M. Vatanparast. 2019. Advances in legume research in the genomics era. Australian Systematic Botany **46**:185-125.

Faircloth, B. C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics **32**:786-788.

Faircloth, B. C., M. G. Branstetter, N. D. White, and S. G. Brady. 2014. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. Molecular Ecology Resources **15**:489-501.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Systematic Biology **61**:717-726.

Faircloth, B. C., L. Sorenson, F. Santini, and M. E. Alfaro. 2013. A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). PLoS One **8**:e65923.

Garrison, E., and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907.

Griesmann, M., Y. Chang, X. Liu, Y. Song, G. Haberer, M. B. Crook, B. Billault-Penneteau, D. Lauressergues, J. Keller, L. Imanishi, Y. P. Roswanjaya, W. Kohlen, P. Pujic, K. Battenberg, N. Alloisio, Y. Liang, H. Hilhorst, M. G. Salgado, V. Hocher, H. Gherbi, S. Svistoonoff, J. J. Doyle, S. He, Y. Xu, S. Xu, J. Qu, Q. Gao, X. Fang, Y. Fu, P. Normand, A. M. Berry, L. G. Wall, J.-M. Ané, K. Pawlowski, X. Xu, H. Yang, M. Spannagl, K. F. X. Mayer, G. K.-s. Wong, M. Parniske, P.-M. Delaux, and S. Cheng. 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science **361**:eaat1743-1713.

Harris, R. S. 2007. Improved pairwise alignment of genomic DNA. The Pennsylvania State University.

Hartvig, I., M. Czako, E. D. Kjær, L. R. Nielsen, and I. Theilade. 2015. The use of DNA barcoding in identification and conservation of rosewood (Dalbergia spp.). PLoS One **10**:doi: 10.1371-journal.pone.0138231.

Hassold, S., P. P. Lowry, M. R. Bauert, A. Razafintsalama, L. Ramamonjisoa, and A. Widmer. 2016. DNA barcoding of Malagasy rosewoods: towards a molecular

identification of CITES-listed *Dalbergia* species. PLoS ONE: doi: 10.1371-journal.pone.0157881.

Hughes, C., and R. Eastwood. 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**:10334-10339.

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

Johnson, M. G., L. Pokorny, S. Dodsworth, L. R. Botigué, R. S. Cowan, A. Devault, W. L. Eiserhardt, N. Epitawalage, F. Forest, J. T. Kim, J. H. Leebens-Mack, I. J. Leitch, O. Maurin, D. E. Soltis, P. S. Soltis, G. K.-s. Wong, W. J. Baker, and N. J. Wickett. 2019. A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any Flowering Plant Designed Using k-Medoids Clustering. Systematic Biology **68**:594-606.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics **24**:1403-1405.

Jombart, T., and I. Ahmed. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics **27**:3070-3071.

Jones, M. R., and J. M. Good. 2016. Targeted capture in evolutionary and ecological genomics. Molecular Ecology **25**:185-202.

Jongkind, C. C. H. 2007. A new species of Dalbergia (Leguminosae, Dalbergieae) from Western Africa. Blumea **52**:11-14.

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

Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ **2**:doi: 10.7717-peerj.7281.

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

Knaus, B. J., and N. J. Grünwald. 2017. vcfr: a package to manipulate and visualize variant call format data in R. Molecular Ecology Resources **17**:44-53.

Koenen, E. J. M., D. I. Ojeda, F. T. Bakker, J. J. Wieringa, C. Kidner, O. J. Hardy, R. T. Pennington, P. S. Herendeen, A. Bruneau, and C. E. Hughes. 2020a. The Origin of the Legumes is a Complex Paleopolyploid Phylogenomic Tangle closely associated with the Cretaceous-Paleogene (K-Pg) Mass Extinction Event. Systematic Biology **syaa041**:https://doi.org/10.1093/sysbio/syaa1041.

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

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources **15**:1179-1191.

Lachenaud, O. 2016. *Dalbergia*. Pages 101-153 *in* M. S. M. Sosef, J. Florence, L. N. Banak, H. P. Bourobou, and P. Bissiengou, editors. Flore du Gabon - Leguminosae - Papilionoideae. Margraf Publishers, Wekersheim.

Lachenaud, O., and L. van der Maesen. 2016. Notes on African *Dalbergia* (Leguminosae–Papilionoideae) with the description of two new species from Atlantic Central Africa. Symbolae Botanicae Upsalienses **38**:167-194.

Lavin, M., R. T. Pennington, B. B. Klitgaard, J. I. Sprent, H. C. de Lima, and P. E. Gasson. 2001. The dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. American Journal of Botany **88**:503-533.

Lemmon, A. R., S. A. Emme, and E. M. Lemmon. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Systematic Biology **61**:727-744.

Lemmon, A. R., and E. M. Lemmon. 2012. High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. Systematic Biology **61**:745-761.

Léveillé-Bourret, É., J. R. Starr, B. A. Ford, E. Moriarty Lemmon, and A. R. Lemmon. 2018. Resolving rapid radiations within Angiosperm families using anchored phylogenomics. Systematic Biology **67**:94-112.

Lewis, G. P., B. D. Schrire, B. A. Mackinder, and M. Lock. 2005. Legumes of the World. Royal Botanical Gardens, Kew.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1302.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**:1754-1760.

Linares, J., and M. Sousa. 2007. Nuevas especies de *Dalbergia* (Leguminosae: Papilionoideae: Dalbergieae) en México y Centroamérica. Ceiba **48**:61-82.

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

LPWG. 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. Taxon **66**:44-77.

Lunter, G., and M. Goodson. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Research **21**:936-939.

Madagascar Catalogue. 2020. Catalogue of the Plants of Madagascar. Missouri Botanical Garden, St. Louis, U.S.A. [http://www.tropicos.org/Project/Madagascar. Accessed: June, 2020].

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

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

McMahon, M., and L. Hufford. 2004. Phylogeny of Amorpheae (Fabaceae: Papilionoideae). American Journal of Botany **91**:1219-1230.

Mirarab, S., R. Reaz, M. S. Bayzid, T. Zimmermann, M. S. Swenson, and T. Warnow. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics **30**:I541-I548.

Moreira, A. S. F. P., A. C. L. Queiroz, F. d. V. Barros, M. F. Goulart, and J. Pires de Lemos-Filho. 2013. Do leaf traits in two *Dalbergia* species present differential plasticity in relation to light according to their habitat of origin? Australian Journal of Botany **61**:592-598.

Mousavi-Derazmahalleh, M., P. E. Bayer, J. K. Hane, B. Valliyodan, H. T. Nguyen, M. N. Nelson, W. Erskine, R. K. Varshney, R. Papa, and D. Edwards. 2018. Adapting legume crops to climate change using genomic approaches. Plant, Cell & Environment **42**:6-19.

Nicholls, J. A., R. T. Pennington, E. J. M. Koenen, C. E. Hughes, J. Hearn, L. Bunnefeld, K. G. Dexter, G. N. Stone, and C. A. Kidner. 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**:10.3389-fpls.2015.00710.

Ojeda, D. I., E. Koenen, S. Cervantes, M. de la Estrella, E. Banguera-Hinestroza, S. B. Janssens, J. Migliore, B. B. Demenou, A. Bruneau, F. Forest, and O. J. Hardy. 2019. Phylogenomic analyses reveal an exceptionally high number of evolutionary shifts in a florally diverse clade of African legumes. Molecular Phylogenetics and Evolution **137**:156-167.

Paradis, E., and K. Schliep. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics **35**:526-528.

Pelser, P. B., A. H. Kennedy, E. T. A. J. of, and C. Chatelain. 2010. Patterns and causes of incongruence between plastid and nuclear Senecioneae (Asteraceae) phylogenies. American Journal of Botany **97**:856-873.

POWO. 2020. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. http://www.plantsoftheworldonline.org/ accessed on Feb 1 2020.

Prain, D. 1904. The species of *Dalbergia* of south-eastern Asia. Annals of the Royal Botanic Garden, Calcutta **10**:1-411.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945-959.

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

Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**:841-842.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Ribeiro, R. A., M. Lavin, J. P. Lemos-Filho, and C. Filho. 2007. The genus *Machaerium* (Leguminosae) is more closely related to *Aeschynomene* Sect. *Ochopodium* than to *Dalbergia*: inferences from combined sequence data. Systematic Botany **32**:762-771.

Rissler, L. J. 2016. Union of phylogeography and landscape genetics. Proceedings of the National Academy of Sciences **113**:8079-8086.

Rudd, V. E. 1955. The American species of *Aeschynomene*. Bulletin of the United States National Museum **32**:1-176.

Safonova, Y., A. Bankevich, and P. A. Pevzner. 2015. dipSPAdes: assembler for highly polymorphic diploid genomes. Journal of Computational Biology **22**:528-545.

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

Sass, C., W. J. D. Iles, C. F. Barrett, S. Y. Smith, and C. D. Specht. 2016. Revisiting the Zingiberales: using multiplexed exon capture to resolve ancient and recent phylogenetic splits in a charismatic plant lineage. PeerJ **4**:10.7717-peerj.1584.

Schuurman, D., and P. P. Lowry. 2009. The Madagascar rosewood massacre. Madagascar Conservation and Development **4**:98-102.

Shahi Shavvon, R., S. Kazempour Osaloo, A. A. Maassoumii, F. Moharrek, S. Karaman Erkul, A. R. Lemmon, E. M. Lemmon, I. Michalak, A. N. Muellner-Riehl, and A. Favre. 2017. Increasing phylogenetic support for explosively radiating taxa: The promise of high-throughput sequencing for *Oxytropis* (Fabaceae). Journal of Systematics and Evolution **55**:385-404.

Siriwat, P., and V. Nijman. Using online media-sourced seizure data to assess the illegal wildlife trade in Siamese rosewood. Environmental Conservation **45**:352-360.

Slater, G. S. C., and E. Birney. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics **6**:31.

Smit, A. F. A., R. Hubley, and P. Green. 2013. RepeatMasker Open-4.0.2013-2015 http://www.repeatmasker.org.

Smith, B. T., M. G. Harvey, B. C. Faircloth, T. C. Glenn, and R. T. Brumfield. 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Systematic Biology **63**:83-95.

Smith, T. F., and M. S. Waterman. 1981. Identification of common molecular sequences. Journal of Molecular Biology **147**:195-197.

Soosairaj, S., P. Raja, S. J. Britto, and B. Balaguru. 2018. A new species of *Dalbergia* (Fabaceae: Dalbergieae) from Tamil Nadu, India. Phytotaxa **360**:282-286.

Sprent, J. I., J. Ardley, and E. K. James. 2017. Biogeography of nodulated legumes and their nitrogen-fixing symbionts. The New Phytologist **215**:40-56.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**:1312-1313.

Tange, O. 2011. GNU parallel—The command-line power tool. Login: The USENIX Magazine **36**:42-47.

UNODC. 2016a. Best practice guide for forensic timber identification. Pages 1-213. United Nations, New York.

UNODC. 2016b. World wildlife crime report: trafficking in protected species. United Nations, New York.

Vardeman, E., and J. V. Runk. 2020. Panama's illegal rosewood logging boom from *Dalbergia* retusa. Global Ecology and Conservation **23**:e01098.

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

Vatanparast, M., A. Powell, J. J. Doyle, and A. N. Egan. 2018. Targeting legume loci: A comparison of three methods for target enrichment bait design in Leguminosae phylogenomics. Applications in Plant Sciences **6**:e1036.

Vences, M., K. C. Wollenberg, D. R. Vieites, and D. C. Lees. 2009. Madagascar as a model region of species diversification. Trends in Ecology and Evolution **24**:456-465.

Viguier, R. 1944. Légumineuses de Madagascar. Archives Botaniques **6**:685-725, non valid. publ.

Villaverde, T., L. Pokorny, S. Olsson, M. Rincón-Barrado, M. G. Johnson, E. M. Gardner, N. J. Wickett, J. Molero, R. Riina, and I. Sanmartín. 2018. Bridging the micro- and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. The New Phytologist **220**:636-650.

Waeber, P. O., D. Schuurman, B. Ramamonjisoa, M. Langrand, C. V. Barber, J. L. Innes, P. P. Lowry, and L. Wilmé. 2019. Uplisting of Malagasy precious woods critical for their survival. Biological Conservation **235**:89-92.

Weitemier, K., S. C. K. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb-Seq: Combining Target Enrichment and Genome Skimming for Plant Phylogenomics. Applications in Plant Sciences **2**:1400042.

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

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.

Wojciechowski, M. F., M. Lavin, and M. J. Sanderson. 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**:1846-1862.

Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T.-Y. Lam. 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**:28-36.

Yuan, H., J. Jiang, F. A. Jiménez, E. P. Hoberg, J. A. Cook, K. E. Galbreath, and C. Li. 2016. Target gene enrichment in the cyclophyllidean cestodes, the most diverse group of tapeworms. Molecular Ecology Resources **16**:1095-1106.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics **19**:1-16.

Zhuang, W., H. Chen, M. Yang, J. Wang, M. K. Pandey, C. Zhang, W.-C. Chang, L. Zhang, X. Zhang, R. Tang, V. Garg, X. Wang, H. Tang, C.-N. Chow, J. Wang, Y. Deng, D. Wang, A. W. Khan, Q. Yang, T. Cai, P. Bajaj, K. Wu, B. Guo, X. Zhang, J. Li, F. Liang, J. Hu, B. Liao, S. Liu, A. Chitikineni, H. Yan, Y. Zheng, S. Shan, Q. Liu, D. Xie, Z. Wang, S. A. Khan, N. Ali, C. Zhao, X. Li, Z. Luo, S. Zhang, R. Zhuang, Z. Peng, S. Wang, G. Mamadou, Y. Zhuang, Z. Zhao, W. Yu, F. Xiong, W. Quan, M. Yuan, Y. Li, H. Zou, H. Xia, L. Zha, J. Fan, J. Yu, W. Xie, J. Yuan, K. Chen, S. Zhao, W. Chu, Y. Chen, P. Sun, F. Meng, T. Zhuo, Y. Zhao, C. Li, G. He, Y. Zhao, C. Wang, P. B. Kavikishor, R.-L. Pan, A. H. Paterson, X. Wang, R. Ming, and R. K. Varshney. 2019. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nature Genetics **51**:1-17.

Appendix I

Supplementary Figures

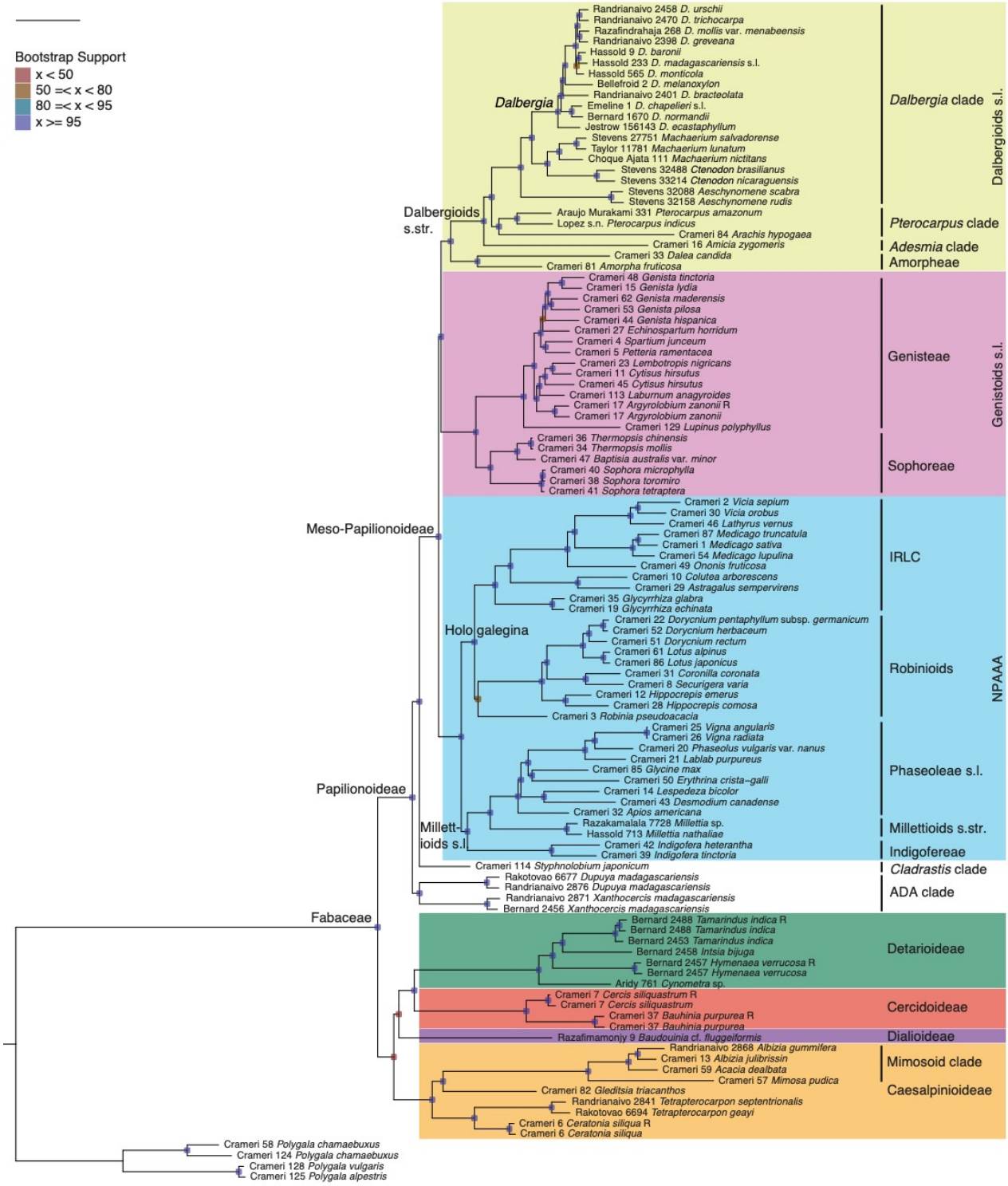

FIGURE S1: Maximum likelihood tree of the subfamily set (n = 110) inferred using RAxML on the alignment supermatrix (concatenation method). The supermatrix had 1,196,506 alignment sites, 1,028,714 distinct alignment patterns (unique columns), and 21.73% missing data.

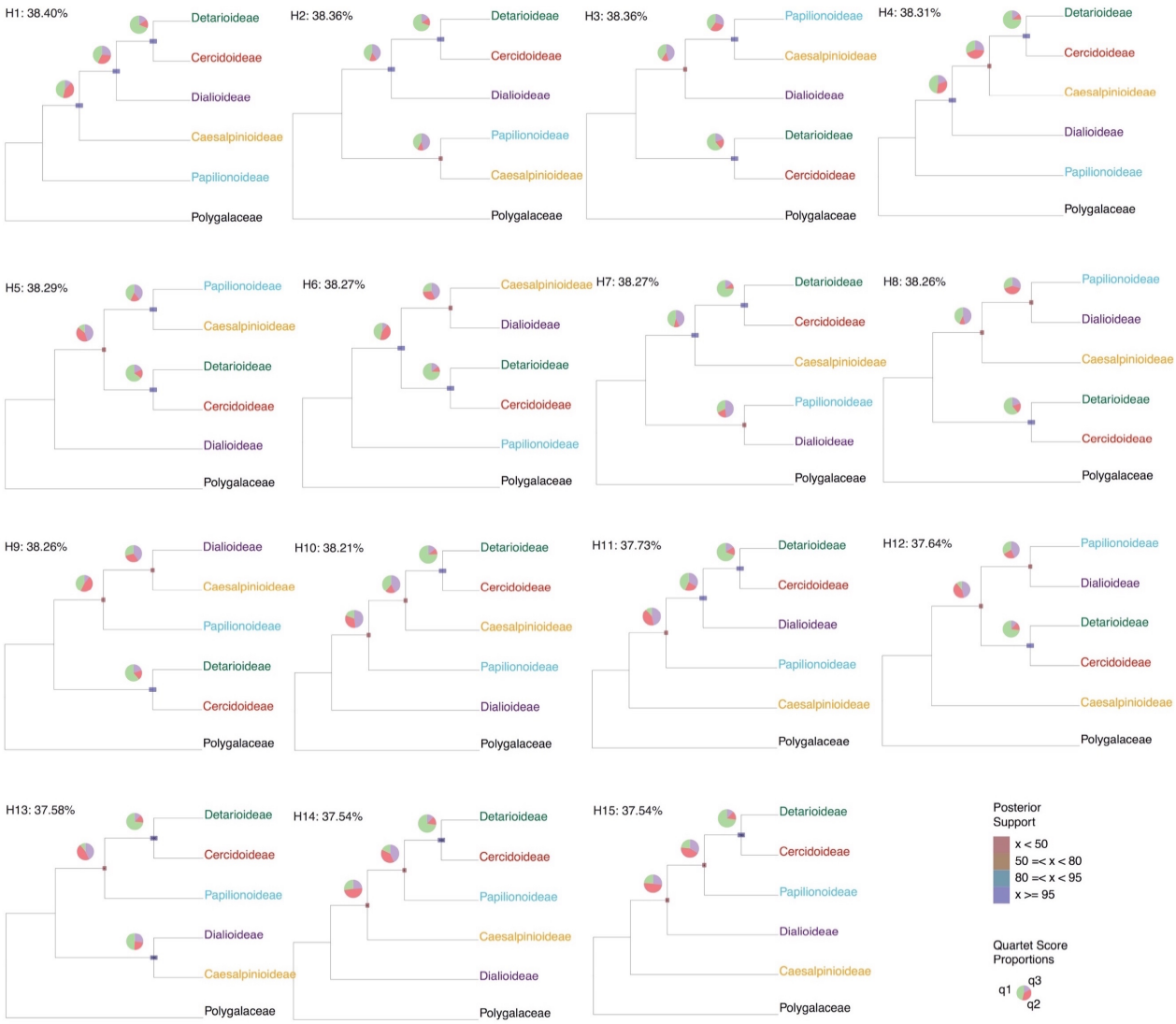

FIGURE S2: Quartet support for fifteen hypotheses (H1 to H15) on relationships between Fabaceae subfamilies. All possible topologies with Caesalpinioideae, Dialioideae, Papilionoideae and the (Cercidoideae, Detarioideae) clade as ingroups were tested. Pie charts denote the fraction of gene trees that are consistent with a given node (green) or with alternative topologies (red, purple). The panels are ranked by decreasing normalized quartet score (i.e., the percentage of input gene tree quartet trees that are consistent with a hypothesis).

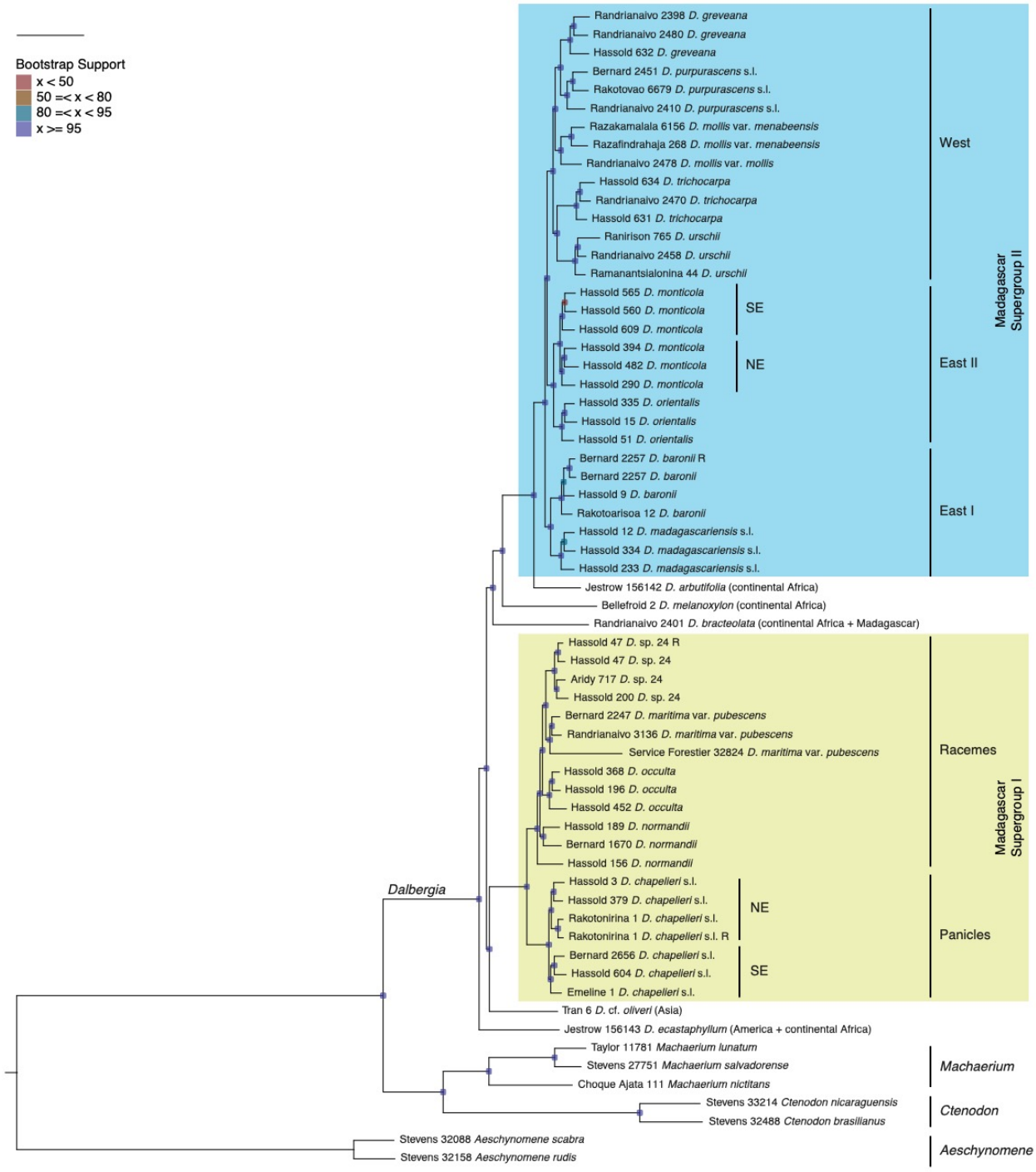

FIGURE S3: Maximum likelihood tree of the species set (n = 63) inferred using RAxML on the alignment supermatrix (concatenation method). The supermatrix had 2,589,455 alignment sites, 1,270,746 distinct alignment patterns (unique columns), and 17.18% missing data.

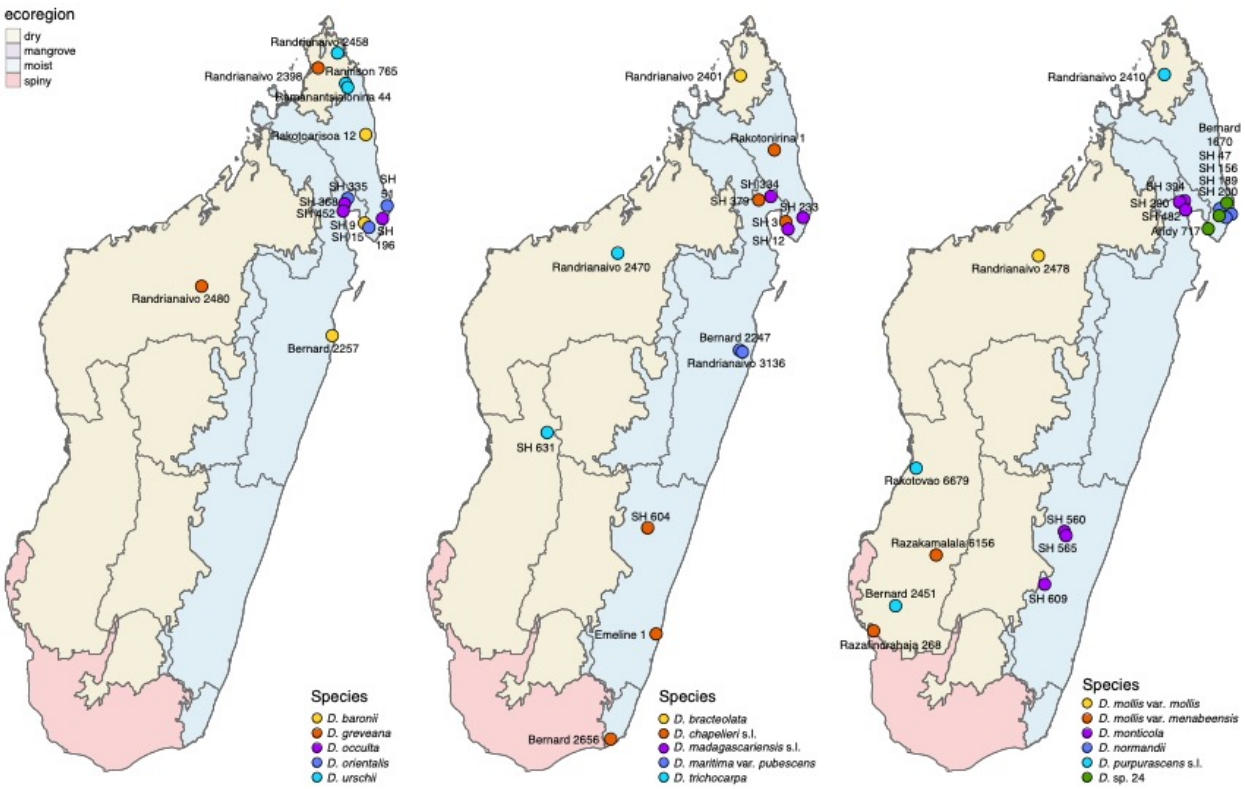

FIGURE S4: Sampling locations of species set specimens. Species are arbitrarily color-coded and assigned to separate panels with color-coded ecoregions. Coordinates are jittered for better readability or omitted if not precisely known. Maps were created using tmap version 3.0 (TENNEKES 2018). SH: S. Hassold.

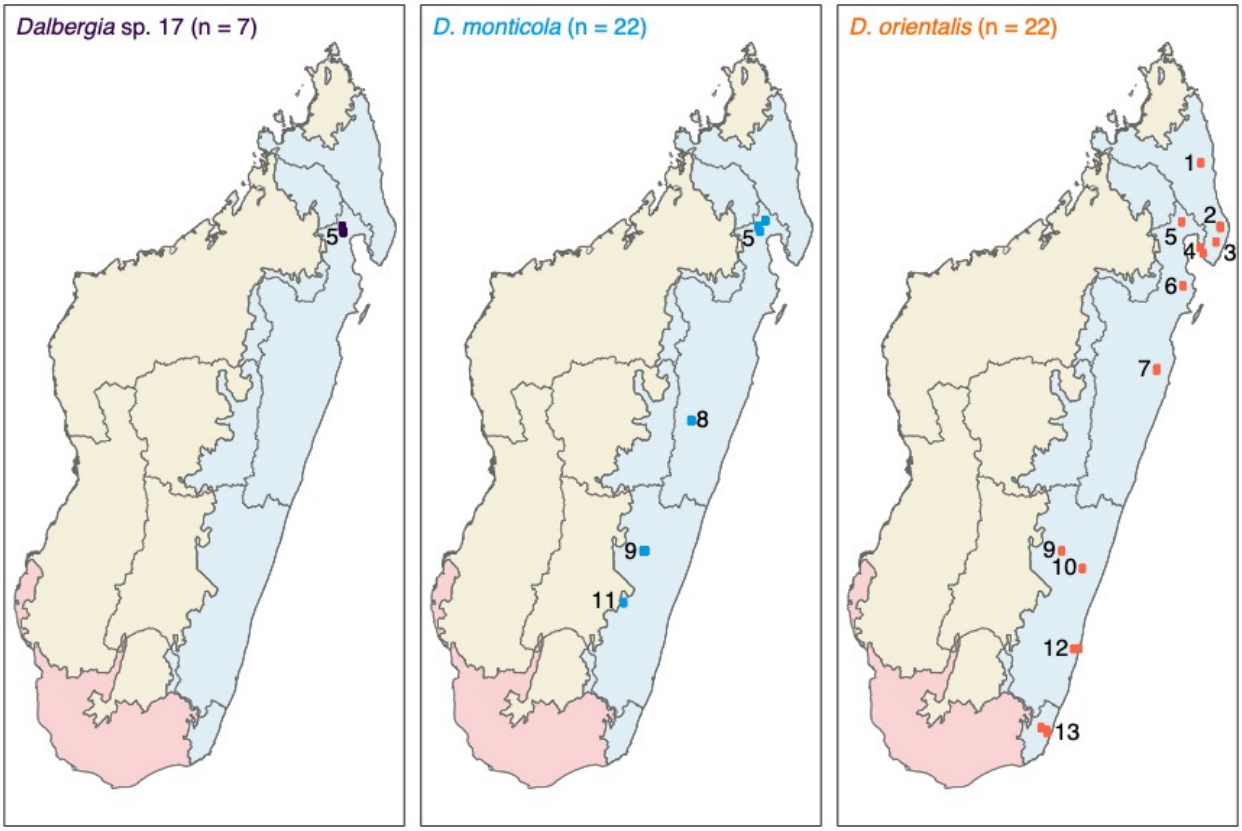

FIGURE S5: Sampling locations (red symbols) of population set specimens (n = 51). Three inferred species are shown as three separate panels with color-coded ecoregions. Locations are numbered as in Table S3. The two described and closely related species *Dalbergia monticola* and *D. orientalis* are wide-spread in humid evergreen forests of East Madagascar. *Dalbergia monticola* occurs mainly at higher elevation (few collections are from below 400 m a.s.l.), while *D. orientalis* occurs mainly at lower elevation (few collections are from above 750 m a.s.l.). Both species co-occur at mid elevation, e.g. in locations 5 and 9. A third species, *Dalbergia* sp. 17, is closely related but genetically distinct from both *D. monticola* and *D. orientalis*, and co-occurs with both species in northeast Madagascar (location 5). Maps were created using tmap version 3.0 (TENNEKES 2018).

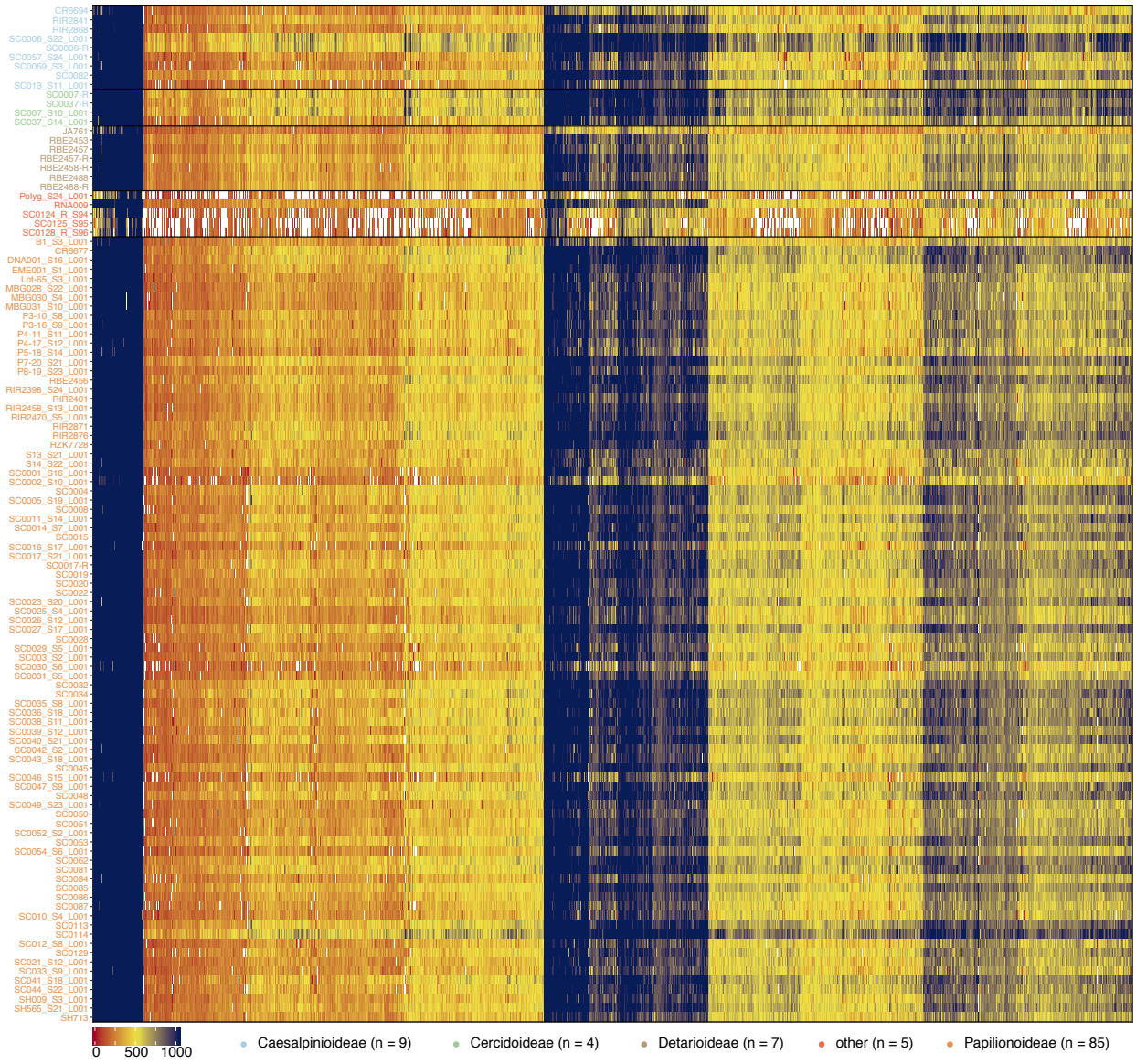

FIGURE S6: Heatmap of covered length in base pairs in 1,917 target regions (x axis) retained after step 2 of the second iteration in the subfamily set specimens (y axis, n = 110). Specimens are sorted according to the color-coded taxon groups used for target region filtering. Target regions are sorted according to hierarchical clusters (ward.D2 method). Values above 1,026 base pairs were capped for better readability. Sample identifiers are as in Table S1.

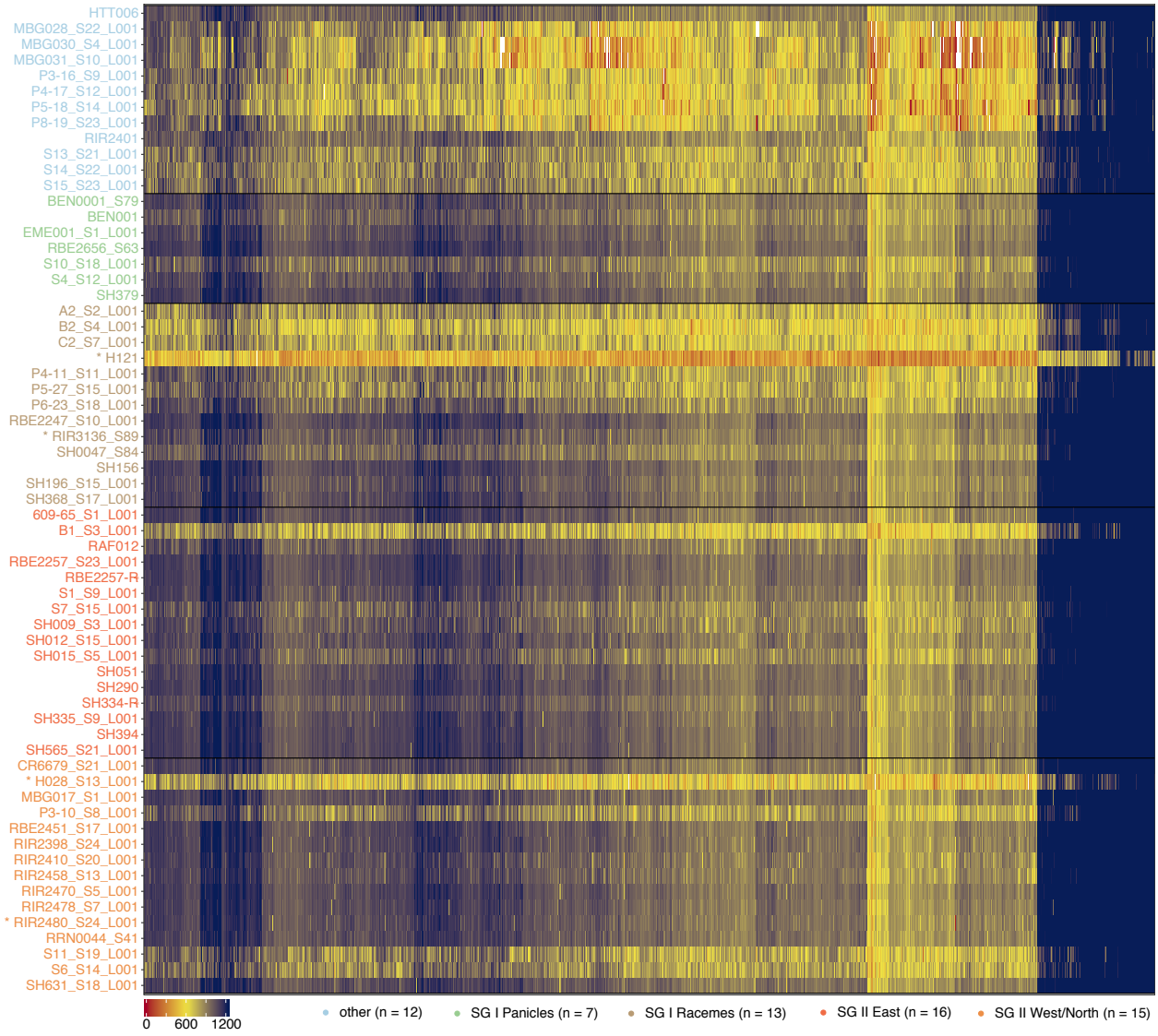

FIGURE S7: Heatmap of covered length in base pairs in 3,418 target regions (x axis) retained after step 2 of the second iteration in the species set specimens (y axis, n = 63). Specimens are sorted according to the color-coded taxon groups used for target region filtering. Target regions are sorted according to hierarchical clusters (ward.D2 method). Values above 1,221 base pairs were capped for better readability. Sample identifiers are as in Table S2. Samples marked with asterisks were extracted from herbarium vouchers.

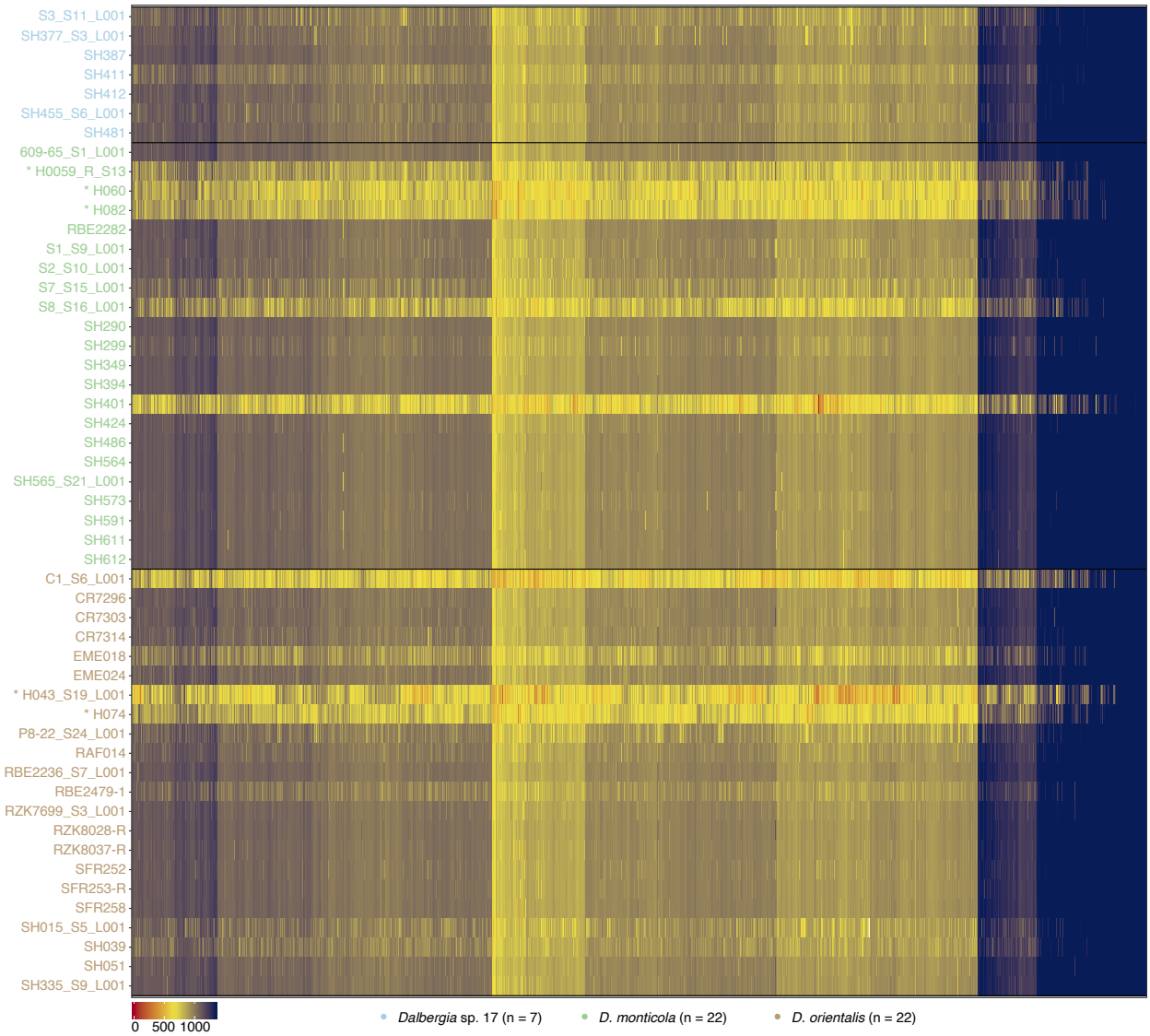

FIGURE S8: Heatmap of covered length in base pairs in 2,396 target regions (x axis) used as mapping targets in the species set specimens (y axis, n = 51). Specimens are sorted and color-coded for species. Values above 1,339 base pairs were capped for better readability. Target regions are sorted according to hierarchical clusters (ward.D2 method). Sample identifiers are as in Table S3. Samples marked with asterisks were extracted from herbarium vouchers.

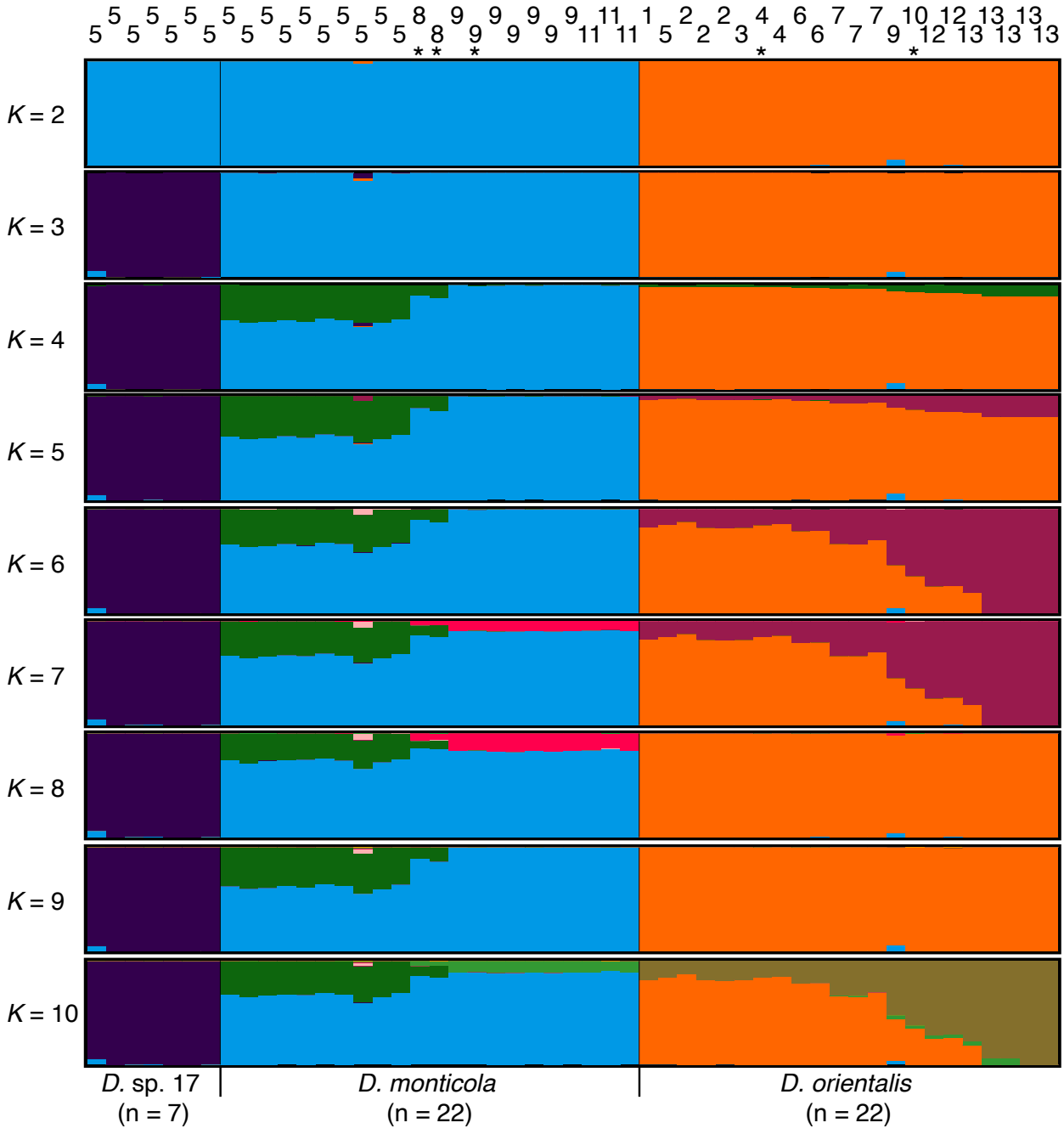

FIGURE S9: Structure results for up to ten assumed clusters *K* in 51 individuals and 7,156 single nucleotide polymorphisms (SNPs). Individuals are sorted by species and then by increasing degrees south latitude. Numbers at the top indicate broad sampling locations as in Table S3 and Figure S5. The major clusters averaged across ten replicate runs using Clumpak (KOPELMAN *et al.* 2015) are shown. The first split at *K* = 2 separated *Dalbergia orientalis* from *D. monticola* and *D.* sp. 17, and the second split at *K* = 2 further separated *D. monticola* from *D.* sp. 17. Admixture proportions were less clear-cut for higher values of *K*, but indicated isolation by distance at a broad geographical scale, dividing specimens from northeast (locations 1 to 6), central-east (locations 7 and 8) and southeast Madagascar (locations 9 to 13) in both *D. monticola* and *D. orientalis*. Samples marked with an asterisk (*) denote were extracted from herbarium vouchers.

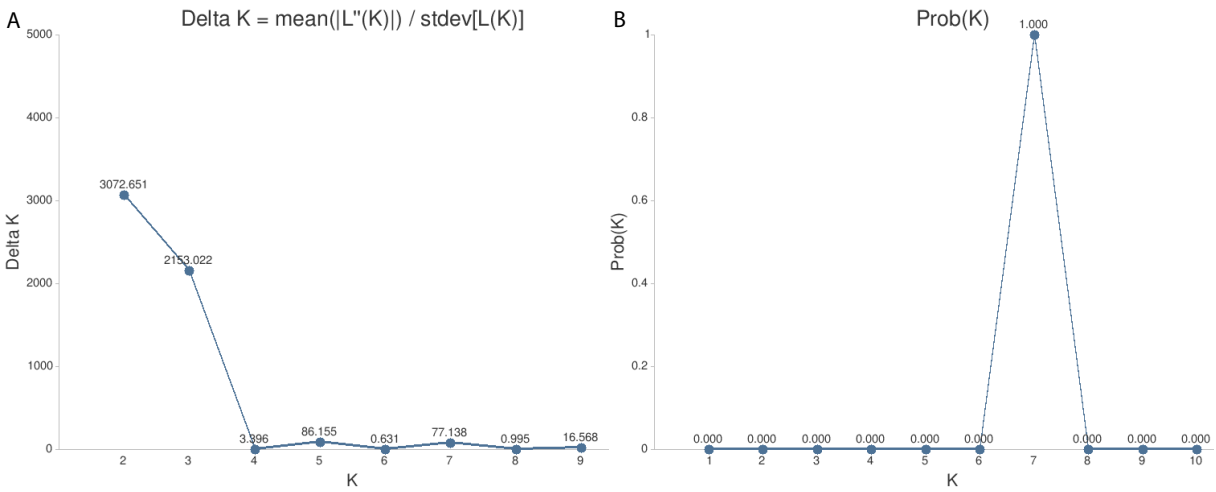

FIGURE S10: Structure probability at different values of *K*. **A)** Delta *K* statistic. **B)** Probability by *K*. Graphs were produced using Clumpak (KOPELMAN *et al.* 2015). Clustering solutions at *K* = 2 (splits *D. orientalis* from the rest), 3 (splits the three species), 5 (additional isolation by distance within both *D. monticola* and *D. orientalis*) and 7 (more pronounced isolation by distance, but includes two ghost clusters) appear biologically meaningful.

Supplementary Tables

Supplementary Tables S1 – S5 are available as Supporting Online Material at https://github.com/scrameri/DalbergiaPhylogenomics.

TABLE S1: Subfamily set specimens (n = 110) and associated collection data, sequencing data, and mapping statistics.

TABLE S2: Species set specimens (n = 63) and associated collection data, sequencing data and mapping statistics.

TABLE S3: Population set specimens (n = 51) and associated collection data, sequencing data and mapping statistics.

TABLE S4: Gene annotations for 1,005 target regions for genome-wide analysis in Fabaceae.

TABLE S5: Gene annotations for 2,396 target regions for genome-wide analysis in *Dalbergia*.

Supplementary Methods

Design of target enrichment baits

The following genome sequences were used for divergent reference capture:

- *Cajanus cajan* (L.) Millsp.) v1.0 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000340665.1)
- *Glycine max* (L.) Merr. v1.0 (http://phytozome.jgi.doe.gov)
- *Lotus japonicus* L. v2.5 (https://lotus.au.dk/data/download)
- *Medicago truncatula* Gaertn. v3.5 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000219495.1)
- *Phaseolus vulgaris* L. v2.1 (DOE-JGI and USDA-NIFA, http://phytozome.jgi.doe.gov)

Assembly of a draft Dalbergia transcriptome

For transcriptome assembly, we collected fresh leaf material of *Dalbergia madagascariensis* Vatke subsp. *antongilensis* Bosser & R. Rabev. from a young plant (progeny of collection Jean Luc Mora 26) cultivated in a greenhouse at ETH research station Lindau-Eschikon, Switzerland. Young leaves and leaf buds were snap-frozen in liquid nitrogen and stored until used. Frozen leaves were later ground with a BeadRuptor (Omni International, USA) and total RNA was isolated using the RNeasy Plant Mini Kit (QIAGEN, USA). We quantified total RNA using the RNA BR (broad range) assay kit (Life Technologies, Carlsbad, CA, USA) for QubitTM 2.0 fluorometer (Invitrogen), checked RNA integrity on a 1.5% agarose gel, and verified RNA quality on an Agilent 2100 Bioanalyzer. Total RNA was further treated with DNase I prior to library preparation with the Illumina TruSeq RNA kit (Illumina, San Diego, CA, USA) at the Functional Genomics Centre, Zurich, Switzerland (FGCZ). Paired-end sequencing was performed on an Illumina HiSeq 2000. The sequence data are deposited in the NCBI Sequence Read Archive (accession number will be added). We obtained 63 million paired-end reads, corresponding to a total of 6.2 gigabases (Gb). The raw reads were cleaned by removing adapter sequences with cutadapt (MARTIN 2011), followed by filtering and trimming of low-quality reads and bases with condetri.pl (SMEDS AND KÜNSTNER 2011). We performed a de novo assembly of the transcriptome using Trinity release 2012-01-25 (GRABHERR *et al.* 2011).

Taxon samples for target enrichment bait validation

The eight species of Caesalpinioideae included four species of the mimosoid clade (LPWG 2017). The 85 Papilionoideae samples were represented by one species of the Cladrastis clade (WOJCIECHOWSKI 2013), two species of the Angylocalyx-Dipterygeae-Amburana (ADA) clade (CARDOSO *et al.* 2012) and 78 species of Meso-Papilionoideae (WOJCIECHOWSKI 2013). Within the Meso-Papilionoideae, we included representatives of three species-rich clades: 19 species of Genistoids s.l. (WOJCIECHOWSKI *et al.* 2004, CARDOSO *et al.* 2012), 25 species of Dalbergioids s.l. (WOJCIECHOWSKI *et al.* 2004), and 34 species of the non-protein-amino-acid-accumulating (NPAAA) clade (WOJCIECHOWSKI *et al.* 2004, CARDOSO *et al.* 2012).

Samples were obtained from various sources. We re-analyzed 46 DNA samples from a study on *Dalbergia* chloroplast variation (HASSOLD *et al.* 2016), analyzed 39 *Dalbergia* samples collected for this study in Madagascar, and 48 samples collected with permission from the Zurich Botanical Garden (Switzerland). We also obtained twelve samples through the Missouri Botanical Garden DNA bank (St. Louis, MO, USA), three herbarium samples from the Conservatoire et Jardin botaniques Genève (CJBG), three herbarium samples from the Muséum National d'Histoire Naturelle (MNHN) Paris, and one sample from the USDA-ARS Tropical Agriculture Research Station (Mayaguez, Puerto Rico). Ten samples were collected in the field in Switzerland by S. Crameri, A. Widmer and M. Baltisberger. Five samples were grown from commercially available seeds at the ETH research station Lindau (Switzerland) and two samples were purchased from a commercial source.

Read quality-trimming and quality-filtering

Raw paired-end reads with the indicated file extensions (-x option) located in $rawreads (-r option) were quality-trimmed and quality-filtered using trimmomatic. Specifically, we used ILLUMINACLIP with an adapter sequence file containing NEBNext, TruSeq and Illumina adaptor sequences (-a option), a seed mismatch of 2, a palindrome clip threshold of 20, a simple clip threshold of 10, a minimum adapter length of 10, while keeping both reads. Leading and trailing bases of each read were removed if the quality was below 5. Sliding window trimming was performed using a window size of 4 and a required average quality of 15. Quality-trimmed reads shorter than 50 bases were removed. The following script executed trimmomatic as specified above, for 20 samples in parallel:

```
trim.fastq.sh -s samples.txt -a illumina.truseq.indexing.adaptors -r $rawreads -x '_R1.fastq.gz,_R2.fastq.gz' -t 20
```

The quality of raw and trimmed reads was assessed with FastQC version 0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc).

Executed command-line scripts and parameter choices for steps 1-7 and iterations 1-2

The following tables and sections track the executed pipeline scripts and chosen parameters at each analysis step. For clarity, executed scripts are only listed once despite having been executed for several taxon sets and iterations. Parameter choices that varied between taxon sets or iterations are denoted as $X and are specified in the respective tables. Pipeline scripts are further documented on the README.md and wiki pages on the accompanying GitHub page (https://github.com/scrameri/CaptureAl).

Var	Parameter	Iteration 1	Iteration 2
$1	Reference sequences	orig.fragments.merged100_6555.fasta	consFabaceae_4c_2468.fasta
$2	Sample / Group file	mapfile.fabaceae.12.txt	mapfile.fabaceae.txt
$3	Min. frac. regions	0.2	0.2
$4	Min. frac. taxa	0.3	0.7
$5	Min. aln. length	1	1
$6	Min. avg. coverage	6	8
$7	Max. avg. coverage	1000	1000
$8	Min. aln. fraction	0	0
$9	Min. frac. conforms	0.3	0.7
$10	Min. normalized score	2	1
$11	Min. frac. regions	0.2	0.2
$12	Min. frac. taxa	0.5	0.75
$13	Max. nb. contigs	2	2
$14	Min. normalized score	2	2
$15	Min. aln. length	80	80
$16	Min. aln. fraction	0	0
$17	Min. score	1	1
$18	Min. contig length	1	1
$19	Min. frac. conforms	0.5	0.5
$20	Min. merging score	0.85	0.85
$21	Max. gap ratio	-	0.35
$22	Max. nucleotide diversity	-	0.35

Parameter choices for Subfamily set.

Var	Parameter	Iteration 1	Iteration 2
$1	Reference sequences	orig.fragments.merged100_6555.fasta	consDalbergia_4c_3736.fasta
$2	Sample / Group file	mapfile.dalbergia.12.txt	mapfile.dalbergia.txt
$3	Min. frac. regions	0.2	0.2
$4	Min. frac. taxa	0.4	0.7
$5	Min. aln. length	1	1
$6	Min. avg. coverage	8	8
$7	Max. avg. coverage	1000	1000
$8	Min. aln. fraction	0	0
$9	Min. frac. conforms	0.4	0.7
$10	Min. normalized score	2	2
$11	Min. frac. regions	0.2	0.7
$12	Min. frac. taxa	0.5	0.85
$13	Max. nb. contigs	2	2
$14	Min. normalized score	2	2
$15	Min. aln. length	80	80
$16	Min. aln. fraction	0	0
$17	Min. score	1	1
$18	Min. contig length	1	1
$19	Min. frac. conforms	0.5	0.7
$20	Min. merging score	0.9	0.95
$21	Max. gap ratio	-	0.3
$22	Max. nucleotide diversity	-	0.15

Parameter choices for Species set.

Step 1: Read mapping

We ran BWA-MEM in the $mappingdir directory, using the quality-trimmed and quality-filtered reads with the indicated file extensions (-e option) located in the $trimmedreads directory (-d option), and the respective reference sequences for each taxon set and iteration (-r option). The script outputs reads with a minimum alignment score of 10 (-T option), marks secondary hits, and only retains reads with a minimum mapping quality of 10 (-Q option) in the final SAM files before compressing them to BAM format. Computations were performed for all samples specified in samples.txt (-s option) using 4 times 5 threads in parallel (-t option) as follows:

```
run.bwamem.sh -s samples.txt -r $1 -e .trim1.fastq.gz,.trim2.fastq.gz -T 10 -Q 10 -d $trimmedreads -t 4
```

We performed coverage analysis on the BAM files filtered for mapping quality equal or above 10 (-Q option) for each sample, and wrote all coverage results to one file as follows:

```
get.coverage.stats.sh -s samples.txt -Q 10 -t 20
collect.coverage.stats.R samples.txt 10
```

We implemented seven filtering criteria to identify target regions with adequate average coverage across the taxon groups specified in $2. The first two filters take absolute thresholds and aim to remove poorly sequenced samples or target regions: $3 minimum fraction of regions with at least one mapped read in a sample (filters samples), $4 minimum fraction of samples with at least one mapped read in a region (filters target regions). The next four filters take thresholds that need to be met in a specified fraction of samples in each considered taxon group: $5 minimum BWA-MEM alignment length, $6 minimum average coverage in the aligned region, $7 maximum average coverage in the aligned region, $8 minimum alignment fraction (BWA-MEM alignment length divided by target region length). $9 is the minimum fraction of samples in each taxon group that need to pass each filter in order to keep a certain target region.

```
filter.visual.coverages.R $2 coverage_stats.txt $1 $3 $4 $5 $6 $7 $8 $9
```

This script visualized the coverage statistics as violin plots and heatmaps, and saved a list of kept samples (samples.txt) as well as a list of kept regions ($regions) for sequence assembly.

Step 2: Sequence assembly

We extracted read pairs from quality-filtered and quality-trimmed reads located in the $trimmedreads directory (-d option). This step was carried out on a local scratch ($extractedreads directory) using 20 parallel threads (-t option). At least one of the two reads per extracted read pair mapped to a retained target region with a minimum mapping quality of 10 (-Q option):

```
extract.readpairs.sh -s samples.txt -l $regions -d $trimmedreads -m $mappingdir -Q 10 -t 20
```

We assembled the extracted reads located in the $extractedreads directory (-r option) into consensus contigs (contigs hereafter) separately for each sample and retained region using dipSPAdes in ‘assembly-only‘ and ‘careful’ mode, with an automatic coverage cutoff. This

step was carried out on a local scratch ($assemblies directory) using 20 parallel threads (-t option):

```
run.dipspades.sh -s samples.txt -r $extractedreads -t 20
```

Step 3: Orthology assessment

We ran exonerate for each sample and each retained target region (-l option) with the 'affine:local' and 'exhaustive' options, using the contigs located in the $assemblies directory (-d option) as query sequences and the target regions (-r option) as target sequences. We stored alignment statistics of all consensus contigs that aligned to the same target region in the $exonerate directory (-d option), but limited the report to the best alignment per contig as follows:

```
select.best.contigs.per.locus.sh -s samples.txt -l $regions -r $1 -d $assemblies -t 20
```

Contigs with a target alignment length of at least the specified threshold (-a option) and a normalized alignment score (defined as the raw exonerate alignment score divided by the target alignment length) of at least the specified threshold (-c option) were considered as potentially homologous and retained. If more than one contig met these requirements, and if none of these contigs physically overlapped based on the alignment statistics, the best-matching contig was combined with the additional contig(s) using an appropriate spacer and by taking the directionality into account as follows:

```
combine.contigs.parallel.sh -s samples.txt -d $exonerate -a $5 -c $10 -t 20
```

We collected the exonerate statistics of each sample and plotted the number of contigs per target region for the different taxon groups as follows:

```
collect.exonerate.stats.R samples.txt $exonerate
plot.contig.numbers.R regions_contignumbers.txt $2
```

Step 4: sample and region filtering

We implemented nine filtering criteria to identify target regions with adequate assembled across the taxon groups specified in $2. The first two filters take absolute thresholds and aim to remove poorly assembled samples or target regions: $11 minimum fraction of regions with at least one contig in a sample (filters samples), $12 minimum fraction of

samples with at least one contig in a region (filters target regions). The next five filters take thresholds that need to be met in a specified fraction of samples in each considered taxon group: $13 maximum number of non-zero (fragments combined) contigs in a target region, $14 minimum normalized exonerate alignment score, $15 minimum exonerate alignment length, $16 minimum alignment fraction (exonerate alignment length divided by target region length), $17 minimum raw exonerate alignment score, $18 minimum contig length. $19 is the minimum fraction of samples in each taxon group that need to pass each filter in order to keep a certain target region.

```
filter.visual.assemblies.R $2 loci_stats.txt $1 $11 $12 $13 $14 $15 $16 $17 $18 $19
```

Step 5: Target region alignment and alignment trimming

We generated multifasta files in the $multifasta directory for all retained target regions, containing all retained contigs and samples as follows:

```
taxa=taxa_kept-$11.txt
regions=regions_kept-$11-$12-$13-$14-$15-$19.txt
create.multifastas.parallel.sh -s $taxa -l $regions -d $exonerate -t 20
```

We generated mafft alignments in the $mafft directory using the 'localpair' and 'adjustdirection' flags, using 1000 maximum iterations:

```
align.multifastas.parallel.sh -d $multifasta -m 'localpair' -t 20
```

Raw alignments were trimmed at both ends until an alignment site had nucleotides in at least 50% of aligned sequences (-c option) and a maximum nucleotide diversity (i.e., the sum of the number of base differences between sequence pairs divided by the number of comparisons) of 0.25 (-n option). The -v flag triggered visualisation of the alignment end trimming procedure. Trimmed alignments were written to the $endtrimmed directory as follows:

```
trim.alignment.ends.parallel.sh -s $2 -d $mafft -c 0.5 -n 0.25 -t 20 -v
```

Internal trimming was carried out by first removing any alignment site with nucleotides in less than 40% of aligned sequences (-c option). Potential mis-assemblies or mis-alignments in each sequence were resolved using a sliding window approach with window size 20 (-z option) and step size 1 (-S option). Specifically, we trimmed windows at contig ends if

more than 50% of the nucleotides in the conserved part of the window deviated from the alignment consensus (-n option). The script defines a conserved part of each window as the alignment sites with nucleotides in at least 20% of samples, and where the frequencies of minor alleles are all below 30% without considering gaps. After window-based trimming, the script also removes sites with sequence data for less than the specified fraction of aligned sequences (-c option) again. The -v flag triggered visualisation of the internal trimming procedure with sliding window approach. Trimmed alignments were written to the $trimmed directory as follows:

```
trim.alignments.parallel.sh -s $2 -d $endtrimmed -c 0.4 -z 20 -n 0.5 -S 1 -t 20 -v
```

Step 6: Merge overlapping alignments

We calculated a consensus sequence for each end-trimmed and internally trimmed alignment located in the directory $trimmed, using a minimum allele frequency of 1 (-m option) to call IUPAC ambiguity and a minimum base frequency of 0.01 (-b option) to return a consensus instead of a gap. This parameter combination ensured that the most frequent allele was called at each alignment site rather than IUPAC ambiguity codes or gaps. If two alleles were equally frequent at any alignment site, one was randomly sampled to represent the consensus. The -g flag ensured that gaps were removed from the final consensus sequence, and the -n flag ensured that completely ambiguous consensus bases (Ns) were removed from the final consensus sequence. The -v flag triggered visualisation of the consensus calculation:

```
get.consensus.from.alignment.parallel.sh -s $taxa -d $trimmed -m 1 -b 0.01 -t 20 -gnv
```

We renamed the sequence names of alignment consensus sequences stored in $cons to dispose of the suffix added during alignment and trimming before identifying the best non-reciprocal Blast hits between alignment consensus sequences as follows:

```
rename.fasta.headers.R $cons ".all.aln.etr.itr" FALSE FALSE
blast.vs.self.sh $cons
```

We then identified Blast hits at alignment ends and stored a list of physically overlapping alignments (names of overlapping target regions on the same line) as follows:

```
cbase=$(basename $cons .fasta)
find.overlapping.alignments.R $cbase.vs.self.blast.filtered
```

We then aligned all contigs of up to five physically overlapping target regions using the same alignment algorithm as before. Merged alignments were written to the $merged directory as follows (visualisation was triggered by default):

```
overlaps=$cbase.list
align.overlapping.contigs.sh -l $overlaps -c $multifasta -m 'localpair' -t 20
```

In cases where contigs of the same sample overlapped with a mismatch, only the base with higher frequency at that alignment site was considered. A success score of each merging procedure was computed based on the number of mismatches in overlapping contigs of the same individuals relative to the total number of bases in the alignment. The success score amounted to 1 if there were no mismatches in any individual. We discarded any merged alignment with a score smaller than 0.85 (subfamily set) or 0.9 (species set).

```
filter.merged.alignments.sh -d $merged -s $20
```

Unsuccessfully merged alignment sets were visually inspected to identify whether some subsets of alignments sufficiently overlapped to allow for merging. The manually selected alignments were merged again and combined with the automatically merged alignments if they showed a sufficient success score. All successfully merged alignments were then trimmed as before and used as replacements for overlapping alignments.

```
trim.alignment.ends.parallel.sh -s $s -d $merged -c 0.5 -n 0.25 -t 20 -v
trim.alignments.parallel.sh -s $s -d $merged -c 0.4 -z 20 -n 0.5 -S 1 -t 20 -v
replace.overlapping.alignments.R $trimmed $merged $overlaps
```

Step 7: Create representative reference sequences

Sets of reference consensus sequences for different taxon groups were generated, combined, aligned, and a group consensus was derived as follows:

```
get.group.consensus.sh -s $2 -d $trimmed -m 1 -b 0.01 -z ".all.aln.etr.itr.cons" -t 20 -gnv
```

The resulting FASTA file $newref was renamed according to the taxon set and number of remaining target regions:

```
rename.fasta.headers.R $newref ".cons.aln" FALSE FALSE
mv $newref cons<TAXON SET NAME>_<NB. of TAXON GROUPS>c_<NB. of REGIONS>.fasta
```

Phylogenetic analyses

Supermatrix (concatenation) approach

We assessed alignments as follows:

```
assess.alignments.parallel.sh -f $trimmed -t 20
```

We filtered out alignments with excessive gap ratio or nucleotide diversity as follows:

```
plot.assessed.alignments.R $trimmed.assess.txt $21 FALSE $22 100
```

We concatenated alignments as follows:

```
concatenate.fastas.R $trimmed
```

We ran maximum likelihood search on the concatenated alignments as follows:

```
raxmlHPC-PTHREADS-SSE3 -f a -m GTRCAT -x 85397 -p 24686 -s $trimmed.phy
-n $trimmed.catBS100.nex -T 20 -N 100 > catBS100.log 2> catBS100.err
```

Gene tree summary approach

We generated gene trees for different alignments as follows:

```
get.gene.trees.parallel.sh -d $trimmed -n 100 -t 20
```

We collapsed branches with low support (below 10) as follows:

```
nw_ed $trimmed.genetrees 'i & b<=10' o > $trimmed.BS10.genetrees
```

We inferred the species tree using ASTRAL as follows:

```
java -Xmx1G -jar astral.5.6.3.jar -i $trimmed.BS10.genetrees -o $trimmed.BS10.single.spectree -t 2 2> $trimmed.BS10.single.log
```

We added 0.5 coalescent units to each zero terminal branch length as follows:

```
add.to.terminal.branches.R $trimmed.BS10.single.spectree 0.5
```

Popluation genetics analyses

We removed PCR duplicates and capped BAM files as follows:

```
bsub < remove.dups.and.cap.lsf
```

We collected a table of percentage of PCR duplicates as follows:

```
grep '^LIBRARY' -A1 ${folder}/stats_dup/*txt --no-group | awk '!seen[$3]++' | cut -f1,9 | cut -f3 -d'/' | sed -e 's/.dupstats.txt-Unknown Library//' > percdup.txt
```

We called SNPs on the EULER cluster as follows:

```
module load gcc/4.8.2 gdc python/2.7.11 perl/5.18.4 samtools/1.3

folder="Chapter1.3_mapsnp-2396_51.nodup.cov500"
ref="consDalbergia_4c_2396.fasta"
maxlen=10000
lenperjob=10000

## Split jobs by regions
samtools faidx ${ref}
fasta_generate_regions.py ${ref} ${maxlen} > regions.txt
split.freebayes.regions.file.pl regions.txt ${lenperjob}
mkdir regions
mv regions_${lenperjob}_* regions

## Create output dir
mkdir vcfs
ls -1 ${folder}/*.bam > bamlist.txt
```

```
## Submit job
bsub < submit.multi.freebayes.cmds.lsf

## Combine .vcf files
combine.vcf.files.sh vcfs
```

We filtered raw SNPs (vcfs.vcf) as follows:

```
filter.snps.sh -v vcfs.vcf -r $ref -n 'Dalbergia_CH1.3_51_'
```

We created a genind object from the filtered SNPs for analyses using the R adegenet package as follows:

```
vcf=Dalbergia_CH1.3_51_116500.filtered.vcf
grep '^>' $ref | cut -f2 -d'>' > regions
subset.vcf.by.region.sh -v $vcf -r regions -t 20
vcf2adegenet.R $(basename $vcf .vcf) $ref 1> get_gi.log 2> get_gi.err
```

References

Cardoso, D., L. P. de Queiroz, R. T. Pennington, H. C. de Lima, E. Fonty, M. F. Wojciechowski, and M. Lavin. 2012. Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. American Journal of Botany **99**:1991-2013.

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

Hassold, S., P. P. Lowry, M. R. Bauert, A. Razafintsalama, L. Ramamonjisoa, and A. Widmer. 2016. DNA barcoding of Malagasy rosewoods: towards a molecular identification of CITES-listed *Dalbergia* species. PLoS ONE: doi:10.1371-journal.pone.0157881.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources **15**:1179-1191.

LPWG. 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. Taxon **66**:44-77.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal **17**:10-12.

Smeds, L., and A. Künstner. 2011. ConDeTri - a content dependent read trimmer for Illumina data. PLoS One **6**:e26314.

Tennekes, M. 2018. tmap: Thematic Maps in R. Journal of Statistical Software **84**:1-39.

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.

Wojciechowski, M. F., M. Lavin, and M. J. Sanderson. 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**:1846-1862.

Chapter II

Species discovery, phylogenomics and integrative species delimitation in *Dalbergia* (Fabaceae) precious woods of Madagascar and the Comoros

Unpublished manuscript co-authored by

Simon Crameri[1], Peter B. Phillipson[2,3], Nicholas Wilding[2,3], Tendro Radanielina[4], Muséum National d'Histoire Naturelle [5], Porter P. Lowry II[2,3], Alex Widmer[1]

[1] ETH Zurich, Institute of Integrative Biology, Universitätstrasse 16, Zürich, Switzerland.
[2] Missouri Botanical Garden, 4344 Shaw Blvd., St. Louis, MO, 63110
[3] Institut de Systématique, Évolution et Biodiversité (ISYEB), Muséum national d'Histoire naturelle (MNHN), Centre National de la Recherche Scientifique, Sorbonne Université, École Pratique des Hautes Études, Université des Antilles, C.P. 39, rue Cuvier 57, 75005 Paris, France.
[4] Département de Biologie et Écologie Végétales, Université d'Antananarivo, Antananarivo, Madagascar.
[5] Muséum National d'Histoire Naturelle, 57 rue Cuvier, 75005 Paris, France.

Author contributions

SC and AW designed the study. SC, PBP and NW carried out taxonomic work. MNHN and PPL provided herbarium material for sequencing. SC analysed the data. SC and AW wrote the manuscript with contributions from NW and PPL.

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.

Abstract

Evolutionary species are often challenging to discover and distinguish from other such species if few stable and diagnostic morphological characters are available. Such a lack of information often translates to taxonomic uncertainties and slows down the setting of conservation priorities, in particular in biodiversity hotspots. The pantropical and taxonomically complex genus *Dalbergia* (Fabaceae) contains numerous sought-after and endangered precious timber species known as rosewoods and palisanders, many of which are frequently confused and likely insufficiently researched. We applied a species discovery and integrative species delimitation approach to address unresolved taxonomic issues in *Dalbergia* species from Madagascar and the Comoros. We used a recently developed set of target enrichment baits to collect sequencing data for 2,396 nuclear genomic regions in 719 individuals. These included all 49 currently accepted taxonomic species, many recent collections from Madagascar, as well as several species from other regions. Species discovery based on principal component and neighbour-joining tree analyses revealed the presence of 94 putative species in Madagascar and one in the Comoros. We then assessed lineage separation by integrating leaf morphometric and ecogeographic data, and by considering genetic differentiation statistics as well as the morphology and phenology of fertile structures. We could confirm the distinction of 46 described species, inferred 3 potential synonyms, 31 unconfirmed candidate species pending further study, and 14 confirmed new candidate species. Phylogenomic analyses confirmed the existence of two diverse Malagasy clades (supergroups) and a biogeographic connection between Madagascar and East Africa. Analyses further revealed an isolated Malagasy lineage, 12 highly supported Malagasy subgroups for infrageneric classification, and a biogeographic connection between the Comoros and West Madagascar. As regions of highest species richness and phylogenetic diversity we identified Northern Madagascar and the area in which the Ankarafantsika national park is located. We attribute this finding to the coincidence of several ecotones in these areas. Our work highlights the value of genome-wide DNA sequence data and integrative approaches to inform taxonomic work and conservation assessments.

Keywords — *Dalbergia*, Madagascar, species delimitation, phylogeny, phylogenomics

Introduction

Species are widely recognised as fundamental units in ecology and evolution (MACE 2004, DE QUEIROZ 2005a). Although there are disputes regarding the ontological status of the species category and its value for conservation efforts (REYDON 2019), species are a defining component of biodiversity (GASTON AND SPICER 2013) and in practice the most commonly used units on lists of threatened, protected or regulated taxa. However, there exists a conceptual difference between evolutionary and taxonomic species (GHISELIN 2001, ZACHOS 2016). An evolutionary species is "a single lineage of ancestral descendant populations of organisms which maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate" (WILEY 1978, p. 18). As such, evolutionary species, or subsets thereof, are often the units of conservation concern (RYDER 1986, CRANDALL *et al.* 2000). Taxonomic species are entities with binomial names, which have been circumscribed by taxonomists based on the best data available at a particular time. Taxonomic species are often used as the counting unit in biodiversity studies and conservation efforts because ideally, they correspond to evolutionary species. However, this is often not the case, especially in insufficiently researched groups of taxonomically complex organisms (pers. obs.). For an effective conservation prioritisation, it is therefore important to assess the diversity and distribution of evolutionary species and to evaluate their correspondence with taxonomic species.

Integrative taxonomy is the combination of complementary sources of information (morphological, ecological or behavioural divergence, population genetics and phylogeography, among others) to discover, delimit and describe species by studying speciation processes (DAYRAT 2005, PADIAL *et al.* 2010, SCHLICK-STEINER *et al.* 2010, CARSTENS *et al.* 2013). This approach is especially promising where poor differentiation in the available morphological characters hinders the formulation of sound primary species hypotheses, and has therefore seen an increasing use in taxonomically complex groups (VIEITES *et al.* 2009, WACHTER *et al.* 2015, PRATA *et al.* 2018, YOUNGER *et al.* 2018). Populations can hereby be regarded as the smallest units of species discovery and the starting point for delimitation hypotheses (CARSTENS *et al.* 2013). The integrated data and analyses can then serve to characterise secondary defining properties of species, which are not necessary criteria but relevant to assess lineage separation (DE QUEIROZ 2005b). The unified species concept views segments of separately evolving lineages as "the only necessary property of species" (DE QUEIROZ 2005b, 2007) and thus provides a conceptual basis for integrative taxonomy, which is needed to assess lineage separation (SCHLICK-STEINER *et al.* 2010).

Advancements in next-generation sequencing have facilitated the integration of population genetics, phylogenetics and phylogeography into species delimitation (CARSTENS *et al.* 2013), and have mediated a paradigm shift, the transition from analysing few – mostly plastid – markers for phylogeny reconstruction to genome-wide analyses of nuclear and plastid DNA variation. Target enrichment sequencing is a powerful approach to such genome-wide analyses because it can provide phylogenetic signals that are informative at various evolutionary time scales, including those relevant for speciation (JONES AND GOOD 2016, CRAMERI *et al.* in prep.-b). Numerous studies have benefitted from target enrichment sequencing, sometimes in combination with other techniques, for species delimitation in vertebrates (SONG *et al.* 2017, ZARZA *et al.* 2017, MUSHER AND CRACRAFT 2018, PIE *et al.* 2019, NATUSCH *et al.* 2020), non-vertebrate animals (ERICKSON *et al.* 2020, GUEUNING *et al.* 2020), and more recently also in plants (KARBSTEIN *et al.* 2020). However, genetic divergence alone is often not sufficient to decide whether two inferred populations or reciprocally monophyletic clades represent separately evolving lineages, because different evolutionary processes and time scales may lead to comparable amounts of divergence, with or without involving speciation (HEY AND PINHO 2012). The integration of ecogeography can reveal speciation processes such as strong reproductive isolation, for instance when two morphologically similar yet genetically distinct entities, also known as cryptic species (BICKFORD *et al.* 2007), co-occur in the same region and habitat.

The pantropical plant genus *Dalbergia* L.f. (Fabaceae) represents a taxonomically complex and species-rich group of mainly shrubs and trees that contains many high-quality timber species, which are in high demand in the often illegal international trade and therefore of major conservation concern (BARRETT *et al.* 2010, CITES 2019, WAEBER *et al.* 2019). Species delimitation in *Dalbergia* is challenging because of the large diversity and a prevalent reliance on inflorescence, flower, and fruit characters to identify and distinguish the different species (e.g., PRAIN 1904, BOSSER & RABEVOHITRA 1996, DE CARVALHO 1997, ADEMA *et al.* 2016). However, flower and fruit characteristics are often not observable on the same individual at a given point in time, or altogether absent on living specimens encountered in the field or in museum collections. This has resulted in large numbers of unidentified herbarium specimens and a high proportion of attempted identifications being incorrect (pers. obs.), which further complicates and slows down the definition and prioritisation of relevant units for conservation. Even when flowers and fruits are available, there exist numerous described species with apparently little differentiation in these characters (BOSSER & RABEVOHITRA 1996, 2002, 2005). For these reasons, conservation efforts in the genus *Dalbergia* require support from research

developing an integrative taxonomic approach. This is especially true in Madagascar, where the diversity is particularly high, as is the pressure on the increasingly few remaining populations as a result of ongoing deforestation (VIEILLEDENT *et al.* 2018) and the illegal timber trade (SCHUURMAN AND LOWRY II 2009, WAEBER *et al.* 2019). The integration of target enrichment sequencing data has become feasible due to a recently developed set of enrichment probes targeting 2,396 nuclear genomic regions, which have been specifically designed and validated for *Dalbergia* species (CRAMERI *et al.* in prep.-b).

The aim of this study was to provide a phylogenetic backbone of the genus *Dalbergia* in Madagascar, to develop an integrative taxonomic approach based on a comprehensive sampling of Malagasy *Dalbergia*, and to review existing hypotheses on species and units of conservation concern. Specifically, we asked (i) what are the phylogenetic relationships among *Dalbergia* species from Madagascar, and what are the biogeographic connections of Malagasy taxa with those from other regions?, (ii) how many candidate species can be discovered in Madagascar based on multi-locus nuclear genetic data?, (iii) which candidate species are coherent and distinct from other such species based on integration of leaf morphology and ecogeography?, (iv) do current taxonomic circumscriptions of Malagasy *Dalbergia* species correspond to the inferred candidate species?, (v) how is the ecogeographic diversity distributed across different phylogenetic groups?, and (vi) where are the hotspots of species richness and phylogenetic diversity, and are these included in Madagascar's protected areas network?

To address these questions, we first collected one to several DNA samples from all 49 currently accepted *Dalbergia* species from Madagascar and the Comoros, supplemented this taxon sampling with recently collected material from Madagascar, and included several species from continental Africa, the Americas and Asia for a broader phylogenetic and phylogeographic perspective. We then performed species discovery based on multivariate analysis of 2,396 targeted nuclear genetic loci, and evaluated the discovered units against leaf morphological characters, ecogeographic distribution patterns and the current taxonomy to formulate revised species hypotheses and direct taxonomic prioritisations. Finally, we performed random forest classification and discriminant analysis to assess the discriminatory power and potential taxonomic relevance of leaf morphological and ecogeographic traits, and used the revised species hypotheses to identify regions of high species richness and phylogenetic diversity to target conservation efforts.

Material and Methods

Plant material

Sampling of plant material aimed at integrating all currently accepted *Dalbergia* species of Madagascar, representatives of the genus from other areas of its pantropical distribution range, as well as samples of closely related genera and suitable outgroup taxa as confirmed by RIBEIRO *et al.* (2007) and CRAMERI *et al.* (in prep.-b). In addition to accepted species, we also sampled collections from Madagascar with doubtful identifications, and collections that could not be clearly assigned to any of the currently accepted species. In total, we included 1,089 individuals in our study (Table S1). Of these, 1,082 belong to the genus *Dalbergia*, three belong to *Machaerium* Pers., two belong to *Aeschynomene* sect. *Ochopodium* Vogel, and two belong to *Aeschynomene* sect. *Aeschynomene* L., which was selected as the outgroup. The *Dalbergia* ingroup consisted of 1,053 individuals from Madagascar and the Comoros, 13 individuals of twelve species from continental Africa, nine individuals of nine species from the Americas, and seven individuals of four species from Asia. Most plant material from Madagascar was collected during fieldwork carried out between 2010 and 2019. A dedicated sampling protocol was applied wherever possible (see HASSOLD *et al.* 2018), which included the collection of silica gel dried leaf tissue for DNA analysis, as well as herbarium vouchers and additional detached leaves for morphometric analysis.

In total, we used voucher specimens of 708 individuals of *Dalbergia* from Madagascar to collect morphometric data, and we included DNA sequence data from 719 individuals (810 samples including replicates), of which 683 individuals (767 samples) originated from Madagascar and the Comoros. New sequence data were collected for 617 individuals (699 samples) and combined with data from 102 individuals (111 samples) from an earlier study (CRAMERI *et al.* in prep.-b). For 338 individuals both morphometric and molecular data were available (Table S1). To establish links between recent collections and collections from before 2006, which are the foundation of the current taxonomy (BOSSER & RABEVOHITRA 1996, 2002, 2005), we sequenced DNA of 195 specimens from the herbaria or DNA banks of the Muséum national d'Histoire naturelle (P, Paris, France, n = 104), the Missouri Botanical Garden (MO, St. Louis, USA, n = 45), the Conservatoire et jardin botaniques de la ville de Genève (G, Geneva, Switzerland, n = 29), the United Herbaria of the University and ETH Zurich (Z+ZT, Zurich, Switzerland, n = 8), and the Herbier du Parc Botanique et Zoologique de Tsimbazaza (TAN, Antananarivo, Madagascar, n = 7), as well as from the living collections of the Masoala rainforest at

Zurich Zoo (n = 2). The historic collections dated back to 1922 and included 20 nomenclatural type specimens of *Dalbergia* species endemic to Madagascar.

DNA extraction, library preparation, sequencing and variant calling

DNA extraction and library preparation were carried out as outlined in CRAMERI *et al.* (in prep.-b). Up to six individually indexed genomic DNA libraries were pooled for in-solution hybridisation and target capture enrichment using 12,049 tiled RNA baits. Because of long time intervals between reception of samples, sequencing was carried out on separate lanes on an Illumina MiSeq (2×300 bp paired-end sequencing, 65 libraries) at the Genetic Diversity Centre (GDC) Zurich, on an Illumina HiSeq 4000 (2×150 bp paired-end sequencing, 361 libraries) at the Functional Genomics Center Zurich (FGCZ) or Fasteris SA (Plan-les-Ouates, Switzerland), or on an Illumina NovaSeq 6000 SP flow cell (2×150 bp paired-end sequencing, 273 libraries) at the FGCZ. We repeated DNA extraction, target enrichment and sequencing for 46 samples of different ages, from silica gel dried or herbarium voucher material, and sequenced these on the same or a different sequencing platform. This served to assess the reproducibility of target enrichment and sequencing across tissue ages, tissue types and sequencing platforms. Whenever possible, we repeated DNA extraction, target enrichment and sequencing for poorly sequenced samples, as well as for a few samples showing signs of possible cross-contaminations.

Raw sequence reads were trimmed and quality-filtered as in CRAMERI *et al.* (in prep.-b). We used multivariate and population genetic approaches on single nucleotide polymorphisms (SNPs) to gain insight into genetic diversity, differentiation, and evolutionary relationships. For variant (SNP) calling, we mapped the quality-filtered sequence reads to available reference sequences of 2,396 genomic regions using BWA-MEM (LI 2013). These target regions were revealed to be consistently recovered with high sensitivity and specificity throughout most *Dalbergia* species analysed in CRAMERI *et al.* (in prep.-b). We then removed PCR duplicates and performed variant (SNP) calling and filtering as outlined in detail in CRAMERI *et al.* (in prep.-b). We then used *vcfR* version 1.10.0 (KNAUS & GRÜNWALD 2017) and *adegenet* version 2.1.1. (JOMBART 2008, JOMBART & AHMED 2011) in *R* version 3.6.3 (R CORE TEAM 2020) to generate *genind* and *genlight* objects that represented the allele table of filtered variants with associated metadata such as individual missingness, species identification and sampling location.

To check for sequencing quality and possible laboratory errors, we first performed exploratory data analysis on all 810 sequenced samples using PCA on the centred allele covariance matrix and pairwise Nei's genetic distances (NEI 1972) on an individual basis for the estimation of neighbour-joining (NJ) trees (SAITOU AND NEI 1987). This led to the

exclusion of 72 (8.9%) poorly sequenced samples (< 1000 target regions with at least 10× average coverage, or > 50% SNP missingness in regular samples and > 75% in type specimen samples) representing 58 (8.1%) individuals, 2% of silica gel dried samples and 22% of herbarium samples, as well as five (0.6%) samples with suspected cross-sample contaminations (see Supplementary Methods for details). Repetitions of DNA extraction, library preparation and sequencing resulted in data recovery for 25 (40%) excluded individuals, which left 733 samples (694 from Madagascar, two from Mozambique, one from the Comoros and 36 samples from elsewhere) representing 681 individuals in the final SNP dataset.

Reproducibility of target enrichment and sequencing

To assess the reproducibility of target enrichment and sequencing, we computed the fraction of identically called variants (identity hereafter) of each pair of replicates, as well as for randomly drawn pairs of the same putative species (see integrative species delimitation) and tissue type (silica gel dried or herbarium specimen) for comparison. We determined the most informative predictors of identity by fitting a linear regression model in *R*, using pair type (conspecifics or replicates), tissue type (silica gel dried or herbarium voucher), a binary variable coding for pairwise difference in sequencing platform, specimen age (mean age for conspecific pairs) and the minimum, maximum and pairwise differences in the number of target regions with at least 10× average coverage as predictors. We performed residual analysis and used first-aid transformations (i.e., arcsin-square root for identity, logarithm for age and square root for numbers of high-coverage target regions) on the predictor and response variables, and second-level interaction terms between pair type, tissue type, difference in sequencing platform and specimen age to achieve a better model fit, and performed step-wise backward model selection using the *step* function. Finally, we visualised the dependence of identity between replicate or conspecific pairs using *ggplot2* (WICKHAM 2016).

Morphometric analyses

Five hundred and nine (72%) of the 708 available voucher specimens were sterile. We hence limited our morphometric analyses to the following leaf characters: shape, colour, size and pubescence of different leaf organs under dry (herbarium voucher) condition. Leaflet shape is often described by terms such as ovate, elliptic or oblong. However, the variation in leaflet shape observed within and among *Dalbergia* species is large, and standard terminology fails to always adequately describe such continuous variation in

shape. We therefore quantified leaflet shape with outline analysis using the *R* package *Momocs* version 1.2.9 (BONHOMME *et al.* 2014). This package allows for the importation of binarised images of closed outlines as outline coordinates, as well as performing elliptical Fourier analysis (KUHL AND GIARDINA 1982) on these outlines, which results in a set of harmonic coefficients that quantitatively describe individual shapes. Images of several leaflets per individual considered to be representative of the observed variation were taken under suitable lighting conditions using a high-resolution scanner (EPSON Expression 11000XL) or a camera (SONY RX10 III) mounted on a tripod. We used white herbarium sheets as background for maximum contrast, arranged individual leaflets with sufficient spacing from each other using a standardised direction, and placed a scale on each image. Streamlined image binarisation and segmentation were performed using the *R* packages *raster* version 3.0.2 (HIJMANS AND VAN ETTEN 2012) and *EBImage* version 4.28.1 (PAU *et al.* 2010). We wrote the wrapper function *extract.segments* (available at https://github.com/scrameri/DalbergiaSpeciesDelimitation), which allows for streamlined analysis of hundreds to thousands of images in sequence or in parallel. The function carries out image reading, cropping, adjustment of brightness and contrast, K-means clustering (HARTIGAN AND WONG 1979) of colours, global or adaptive thresholding for image binarisation (i.e., the definition of foreground and background using black and white colours, respectively), as well as image segmentation (i.e., the isolation of all segments of interest as different binary images) using one out of two methods (*bwlabel* for non-overlapping objects and *watershed* for touching or overlapping objects) implemented in *EBImage*. Because our outlines of interest were contrasting and non-overlapping, we used the default K-means clustering method for binarisation and the *bwlabel* method for segmentation. Non-standard parameters included 0.3 for brightness and 1.25 for contrast adjustment, a segment size filtering using 0.01 and 0.5 as the minimum and maximum allowed segment sizes relative to the number of rows in the image array, and a removal of segments intersecting with the cropped image edges. We also saved the mean red, green and blue colour components of the foreground (i.e., leaflet), which can be extracted by the *extract.segments* function after adjusting brightness and contrast.

The leaflet shape analysis was based on the segmented binary images of leaflet outlines, which we imported as outline coordinates in *Momocs*. We visually inspected all outlines and manually excluded damaged leaflet outlines from downstream analyses. We also digitally removed petiolule fragments from the leaflet base, which were initially frequently present in leaflet outlines and were found to bias shape analysis in preliminary analyses. For this step we used the specifically developed *find.petiolule* function (available at https://github.com/scrameri/DalbergiaSpeciesDelimitation). We then centered and

aligned all remaining outlines to the same origin and orientation and used available *Momocs* functions to annotate each outline with its length, width, and perimeter (i.e., the total outline length) as additional leaflet variables, and converted pixels to millimetres using the scale bar. We then scaled the outlines with respect to centroid size and identified the optimal number of harmonics using *Momocs*' *calibrate_harmonicpower_efourier* function, which resulted in 21 harmonics. The leaflet outlines were then subjected to elliptical Fourier analysis (*efourier* function) using the optimal number of harmonics and no normalisation. We then calculated mean shapes (*MSHAPES* function) to obtain a representative shape per individual and species, and subjected the resulting coefficient matrix to principal component analysis (PCA). Leaflet shape variation along the first six principal components, which explained 98.5% of the total observed variance, was plotted using the *PCcontrib* function and visually assessed to identify five biologically meaningful principal components of leaflet shape (Figure S18A).

We further selected four ordinal, one discrete and two quantitative continuous characters to describe the pubescence, leaflet number and other leaf attributes. Pubescence was coded separately for the rachis, the petiolule, as well as the upper and lower leaflet lamina, using a three-level ordinal scoring system (1 = glabrous, 2 = scarcely pubescent, 3 = densely pubescent). Scoring of the sometimes scarcely discernible pubescence was performed using a hand lens (10× magnification) or binocular (LEICA MZ8, 6.3–50× magnification). The discrete number of leaflets per leaf was assessed on several leaves per individual. In cases where specimens showed disarticulated leaves, we counted the number of leaflet insertion scars on the rachis. The quantitative continuous variables leaf length, here defined as the combined length of petiole and rachis of mature leaves, and petiolule length were recorded on the basis of pictures of whole leaves or the whole herbarium specimens. We used the scale on each image and the segmented line tool implemented in *ImageJ* version 1.47 software (SCHNEIDER *et al.* 2012) to perform digital measurements. We also added the following four ratios between mean character values to the morphological data set: leaflet length relative to leaflet width, number of leaflets relative to leaf length, petiolule length relative to leaflet length, and petiolule length relative to leaf length. The leaf colour componets were reduced to the red component and a greenish component computed as green / (red + blue). This resulted in 21 leaf morphological characters (Table 1).

Repeated measurements per individual were summarised using the mean or the median. We assessed Spearman's rank correlations between all morphological variables (Figure S19) and excluded a variable with a correlation coefficient above 95% (first

principal component of leaflet shape) to another, kept variable. This resulted in the final morphological dataset consisting of 708 individuals and 20 variables.

TABLE 1: Twenty-one assessed leaf morphological characters. One shape component (marked with an *) was removed from analysis due to high correlation.

Code	Character	Type
Leaflet size		
LftLen	Median leaflet length from base to apex	Continuous
LftWid	Median leaflet width at widest point	Continuous
PetuLen	Median petolule length	Continuous
Leaflet shape		
Shp1*	Mean leaflet shape component 1 (overall shape ovate to orbicular vs. narrowly oblong)	Continuous
Shp2	Mean leaflet shape component 2 (rounded or retuse vs. acute or acuminate apex)	Continuous
Shp4	Mean leaflet shape component 4 (rounded apex / truncate base vs. obtuse apex / cuneate base)	Continuous
Shp5	Mean leaflet shape component 5 (tapering apex)	Continuous
Shp6	Mean leaflet shape component 6 (rounded apex / cuneate vs. obtuse apex / rounded base)	Continuous
LftPeri	Median leaflet perimeter	Continuous
Leaf		
LeafLen	Median leaf length (petiole + rachis)	Continuous
LeafNb	Median leaflet number	Discrete
Pubescence		
PubRac	Pubescence on rachis (1: glabrous, 2: scarce, 3: dense)	Ordinal
PubPetu	Pubescence on petiolule (1: glabrous, 2: scarce, 3: dense)	Ordinal
PubUp	Pubescence on upper leaflet lamina (1: glabrous, 2: scarce, 3: dense)	Ordinal
PubLow	Pubescence on lower leaflet lamina (1: glabrous, 2: scarce, 3: dense)	Ordinal
Colour		
ColRed	Red colour component	Continuous
ColGreenish	Greenish colour component: green / (red + blue)	Continuous
Ratios		
RatioLftLenWid	Mean leaflet length / leaflet width	Continuous
RatioNbLeafLen	Mean leaflet number / leaf length	Continuous
RatioPetuLftLen	Mean petiolule length / leaflet length	Continuous
RatioPetuLeafLen	Mean petiolule length / leaf length	Continuous

Ecological characteristics

Historic botanical collections from Madagascar can often be georeferenced *post facto* with relatively high precision using the Gazetteer to Malagasy Botanical Collecting Localities (http://www.mobot.org/mobot/research/madagascar/gazetteer). However, older *Dalbergia* collections often contain little or no habitat information. By contrast, recent collections are most often precisely georeferenced and contain useful ecological information but frequently lack information on the soil type. To consistently characterise the vegetation ecology of different putative species, we downloaded all records of Malagasy *Dalbergia* with available geographic coordinates from the Tropicos specimen database (https://tropicos.org, accessed on October 4, 2020), and identified adequately georeferenced collections. These had to meet one of the following criteria: (i) available GPS coordinates, or (ii) georeferenced to a clearly delimitable locality based on the field notes and a map such as Google Earth, or (iii) georeferenced *post facto* with high precision, often after consulting with the collectors. Inadequately georeferenced collections used in this study were reviewed and georeferenced if possible. We then extracted the corresponding ecological features from all adequately georeferenced Tropicos records using various ecological raster or polygon layers and the *raster* and *sp* (BIVAND *et al.* 2013) packages in *R*. The spatial layers included the seven terrestrial ecoregions of Madagascar following DINERSTEIN *et al.* (2017), altitude (provided by Conservation International), forest cover density in 2005 from (VIEILLEDENT *et al.* 2018), vegetation class following MOAT AND SMITH (2007), surface lithology (available for the African continent at https://servirglobal.net/Data-and-Maps), as well as the annual mean temperature (Bio1), isothermality (Bio3), temperature seasonality (Bio4), annual precipitation (Bio 12) and precipitation seasonality (Bio15) from the CHELSA version 1 Bioclim database (https://chelsa-climate.org/bioclim). We used the wrapper function *extract2* (available at https://github.com/scrameri/DalbergiaSpeciesDelimitation), which allows for flexibility regarding different spatial representations and coordinate reference systems, and which by default extracts the class of categorical raster or polygon layers (surface lithology, vegetation class, ecoregion) where a coordinate falls into, or a summary value interpolated from the values of the four nearest raster cells for numerical raster layers (altitude, forest cover density, CHELSA Bioclim variables). In cases where coordinates lie just outside of any mapped area, it determines the closest features within 200 meters. We combined two highly similar vegetation classes (Western dry forest with codes 5 and 10) and three classes of surface lithology (different classes of alluvium deposits coded as 14, 15 and 18), converted nominal factors to binary dummy variables and removed the following classes due to absence or low occurrence frequencies across the Tropicos records: the Madagascar

mangrove and ericoid ecoregions, degraded south western dry spiny forests, mangroves, as well as ultrabasic and hydric organic surface lithologies. This resulted in an ecological dataset of 33 variables (14 vegetation classes, seven classes of surface lithology, five ecoregions, five climate variables, one forest density variable and elevation), which we merged with the morphological dataset and the geographic coordinates to create a combined dataset of 55 variables for the 708 individuals.

Integrative species delimitation

We combined PCA and NJ trees on the multi-locus nuclear SNP allele table to perform species discovery (sensu CARSTENS *et al.* 2013), considered the discovered units as putative species (PS), and evaluated each PS against morphology and ecogeography as secondary species properties to assess lineage separation, which is consistent with an integrative taxonomic approach (DAYRAT 2005) and the unifying species concept (DE QUEIROZ 2005b, 2007). Specifically, we conducted PCA on the centred allele covariance matrix and calculated pairwise Nei's genetic distances on an individual basis for the estimation of NJ trees. For calculation of Nei's genetic distances, we used the *nei.dist* function implemented in *poppr* version 2.8.5 (KAMVAR *et al.* 2014) after imputing missing alleles with mean allele counts, and estimated NJ trees using the *nj* function from the *ape* package version 5.3 (PARADIS AND SCHLIEP 2018). The large number of analysed samples, SNPs and covariables of secondary species properties demanded the use of efficient computing and advanced graphical methods. We used *adegenet*'s *glPca* function on binary SNPs represented as *genlight* objects for PCA, and *ggplot2*, *ggrepel* (SLOWIKOWSKI 2020), and additional wrapper functions (available at https://github.com/scrameri/DalbergiaSpeciesDelimitation) for a graphical display of the first few principal components. This served to prevent label overlaps and to visualize relevant information as points of different size for SNP missingness and other continuous covariables, or points of different colour or shape for inferred PS and other categorical covariables. Because the informative historical collections often showed higher levels of missingness compared to recently collected and silica gel dried samples, we conducted PCA and NJ tree estimation with 0% allowed missingness in the considered samples at a deeper evolutionary scale, and 1% to 5% allowed missingness for analyses involving closely related PS or populations.

We started with a PCA including all *Dalbergia* samples from Madagascar and the Comoros (n = 695), as well as two samples of *D. bracteolata* Baker from Mozambique. A corresponding NJ tree was estimated on the same set of samples but including *Machaerium* spp., *Aeschynomene* sect. *Ochopodium* spp., as well as *Aeschynomene* sect. *Aeschynomene* spp. as the outgroup for rooting. We then inspected the first four principal components and

the major NJ tree branches to identify genetic clusters and lineages at a deep evolutionary scale, and assigned samples to the resulting major groups (supergroups hereafter). We then carried out nested PCA and NJ tree analyses within supergroups to identify genetic structure at a shallower evolutionary scale, assessed the resulting genetic clusters and lineages against morphology, ecogeography and syntopic occurrence, and continued the procedure of successive nested analyses until the remaining samples constituted genetically coherent units that were at the same time morphologically and ecogeographically coherent, but not necessarily distinct segments of putatively separately evolving lineages. These units represented the PS, which we subjected to integrative species delimitation.

We considered NJ trees, genetic differentiation statistics, the morphology of leaves and fertile structures, reproductive phenology, and ecogeography with special emphasis on syntopic occurrence to infer the strength of reproductive barriers and lineage separation between closely related PS. The scarcity or absence of fertile material for many PS precluded a systematic quantitative analysis of fertile characters and phenology. We therefore limited quantitative assessments of non-genetic data to leaf morphological and ecogeographic differentiation using PCA on scaled and centered subsets of the morphological and the combined dataset, after replacing missing values in the morphological dataset (3.7%) with means. We also could not integrate morphometric data for all PS due to delays in dispatch and shipment of voucher specimens. For a visualisation of morphological, ecological or genetic overlaps between PS, we used the *autoplot* function of *ggplot2* with convex hulls. To identify characters with high variation among PS and of potentially diagnostic importance, we displayed principal component loadings as arrows on the same graphs (biplots) if they exceeded the 70% quantile of the arrow length distribution in the plane defined by the first two principal components. Levels of genetic differentiation were assessed based on Weir & Cockerham's F_{st} values (WEIR AND COCKERHAM 1984) using the *snpgdsFst* function of the *SNPRelate* package version 1.20.1 (ZHENG *et al.* 2012). We also identified and visualised ten potentially diagnostic SNP loci to distinguish among PS in each PCA of the SNP dataset. These represented the SNPs with the highest absolute principal component loadings in the first ten principal components.

The often limited availability of good-quality (fertile) collections and low sample numbers with genetic data per PS precluded a conclusive species delimitation in all cases. We therefore took a conservative approach, as recommended by CARSTENS *et al.* (2013), and followed a modified version of the candidate species approach of VIEITES *et al.* (2009) to distinguish between confirmed species and unconfirmed species pending further sampling and investigations. Specifically, we assigned each PS to one of four classes for

taxonomic prioritisation, under special consideration of the sequenced type specimens and topotypic collections (BELL *et al.* 2020): (i) confirmed described species (CDS) for PS with high levels of morphological distinction from other PS, or complete ecogeographical separation, or evidence for syntopic occurrence without evident admixture with closely related PS, if it corresponded to one or more currently accepted species, (ii) confirmed candidate species (CCS) for PS with the same distinction attributes as CDS, if a PS could not be assigned to any currently accepted species, (iii) unconfirmed candidate species (UCS) for PS where morphological and ecogeographical distinction or reproductive barriers could not be thoroughly assessed because of an insufficient number of samples or specimen quality, and (iv) a putative hybrid (PH) class for specimens with indications of genetic admixture between distinct PS. The PH class was only applied if the available genetic, morphological and ecogeographic data allowed for a hypothesis on possible parent species.

We labelled each CDS using the only accepted taxonomic species name, or according to the rule of priority in botanical nomenclature (TURLAND *et al.* 2018) in cases where a CDS contained more than one currently accepted taxonomic species. CCS were labelled using validly published names of currently unaccepted taxonomic species if there was correspondence. The remaining CCS, UCS and PH were labelled using a unique alphanumeric keyword (e.g. *D.* sp. 01). We updated species identifications on Tropicos accordingly, where identifications of inspected collections continue to be updated, using the appropriate taxonomic species names or a genus-level identification with the alphanumeric keyword.

Phylogenetic analyses

We generated alignments for phylogenetic analysis using a dedicated bioinformatics pipeline available on GitHub https://github.com/scrameri/CaptureAl for read mapping, sequence assembly, orthology assessment, sample filtering, target region alignment and alignment trimming, merging of overlapping alignments, alignment assessment and filtering (see CRAMERI *et al.* (in prep.-b) for details on programs and parameters used). As for variant calling, we used the 2,396 available reference sequences for mapping. However, we limited phylogenetic analysis to 296 individuals to considerably reduce computing resources and time, as well as to avoid long-branch attraction, which occurred when too many herbarium samples with possibly accumulated post-mortem modifications were included in the alignments. The 296 individuals comprised one to several accessions for every CDS, CCS or UCS of Malagasy or Comorian *Dalbergia* that was inferred during integrative species delimitation, 23 *Dalbergia* species from other regions (excluding a

poorly sequenced *D. bignonae* Berhaut from the Ivory Coast), and seven species of *Machaerium* and *Aeschynomene*, including the outgroup.

Phylogenetic trees were constructed using *RAxML* version 8.2.11 (STAMATAKIS 2014) to estimate gene trees for each target region, and the gene tree summary method *ASTRAL-III* version 5.6.3 (MIRARAB *et al.* 2014, ZHANG *et al.* 2017). Specifically, we used the workflow and parameters as outlined in detail in CRAMERI *et al.* (in prep.-b). In addition to a summary tree of all 296 individuals, we inferred a putative species tree of the currently accepted species and new candidate species using the -a option and a file that mapped individuals to taxa. We displayed phylogenetic trees with *ggtree* version 2.0.2 (YU *et al.* 2016) and identified well-supported clades and evolutionary grades (subgroups hereafter) with corresponding morphological synapomorphies.

Random forest classification and variable importance

We used classification models to identify diagnostic morphological and ecogeographic characteristics of different supergroups, subgroups and PS, and to create a statistical tool for assistance in identification of dried voucher specimens, from sterile collections in particular. We considered the following groups as the different units for classification: (i) supergroups, (ii) subgroups, (iii) all PS, and (iv) PS within subgroups, classified within the corresponding subsets of the combined dataset. We performed classifications on only the leaf morphological characters (20 variables) and on the combined dataset consisting of leaf morphological and ecogeographic characters for comparison. We discarded the five ecoregions from the ecological predictors due to high correlations between ecoregions and climate variables, and excluded seven vegetation classes represented by less than 1% of individuals with morphological data, which left 21 ecological variables and two geographic coordinates, or 43 variables for classification of the combined dataset. For each classification, we reduced the dataset to only contain classes with at least three available collections. We then divided the dataset into a training set for model fitting and a validation set for model evaluation using a stratified random sampling approach. The training sets had to contain 67% of samples per class, except for classes with many available samples, where we limited the number of training samples to 33% of the largest class size. This served to create a training set with a more equal representation of the various classes, and a validation set (i.e., samples that were not in the training set) with a more natural frequency of classes. We chose the random forest classificator (BREIMAN 2001) for model fitting, an ensemble machine learning method implemented in the *randomForest* package version 4.6.14 (LIAW AND WIENER 2002), because of its flexibililty and scalability to large datasets involving many samples and variables. Multiple hyperparameters influence model

performance, of which the *mtry* hyperparameter (i.e., the number of randomly drawn variables used to define splits during growth of a large number of binary decision trees, which constitute the random forest) is frequently the most influential in terms of prediction performance (PROBST *et al.* 2018). We therefore tuned the *mtry* hyperparameter on the training set by repeated *K*-fold cross-validation, using 10 folds, five repetitions, and used the overall prediction accuracy (i.e., the percentage of correct predictions of validation samples) as the criterion to select an optimal *mtry* parameter value for each classification. We then evaluated the performance of each classification based on the validation set and overall prediction accuracy, mean sensitivity, specificity and precision per class as performance metrics using the *confusionMatrix* function from the *caret* package version 6.0.86 (KUHN 2020). We finally visualised the confusion matrices using the *table.value* function implemented in *adegenet*.

To identify morphological characters of possible taxonomic significance, we applied disciminant analysis of principal components (JOMBART *et al.* 2010) and evaluated the discriminant coefficients of morphological and ecogeographic variables. Specifically, we used the same datasets and classes as for the random forest classifications, but did not separate the data into a training and a validation set. We first applied the *xvalDapc* function from the *adegenet* package to define an optimal number of retained principal components via cross-validation. The function was carried out with default parameters except for a scaling of the input data and a testing of all possible numbers of retained principal components. We then chose the number of retained principal components yielding the lowest mean squared error as the optimal value, fitted a corresponding discrimination model, and saved the discriminant coefficients of each original variable and discriminant function. We then plotted a heat map of the normalised discriminant coefficients versus the discriminant analyses.

Ecogeographic distributions of verified collections

We combined the geographic coordinates of genotyped and morphotyped collections with additional coordinates from the Tropicos database to assess the ecogeographic distribution and vegetation ecology of inferred evolutionary lineages (supergroups, subgroups and species) based on a more inclusive dataset. Due to the frequency of wrongly identified material, and given that we could not examine all databased collections, we applied the following filters to define a set of verified collections: (i) genus *Dalbergia*, and (ii) country Madagascar, and (iii) adequate georeferencing, and (iv) not cultivated, and one of the following conditions: (a) type specimen, or (b) collection inspected and identified to an accepted described species or a new candidate species by SC, NW or PBP. We then

assigned the verified collections (n = 1,846) to their respective supergroup and subgroup based on the phylogenetic placement of conspecific sequenced collections, and drew static and interactive distribution maps with the ecogeographic characteristics using the *R* packages *tmap* version 3.0 (TENNEKES 2018) and *leaflet* version 2.0.3 (CHENG *et al.* 2019). Alpha hulls were drawn around collections of the same subgroup using the *EOO.computing* function from the *R* package *conR* version 1.3.0 (DAUBY *et al.* 2017), a shape file of Madagascar's land surface (distributed by Conservation International), an appropriate alpha value and a buffer of 0.25 degrees.

We computed occurrence frequencies for each combination of species-level taxon and ecological factor level based on the extracted ecological data of verified collections. To prevent the frequency estimates from being biased towards locations where a species has been collected multiple times, we reduced the dataset of verified collections to only include 994 records that were at least half a degree minute apart from conspecific records. We divided the ranges of continuous ecological variables (elevation and bioclimate) to a low, medium and high category based on thresholds appropriate for Madagascar, and merged the occurrence frequencies of the low and high classes with the frequencies of categorical ecological factor levels to compile an ecological resource matrix of occurrence frequencies. We then visualised the resource matrix as a heatmap, and ordered rows and columns using Ward's hierarchical agglomerative clustering method (ward.D2) on Euclidean distances. In addition, we estimated the ecological niche breadth of each species using the *niche.width* function from the *spaa* package version 0.2.2 (ZHANG 2016), and determined the frequency of occurrence in protected areas using the *extract2* function and a corresponding shape file distributed by Conservation International.

Species richness and community diversity

We divided Madagascar's surface into a regular grid of 0.5 degree cell width and counted the number of individuals as well as the number of CDS, CCS, and UCS present in each grid cell (i.e., the abundance matrix) using the dataset of verified collections (n = 1832 excluding PH). We then applied rarefaction and extrapolation to the abundance matrix using the *iNEXT* function from the *iNEXT* package version 2.0.20 (CHAO *et al.* 2014), 100 knots and 100 bootstrap replicates, which yielded asymptotic estimates of species richness per grid cell. In addition, we computed Faith's (unrooted) phylogenetic diversity (PD, FAITH 1992) in each grid cell using the *picante* package version 1.8.2 (KEMBEL *et al.* 2010) and the ASTRAL-III putative species tree. We then visualised the species richness and community structure metrics on annotated maps of Madagascar.

Results

Sequencing, alignment and variant calling

Sequencing resulted in an average of 3.10 million read pairs per sample. After trimming and quality-filtering, 93% of reads were retained on average. Across all samples, 39% of trimmed reads mapped against reference sequences of the 2,396 target regions (Table S2). Coverage analysis revealed that 94% of these target regions had at least 10× coverage on average. Coverage was generally higher across silica gel dried samples (98%) than across herbarium samples (87%, Table S2, Figure S1). The bioinformatics pipeline for target region alignment resulted in 2,370 alignments for phylogenetic analysis. Variant calling and filtering resulted in 925,216 filtered variants distributed across all 2,396 target regions in the 810 analysed samples (mean of 386 variants per target region) and 2,080,439 different alleles. The majority of variants (77%) were biallelic, 20% were triallelic, and 3% had more than three alleles.

The assessment of reproducibility of target enrichment and sequencing resulted in a backward-selected regression model with two interaction terms, an adjusted R^2 of 76% and residuals that met the model assumptions reasonably well (Figure S2A – B, Table S5). Identity between pairs was found to be significantly higher in replicates (mean = 99.6%) as compared to conspecific pairs (mean = 98.5%, p = 2.93e-12, Figure S2C). A high coverage of target regions in one of the pairs had a significant positive effect on pairwise identity (p = 1.63e-03). Greater specimen age (p = 2.26e-07), the herbarium tissue type (p = 5.94e-08) and a higher difference in coverage of target regions between pairs (p = 1.98e-04) had a significant negative effect on pairwise identity. By contrast, sequencing platform and low coverage of target regions in one of the pairs did not have an effect on pairwise identity, and were removed during step-wise model reduction.

Integrative species delimitation

Principal component analysis on 697 samples of *Dalbergia* (694 from Madagascar, two from Mozambique and 1 from the Comoros) and 168,278 biallelic loci with no allowed missingness revealed four major genetic groups separated along the first (49% explained variance) and second (6% explained variance) principal components (Figure S3A). The corresponding rooted NJ tree (Figure S3C) revealed four major lineages, which correspond to (i) a diverse lineage of exclusively Malagasy species (Madagascar Supergroup I; "Supergroup I" hereafter), (ii) a diverse lineage of species from Madagascar and the Comoros (Madagascar Supergroup II; "Supergroup II" hereafter), (iii) *D. bracteolata*

specimens from Mozambique and Madagascar, and (iv) *D. xerophila* Bosser & R. Rabev. from Madagascar. The second principal component indicated substructure within Supergroup I, the third principal component (5% explained variance) indicated substructure within Supergroup II, and the fourth principal component (3% explained variance) separated *D. bracteolata* and *D. xerophila* from the two supergroups (Figure S3B). These four major lineages were also revealed in a corresponding NJ tree (Figure S3C). Substructure within both supergroups was further investigated through nested PCA, which revealed separate clusters of individuals within Supergroup I (Figure S4A) and Supergroup II (Figures S4B). Nested analyses within three genetic clusters of Supergroup I (Figure S4C, S5-S8) and four genetic clustes of Supergroup II (Figure S4D, S9-S16) and evaluation against morphology, ecogeography and genetic distinction in syntopy led to the delimitation of 95 PS, including the two isolated lineages of *D. bracteolata* and *D. xerophila* (Table 2). Of these 95 PS, we assigned 46 (48%) to CDS, 31 (33%) to UCS, 14 (15%) to CCS, and 4 (4%) to PH. Putative hybrids (*D. sp.* 13, *D. sp.* 45, *D. sp.* 50, and *D.* sp. 59) were identified based on PCA and NJ trees, morphology and ecogeographic origin (Figures S4C, S4D, and S9).

TABLE 2: Number of currently accepted (BOSSER & RABEVOHITRA 1996, 2002, 2005) and delimited new candidate species of Dalbergia from Madagascar and the Comoros. CDS: confirmed described species; CCS: confirmed candidate species; UCS: unconfirmed candidate species; PH: putative hybrid lineage.

Subgroup	Accepted spp.	CDS	CCS	UCS	PH	Sum
Supergroup I						
Chapelieri	1	1	1	5	-	7
Maritima	4	4	2	5	-	11
Pervillei	2	2	2	1	1	6
Tricolor	6	5	1	3	-	9
Supergroup II						
Arenosa	-	-	1	1	-	2
Baronii	10	9	1	4	-	14
Chlorocarpa	6	6	1	1	-	8
Greveana	4	4	1	5	1	11
Humbertii	3	3	-	-	-	3
incertae sedis	2	1	1	1	-	3
Mollis	2	2	2	-	-	4
Monticola	3	3	1	3	2	9
Peltieri	4	4	-	2	-	6
Separate lineages						
Bracteolata	1	1	-	-	-	1
Xerophila	1	1	-	-	-	1
Sum	49	46 (48%)	14 (15%)	31 (33%)	4 (4%)	95 (100%)

Phylogeny

The 296 individuals selected for phylogenetic analysis represented 91 inferred PS of *Dalbergia* from Madagascar and the Comoros (excluding 4 PH), 23 *Dalbergia* species from other regions, and seven species of *Machaerium* and *Aeschynomene*, including the outgroup. The gene tree summary (ASTRAL-III) topologies achieved an overall normalised quartet score of 91.1% for the 296 individuals, and 88.9% for the putative species tree. Both gene tree summary topologies recovered the genus *Dalbergia*, as well as supergroups I and II as monophyletic with 100% support, and confirmed the two Malagasy species *D. bracteolata* and *D. xerophila* as two separate lineages (Figure 1A), with *D. bracteolata* from Mozambique as sister to the Malagasy populations (Figure S17).

We identified twelve highly supported subgroups in total, four subgroups within Supergroup I and eight subgroups and four PS with uncertain phylogenetic placement within Supergroup II (Figure 1A, Figure S17). Supergroup I can be separated into two reciprocally monophyletic and highly supported clades, one of which we named subgroup Maritima (100% posterior support). The other clade consists of a grade containing various species, which we subsumed under subgroup Tricolor because of shared morphological characters, as well as the two monophyletic and highly supported subgroups Chapelieri (100%) and Pervillei (100%). Supergroup II can be separated into a stem group consisting of two highly supported subgroups, which we named Baronii (100%) and Monticola (100%). The crown group of Supergroup II consists of six well-supported subgroups, which we named Arenosa (100%), Chlorocarpa (100%), Greveana (94%), Humbertii (100%), Mollis (100%) and Peltieri (99%), as well as four PS with uncertain phylogenetic placement and subgroup membership. Within the crown group, subgroups Chlorocarpa and Humbertii were recovered as reciprocally monophyletic with 100% posterior support, while subgroup Mollis formed a grade with subgroup Arenosa, subgroup Peltieri and a weakly supported (65%) clade consisting of subgroup Greveana and the four PS with uncertain subgroup membership.

All analysed specimens from outside Madagascar, the Comoros and Mozambique were not associated with any of the two Malagasy supergroups, but formed distinct lineages or clades. *Dalbergia miscolobium* Benth. from Brazil formed a sister lineage to the remaining analysed *Dalbergia* species. Three accessions from Central-West Africa (*D. louisii* Cronquist and *D. obliquifoliolata* O. Lachenaud from Gabon, *D. bakeri* from Equatorial Guinea) formed a clade with *D. ecastaphyllum* (L.) Taub. sampled in Florida. Two accessions from Asia (*D. lanceolaria* L. f. subsp. *paniculata* (Roxb.) Thoth. and *D. oliveri* Gamble ex Prain) formed a clade, which was sister to two species from continental Africa (*D. lactea* Vatke from Uganda and *D. saxatilis* Hook f. from Cameroon).

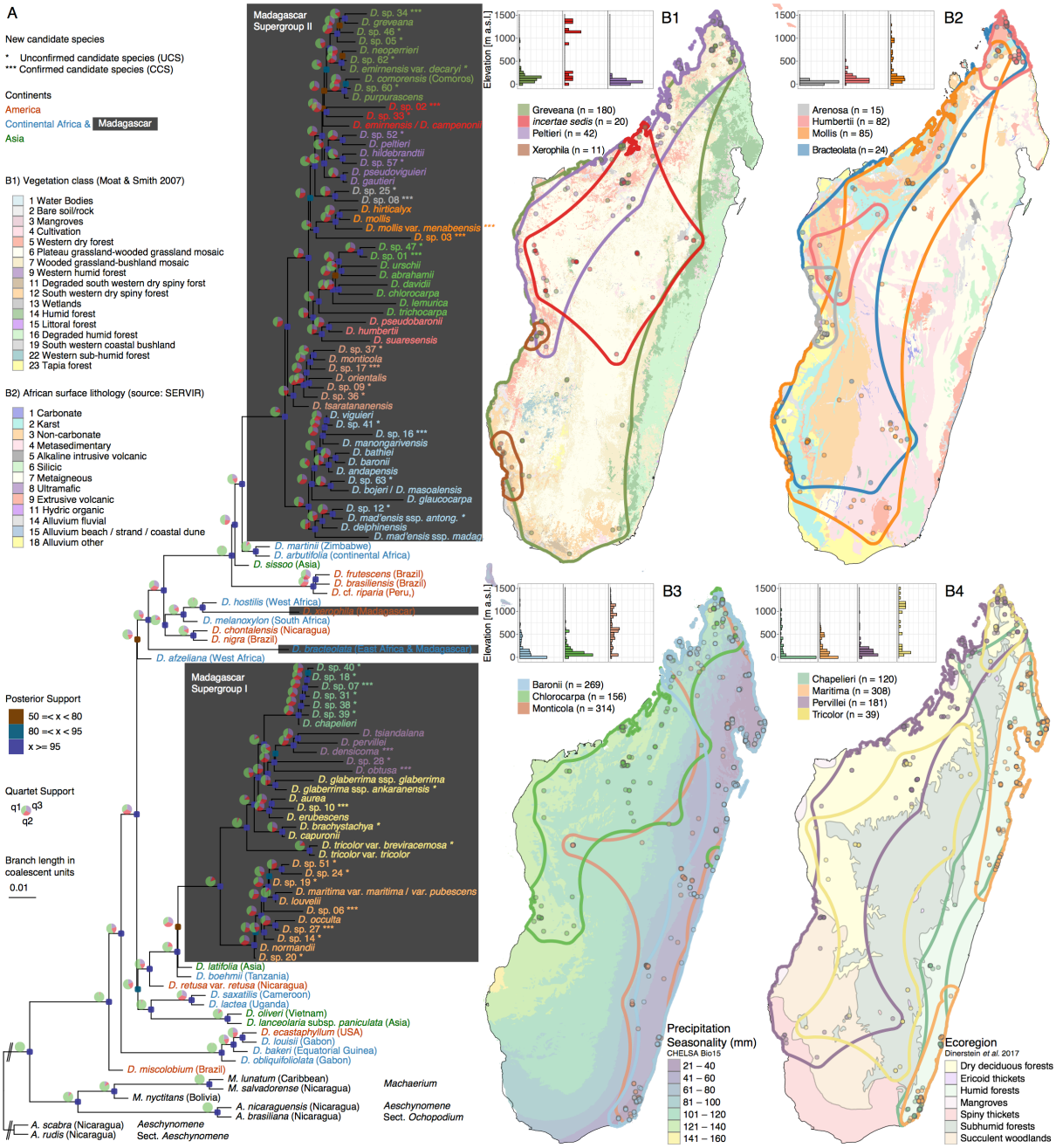

FIGURE 1: **A)** Coalescent-based phylogeny of *Dalbergia* (91 putative species from Madagascar and the Comoros, 23 species from other regions), *Machaerium* (three species) and *Aeschynomene* (four species, two outgroup taxa with shortened branch lengths) inferred using ASTRAL-III on 2,370 gene trees with contracted low-support branches (< 10% bootstrap support). 1,570 gene trees (66%) had missing taxa. The overall normalised quartet score was 88.9%. Dark boxes highlight the Malagasy and Comorian species. Colour-coded tip labels denote subgroup membership, or the continent of origin for species from outside Madagascar and the Comoros. Pie charts represent quartet support, i.e. the fraction of gene trees that are consistent with a given node (green) or with alternative topologies (red, purple). **B)** Ecogeographic distribution of each subgroup, represented as colour-coded alpha hulls in four subpanels. Colour-coded points show slightly jittered locations of sequenced, morphotyped or otherwise verified collections with adequate georeferencing (n = 1,846). Horizontal histograms show elevation profiles for each subgroup.

An accession of *D. retusa* Hemsl. var. *retusa* from Nicaragua represented a distinct lineage, which was sister to *D. boehmii* Taub. from Tanzania, *D. latifolia* Roxb. from Asia and the Malagasy Supergroup I. *Dalbergia afzeliana* G. Don from Central-West Africa and *D. bracteolata* from Mozambique and Madagascar also formed distinct lineages. The Malagasy endemic species *D. xerophila* was recovered as sister to *D. hostilis* Benth. from Central-West Africa, with which it formed a sister group to an accession of *D. melanoxylon* Guill. & Perr. from South Africa. That clade was in turn sister to two species of Central and South America (*D. chontalensis* Standl. & L.O. Williams from Nicaragua and *D. nigra* (Vell.) Allemão ex Benth. from Brazil). Three further accessions from South America (*D. brasiliensis* Vogel and *D. frutescens* (Vell.) Britton from Brazil, *D.* cf. *riparia* (Mart.) Benth. from Peru) formed another distinct clade, which was sister to a grade composed of *D. sissoo* Roxb. ex. DC. from India, two accessions from Central-East Africa (*D. arbutifolia* Baker and *D. maritinii* F. White) and the Malagasy Supergroup II.

Morphological variation

Leaf morphological data are presented in Table S3. The visualisation and interpretation of the first six principal components of leaflet shape are given in Figure S18A. The first principal component (70.2% explained variation) separates broadly ovate or orbicular leaflets from narrowly oblong leaflets. The second component (18.0%) describes the leaflet apex and separates leaflets with rounded or retuse apices from leaflets with acute or acuminate apices. The fourth component (2.4%) is associated with both the leaflet apex and base, and separates leaflets with a rounded apex and truncate base from leaflets with a more obtuse apex and cuneate base. The fifth component (1.1%) describes the tapering of the leaflets towards the apex, and the sixth component (0.5%) separates leaflets with a rounded apex and cuneate base from leaflets with a more obtuse apex and more rounded base. We excluded the third principal component of leaflet shape (6.3%) because it described variation within the same leaf, namely the leaflet curvature as a function of its insertion along the leaf rachis (Figure S18A). Representations of mean leaflet shapes for all PS with available morphometric data are given in Figure S18B.

Spearman's correlations among morphological variables were often above 70%, especially among variables describing leaf pubescence or leaflet size (Figure S19). Generally low correlations were found between and among variables describing leaflet shape and leaflet colour. We excluded the first principal component of leaflet shape from downstream analyses, due to a high correlation with the leaflet length-to-width ratio (-0.99, $p < 2.2e\text{-}16$).

Random forest classification and variable importance

The overall accuracy of validation sample predictions was 80.2% when classifying all 39 PS with at least three available samples in the morphological dataset, and when only considering leaf morphological characters (Table S6). If samples were classified and predicted within the nine subgroups with two or more classes, the accuracy increased to values between 92% and 100% in all subgroups except for subgroups Chapelieri (81.0%) and Peltieri (80.0%). Similarly, overall prediction accuracy increased to 84.9% when the twelve available subgroups were classified instead of PS, and to 95.1% when classifying at the level of supergroups. Random forest classification was at least as accurate, or markedly more accurate when using the combined dataset of leaf morphological and ecogeographic data compared to classifications that only considered leaf morphological characters. Specifically, the overall accuracy increased to 95.9% when classifying among PS, to 92.9% when classifying among subgroups, and to 95.9% when classifying at the level of supergroups based on the combined dataset (Table 3).

The average classification specificity across classes ranged between 90% and 100% in all classifications of the morphological dataset, except for subgroup Peltieri (83.3%). The corresponding average classification sensitivity ranged between 60% and 100% depending on the number of training samples per class, or depending on the similarity to classes with more training samples. Specificity in classifications of the combined dataset ranged between 96% and 100%, while sensitivity ranged between 75% and 100% except for the classifications among PS of subgroups Greveana (60.0%) and Pervillei (66.7%).

The confusion matrix of classification on the morphological dataset and the 39 PS revealed that only 11 (19%) of misclassifications involved PS of the same subgroup, while 47 (81%) involved PS of different subgroups (Figure S21). Similarly, 3 (25%) of misclassifications based on the combined dataset were within subgroups, while 9 (75%) were between subgroups (Figure S22).

Leaflet size variables (length, width, perimeter) discriminated between PS within subgroups Baronii, Chlorocarpa and Maritima, but not among subgroups (Figure S23). Leaflet shape component 2 showed discriminatory power at the subgroup level and within subgroups Baronii (notably for discrimination of *D. baronii* Baker from other PS) and Greveana (notably for discrimination of *D. purpurascens* Baill., *D.* sp. 05 and *D.* sp. 60 from other PS). Similarly, leaflet shape component 4 showed discriminatory power at the subgroup level and within subgroup Greveana (notably for discrimination of *D.* sp. 03 from other PS). Leaflet shape component 6 showed discriminatory power at the subgroup level and within subgroup Mollis (notably for discrimination of *D.* sp. 03 from other PS).

TABLE 3: Random forest classification of combined leaf morphological and ecogeographic data. Overall prediction accuracy, mean specificity and sensitivity per class are based on predictions of validation samples. PS = putative species, SD = standard deviation.

Grouping / subgroup	*mtry*	Classes	Training samples	Validation samples	Prediction accuracy	Mean specificity (± SD)	Mean sensitivity (± SD)
Classification of PS, subgroups and supergroups considering all samples							
PS	8	39	387	293	95.90%	99.9 (0.3)	88.8 (27.4)
Subgroups	4	12	370	338	92.90%	99.4 (0.7)	78.8 (29.9)
supergroups	5	3	298	410	95.85%	96.9 (3)	75.2 (36.2)
Classification of PS within subgroups							
Baronii	2	3	24	35	100%	100 (0)	100 (0)
Chapelieri	33	4	18	21	100%	100 (0)	100 (0)
Chlorocarpa	27	4	28	34	97.06%	98.9 (2.2)	98.5 (2.9)
Greveana	21	5	39	56	96.43%	96.3 (7.3)	60 (54.8)
Maritima	10	7	89	79	100%	100 (0)	100 (0)
Mollis	1	3	9	9	100%	100 (0)	100 (0)
Monticola	6	5	62	80	100%	100 (0)	100 (0)
Peltieri	1	2	10	5	100%	100 (0)	100 (0)
Pervillei	23	3	20	26	96.15%	91.7 (14.4)	66.7 (57.7)

Leaflet number was not revealed to discriminate between subgroups, but showed discriminatory power within subgroups Greveana (notably for discrimination of *D. purpurascens*, *D.* sp. 05 and *D.* sp. 60 from other PS), Maritima (notably for discrimination of *D.* sp. 19 and *D.* sp. 24 from other PS) and Pervillei (notably for discrimination of *D. tsiandalana* R. Vig. from other PS). Leaf length did not result in high discriminant coefficients across discriminations, except for discrimination within subgroup Chlorocarpa (notably for discrimination of *D. trichocarpa* Baker from other PS). Levels of pubescence on the leaflet rachis, petiolule and the upper and lower leaflet lamina were discriminating between subgroups such as the similar subgroups Chapelieri, where all species appear to have glabrous leaves, and Pervillei, where most species show a scarce to dense pubescence on certain leaf surfaces. Pubescence also showed discriminatory power within subgroups Chlorocarpa (notably for discrimination of *D. abrahamii* Bosser & R. Rabev. from other PS), Maritima (notably for discrimination of *D.* sp. 19 and *D.* sp. 24 from other PS, as well as to discriminate between two described varieties of *D. maritima* R. Vig.), and Monticola (notably for discrimination of *D. orientalis* Bosser & R. Rabev. from other PS). The various assessed ratios between leaf morphological variables were revealed to be of some importance in certain instances, but the added discriminatory power was revealed to be low in comparison to the variables on which the ratios were based. By contrast, the greenish colour component ratio of dried hebarium vouchers was revealed to be of potentially high

taxonomic relevance to distinguish species at the supergroup and subgroup levels, respectively (Figure S23).

Ecogeographic distributions

The ecogeographic characteristics extracted from 1,846 taxonomically verified and adequately georeferenced collections are presented in Table S4, Figure 1B and Figure 2. Species richness was found to be highest in dry deciduous forests of west and north Madagascar, followed by humid forests in east Madagascar and subhumid forests on the Central Plateau. Succulent woodlands were revealed to be less species-rich, and species counts were lowest in spiny thickets and mangroves. No confirmed *Dalbergia* collection was recorded in ericoid thickets, which occur above 1,800 m a.s.l. (Table 4). Collection sites of Malagasy *Dalbergia* were located between sea level and around 1,700 m a.s.l. (*D. tsaratananensis* Bosser & R. Rabev., MADAGASCAR CATALOGUE 2020).

Both Malagasy *Dalbergia* supergroups are represented with several species in the four largest ecoregions of Madagascar (dry deciduous forests, humid forests, subhumid forests, and succulent woodlands, Figure 1B, Figure 2, Table 2). Subgroup Maritima (Supergroup I) appears to be restricted to the eastern humid ecoregion. Subgroups Baronii, Monticola (both Supergroup II) and Chapelieri (Supergroup I) are each predominantly distributed in the eastern humid ecoregion, but some species occur in, or extend into the subhumid Northwest (*D. madagascariensis*, *D.* sp. 37), the high Central Plateau (*D. monticola* Bosser & R. Rabev., *D.* sp. 38, *D. tsaratananensis* Bosser & R. Rabev.), or the drier North (*D. viguieri* Bosser & R. Rabev., *D.* sp. 16, *D.* sp. 41). Subgroups Greveana, Mollis (both Supergroup II) and Pervillei (Supergroup I) are predominantly distributed in dry deciduous forests in west and north Madagascar, but some species also occur in the succulent woodlands or spiny thickets (*D. emirnensis* var. *decaryi* Bosser & R. Rabev., *D. mollis* var. *menabeensis* Bosser & R. Rabev., *D. pervillei* Vatke) in southwest and south Madagascar. Subgroups Chlorocarpa, Humbertii and Peltieri (all Supergroup II) also show a western and northern distribution, but they do not appear to reach the spiny thickets in the far South. Subgroup Chlorocarpa appears to have a disjunct distribution due to *D. trichocarpa*, which is predominantly found in west Madagascar but which was also recorded from a littoral forest in central-east Madagascar. Subgroup Arenosa only contains two species, which both occur in west Madagascar, on sandy soil. Subgroup Tricolor contains multiple species that appear to be restricted to the Central Plateau. All subgroups except for subgroup Arenosa, Chapelieri and Maritima are represented in north Madagascar.

TABLE 4: Observed species richness of the four main lineages of *Dalbergia* species in seven ecoregions of Madagascar (DINERSTEIN *et al.* 2017). Counts are based on 1,846 verified collections and represent 90 putative species (45 CDS, 14 CCS, 31 UCS, excluding *D. comorensis* and four PH). Most putative species occur in a single or a few ecoregions and are counted multiple times.

Ecoregion	Supergroup I	Supergroup II	Bracteolata	Xerophila	Sum
Species-rich					
Madagascar dry deciduous forests	10	37	1	0	48
Madagascar humid forests	18	18	0	0	36
Madagascar subhumid forests	11	20	1	0	32
Madagascar succulent woodlands	2	10	1	1	14
Species-poor					
Madagascar mangroves	1	6	1	0	8
Madagascar spiny thickets	1	4	1	1	7
Absent					
Madagascar ericoid thickets	0	0	0	0	0

Species richness and community diversity

Asymptotic estimates of species richness ranged from zero to 40 and were highest in north Madagascar, Ankarafantsika and the regions surrounding Morondava (Figure 3C). A correlation between species richness and sampling effort was also evident (Figure 3A – C). In addition, Ankarafantsika and north Madagascar (Loky-Manambato [Daraina], the Ankarana and Analamerana limestone plateaus and the regions to the north) showed exceptionally high phylogenetic diversity (Figure 3D). Other regions with elevated phylogenetic diversity (≥ 25) included Masoala, the surroundings of Morondava, Namaroka and the Sambirano region in northwest Madagascar.

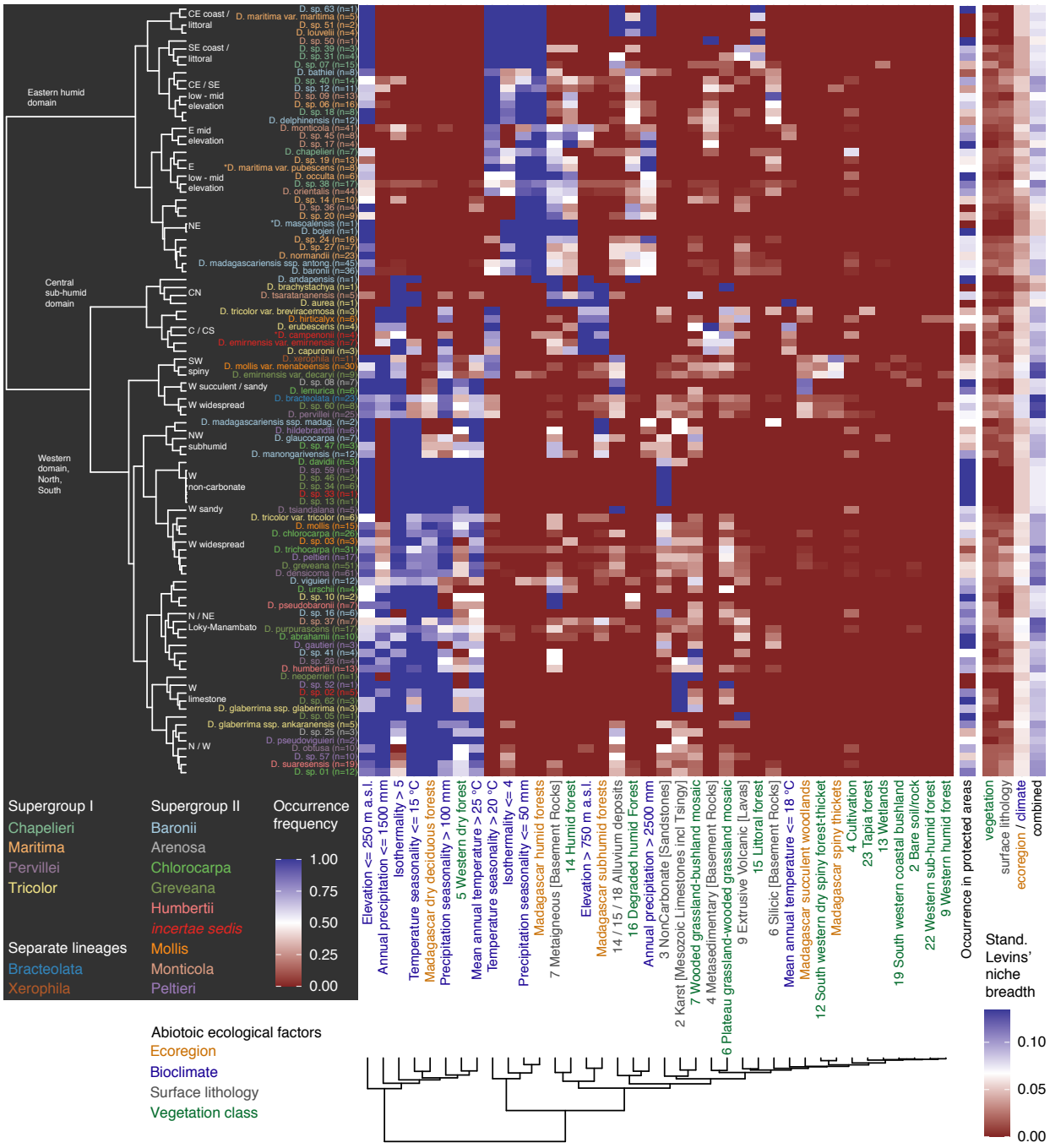

FIGURE 2: Ecological resource matrix of adequately georeferenced and taxonomically verified *Dalbergia* collections from Madagascar (n = 994). Rows represent accepted taxa or new candidate species (n = 97 excluding *D. comorensis* and including three potentially synonymous taxa marked with asterisks, see Discussion), which are colour-coded according to phylogenetic subgroups (see inset legend). Columns represent different levels of abiotic ecological factors (ecoregion following DINERSTEIN *et al.* (2017), CHELSA bioclim variables, surface lithology and vegetation class following MOAT AND SMITH (2007)). Heatmap colours represent the frequency of occurrence for each combination of species-level taxon and ecological factor level. Dendrograms and the order of rows and columns represent the ecological relatedness of species and abiotic factors, as inferred from Ward's hierarchical agglomerative clustering of Euclidean distances. The single isolated column shows the occurrence frequency in protected areas. The four isolated columns on the right represent standardised niche breadth estimates following LEVINS (1968), separately for the vegetation, surface lithology, climate and the combined niche. C: central; E: east, N: north; S: south; W: west.

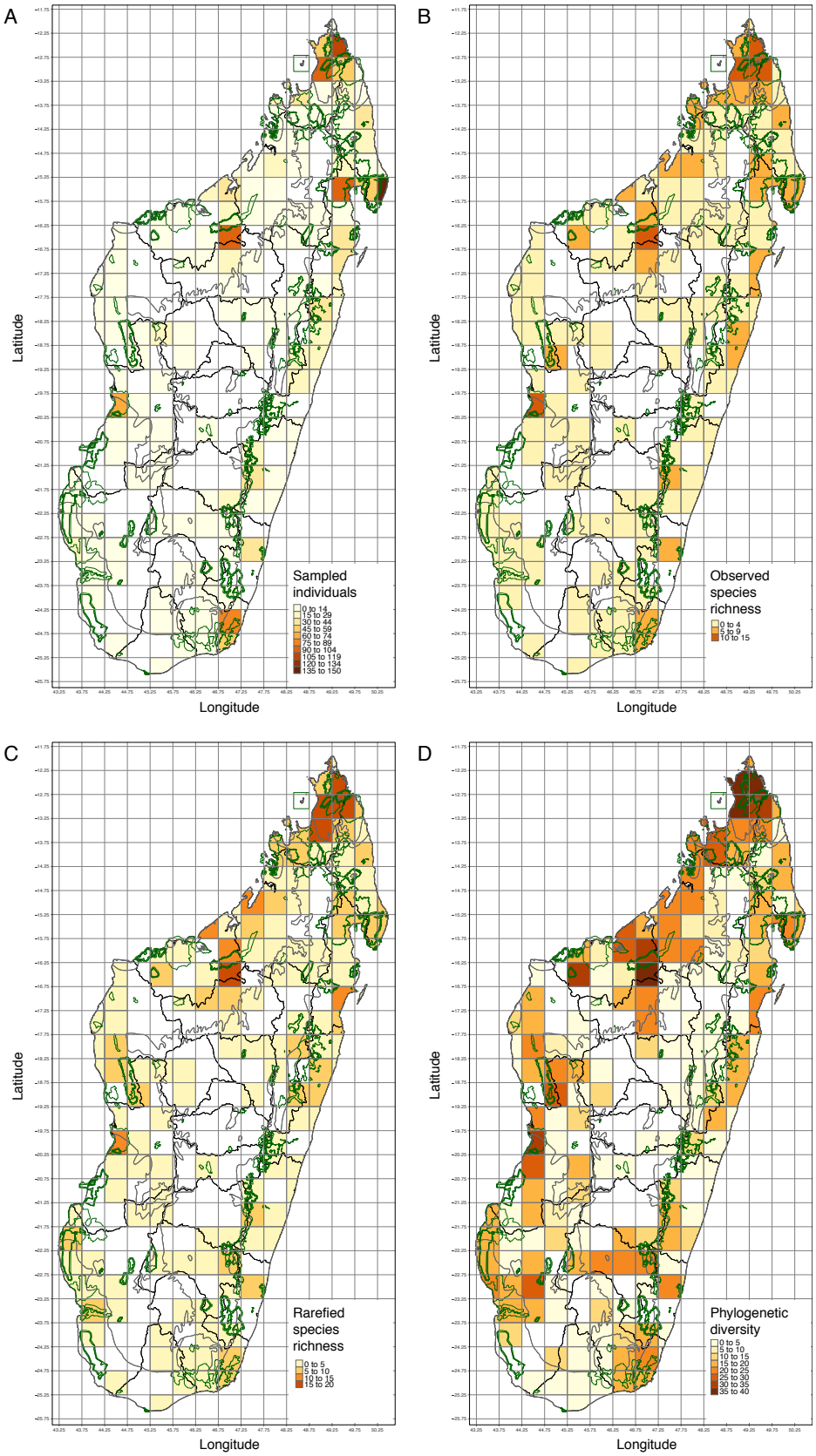

FIGURE 3: Sampling and species diversity patterns. **A)** number of taxonomically verified and adequately georeferenced collections (n = 1846), **B)** number of putative species (PS) present in each grid cell (excluding PH), **C)** number of PS after rarefication within grid cells, **D)**, unrooted phylogenetic diversity inferred from the ASTRAL-III gene tree summary topology with a single tip per PS.

Discussion

We integrated target enrichment sequencing data for 683 *Dalbergia* individuals from Madagascar and the Comoros with leaf morphological and ecogeographic data for 708 individuals, and used this dataset for species discovery and species delimitation. The target enrichment approach that interrogates 2,396 nuclear genomic regions (CRAMERI *et al.* in prep.-b) was found to represent a highly valuable genomic resource for species discovery and delimitation in *Dalbergia*, and provides sufficient phylogenetic resolution to distinguish even among closely related PS. This stands in stark contrast to the standard DNA barcoding approach adopted by HASSOLD *et al.* (2016), who used three plastid markers on an overlapping set of sequenced individuals. The DNA barcoding approach yielded sufficient resolution to distinguish between subgroups Chapelieri and Maritima, but insufficient resolution to distinguish among subgroups of Supergroup II, and no resolution to distinguish between closely related species. The present study reveals that *Dalbergia* species diversity in Madagascar has been vastly underestimated. This is in part due to known taxonomic problems in particularly challenging groups, such as subgroup Pervillei, where the genetic data provides valuable insights and supports the resolution of several existing taxonomic uncertainties, e.g. on the recognition of *D. densicoma* Baill. as a separate species. On the other hand, the intensive collecting efforts made over recent years by members of an international consortium for research in Malagasy precious woods also brought to light new candidate species that had never been collected before (e.g., *D.* sp. 17), that were present as overlooked specimens within existing species concepts (e.g., *D.* sp. 08), or that had already been collected before 2005, but were left undetermined due to a sterile condition (e.g., *D.* sp. 03).

Phylogeny and biogeographic connections

Phylogenetic analysis based on the currently most comprehensive sampling of *Dalbergia* species from Madagascar and the Comoros, and integration of 23 species from continental Africa, the Americas and Asia, confirmed the presence of two species-rich monophyletic Malagasy *Dalbergia* clades, Supergroup I and Supergroup II, which were recovered in previous phylogenetic studies (VATANPARAST *et al.* 2013, HASSOLD *et al.* 2016, CRAMERI *et al.* in prep.-b). Increased taxon sampling and phylogenetic resolution in the present study revealed interesting new findings. First, *D. xerophila* Bosser & R. Rabev. represents an isolated lineage and is not associated with the two Malagasy supergroups, but is instead related to *D. hostilis* Benth. from West Africa and, more distantly, to *D. melanoxylon* Guill. & Perr. from South Africa. This finding can be reconciled with its unique morphology

(formation of brachyblasts) compared to other Malagasy taxa, which possibly represents an adaptation to the arid conditions in the spiny and succulent ecoregions of Madagascar (BOSSER & RABEVOHITRA 2002). Secondly, a topotypic collection of *D. aurea* Bosser & R. Rabev. is here revealed as part of subgroup Tricolor in Supergroup I, contrary to a phylogeny based on internal transcribed spacer sequence variation (VATANPARAST *et al.* 2013), where the type specimen of *D. aurea* was recovered as sister to *D. lactea* Vatke with high support. We included two accessions of *D. lactea* in our study, which are here revealed as sisters to another species from continental Africa, *D. saxatilis* Hook f. (Figure 1, Figure S17). Our finding agrees with flower and fruit charaters of *D. aurea* (8 – 10 mm long flowers, purple calyx, glabrous gynoecium, thinly coriaceous fruit pericarp), which are consistent with other species of Supergroup I. Thirdly, we found that there is a biogeographic connection between *Dalbergia* species from Madagascar and the Comoros. We recovered the only described *Dalbergia* species of the Comoros, *D. comorensis* Bosser & R. Rabev., as closely related to *D. purpurascens* and an unconfirmed new candidate species (*D.* sp. 60, Figure S15), which represents material from west and southwest Madagscar previously included in *D. purpurascens*. *Dalbergia comorensis* is known from Grande Comore as well as from Mayotte, which is located c. 300km from the nearest Malagasy coast. The present data suggest that the ancestor of *D. comorensis* might have colonised the Comoros (including Mayotte) from western Madagascar. This contrasts to an earlier hypothesis involving *D. hildebrandtii* Vatke, whose type locality is on Nosy Be in north-west Madagascar, as the closest species associated with *D. comorensis* (BOSSER & RABEVOHITRA 1996). Biogeographic connections between western Madagascar and the Comoros (including Mayotte) have been observed in other plant genera, and there exist plant species that occur in Madagascar and the Comoros (e.g., LE PÉCHON *et al.* 2020).

Our analyses further confirm two proposed biogeographic connections between Madagascar and continental Africa. The first connection is reprented by a confirmed clustering of *D. bracteolata* from Madagascar and Mozambique (Figure S17, BOSSER & RABEVOHITRA 2002). The second is given by *D. martinii* F. White from Zimbabwe and *D. arbutifolia* Baker from an unspecified location in continental Africa, which form the sister group to Supergroup II in our analysis. Both these species have also been analysed by VATANPARAST *et al.* (2013), who recovered *D. martinii*, but not *D. arbutifolia*, as sister to their M3 clade, which contains species here placed in Supergroup II.

Supergroup I is here recovered as sister to an Asian species (*D. latifolia* Roxb.), but that relationship is weakly supported (74%) because a similar number of gene trees supported an alternative topology involving *D. bohemii* Taub. from Tanzania as the sister species of Supergroup I (Figure 1A), a hypothesis that is strongly indicated by the ITS

phylogeny, see clade M1 in VATANPARAST *et al.* (2013). The apparent lack of gene trees supporting *D. bohemii* Taub as sister to Supergroup I in our analysis could be related to the suboptimal sequencing data quality for this herbarium collection from 1968, which passed the 1,000 regions with sufficient coverage threshold for phylogenetic analysis, but which was excluded from the SNP dataset owing to a missingness of over 50%. Additional sequencing data would therefore be needed to confirm the hypothesis that there exists a fourth biogeographical connection between Madagascar and East continental Africa.

Overall, the biogeography of *Dalbergia* appears to be complex, with multiple African, American and Asian clades, as already shown by VATANPARAST *et al.* (2013). We anticipate that a more comprehensive sampling of taxa from all continents, in combination with phylogenomic analysis of nuclear sequence variation as presented here, would reveal many more highly supported clades and biogeographic connections, and significantly improve our understanding of the evolutionary history and the morphological and ecological variation in *Dalbergia* species worldwide.

Species discovery

To estimate the number of PS in Madagascar, we sequenced accessions assignable to all currently accepted species together with a substantial number of recent collections with uncertain taxonomic identity and/or from understudied regions. Using our multi-locus nuclear genetic dataset and a combination of PCA and NJ trees, we identified 95 PS that were genetically, morphologically and ecogeographically coherent, but not necessarily distinct. The present geographical sampling (Figure 3) is based on the current availability of DNA samples, and we strongly expect that more field collections from understudied regions of Madagascar will reveal additional PS.

Species validation

Of the 95 PS, we could confirm 60 (46 CDS and 14 CCS) Malagasy and Comorian *Dalbergia* (Figures S5 – S16) as separately evolving lineages. Leaf morphological characters showed a considerable degree of distinction among many pairs of assessed and closely related PS, notably for *D. baronii* and *D. glaucocarpa* Bosser & R. Rabev. in subgroup Baronii, *D.* sp. 07 in subgroup Chapelieri, *D. abrahamii* in subgroup Chlorocarpa, *D. normandii* Bosser & R. Rabev. in subgroup Maritima, *D. hirticalyx* Bosser & R. Rabev. in subgroup Mollis and *D. tsiandalana* in subgroup Pervillei.

Overall, we found high levels of congruence between the genetic, ecogeographic and morphological datasets (Figures S5 – S16), although the different datasets yielded

contrasting levels of resolution depending on the particular case (Figure S23). The assessment of 20 leaf morphological characters within genetic subgroups yielded a generally high discrimination success based on the included PS, except for subgroups Peltieri and Chapelieri (Table S6). These two subgroups contain one to several UCS, which can be distinguished with considerably higher accuracy by integrating leaf morphology with ecogeography (Table 3).

Taxonomic implications

We could confirm 46 of the 49 currently accepted taxonomic species of Malagasy and Comorian *Dalbergia* as separately evolving lineags (evolutionary species). Three species that represent currently accepted taxa were revealed as potential synonyms of other taxa accepted by BOSSER & RABEVOHITRA (1996, 2002, 2005). Firstly, *D. campenonii* Bosser & R. Rabev. most likely represents a synonym of *D. emirnensis* Benth. The two taxonomic species are hardly distinguishable morphologically (RAKOTONIRINA *et al.* in prep.) and are indistinguishable genetically, forming a grade of two collections each, and their respective ecogeographic distributions overlap (Figure 2, Figure S16). Secondly, *D. masoalensis* possibly represents a synonym of *D. bojeri* Drake. The NJ tree as well as the individual-based ASTRAL summary topology revelaed the type specimen of *D. masoalensis* to be nested within a *D. bojeri* clade. Both taxonomic species share a shrubby growth habit (often lianescent in *D. bojeri*), have very similar leaves, and overlap in their ecogeography (Figure 2). Thirdly, the type specimen of *D. brachystachya* Bosser & R. Rabev., which is the only known collection of this taxon, cannot be confirmed as distinct from the type specimen of *D. capuronii* Bosser & R. Rabev. based on the current sampling (Figure S7). The two taxonomic species were described and distinguished based on the unusually short and spicate inflorescence of the *D. brachystachya* type, which was considered to be a unique characteristic among Malagasy *Dalbergia* by BOSSER & RABEVOHITRA (2002). However, *D. brachystachya* and *D. capuronii* have very similar flowers, and *D. brachystachya* has not yet been confirmed by any other collection. Moreover, there is a lack of genetic differentiation between the *D. brachystachya* type specimen and *D. capuronii* (Figure S7). Therefore, our current hypothesis is that the inflorescence seen on the type of *D. brachystachya* is an immature or malformed inflorescence of *D. capuronii*.

We identified 14 confirmed new candidate species in our study. Nine of these were represented by one or just a few specimens within six broadly circumscribed species recognised by BOSSER & RABEVOHITRA (1996, 2005), but they are distinct from the entities corresponding to the type collections. These six broadly circumscribed taxonomic species might therefore need emended descriptions: *D. greveana* Baill. (includes *D.* sp.

34), *D. lemurica* Bosser & R. Rabev. (includes *D.* sp. 01), *D. maritima* (includes *D.* sp. 06 and *D.* sp. 27, as well as *D.* sp. 07 from another subgroup), *D. madagascariensis* Vatke (includes *D.* sp. 16), *D. neoperrieri* Bosser & R. Rabev. (includes *D.* sp. 08 from subgroup Arenosa), and *D. pervillei* Vatke (currently includes *D. densicoma* and *D. obtusa* Lecomte, two separately evolving lineages, as synonyms), see Figure S20. A more limited number of available collections and the absence of a genetic perspective likely explains why these nine confirmed new candidate species were not previously recognised as distinct. Four confirmed new candidate species (*D.* sp. 02, *D.* sp. 03, *D.* sp. 10 and *D.* sp. 17) were either never collected before 2005, or the few available collections were not seen by Bosser & R. Rabev., or doubtfully assigned to other species. Finally, *D.* sp. 58 was recognised as a distinct taxon by BOSSER & RABEVOHITRA (1996), but it was treated as a variety of *D. mollis* rather than separate species, which necessitates taxonomic changes (WILDING *et al.* in prep.-b).

A comparatively large number of PS (31 or 33%) remain unconfirmed. These UCS probably include populations showing isolation by distance from related populations, such as *D. monticola* from central-north Madagascar and the morphologically similar *D.* sp. 37 from north Madagascar. More collections from between the respective distributions are likely required to conclusively assess the lineage separation in such cases. However, we anticipate that multiple UCS will be upgraded to CCS once more collections and fertile material becomes available.

Ecogeographic diversity

Clustering of the ecological resource matrix revealed three main ecological groups of species (Figure 2), which largely correspond to the humid, subhumid and dry ecoregions. Occurrence frequencies in the succulent and spiny ecoregions were generally low (Table 4), except for *D. lemurica*, and *D.* sp. 08 (succulent woodlands) and for *D. emirnensis* var. *decaryi*, *D. mollis* var. *menabeensis* and *D. xerophila* (spiny thickets). These species belong to different genetic subgroups, indicating independent adaptations to the arid conditions in south-west and south Madagascar. In other cases, we found correlations between the genetic and ecological groups. Firstly, subgroup Tricolor contains multiple species that appear to be restricted to the subhumid ecoregion. Secondly, all of subgroup Maritima appears to be restricted to the eastern humid ecoregion. Lastly, subgroup Chlorocarpa mainly occurs in dry deciduous forests, a notable exception being *D. trichocarpa*, which appears to mainly occur in West Madagascar, but which was recently and unexpectedly collected in the littoral forest of Vohibola in central-east Madagascar (Figure 1B3).

Hotspots of species richness and phylogenetic diversity

The region between Ankarana and Loky-Manambato (Daraina), and the northern tip of Madagascar represent centres of high species richness and phylogenetic diversity (Figure 3). Several protected areas are located in this region and include humid forest (e.g. in the Montagne d'Ambre National Park) and dry deciduous forest (e.g. in the Ankarana Special Reserve). Particularly the protected area of Loky-Manambato is characterised by a mosaic of communities typical of the eastern humid, central subhumid and western dry ecoregions, and this region is known to harbour an exceptional plant diversity (NUSBAUMER *et al.* 2010). Another centre of high species richness and phylogenetic diversity is a region in west Madagascar, where the Ankarafantsika national park is located. This region lies within a single ecoregion, but it is characterised by a diverse surface lithology (Figure 1B2), and mosaics of dense western dry forest and wooded grassland are also present in the area. Together this suggests that the hotspots of species richness and phylogenetic diversity in Malagasy *Dalbergia* coincide with the presence of multiple ecotones.

The taxonomic validation and georeferencing of 1,846 collections revealed that several species cannot currently be confirmed to occur in Madagascar's network of protected areas, notably *Dalbergia aurea*, *D. louvelii* R. Vig., and *D. neoperrieri* Bosser & R. Rabev. (Figure 2). These species are only known from a few confirmed collections, and targeted searches for them are needed to evaluate their risks of extinction.

Conclusions and perspectives

The integrative taxonomic workflow continues (YEATES *et al.* 2010). Ongoing field work is producing collections from poorly studied regions, improving the understanding of plastic versus heritable or clinal variation in morphological characters, and providing new insights through fertile and informative material. This new information will provide opportunities to systematically collect morphological and genetic data of more taxa, to test existing species hypotheses or to formulate new ones. Also, the present phylogenetic backbone supports the discovery of possible synapomorphies, which should facilitate the development of effective identification keys.

To have a conservation impact, it is important to complete the workflow from species discovery to species delimitation and species description (CARSTENS *et al.* 2013, PANTE *et al.* 2015). We and our colleagues from the University of Antananarivo, Missouri Botanical Garden and the Muséum national d'Histoire naturelle in Paris are currently preparing formal species descriptions for several confirmed new candidate species, as well as emended descriptions for currently accepted taxonomic species that were here found to

be polyphyletic and/or contain multiple confirmed evolutionary species (CRAMERI *et al.* in prep.-a, RAKOTONIRINA *et al.* in prep., WILDING *et al.* in prep.-a). These formalised species hypotheses ultimately form the basis of Red List assessments (IUCN 2012) and conservation efforts. Unfortunately, Red Listing and the designation of protected areas does not prevent illegal logging (PATEL 2007, WAEBER *et al.* 2019). Our goal is therefore to apply novel taxonomic hypotheses to improve the quality of a growing reference database of Malagasy *Dalbergia* timber species for forensic timber identification. This reference database is being built and curated by an international consortium for research in precious woods from Madagascar, and consists of genetic and morphological data as well as anatomical, spectroscopic and spectrometric characteristics of the heartwood (see DORMONTT *et al.* 2015). New insights into species diversity brought about by the present study therefore have the potential to support ongoing conservation impact.

Acknowledgements

We are greatful to the botanists of the Missouri Botanical Garden Madagascar program (notably to Roger Bernard, Charles Rakotovao, Richard Randrianaivo and Richardson Razakamalala), Sonja Hassold, as well as to undergraduate and doctroal students of the University of Antananarivo for conducting a staggering amount of field collecting throughout Madagascar. We also thank the G3D team for the organisation of field work and sample dispatch. We are further greatful to Claudia Michel for conducting an enormous amount of laboratory work, to Michael Zehnder and Heidi Amstalden for help with morphometric measurements, and to Niklaus Zemp and Stefan Zoller for bioinformatics support. Sonja Hassold, Martin Fischer, and Simone Fior provided helpful inputs and comments on this manuscript.

Data accessibility

Raw sequence reads and a VCF file containing SNP data for 810 genotyped samples, and alignments for 296 individuals will be added on the European Nucleotide Archive (ENA) under the primary accession number PRJEB42629. Voucher specimen information for 1,089 individuals (Table S1), sequencing and mapping statistics for 810 samples (Table S2), morphometric measurements (Table S3) and ecogeographic characteristics (Table S4) for 1,846 individuals are available as Supporting Online Material at https://github.com/scrameri/DalbergiaSpeciesDelimitation.

References

Adema, F., H. Ohashi, and B. Sunarno. 2016. Notes on Malesian Fabaceae (Leguminosae-Papilionoideae) 17. The genus *Dalbergia*. Blumea **61**:186-206.

Barrett, M. A., J. L. Brown, M. K. Morikawa, J.-N. Labat, and A. D. Yoder. 2010. Conservation. CITES designation for endangered rosewood in Madagascar. Science **328**:1109-1110.

Bell, R. C., D. G. Mulcahy, S. W. Gotte, A. J. Maley, C. Mendoza, G. Steffensen, J. C. Barron II, O. Hyman, W. Flint, A. Wynn, R. W. Mcdiarmid, and D. S. Mcleod. 2020. The Type Locality Project: collecting genomic-quality, topotypic vouchers and training the next generation of specimen-based researchers. Systematic and Biodiveristy **18**:557-572.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology and Evolution **22**:148-155.

Bivand, R. S., E. J. Pebesma, and V. Gomez-Rubio. 2013. Applied Spatial Data Analysis with R. 2 edition. Springer, New York.

Bonhomme, V., S. Picq, and C. Gaucherel. 2014. Momocs: outline analysis using R. Journal of Statistical Software **56**:1-24.

Bosser, J., and R. Rabevohitra. 1996. Taxa et noms nouveaux dans le genre *Dalbergia* (Papilionaceae) à Madagascar et aux Comores. Bull. Mus. Natl. Hist. Nat. B, Adansonia **18**:171-212.

Bosser, J., and R. Rabevohitra. 2002. Tribe Dalbergieae. Pages 321-361 *in* D. J. Du Puy, J. N. Labat, R. Rabevohitra, J. F. Villiers, J. Bosser, and J. Moat, editors. The Leguminosae of Madagascar. Royal Botanical Gardens, Kew.

Bosser, J., and R. Rabevohitra. 2005. Espèces nouvelles dans le genre *Dalbergia* (Fabaceae, Papilionoideae) à Madagascar. Adansonia **27**:209-216.

Breiman, L. 2001. Random forests. Machine Learning **45**:5-32.

Carstens, B. C., T. A. Pelletier, N. M. Reid, and J. D. Satler. 2013. How to fail at species delimitation. Molecular Ecology **22**:4369-4383.

Chao, A., N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, and A. M. Ellison. 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecological Monographs **84**:45-67.

Cheng, J., B. Karambelkar, and Y. Xie. 2019. leaflet: Create Interactive Web Maps with the JavaScript „Leaflet“ Library. R package version 2.0.3. https://CRAN.R-project.org/package=leaflet.

CITES. 2019. Appendices I, II and III, https://cites.org/eng/app/appendices.php accessed on Feb 13 2020.

Crameri, S., P. B. Phillipson, N. H. Rakotonirina, N. Wilding, R. L. Andriamiarisoa, P. P. Lowry II, and A. Widmer. in prep.-a. Two new species from southeastern Madagascar and an emended description of the rosewood species *Dalbergia maritima* R. Vig.

Crameri, S., S. Zoller, and A. Widmer. in prep.-b. A target enrichment approach for phylogenomic analyses at multiple evolutionary timescales in rosewoods (*Dalbergia* spp.) and the legume family (Fabaceae).

Crandall, K. A., O. Bininda-Emonds, G. M. Mace, and R. K. Wayne. 2000. Considering evolutionary processes in conservation biology. Trends in Ecology and Evolution **15**:290-295.

Dauby, G., T. Stévart, V. Droissart, A. Cosiaux, V. Deblauwe, M. Simo-Droissart, M. S. M. Sosef, P. P. Lowry II, G. E. Schatz, R. E. Gereau, and T. L. P. Couvreur. 2017. ConR: An R package to assist large-scale multispecies preliminary conservation assessments using distribution data. Ecology and Evolution **7**:11292-11303.

Dayrat, B. 2005. Towards integrative taxonomy. Biological Journal of the Linnean Society **85**:407-415.

de Carvalho, A. M. 1997. A synopsis of the genus *Dalbergia* (Fabaceae: Dalbergieae) in Brazil. Brittonia **49**:87-109.

De Queiroz, K. 2005a. Ernst Mayr and the modern concept of species. Proceedings of the National Academy of Sciences of the United States of America **102**:6600-6607.

de Queiroz, K. 2005b. A Unified Concept of Species and Its Consequences for the Future of Taxonomy. Proceedings of the California Academy of Sciences **56**:196-215.

De Queiroz, K. 2007. Species concepts and species delimitation. Systematic Biology **56**:879-886.

Dinerstein, E., D. Olson, A. Joshi, C. Vynne, N. D. Burgess, E. Wikramanayake, N. Hahn, S. Palminteri, P. Hedao, R. Noss, M. Hansen, H. Locke, E. C. Ellis, B. Jones, C. V. Barber, R. Hayes, C. Kormos, V. Martin, E. Crist, W. Sechrest, L. Price, J. E. M. Baillie, D. Weeden, K. Suckling, C. Davis, N. Sizer, R. Moore, D. Thau, T. Birch, P. Potapov, S. Turubanova, A. Tyukavina, N. de Souza, L. Pintea, J. C. Brito, O. A. Llewellyn, A. G. Miller, A. Patzelt, S. A. Ghazanfar, J. Timberlake, H. Klöser, Y. Shennan-Farpón, R. Kindt, J.-P. B. Lillesø, P. van Breugel, L. Graudal, M. Voge, K. F. Al-Shammari, and M. Saleem. 2017. An ecoregion-based approach to protecting half the terrestrial realm. BioScience **67**:534-545.

Dormontt, E. E., M. Boner, B. Braun, G. Breulmann, B. Degen, E. Espinoza, S. Gardner, P. Guillery, J. C. Hermanson, G. Koch, S. L. Lee, M. Kanashiro, A. Rimbawanto, D. Thomas, A. C. Wiedenhoeft, Y. Yin, J. Zahnen, and A. J. Lowe. 2015. Forensic timber identification: it's time to integrate disciplines to combat illegal logging. Biological Conservation **191**:790-798.

Erickson, K. L., A. Pentico, A. M. Quattrini, and C. S. McFadden. 2020. New approaches to species delimitation and population structure of anthozoans: Two case studies of octocorals using ultraconserved elements and exons. Molecular Ecology Resources **57**:doi:10.1111/1755-0998.13241.

Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation **61**:1-10.

Gaston, K. J., and J. I. Spicer. 2013. Biodiversity: an introduction. John Wiley & Sons, Chicester, UK.

Ghiselin, M. T. 2001. Species concepts. Pages 1-5 Encyclopedia of life sciences. Wiley, New York.

Gueuning, M., J. E. Frey, and C. Praz. 2020. Ultraconserved yet informative for species delimitation: Ultraconserved elements resolve long-standing systematic enigma in Central European bees. Molecular Ecology **20**:165-118.

Hartigan, J. A., and M. A. Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society Series C Applied Statistics **28**:100-109.

Hassold, S., S. Crameri, H. Ravaomanalina, P. P. Lowry II, and A. Widmer. 2018. Creating a base for rosewood identification. ITTO Tropical Forest Update **27**:4-7.

Hassold, S., P. P. Lowry II, M. R. Bauert, A. Razafintsalama, L. Ramamonjisoa, and A. Widmer. 2016. DNA barcoding of Malagasy rosewoods: towards a molecular identification of CITES-listed *Dalbergia* species. PLoS ONE: doi: 10.1371-journal.pone.0157881.

Hey, J., and C. Pinho. 2012. Population genetics and objectivity in species diagnosis. Evolution **66**:1413-1429.

Hijmans, R. J., and J. van Etten. 2012. raster: Geographic analysis and modeling with raster data. R package version 2.0-12. http://CRAN.R-project.org/package=raster.

IUCN. 2012. IUCN Red List Categories and Criteria. Version 3.1. Second edition. IUCN Species Survival Commission, Gland & Cambridge.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics **24**:1403-1405.

Jombart, T., and I. Ahmed. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics **27**:3070-3071.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics **11**:doi: 10.1186-1471-2156-1111-1194.

Jones, M. R., and J. M. Good. 2016. Targeted capture in evolutionary and ecological genomics. Molecular Ecology **25**:185-202.

Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ **2**:doi: 10.7717-peerj.7281.

Karbstein, K., S. Tomasello, L. Hodac, F. G. Dunkel, M. Daubert, and E. Hörandl. 2020. Phylogenomics supported by geometric morphometrics reveals delimitation of sexual species within the polyploid apomictic *Ranunculus auricomus* complex (Ranunculaceae). bioRxiv: doi.org/10.1101/2020.1101.1107.896902.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics **26**:1463-1464.

Knaus, B. J., and N. J. Grünwald. 2017. vcfr: a package to manipulate and visualize variant call format data in R. Molecular Ecology Resources **17**:44-53.

Kuhl, F. P., and C. R. Giardina. 1982. Elliptic Fourier features of a closed contour. Computer Graphics and Image Processing **18**:236-258.

Kuhn, M. 2020. caret: classification and regression training. R package version 6.0-86. https://CRAN.R-project.org/package=caret.

Le Péchon, T., S. Traclet, F. Barthelat, A. Dimassi, L. D. B. Gigord, A. Hamidou, L. Jourdain-Fievet, G. Viscard, and V. Boullet. 2020. Bridging the Mozambique Channel: first record of *Dombeya rosacea* (Dombeyoideae, Malvaceae) for Mayotte (Comoros Archipelago, Indian Ocean). Phytotaxa **460**:75-88.

Levins, R. 1968. Evolution in changing environments. Princeton University Press, Princeton, New Jersey, USA.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1302.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News **2**:18-22.

Mace, G. M. 2004. The role of taxonomy in species conservation. Philosophical Transactions of the Royal Society B Biological Sciences **359**:711-719.

Madagascar Catalogue. 2020. Catalogue of the Plants of Madagascar. Missouri Botanical Garden, St. Louis, USA. http://www.tropicos.org/Project/Madagascar accessed on June 6 2020.

Moat, J., and P. Smith, editors. 2007. Atlas of the Vegetation of Madagascar. Kew Publishing, Royal Botanic Gardens, Kew.

Musher, L. J., and J. Cracraft. 2018. Phylogenomics and species delimitation of a complex radiation of Neotropical suboscine birds (*Pachyramphus*). Molecular phylogenetics and evolution **118**:204-221.

Natusch, D. J. D., D. Esquerré, J. A. Lyons, A. Hamidy, A. R. Lemmon, E. M. Lemmon, A. Riyanto, J. S. Keogh, and S. Donnellan. 2020. Species delimitation and systematics of the green pythons (*Morelia viridis* complex) of melanesia and Australia. Molecular Phylogenetics and Evolution **142**:106640.

Nei, M. 1972. Genetic distance between populations. The American Naturalist **106**:283-292.

Nusbaumer, L., P. Ranirison, L. Gautier, C. Chatelain, P.-A. Loizeau, and R. Spichiger. 2010. Loky-Manambato: point de rencontre des principales unités phytogéographiques de Madagascar. Pages 253-264 *in* X. van der Burgt, J. van der Maesen, and J. M. Onana, editors. Royal Botanic Gardens, Kew.

Padial, J. M., A. Miralles, I. De la Riva, and M. Vences. 2010. The integrative future of taxonomy. Frontiers in Zoology **7**:1-14.

Pante, E., C. Schoelinck, and N. Puillandre. 2015. From integrative taxonomy to species description: one step beyond. Systematic Biology **64**:152-160.

Paradis, E., and K. Schliep. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics **35**:526-528.

Patel, E. R. 2007. Logging of rare rosewood and palisandre (*Dalbergia* spp.) within Marojejy National Park, Madagascar. Madagascar Conservation and Development.

Pau, G., F. Fuchs, O. Sklyar, M. Boutros, and W. Huber. 2010. EBImage -an R package for image processing with applications to cellular phenotypes. Bioinformatics **26**:979-981.

Pie, M. R., M. R. Bornschein, L. F. Ribeiro, B. C. Faircloth, and J. E. McCormack. 2019. Phylogenomic species delimitation in microendemic frogs of the Brazilian Atlantic Forest. Molecular phylogenetics and evolution **141**:106627.

Prain, D. 1904. The species of *Dalbergia* of south-eastern Asia. Annals of the Royal Botanic Garden, Calcutta **10**:1-411.

Prata, E. M. B., C. Sass, D. P. Rodrigues, F. M. C. B. Domingos, C. D. Specht, G. Damasco, C. C. Ribas, P. V. A. Fine, and A. Vincentini. 2018. Towards integrative taxonomy in Neotropical botany: disentangling the Pagamea guianensis species complex (Rubiaceae). Biological Journal of the Linnean Society **188**:233-231.

Probst, P., M. N. Wright, and A. L. Boulesteix. 2018. Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **9**:281-215.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Rakotonirina, N. H., P. B. Phillipson, and et al. in prep. The taxonomy of *Dalbergia emirnensis* (Fabaceae) and related species from Madagascar.

Reydon, T. A. C. 2019. Are species good units for biodiversity studies and conservation efforts? Pages 167-193 *in* E. Casetta, J. M. da Silva, and D. Vecchi, editors. Springer Nature, Cham, Switzerland.

Ribeiro, R. A., M. Lavin, J. P. Lemos-Filho, and C. Filho. 2007. The genus *Machaerium* (Leguminosae) is more closely related to *Aeschynomene* Sect. *Ochopodium* than to *Dalbergia*: Inferences from combined sequence data. Systematic Botany **32**:762-771.

Ryder, O. A. 1986. Species conservation and systematics: the dilemma of subspecies. Trends in Ecology and Evolution **1**:9-10.

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

Schlick-Steiner, B. C., F. M. Steiner, B. Seifert, C. Stauffer, E. Christian, and R. H. Crozier. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. Annual Review of Entomology **55**:421-438.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods **9**:671-675.

Schuurman, D., and P. P. Lowry II. 2009. The Madagascar rosewood massacre. Madagascar Conservation and Development **4**:98-102.

Slowikowski, K. 2020. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.8.2. https://CRAN.R-project.org/package=ggrepel.

Song, S., J. Zhao, and C. Li. 2017. Species delimitation and phylogenetic reconstruction of the sinipercids (Perciformes: Sinipercidae) based on target enrichment of thousands of nuclear coding sequences. Molecular phylogenetics and evolution **111**:44-55.

Tennekes, M. 2018. tmap: Thematic Maps in R. Journal of Statistical Software **84**:1-39.

Turland, N. J., J. H. Wiersema, F. R. Barrie, W. Greuter, D. L. Hawksworth, P. S. Herendeen, S. Knapp, W.-H. Kusber, D.-Z. Li, K. Marhold, M. T. W., J. McNeill, A. M. Monro, J. Prado, M. J. Price, and G. F. Smith. 2018. International Code of Nomenclature for algae, fungi, and plants (Shenzhen code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Regnum Vegetabile 159. Glashütten: Koeltz Botanical Books, https://doi.org/10.12705/Code.2018.

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

Vieilledent, G., C. Grinand, F. Rakotomalala, R. Ranaivosoa, J.-R. Rakotoarijaona, T. F. Allnutt, and F. Achard. 2018. Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. Biological Conservation **222**:189-197.

Vieites, D. R., K. C. Wollenberg, F. Andreone, J. Köhler, F. Glaw, and M. Vences. 2009. Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. Proceedings of the National Academy of Sciences of the United States of America **106**:8267-8272.

Wachter, G. A., C. Muster, W. Arthofer, G. Raspotnig, P. Föttinger, C. Komposch, F. M. Steiner, and B. C. Schlick-Steiner. 2015. Taking the discovery approach in integrative taxonomy: decrypting a complex of narrow-endemic Alpine harvestmen (Opiliones: Phalangiidae: Megabunus). Molecular Ecology **24**:863-889.

Waeber, P. O., D. Schuurman, B. Ramamonjisoa, M. Langrand, C. V. Barber, J. L. Innes, P. P. Lowry, and L. Wilmé. 2019. Uplisting of Malagasy precious woods critical for their survival. Biological Conservation **235**:89-92.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution **38**:1358-1370.

Wickham, H. 2016. ggplot2: elegant graphics for data analysis. 2 edition. Springer-Verlag, New York.

Wilding, N., P. B. Phillipson, S. Andriambololonera, R. Bernard, S. Crameri, N. Rakotonirina, C. Rakotovao, R. Randrianaivo, R. Razakamalala, and P. P. Lowry II. in prep.-a. Taxonomic studies on Malagasy Dalbergia L.f. (Fabaceae) I: Two new species from Northern Madagascar, and an emended description for a third.

Wilding, N., P. B. Phillipson, and S. Crameri. in prep.-b. Untangling the taxonomy of two Malagasy *Dalbergia* (Fabaceae): *D. chermezonii* R. Vig. & *D. mollis* Bosser & R. Rabev.

Wiley, E. O. 1978. The evolutionary species concept reconsidered. Systematic Biology **27**:17-26.

Yeates, D. K., A. Seago, L. Nelson, S. Cameron, L. Joseph, and J. W. H. Trueman. 2010. Integrative taxonomy, or iterative taxonomy? Systematic Entomology **36**:209-217.

Younger, J. L., L. Strozier, J. Dylan Maddox, Á. S. Nyári, M. T. Bonfitto, M. J. Raherilalao, S. M. Goodman, and S. Reddy. 2018. Hidden diversity of forest birds in Madagascar revealed using integrative taxonomy. Molecular Phylogenetics and Evolution **124**:16-26.

Zachos, F. E. 2016. Species concepts in Biology. Springer, Switzerland.

Zarza, E., E. M. Connors, J. M. Maley, W. L. E. Tsai, P. Heimes, M. Kaplan, and J. E. McCormack. 2017. Bridging multilocus species delimitation and DNA barcoding through target enrichment of UCEs: A case study with Mexican highland frogs. bioRxiv: doi.org/10.1101/153601.

Zhang, J. 2016. spaa: SPecies Association Analysis. R package version 0.2.2. https://CRAN.R-project.org/package=spaa.

Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics **28**:3326-3328.

Appendix II

Supplementary Figures

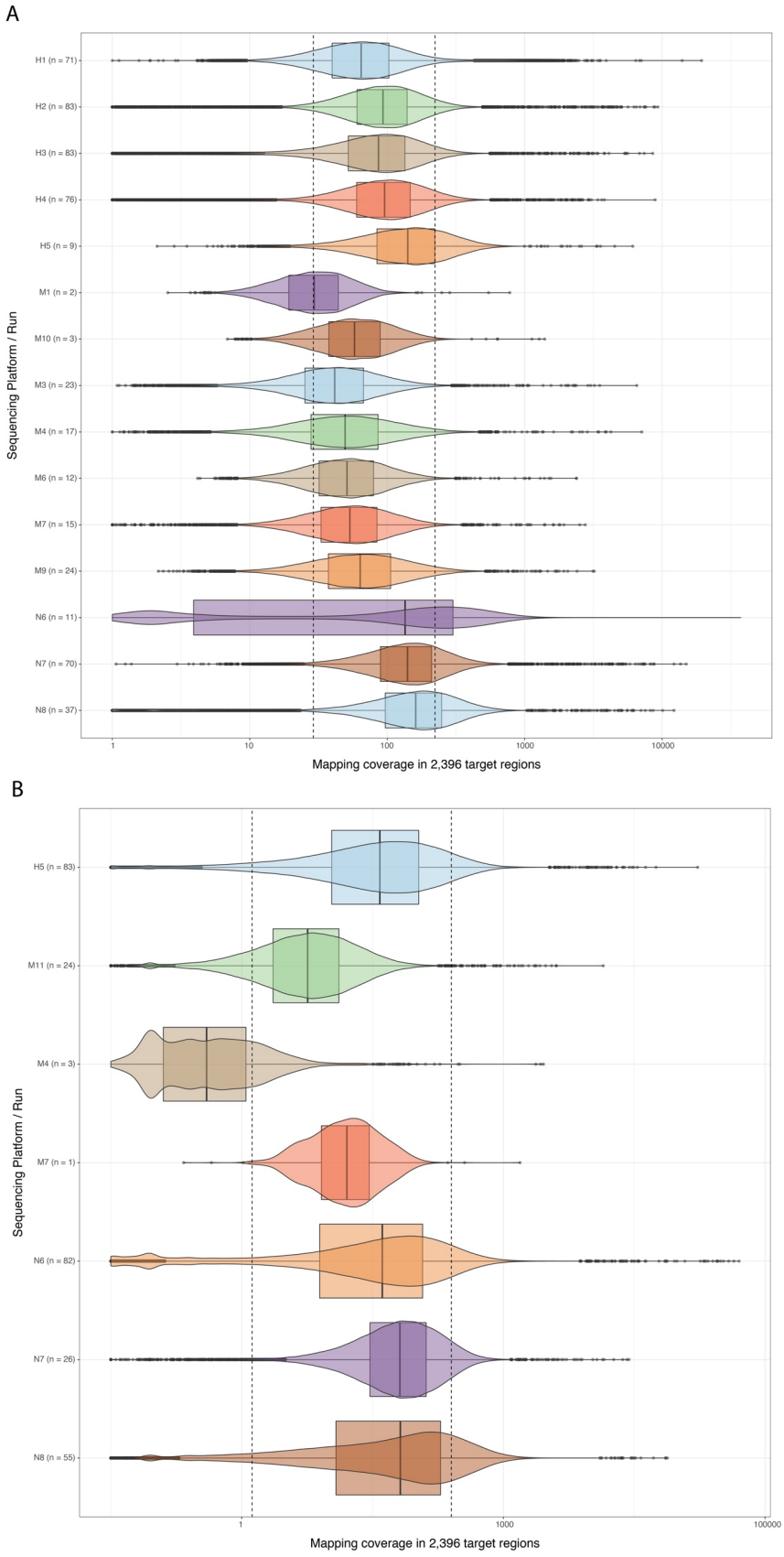

FIGURE S1: Mapping coverage heatmaps. **A)** 536 sequenced silicagel-dried samples, **B)** 274 sequenced herbarium samples. Vertical dotted lines denote the 10% and 90% quantiles, respectively. Sequencing platforms H = HiSeq; M = MiSeq; N = NovaSeq. Note the logarithmic scales.

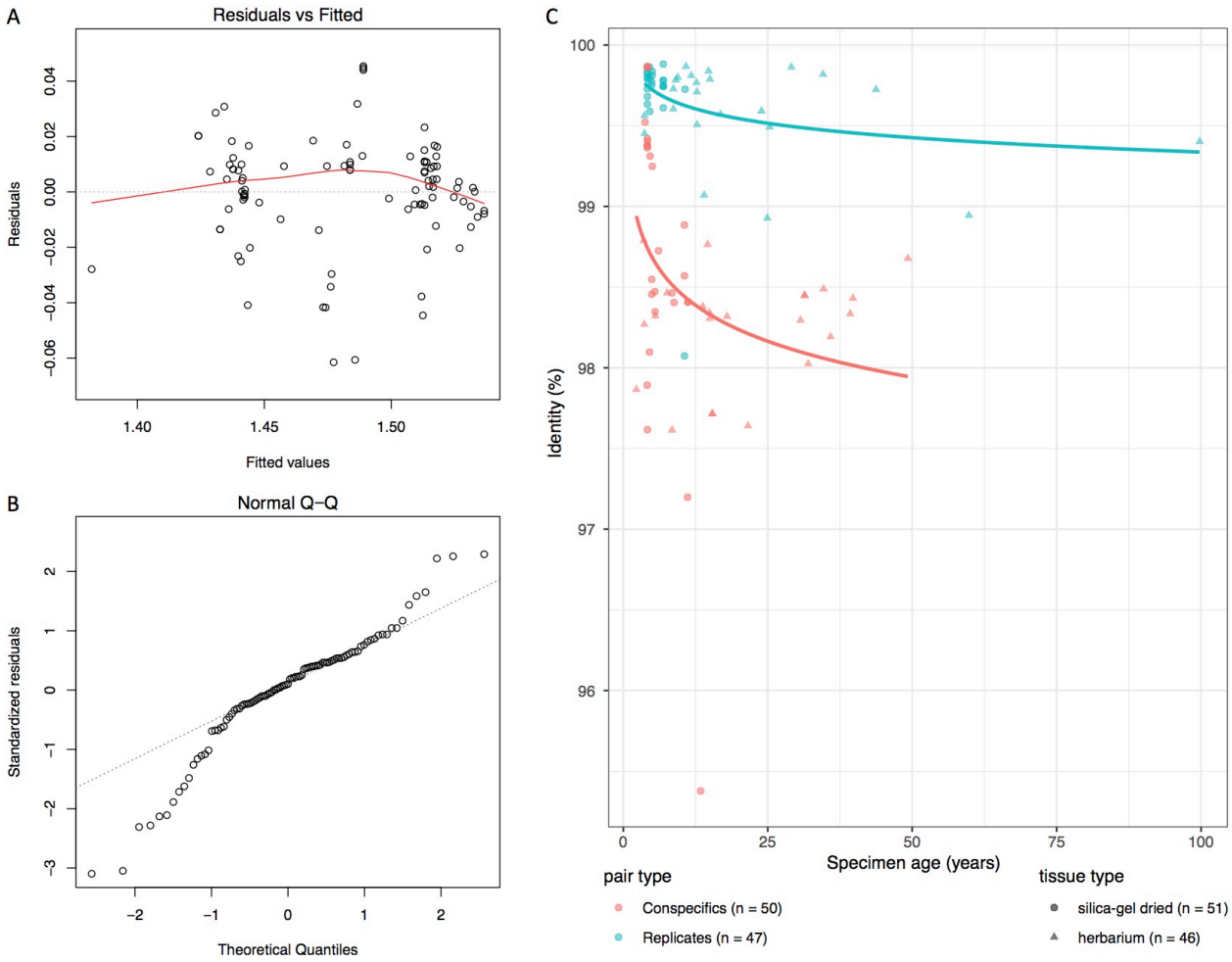

FIGURE S2: Assessment of reproducibility of target enrichment and sequencing. **A)** Tukey-Anscombe plot showing homogeneity of model residual variance. **B)** Normal quantile-quantile plot showing an approximate normal distribution of model residuals. **C)** Relationship between pairwise identity, specimen age, pair type (conspecific or replicate) and tissue type (silica-gel dried or herbarium).

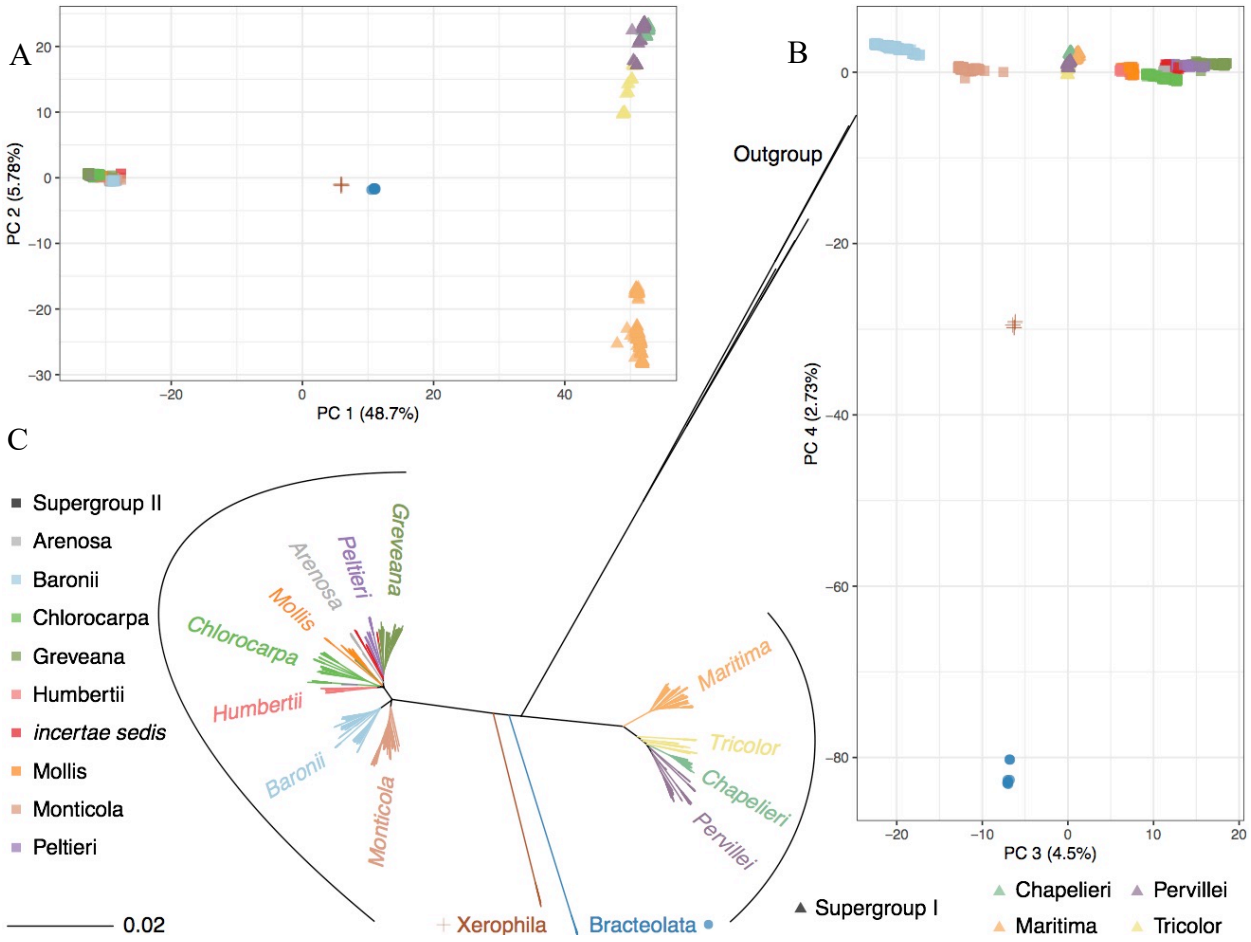

FIGURE S3: Species discovery. **A)** PCA scatterplot of the first and second principal components of 168,278 bi-allelic SNPs (no missingness allowed) in 697 samples of *Dalbergia* (694 from Madagascar, two from Mozambique and one from the Comoros). **B)** PCA scatterplot of the third and fourth principal components. **C)** NJ tree of 162,631 bi-allelic SNPs of the same ingroup, and *Machaerium* spp. and *Aeschynomene* Sect. *Ochopodium* spp. as outgroups for rooting.

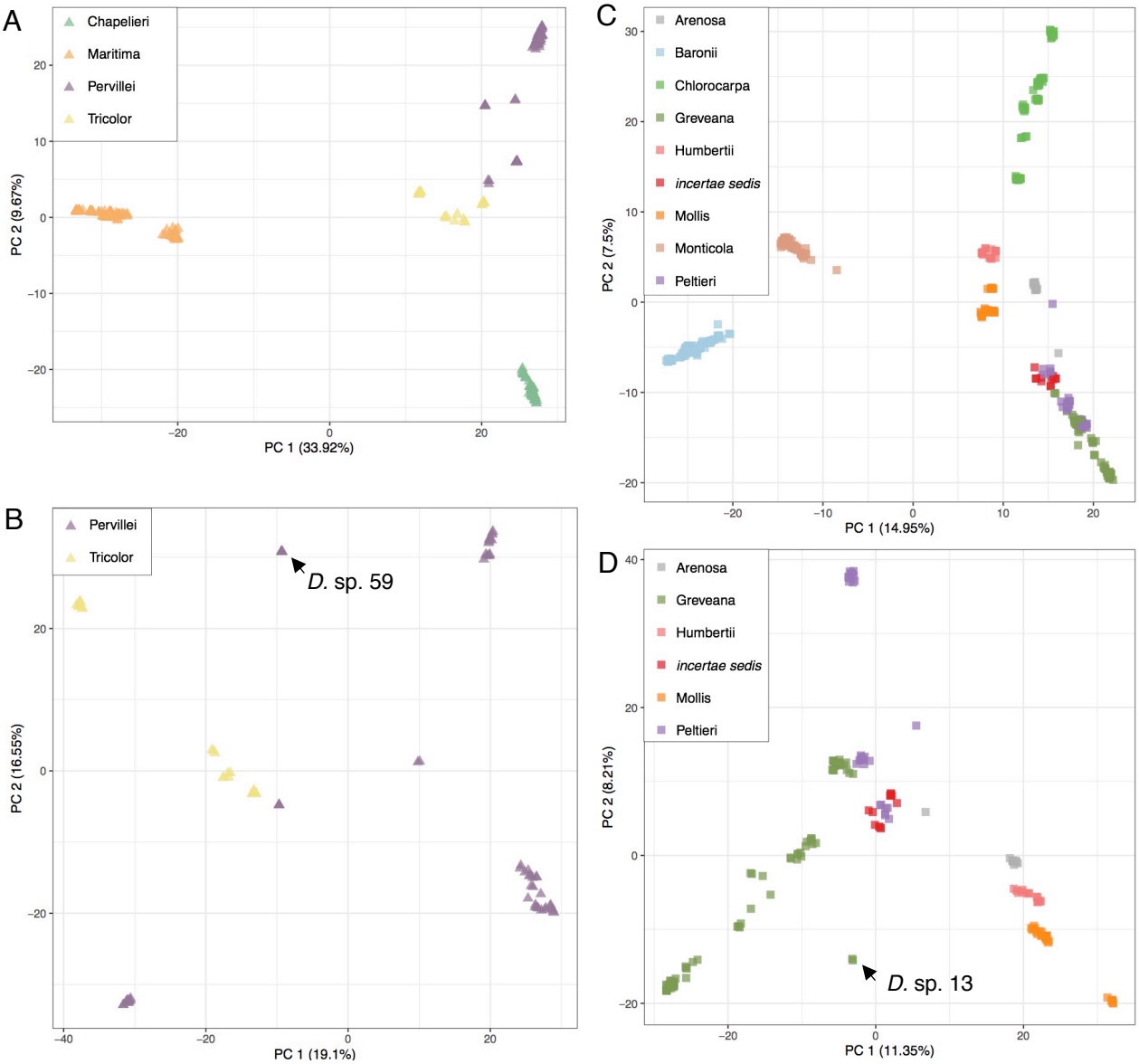

FIGURE S4: PCA of subsets of the SNP dataset, with no missingness allowed. **A)** supergroup I, **B)** subgroups Pervillei and Tricolor (supergroup I), **C)** supergroup II, **D)** crown group of supergroup II. Putative hybrid (PH) individuals between parent species from different subgroups, and showing an intermediate phenotype, are marked with labelled black arrowheads.

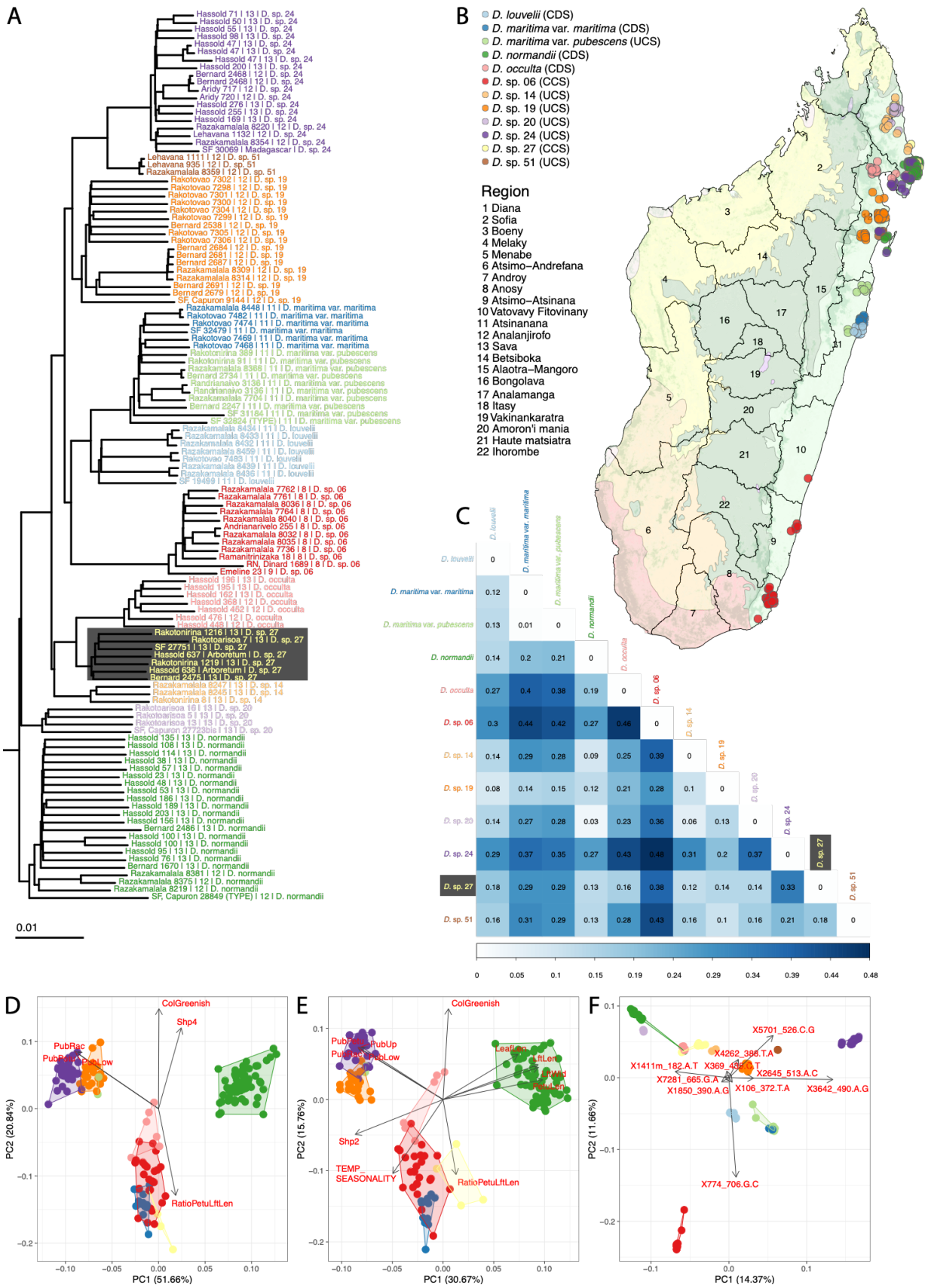

FIGURE S5: Integrative species delimitation in subgroup Maritima. **A)** NJ tree of 90,087 biallelic SNPs (1% allowed missingness), rooted on the outgroup *D. tricolor* (root not shown). Integers after the collection name denote the region of origin (see panel B). **B)** Distribution of verified collections. Coordinates are slightly jittered to reduce overlaps. **C)** Weir & Cockerham's F_{st} between all PS. **D)** PCA of scaled morphological variables. **E)** PCA of scaled morphological and ecological variables. **F)** PCA of 93,969 biallelic SNPs excluding the outgroup, with a minimum minor allele count of 2 and 5% allowed missingness.

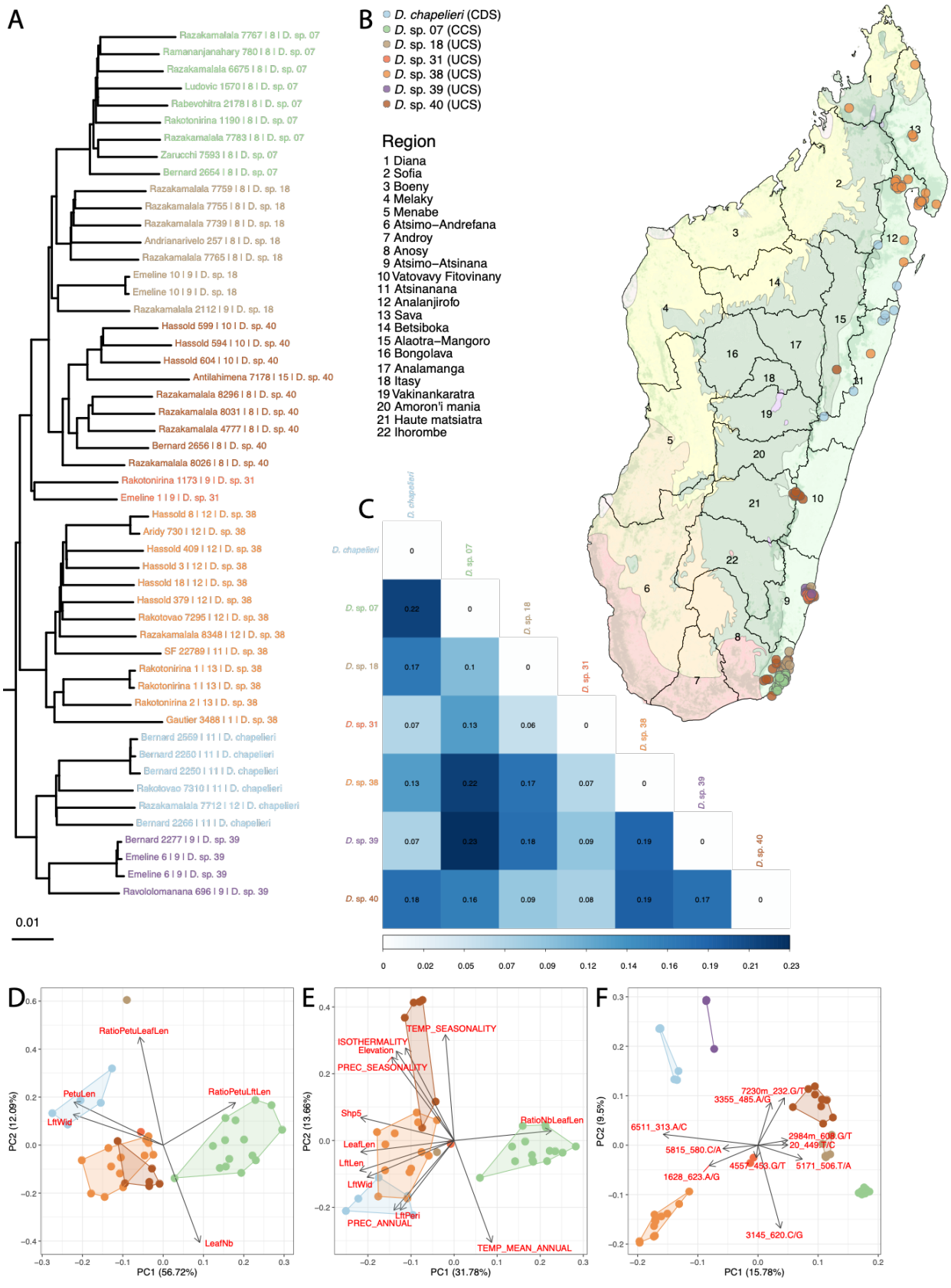

FIGURE S6: Integrative species delimitation in subgroup Chapelieri. **A)** NJ tree of 54,136 biallelic SNPs (1% allowed missingness), rooted on the outgroup *D. obtusa* (root not shown). Integers after the collection name denote the region of origin (see panel B). **B)** Distribution of verified collections. Coordinates are slightly jittered to reduce overlaps. **C)** Weir & Cockerham's F_{st} between all PS. **D)** PCA of scaled morphological variables. **E)** PCA of scaled morphological and ecological variables. **F)** PCA of 41,581 biallelic SNPs excluding the outgroup, with a minimum minor allele count of 2 and 5% allowed missingness.

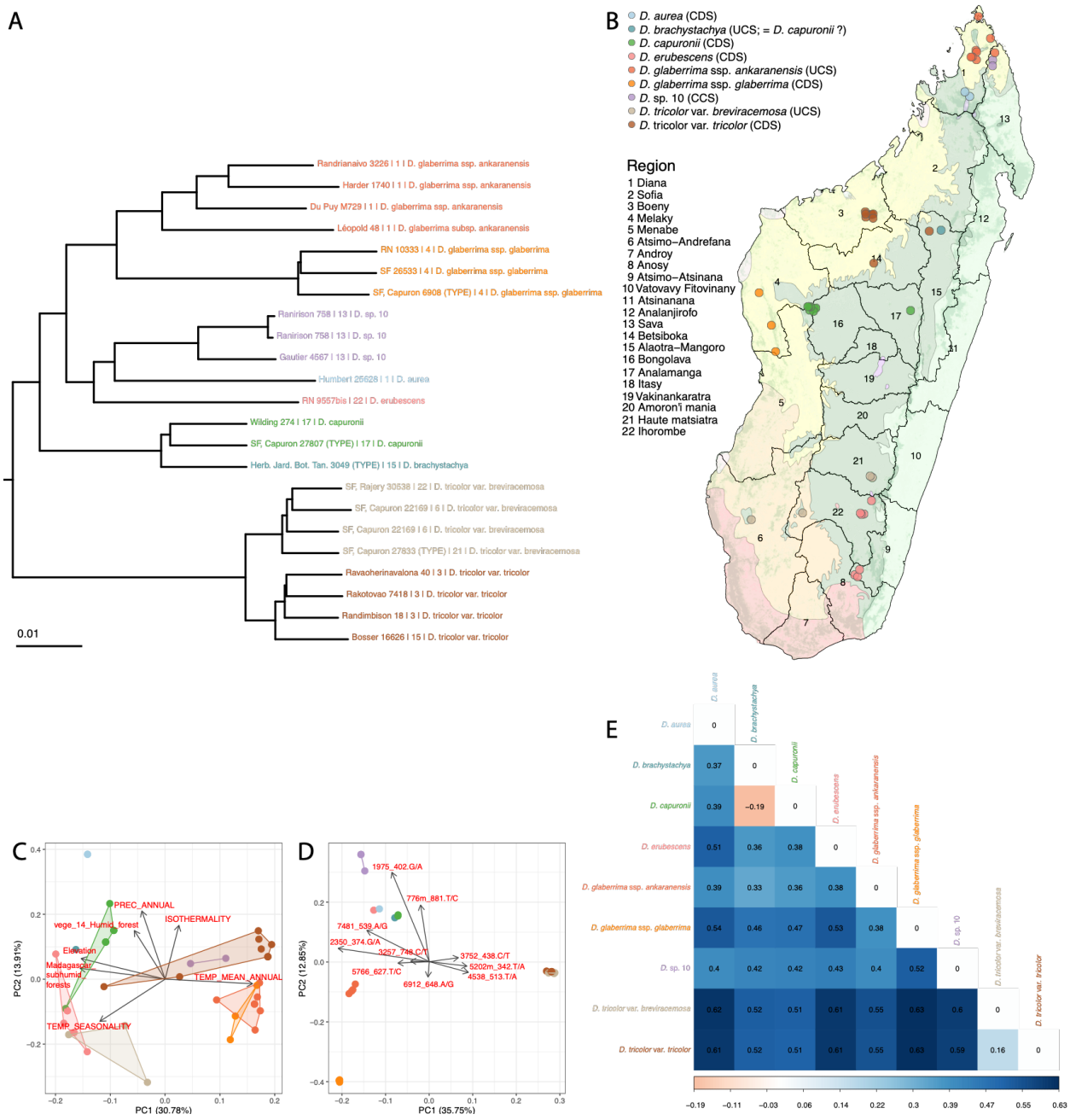

FIGURE S7: Integrative species delimitation in subgroup Tricolor. **A)** NJ tree of 61,773 biallelic SNPs (1% allowed missingness), rooted on the outgroup *D. normandii* (root not shown). Integers after the collection name denote the region of origin (see panel B). **B)** Distribution of verified collections. Coordinates are slightly jittered to reduce overlaps. **C)** PCA of scaled ecological variables. **D)** PCA of 44,401 biallelic SNPs excluding the outgroup, with a minimum minor allele count of 2 and 5% allowed missingness. **E)** Weir & Cockerham's F_{st} between all PS.

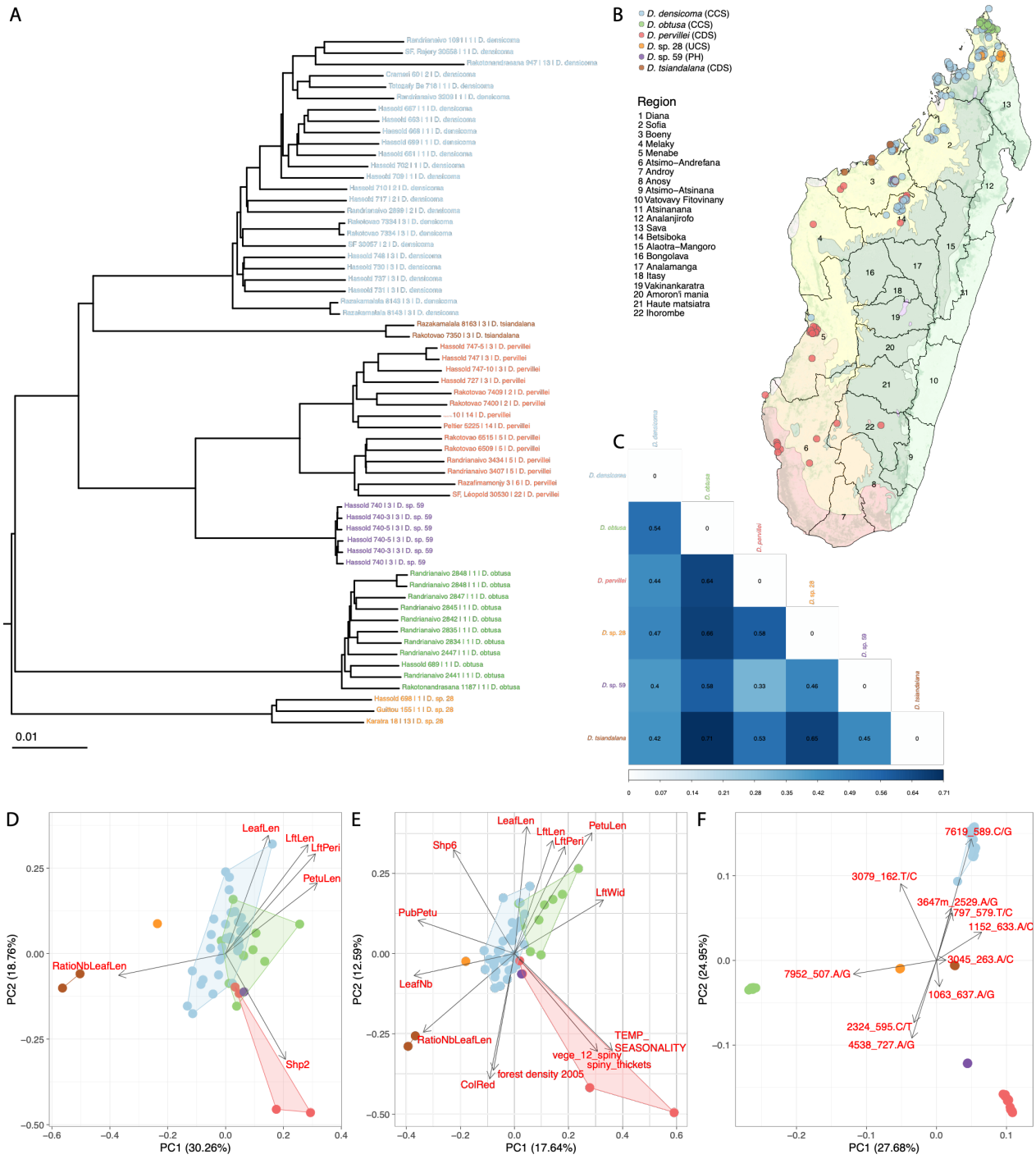

FIGURE S8: Integrative species delimitation in subgroup Pervillei. **A)** NJ tree of 94,695 biallelic SNPs (1% allowed missingness), rooted on the outgroup *D. glaberrima* (root not shown). Integers after the collection name denote the region of origin (see panel B). **B)** Distribution of verified collections. Coordinates are slightly jittered to reduce overlaps. **C)** Weir & Cockerham's F_{st} between all PS. **D)** PCA of scaled morphological variables. **E)** PCA of scaled morphological and ecological variables. **F)** PCA of 98,308 biallelic SNPs excluding the outgroup, with a minimum minor allele count of 2 and 5% allowed missingness.

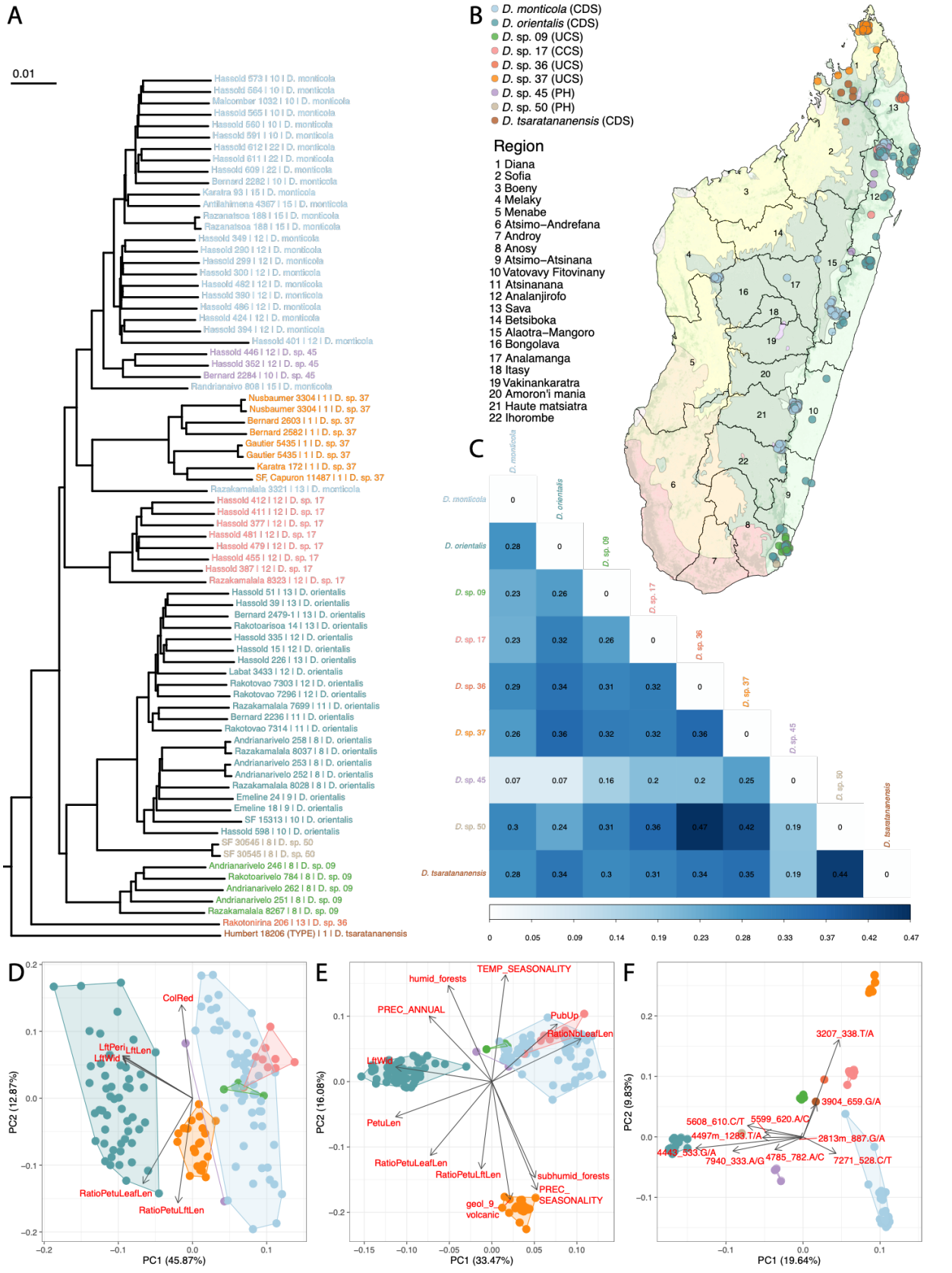

FIGURE S9: Integrative species delimitation in subgroup Monticola. **A)** NJ tree of 57,006 biallelic SNPs (1% allowed missingness), rooted on the outgroup *D. glaucocarpa* (root not shown). Integers after the collection name denote the region of origin (see panel B). **B)** Distribution of verified collections. Coordinates are slightly jittered to reduce overlaps. **C)** Weir & Cockerham's F_{st} between all PS. **D)** PCA of scaled morphological variables. **E)** PCA of scaled morphological and ecological variables. **F)** PCA of 63,671 biallelic SNPs excluding the outgroup, with a minimum minor allele count of 2 and 5% allowed missingness.

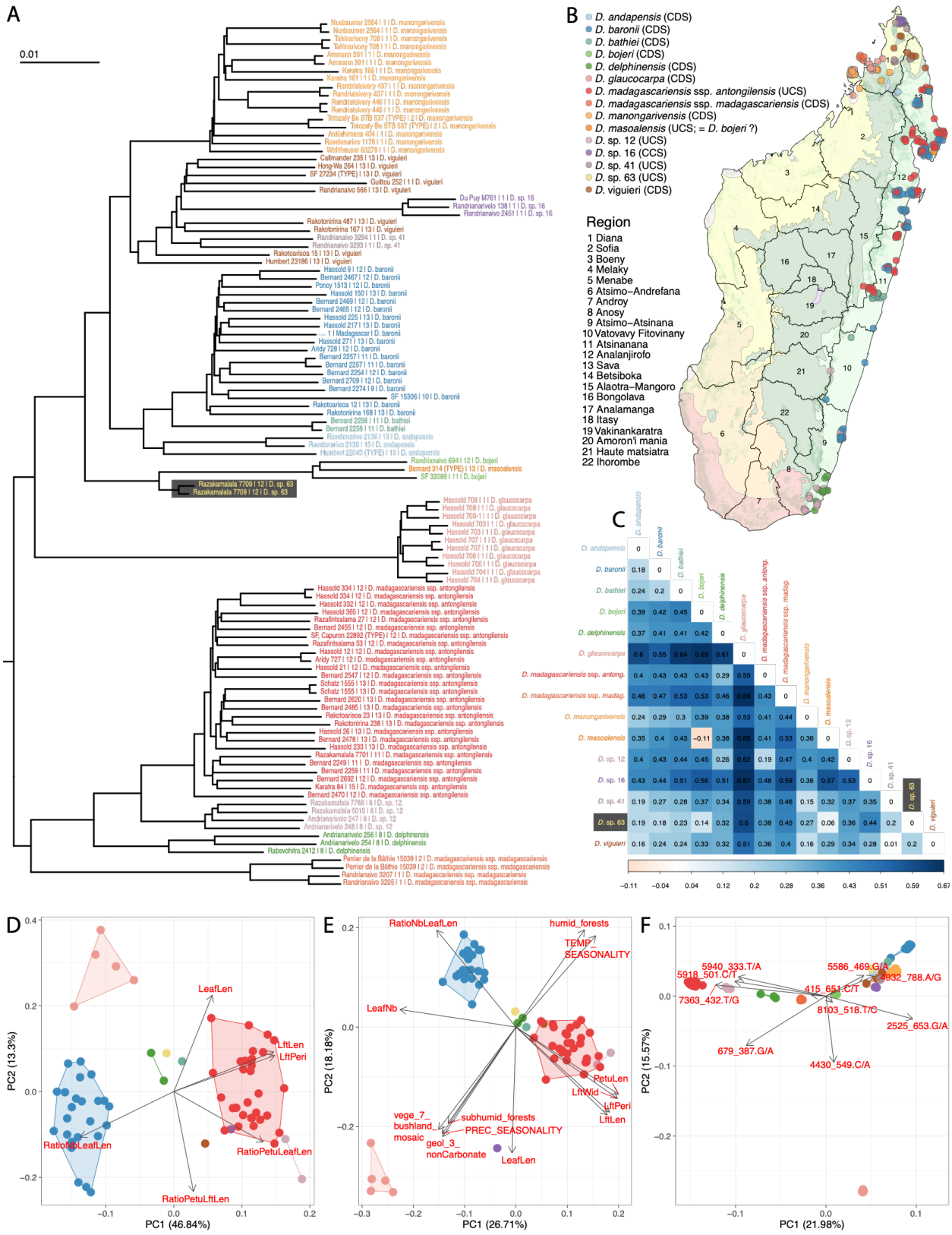

FIGURE S10: Integrative species delimitation in subgroup Baronii. **A)** NJ tree of 79,119 biallelic SNPs (1% allowed missingness), rooted on the outgroup *D. tsaratananensis* (root not shown). Integers after the collection name denote the region of origin (see panel B). **B)** Distribution of verified collections. **C)** Weir & Cockerham's F_{st} between all PS. **D)** PCA of scaled morphological variables. **E)** PCA of scaled morphological and ecological variables. **F)** PCA of 75,387 biallelic SNPs excluding the outgroup, with a minimum minor allele count of 2 and 5% allowed missingness.

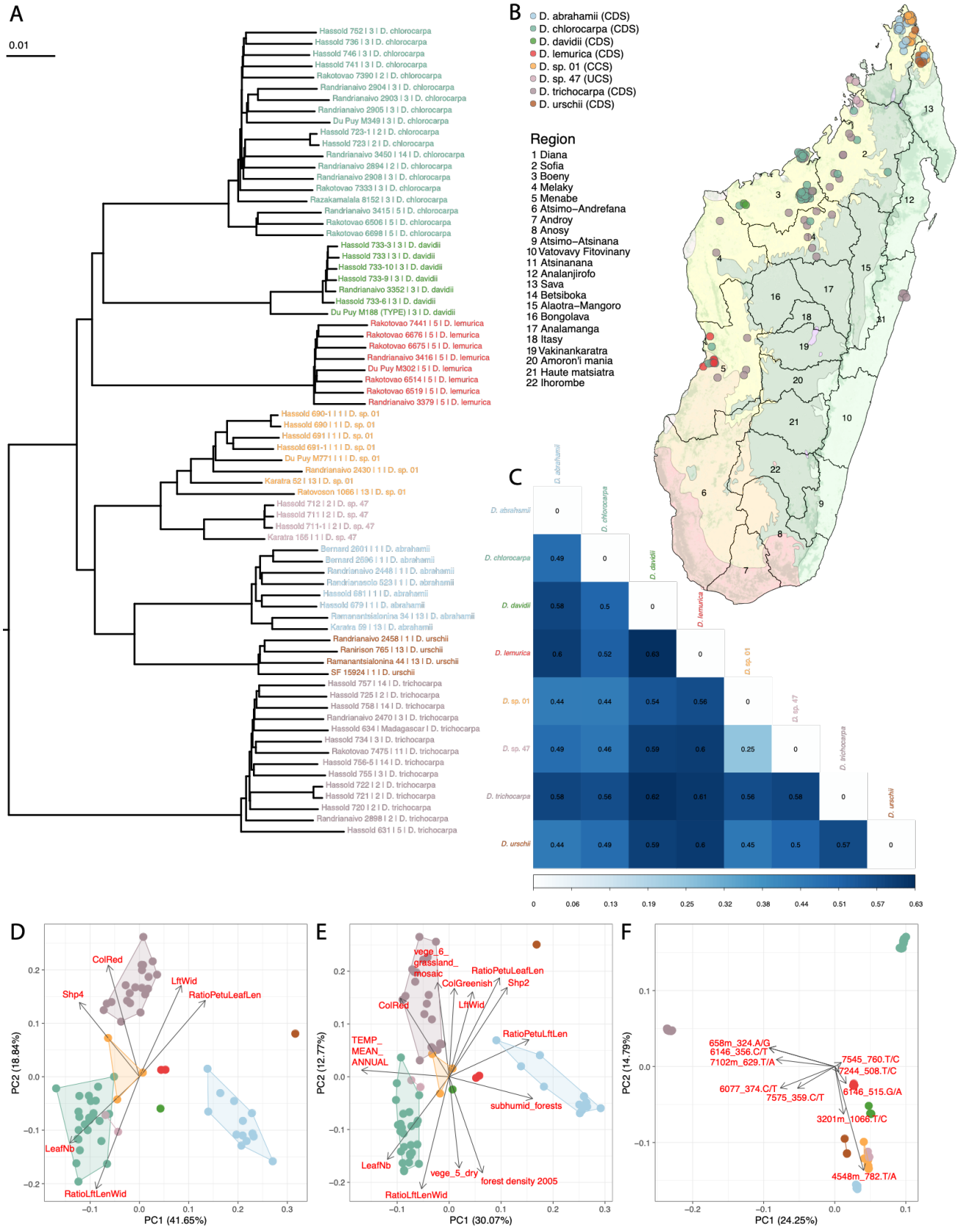

FIGURE S11: Integrative species delimitation in subgroup Chlorocarpa. **A)** NJ tree of 95,904 biallelic SNPs (1% allowed missingness), rooted on the outgroup *D. suaresensis* (root not shown). Integers after the collection name denote the region of origin (see panel B). **B)** Distribution of verified collections. **C)** Weir & Cockerham's F_{st} between all PS. **D)** PCA of scaled morphological variables. **E)** PCA of scaled morphological and ecological variables. **F)** PCA of 87,837 biallelic SNPs excluding the outgroup, with a minimum minor allele count of 2 and 5% allowed missingness.

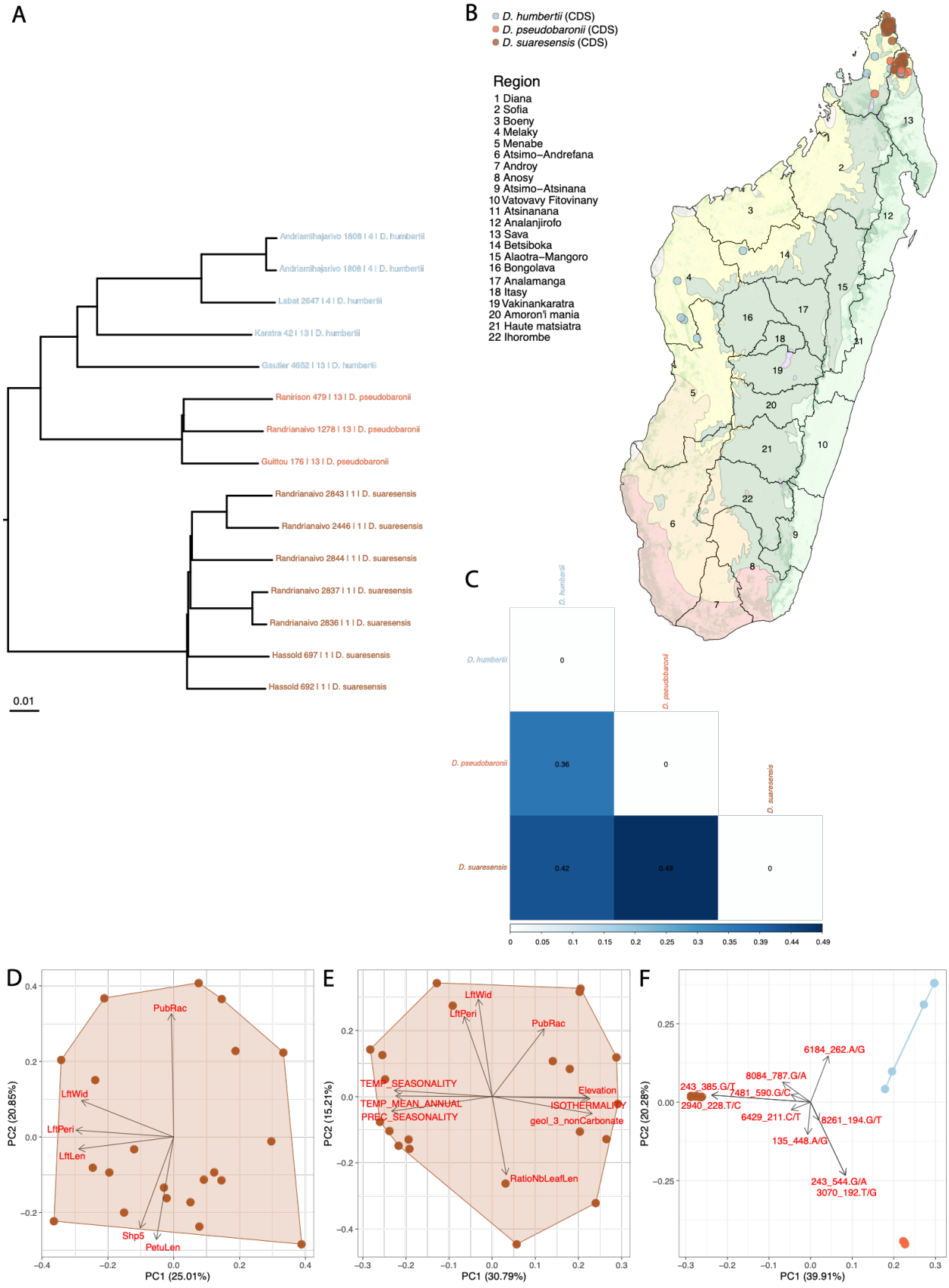

FIGURE S12: Integrative species delimitation in subgroup Humbertii. **A)** NJ tree of 71,915 biallelic SNPs (1% allowed missingness), rooted on the outgroup *D. trichocarpa* (root not shown). Integers after the collection name denote the region of origin (see panel B). **B)** Distribution of verified collections. **C)** Weir & Cockerham's F_{st} between all PS. **D)** PCA of scaled morphological variables. **E)** PCA of scaled morphological and ecological variables. **F)** PCA of 40,990 biallelic SNPs excluding the outgroup, with a minimum minor allele count of 2 and 5% allowed missingness.

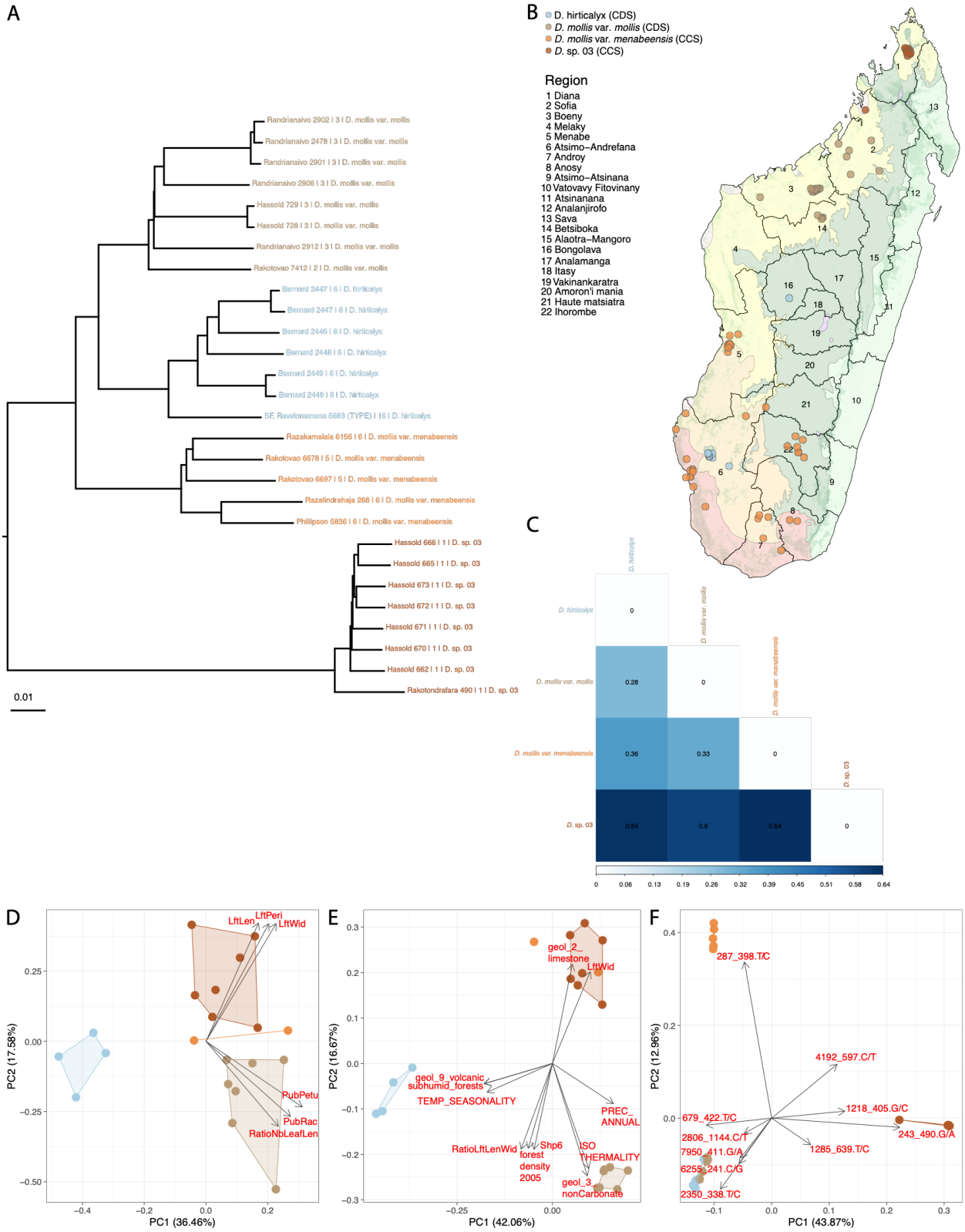

FIGURE S13: Integrative species delimitation in subgroup Mollis. **A)** NJ tree of 71,130 biallelic SNPs (1% allowed missingness), rooted on the outgroup *D.* sp. 08 (root not shown). Integers after the collection name denote the region of origin (see panel B). **B)** Distribution of verified collections. **C)** Weir & Cockerham's F_{st} between all PS. **D)** PCA of scaled morphological variables. **E)** PCA of scaled morphological and ecological variables. **F)** PCA of 58,384 biallelic SNPs excluding the outgroup, with a minimum minor allele count of 2 and 5% allowed missingness.

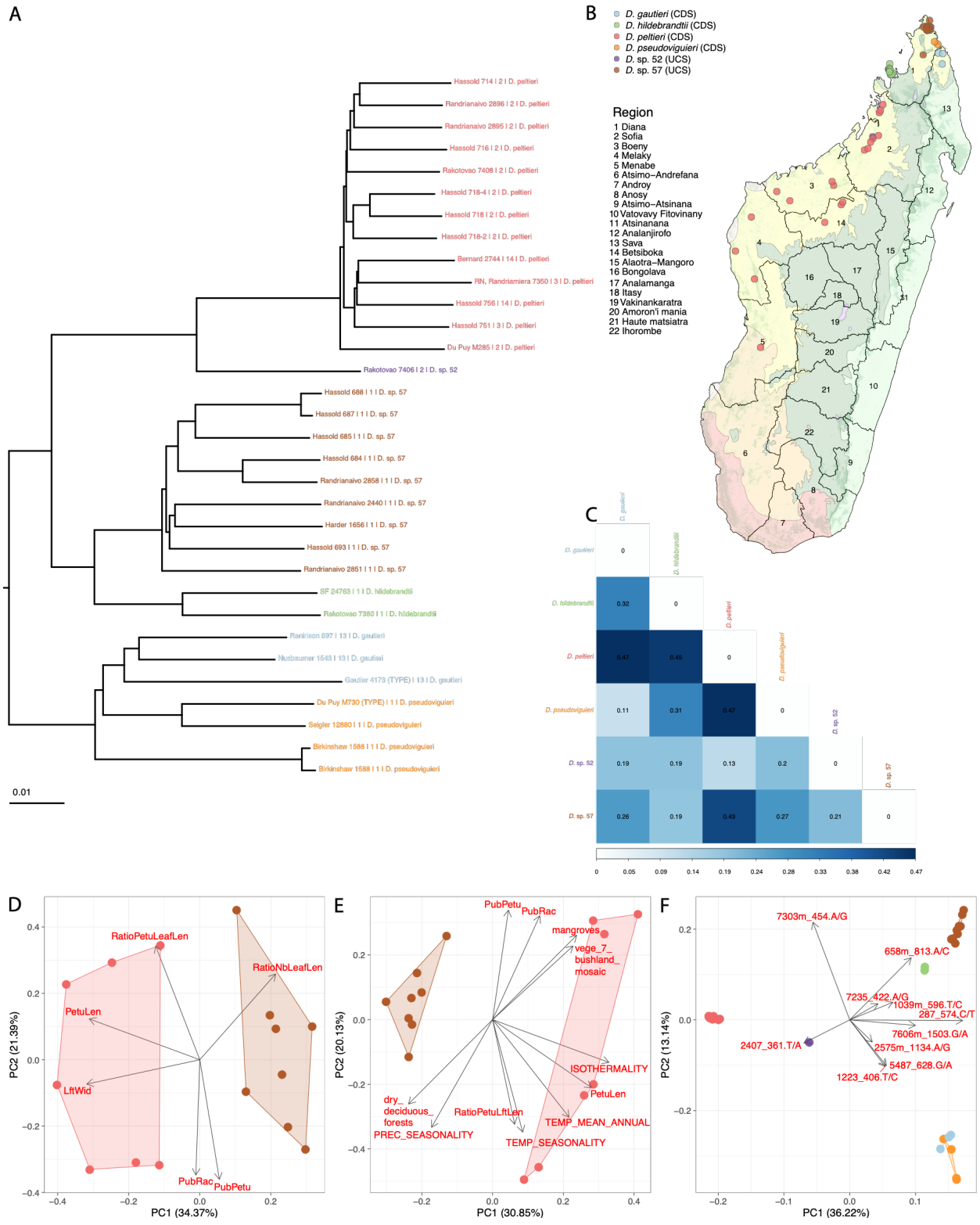

FIGURE S14: Integrative species delimitation in subgroup Peltieri. **A)** NJ tree of 79,419 biallelic SNPs (1% allowed missingness), rooted on the outgroup *D. emirnensis* and *D. campenonii* (root not shown). Integers after the collection name denote the region of origin (see panel B). **B)** Distribution of verified collections. **C)** Weir & Cockerham's F_{st} between all PS. **D)** PCA of scaled morphological variables. **E)** PCA of scaled morphological and ecological variables. **F)** PCA of 54,176 biallelic SNPs excluding the outgroup, with a minimum minor allele count of 2 and 5% allowed missingness.

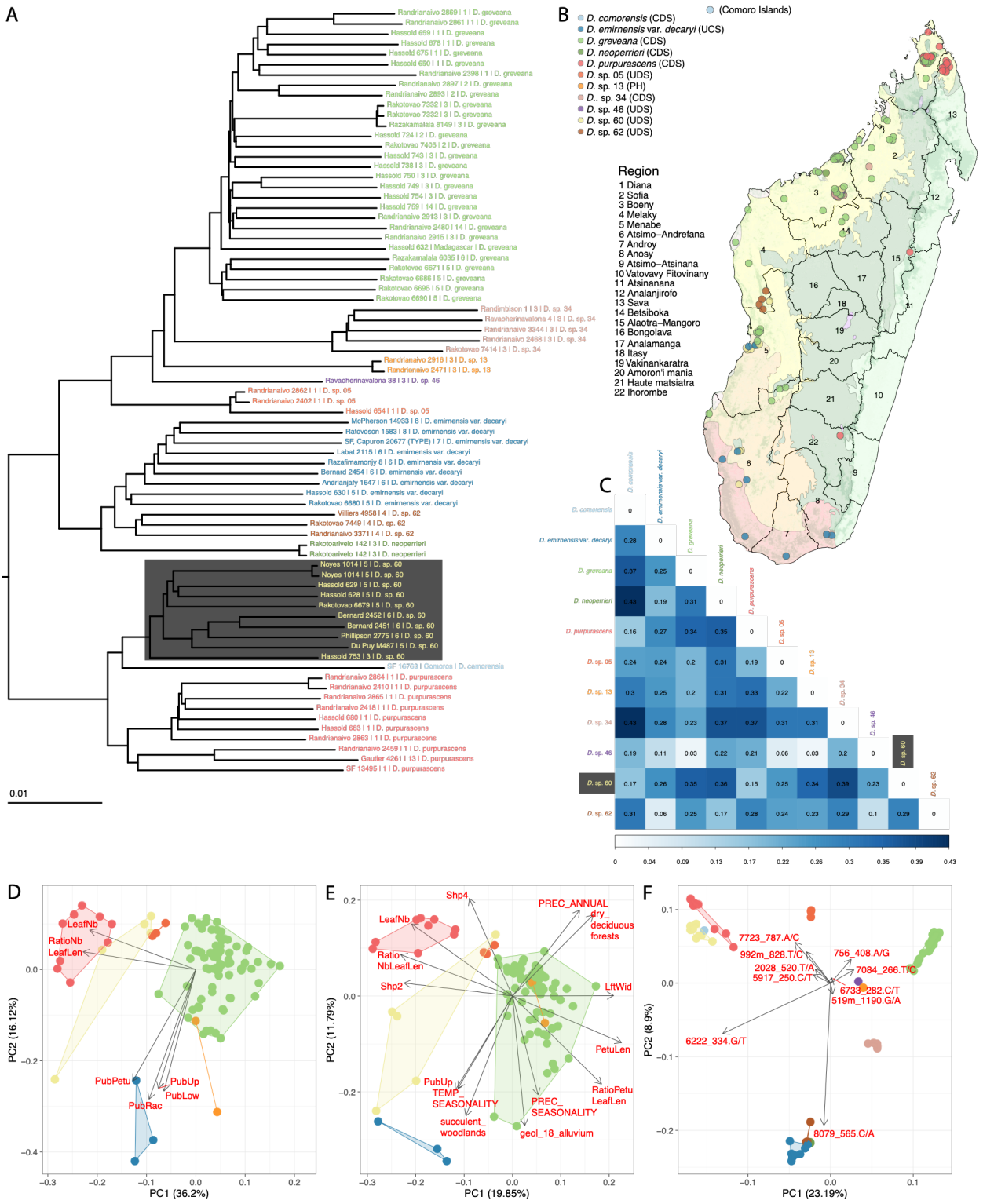

FIGURE S15: Integrative species delimitation in subgroup Greveana. **A)** NJ tree of 99,172 biallelic SNPs (1% allowed missingness), rooted on the outgroup *D. emirnensis* and *D. campenonii* (root not shown). Integers after the collection name denote the region of origin (see panel B). **B)** Distribution of verified collections. **C)** Weir & Cockerham's F_{st} between all PS. **D)** PCA of 20 scaled morphological variables. **E)** PCA of 63 scaled morphological and ecological variables. **F)** PCA of 88,506 biallelic SNPs excluding the outgroup, with a minimum minor allele count of 2 and 5% allowed missingness.

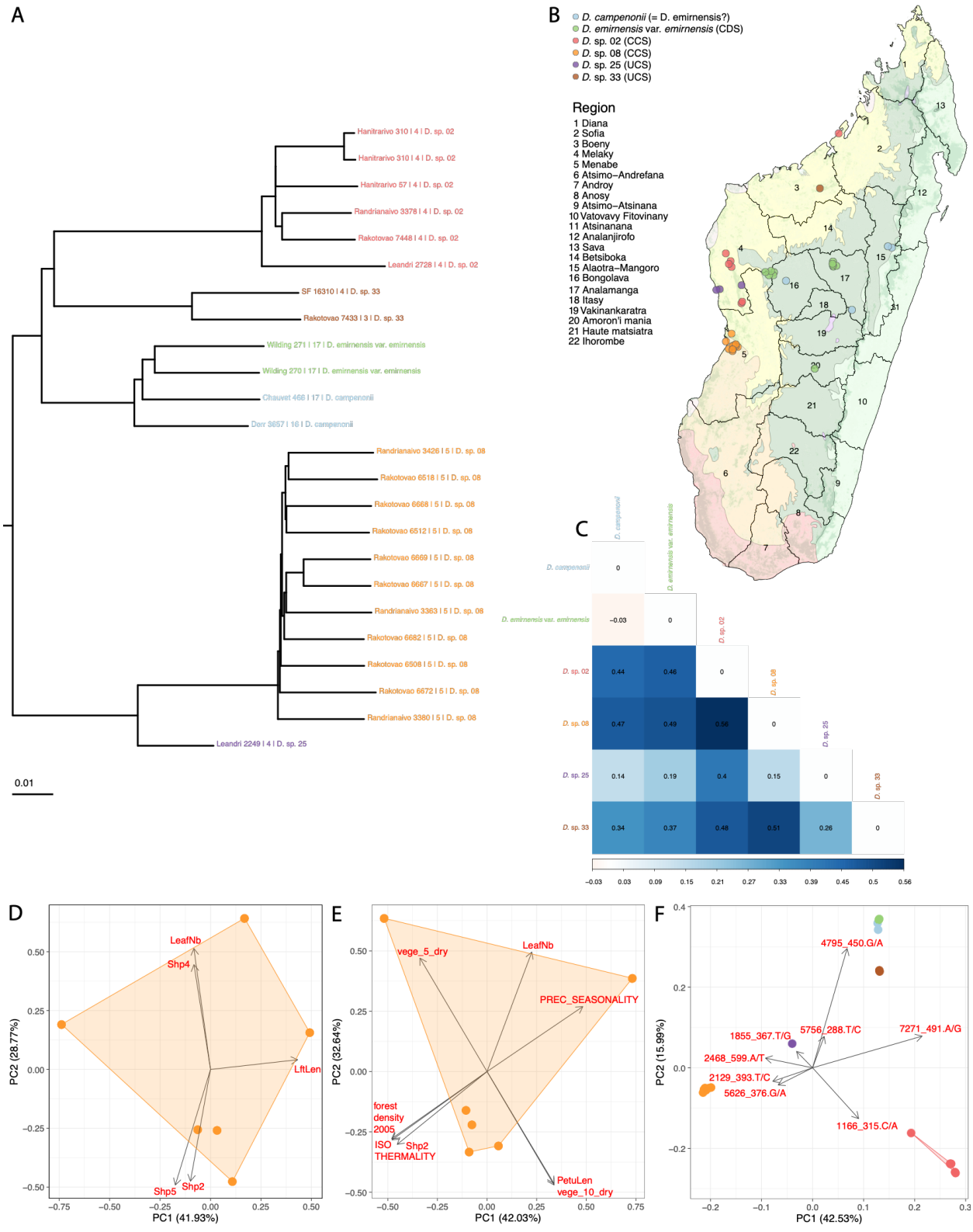

FIGURE S16: Integrative species delimitation in species of subgroup Arenosa and species with uncertain phylogenetic placement relative to other subgroups (*incertae sedis*). **A)** NJ tree of 75,620 biallelic SNPs (1% allowed missingness), rooted on the outgroup *D. mollis* (root not shown). Integers after the collection name denote the region of origin (see panel B). **B)** Distribution of verified collections. **C)** Weir & Cockerham's F_{st} between all PS. **D)** PCA of scaled morphological variables. **E)** PCA of scaled morphological and ecological variables. **F)** PCA of 57,240 biallelic SNPs excluding the outgroup, with a minimum minor allele count of 2 and 5% allowed missingness.

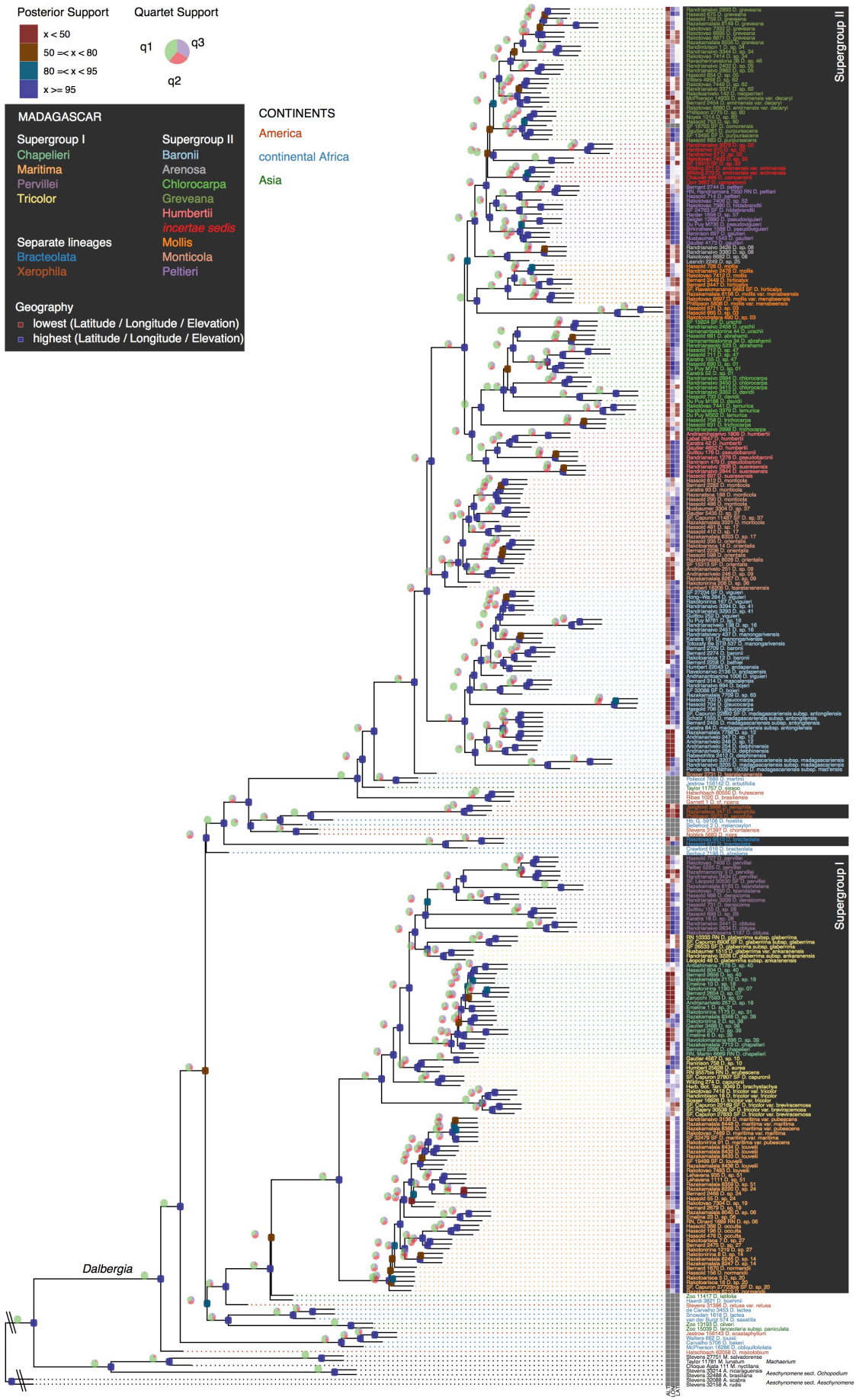

FIGURE S17: ASTRAL-III summary tree including 296 individuals from Madagascar and the Comoros (dark boxes) or elsewhere (see inset legend). The overall normalized quartet score was 91.1%. See Figure 1 for further explanations.

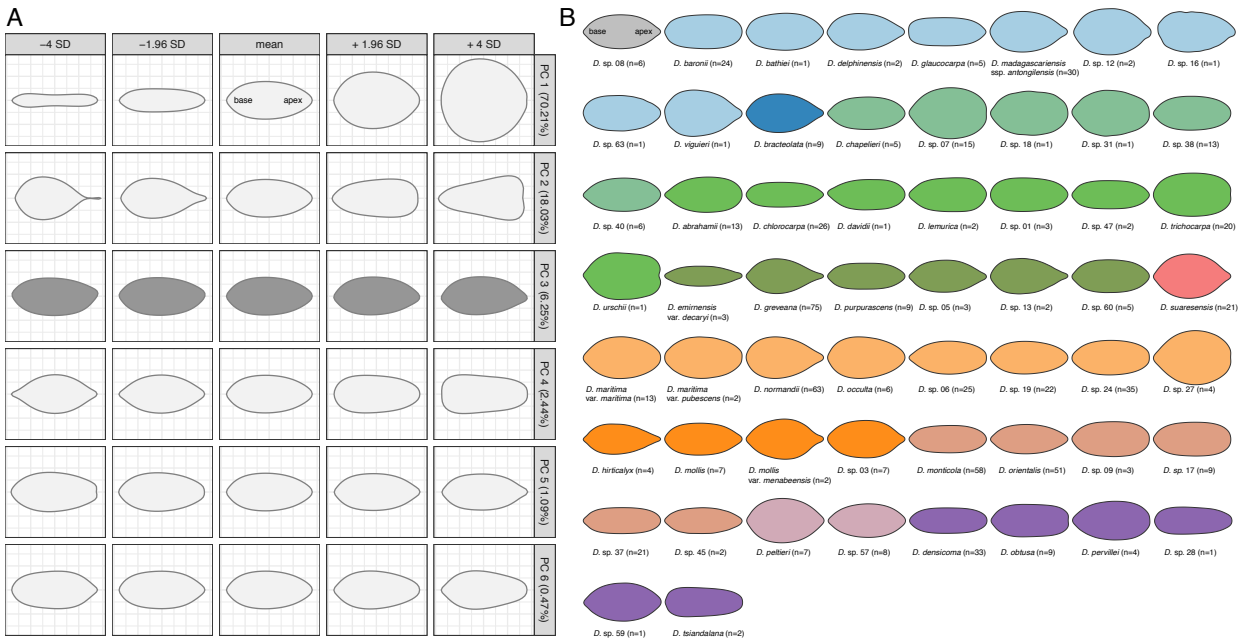

FIGURE S18: Analysis of leaflet shape. **A)** shows the shape variation along the first six principal components of leaflet shape in Malagasy *Dalbergia*. 95% of the variation in each principal component is covered by the three central columns, while columns at the edge represent extremes of each distribution for a better comparison between components. Components 1, 2, 4, 5, 6 were considered as biologically meaningful, while component 3 (shaded in grey) was discarded as an artefact of the position of a leaflet on the leaf rachis. **B)** Reconstructed mean leaflet shapes for all putative species in the morphological dataset (58 out of 95 identified in this study). Colours represent subgroups as in Figures 1-2. *D. maritima* as delimited here represents a single putative species, but is represented separately for two currently accepted varieties. SD = standard deviation.

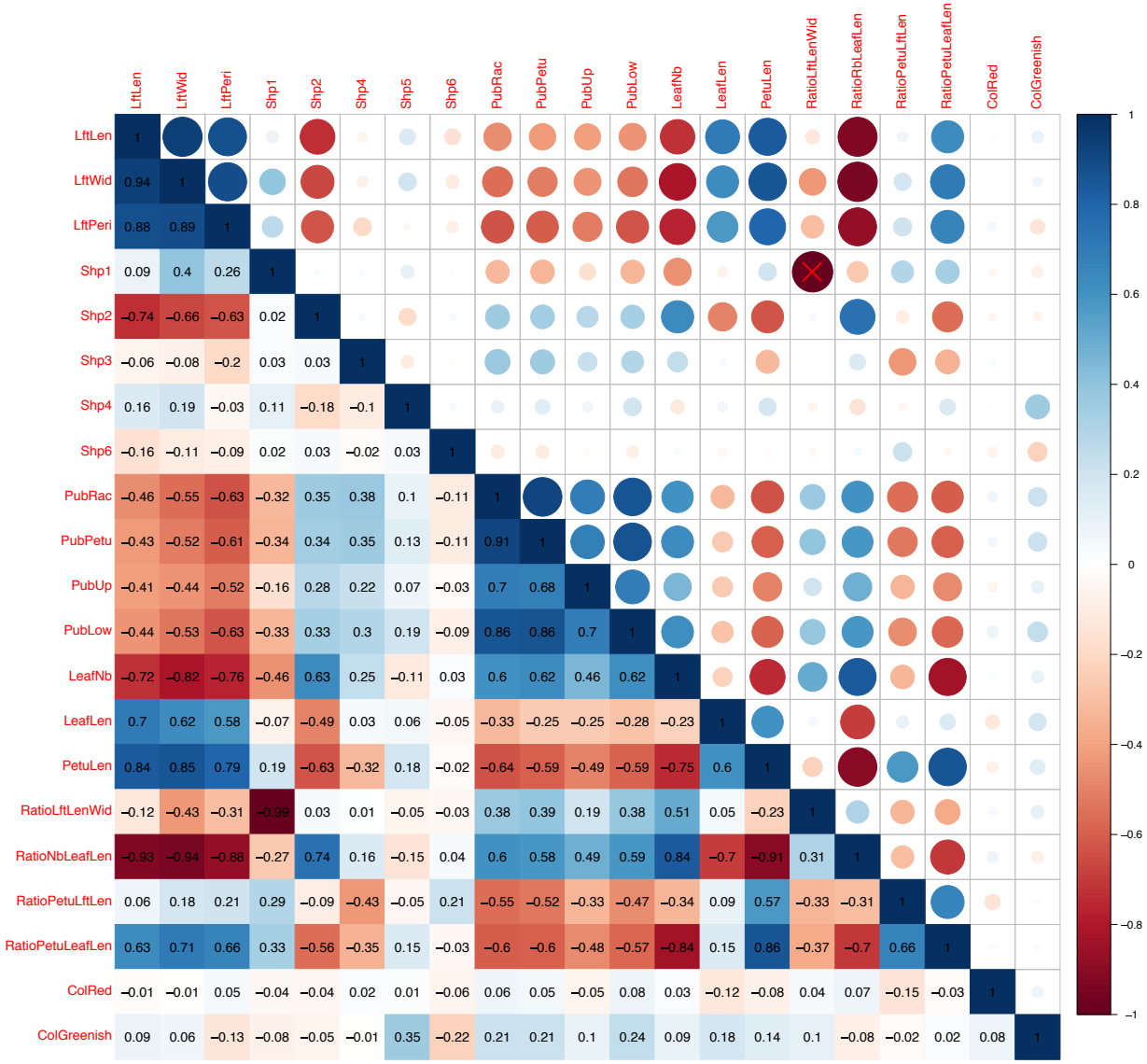

FIGURE S19: Spearman rank correlations among 21 morphological characters. Numbers, colour intensities and circle sizes represent correlation coefficients. One variable (shp1_orbicular, the first principal component of leaflet shape) showed an absolute correlation coefficient above 95% with other variables (see red cross) and was excluded from random forest classification and discriminant analysis.

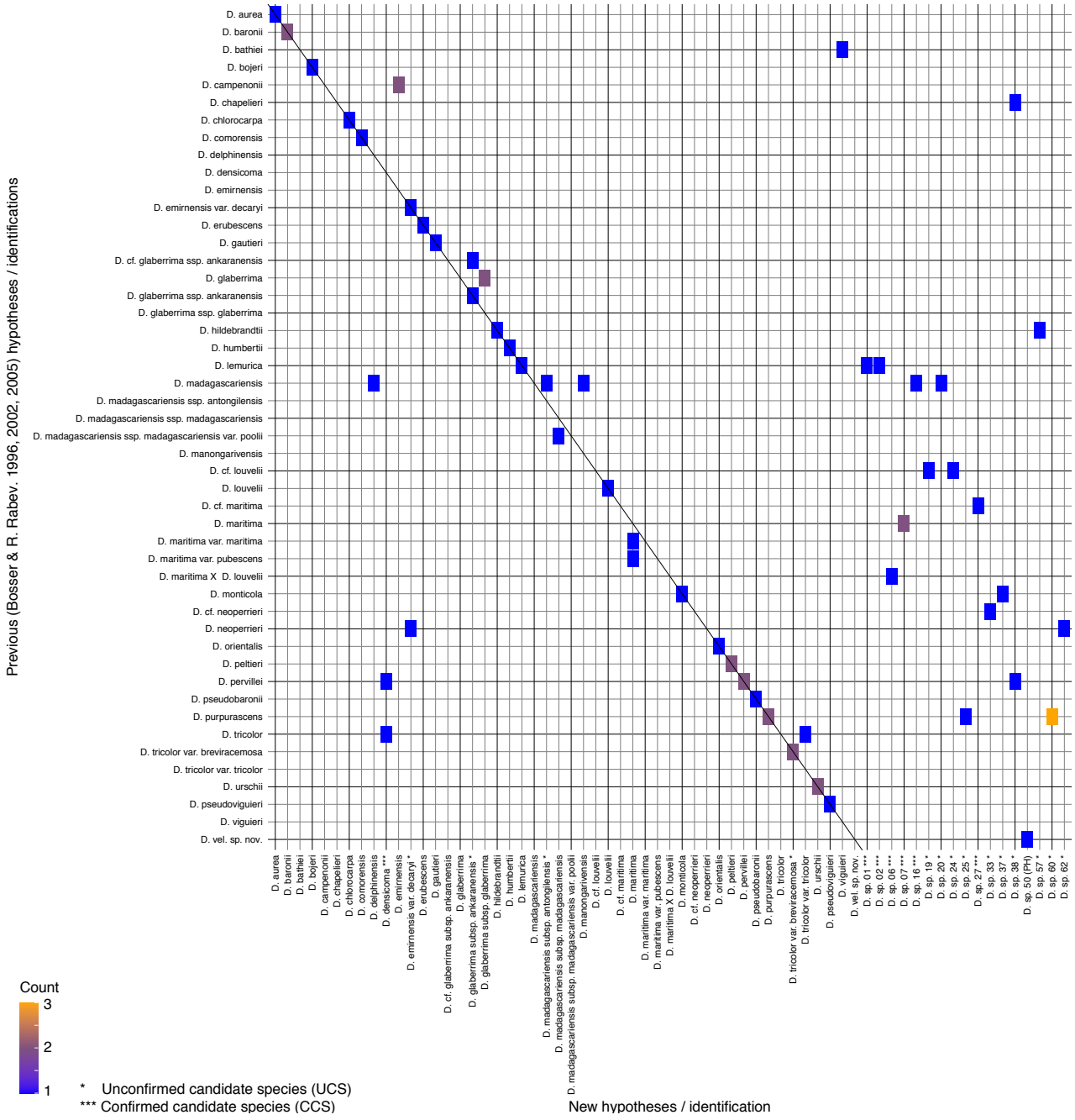

FIGURE S20: Determination history of 65 collections with available sequence data and identifications by BOSSER & RABEVOHITRA (1996, 2002, 2005). In total, we confirmed species identities of 28 collections (43%), resolved confusions between confirmed described species in 15 (23%) cases, and placed 22 (34%) collections to new candidate species. These numbers are based on a targeted sampling of topotypic collections and specimens with doubtful identifications, and are therefore not representative of all identifications by BOSSER & RABEVOHITRA (1996, 2002, 2005).

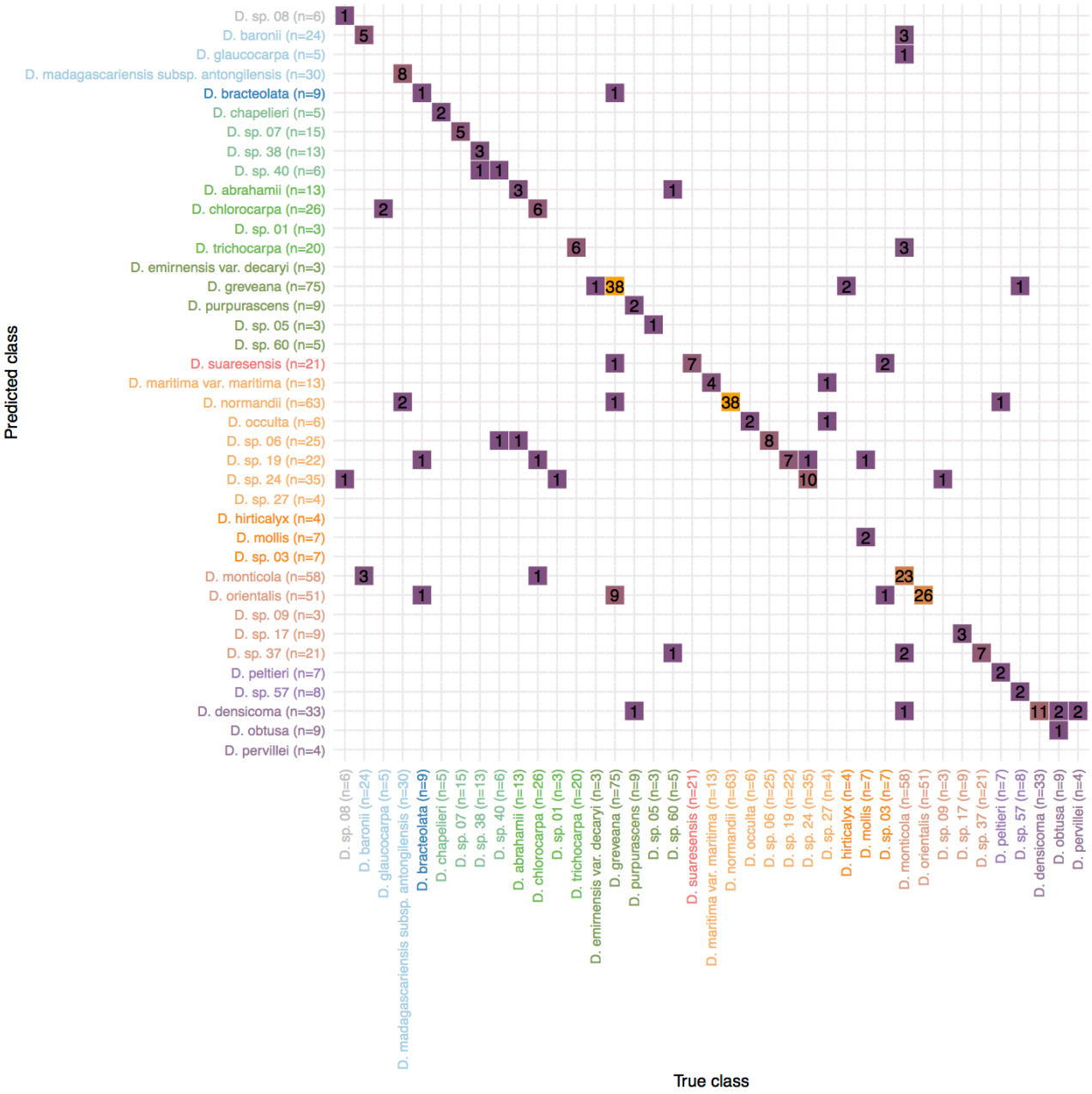

FIGURE S21: Confusion matrix of 293 validation samples classified with random forest using 20 leaf morphological variables, 387 training samples, 39 classes and an *mtry* hyperparameter of 2. Rows represent predictions and columns represent the true validation sample classes. Numbers denote the number of predictions (few = purple, many = orange). The overall accuracy of validation sample predictions using the leaf morphological data was 80.2%.

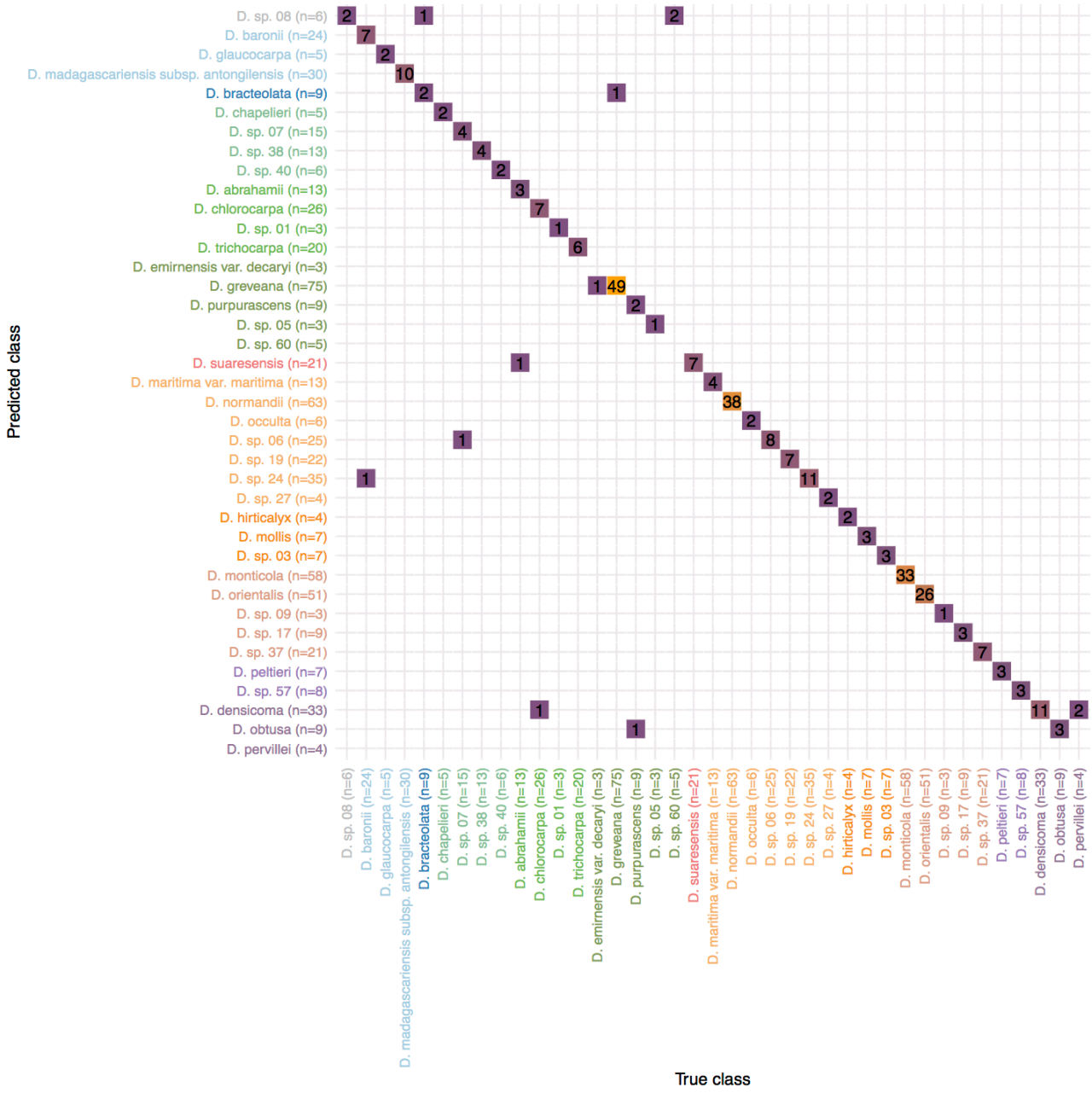

FIGURE S22: Confusion matrix of 293 validation samples classified with random forest using a combination of 20 leaf morphological and 23 eco-geographic variables, 387 training samples, 39 classes and an *mtry* hyperparameter of 8. Rows represent predictions and columns represent the true validation sample classes. Numbers denote the number of predictions (few = purple, many = orange). The overall accuracy of validation sample predictions using the leaf morphological data was 95.9%.

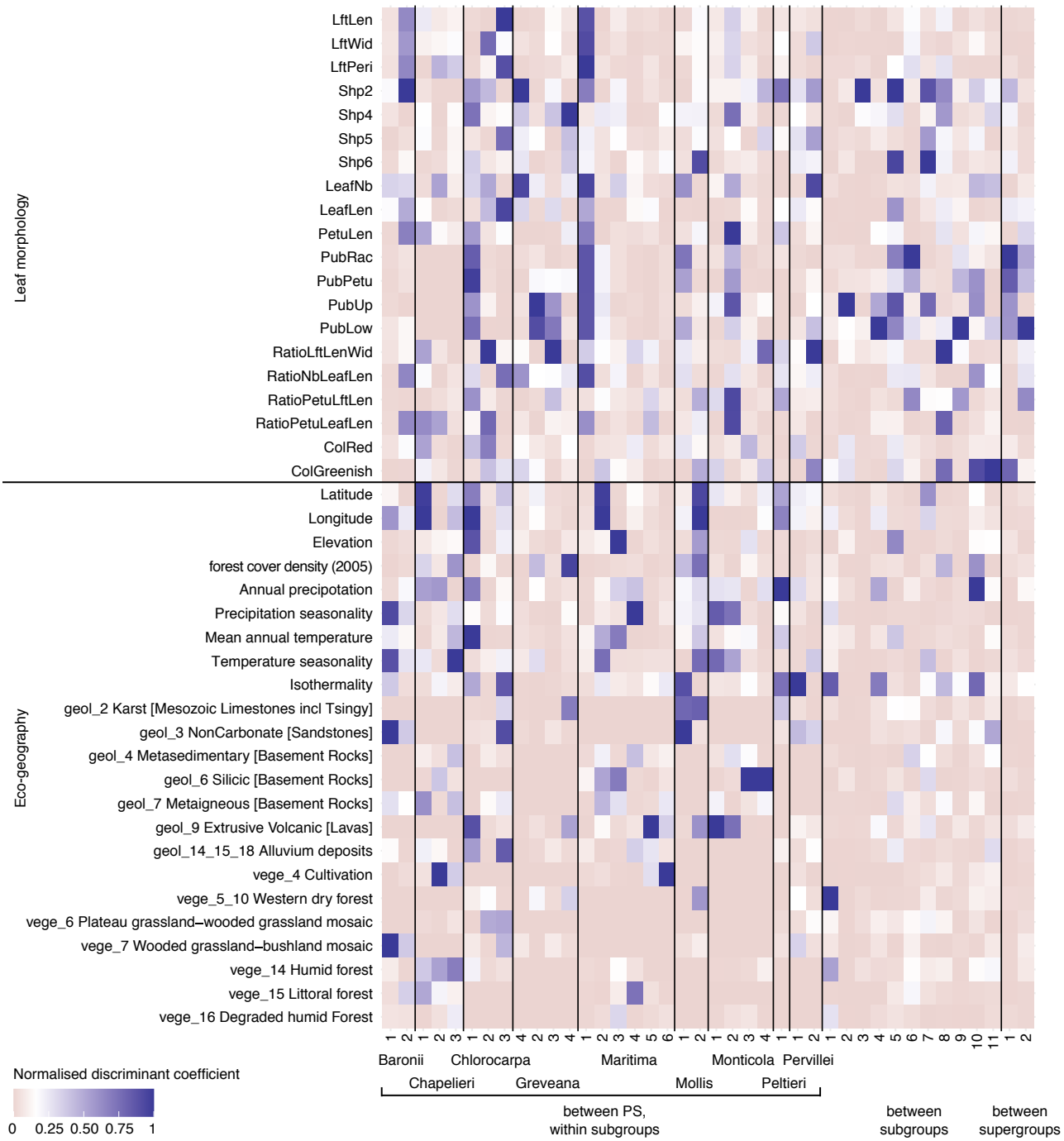

FIGURE S23: Normalised discriminant coefficients (variable importance) of morphological and eco-geographic variables inferred from discriminant analysis of principal components at the level of PS within subgroups, between subgroups, and between supergroups. Vertical black lines group columns by the different conducted discriminant analyses. Numbers below each column denote different discriminant functions within the same discriminant analysis. Rows are separated by variable type. See Table 1 for information on the leaf morphological characters.

Supplementary Tables

Supplementary Tables S1 – S4 are available as Supporting Online Material at https://github.com/scrameri/DalbergiaPhylogenomics.

TABLE S1: Specimen information for 1,089 analysed collections (Supporting Online Material).

TABLE S2: Sequencing and mapping statistics for 810 sequenced samples of *Dalbergia*, *Machaerium* and *Aeschynomene*.

TABLE S3: Leaf morphological measurements and leaflet morphometric data for 708 *Dalbergia* collections from Madagascar.

TABLE S4: Eco-geographic characteristics for 1846 taxonomically verified and adequately georeferenced *Dalbergia* collections from Madagascar.

TABLE S5: R console output of the linear regression model of identity between pairs of target enrichment sequencing samples.

```
Call:
lm(formula = asin(sqrt(identity)) ~ pair + log(age) + tissue +
sqrt(cov.10.2396.dif) + sqrt(cov.10.2396.max) + pair:tissue +
log(age):tissue, data = d.comp)

Residuals:
      Min        1Q    Median        3Q       Max
-0.061479 -0.006236  0.002031  0.009914  0.045380

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               -1.635e+01  6.282e+00  -2.603 0.010836 *
pairReplicates             4.759e-02  5.887e-03   8.083 2.93e-12 ***
log(age)                  -4.635e-02  8.263e-03  -5.610 2.26e-07 ***
tissueherb                -1.099e-01  1.857e-02  -5.919 5.94e-08 ***
sqrt(cov.10.2396.dif)     -1.509e-03  3.886e-04  -3.882 0.000198 ***
sqrt(cov.10.2396.max)      3.658e-01  1.284e-01   2.849 0.005447 **
pairReplicates:tissueherb  2.786e-02  8.573e-03   3.250 0.001629 **
log(age):tissueherb        4.525e-02  8.991e-03   5.032 2.50e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02048 on 89 degrees of freedom
Multiple R-squared:   0.78,     Adjusted R-squared:  0.7627
F-statistic: 45.08 on 7 and 89 DF, p-value: < 2.2e-16
```

Appendix II

TABLE S6: Random forest classification of leaf morphological data. Overall prediction accuracy, mean specificity and sensitivity per class are based on predictions of validation samples. PS = putative species, SD = standard deviation.

Grouping / subgroup	*mtry*	#Clas-ses	Training samples	Validation samples	Prediction accuracy	Mean specificity (±SD)	Mean sensitivity (± SD)
Classification of PS, subgroups and supergroups considering all samples							
PS	2	39	387	293	80.21%	99.5 (0.8)	62.5 (39.3)
subgroups	2	12	370	338	84.91%	98.6 (1.4)	65.4 (33.3)
supergroups	2	3	298	410	95.12%	95.9 (4.8)	63.1 (54.7)
Classification of PS within subgroups							
Baronii	4	3	24	35	100%	100 (0)	100 (0)
Chapelieri	15	4	18	21	80.95%	94.8 (7.4)	87.5 (25)
Chlorocarpa	1	4	28	34	97.06%	98.5 (2.9)	75 (50)
Greveana	4	5	39	56	96.43%	96.3 (7.3)	60 (54.8)
Maritima	3	7	89	79	93.67%	99 (1.3)	78 (37.4)
Mollis	1	3	9	9	100%	100 (0)	100 (0)
Monticola	2	5	62	80	98.75%	99.5 (1.1)	80 (44.7)
Peltieri	1	2	10	5	80%	83.3 (23.6)	83.3 (23.6)
Pervillei	1	3	20	26	92.31%	90.2 (13.4)	65.2 (56.5)

References

Bosser, J., and R. Rabevohitra. 1996. Taxa et noms nouveaux dans le genre *Dalbergia* (Papilionaceae) à Madagascar et aux Comores. Bull. Mus. Natl. Hist. Nat. B, Adansonia **18**:171-212.

Bosser, J., and R. Rabevohitra. 2002. Tribe Dalbergieae. Pages 321-361 *in* D. J. Du Puy, J. N. Labat, R. Rabevohitra, J. F. Villiers, J. Bosser, and J. Moat, editors. The Leguminosae of Madagascar. Royal Botanical Gardens, Kew.

Bosser, J., and R. Rabevohitra. 2005. Espèces nouvelles dans le genre *Dalbergia* (Fabaceae, Papilionoideae) à Madagascar. Adansonia **27**:209-216.

Chapter III

Two New Species from Southeastern Madagascar and an Emended Description of the Rosewood Species *D. maritima* R. Vig.

Manuscript submitted to Systematic Botany and co-authored by

Simon Crameri[1], Peter B. Phillipson[2,3], Nivohenintsoa Rakotonirina[4], Nicholas Wilding[2,3], Roger Lala Andriamiarisoa[5], Porter P. Lowry II[2,3], Alex Widmer[1]

[1] ETH Zurich, Institute of Integrative Biology, Universitätstrasse 16, Zürich, Switzerland.
[2] Missouri Botanical Garden, 4344 Shaw Blvd., St. Louis, MO, 63110
[3] Institut de Systématique, Évolution et Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, Centre National de la Recherche Scientifique, Sorbonne Université, École Pratique des Hautes Études, Université des Antilles, C.P. 39, rue Cuvier 57, 75005 Paris, France.
[4] Département de Biologie et Écologie Végétales, Université d'Antananarivo, Antananarivo, Madagascar.
[5] Missouri Botanical Garden, B.P. 3391, Antananarivo 101, Madagascar.

Author contributions

SC, PBP, NR and NW carried out the taxonomic work, SC, PBP and NR performed the morphological measurements, SC analyzed the morphological and ecological data and prepared the distribution map, NR and RLA performed field work, RLA prepared the illustrations, and SC wrote the manuscript with contributions from AW, PPL, PBP, and NW.

Funding information

This work was supported by ETH Zurich and a grant from the Rübel Foundation to AW. The participation of PBP, NW, and PPL was supported by a grant from the Fondation Franklinia to PPL. Field work and NR's participation were funded by the Délégation de l'Union Européenne à Madagascar (DEUM). The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

Abstract

The Malagasy rosewood species *Dalbergia maritima* has a long history of unsustainable exploitation for its beautiful, burgundy-colored heartwood. As currently circumscribed, *D. maritima* has a wide geographic distribution in eastern Madagascar and exhibits significant morphological, ecological, and genetic variation, suggesting it may comprise more than a single entity. Multivariate analyses of leaf, flower, and inflorescence characters as well as eco-geographic features reveal several morphologically well delimited entities with distinct habitat preferences and/or geographic ranges, which are consistent with results from recent phylogenomic and population genomic studies of Malagasy *Dalbergia*. Based on these findings, we describe and illustrate two new species from southeastern Madagascar comprising material previously assigned to *D. maritima*, viz. *D. pseudomaritima* sp. nov., characterized by paniculate inflorescences and small, broadly elliptic to orbicular, glabrous leaflets, and *D. razakamalalae* sp. nov., distinguished by racemose inflorescences with large flowers, and narrowly ovate to narrowly elliptic, glabrous leaflets. *Dalbergia maritima* is consequently re-circumscribe to include only populations from east-central Madagascar, within which we recognize two subspecies, *D. maritima* subsp. *maritima*, with glabrous leaves, inflorescence axes, and gynoecia, occurring in littoral forest habitats, and *D. maritima* subsp. *pubescens*, with indument on these structures, and growing in evergreen humid forest farther inland. Photos are provided for each taxon, along with line drawings for the two new species. Provisional IUCN Red List assessments indicate that all three species are Endangered, *D. maritima* and *D. razakamalalae* mainly because of selective logging for trade in their high-quality heartwood, and *D. pseudomaritima* mainly because of habitat degradation due to land clearing and fire for subsistence agriculture, which has important implications for their conservation and sustainable management.

Keywords — *Dalbergia,* Fabaceae, IUCN Red List, Leguminosae, Madagascar, new species, rosewood

Introduction

The genus *Dalbergia* L.f. (Fabaceae) encompasses 270 currently accepted species (World Checklist of *Dalbergia*, B. Klitgaard, pers. comm. 2021; POWO 2021) and includes woody lianas, shrubs and trees, which grow in a wide range of habitats throughout the tropics (KLITGAARD & LAVIN 2005). Numerous arborescent species are known to form durable and beautifully colored heartwood (PRAIN 1904; DE CARVALHO 1997; BOSSER & RABEVOHITRA 2002; LACHENAUD & VAN DER MAESEN 2016; CERVANTES *et al.* 2019). Several species are highly sought-after, especially for the production of high-quality furniture and musical instruments (BARRETT *et al.* 2010). Increasing demand for their desirable wood in both the national and international markets has subjected these species to intense, unsustainable logging, most of which is illegal (SCHUURMAN & LOWRY 2009; BARRETT *et al.* 2010; MASON *et al.* 2016; UNODC 2016; WAEBER *et al.* 2019).

The diversity of *Dalbergia* is exceptionally high in Madagascar. The last major revision of the Malagasy species was done by BOSSER & RABEVOHITRA (2002, 2005), who described 25 new species, reinstated one species, and placed three species in synonymy, thereby increasing the number of species to 48 from the 25 previously recognized by VIGUIER (ined., 1944), including 11 species that were posthumously described as new (VIGUIER 1952). Recent collection efforts, integrative analyses, and additional taxonomic work have revealed numerous problems in the application of Bosser and Rabevohitra's species-level taxonomy (CRAMERI *et al.* in prep.), despite the fact that it is relatively recent. Work towards a refined taxonomy of *Dalbergia* species from Madagascar is currently in progress, and has so far led to the description of two new species from northern Madagascar and an emended description for a third species (WILDING *et al.* submitted-a), a new species from central Madagascar (RAKOTONIRINA *et al.* in prep.), and two instances in which infraspecific taxa represent morphologically and eco-geographically distinct species (Rakotonirina *et al.* in prep.; WILDING *et al.* submitted-b). These taxonomic changes are supported by phylogenomic and population genomic analyses based on more than 600 accessions of Malagasy *Dalbergia* (CRAMERI *et al.* in prep.).

Integrated analyses of genetic, morphometric, and eco-geographic data have revealed or confirmed further taxonomic problems, including three polyphyletic species concepts (*D. madagascariensis* Vatke, *D. maritima* R. Vig. and *D. neoperrieri* Bosser & R. Rabev.), several cases in which apparently widely distributed species are separable into two or more morphologically and eco-geographically distinct entities, at least two cases of synonymy involving species recognized by BOSSER & RABEVOHITRA (2002, 2005), and several potentially new species, specimens of which have questionably been included in known species or collected only recently (CRAMERI *et al.* in prep.; MADAGASCAR

CATALOGUE 2021). Taken together, these findings suggest that the diversity of Malagasy *Dalbergia* has been vastly underestimated, and that in reality it is comparable to or more likely exceeds the current number of recognized species in the much larger tropical zones of continental Africa, the Americas, and Southeast Asia, each of which has ca. 60–70 species (World Checklist of *Dalbergia*, B. Klitgaard, pers. comm. 2021; POWO 2021).

The taxonomic inadequacies and uncertainties regarding Malagasy *Dalbergia* can in part be attributed to the previous lack of phylogenetic information, but they can also be explained by a shortage of high-quality representative collections documenting the full range of morphological and eco-geographic diversity found in the genus. A total of ca. 1400 databased collections were available prior to 2006, of which ca. 1000 were examined and identified by Bosser & Rabevohitra, while ca. 400 collections were either left undermined because they are sterile or in poor condition, or had not been seen (Madagascar Catalogue 2021). BOSSER & RABEVOHITRA (1996) noted that the comparison of flowering collections of *Dalbergia* with leafy sterile material is often difficult because some species flower before the leaves emerge. Consequently, several species were only known from and described on the basis of one or a few fertile specimens, resulting in a limited understanding of their range and variability. As part of an ongoing effort over the last several decades to expand our knowledge of the Malagasy flora (LOWRY *et al.* 2018), the number of collections of *Dalbergia* has considerably increased in recent years and amounts to over 4100 available as of January 2021, more than 2000 of which have been made since 2014 as part of a collaborative effort to better document the diversity, occurrence, growth habit, and morphological variation of leaf, flower, fruit and bark characters (Hassold 2015; HASSOLD *et al.* 2016; MADAGASCAR CATALOGUE 2021). These collections comprise herbarium specimens as well as leaf material for genetic analysis along with heartwood samples for anatomical, spectroscopic, and spectrometric characterization. Altogether this material is being used to build a reference collection of Malagasy *Dalbergia*, which can serve as a basis for forensic timber identification (DORMONTT *et al.* 2015) and which provides an important resource for taxonomic studies.

Dalbergia maritima R. Vig. is one of several Malagasy species that produce high-quality rosewood (also known as *bois de rose*) and have a long history of exploitation (NORMAND 1988; RICHTER *et al.* 2014). The species was first described in VIGUIER (1944) and validly published several years later (Viguier 1952) based on a collection from littoral forest on sand at Tampina in east-central Madagascar (*Louvel 200*), and a second collection (*Louvel 79*) without precise locality information ("*forêts côtières de l'Est*"), which most likely also originates from Tampina (Fig. 3D). The two collections have glabrous leaves with small to medium-sized leaflets (Figs. 1A, 2A) and a racemose inflorescence structure

(Fig. 2A). The delimitation of *D. maritima* was later broadened by BOSSER & RABEVOHITRA (1996) to include material with pubescent leaves from the Betampona Special Reserve and the area surrounding Mahavelona (Foulpointe), respectively ca. 65 and 100 km north and further inland from Tampina, which they recognized as a new variety, *D. maritima* var. *pubescens* Bosser & R. Rabev. (Figs. 1B, 2B, 3D). In their treatment of the genus for the *Leguminosae of Madagascar*, BOSSER & RABEVOHITRA (2002) included within an expanded delimitation of the typical variety, material with paniculate inflorescences from littoral forests on sand at Mandena and Sainte Luce (e.g. *Rabevohitra 2178*), over 700 km to the south at the southeastern extremity of the island (Figs. 1F, 2F, 3D). They further included a collection with racemose inflorescences (*Service Forestier 22334*, to be compared with Figs. 1E, 2D) from the same general area but from low-elevation evergreen humid forest on laterite farther inland (Fig. 3D). Lastly, they also included within the typical variety two collections from the SAVA Region in northeastern Madagascar (*Service Forestier 2591*, *27751*), thereby increasing its distribution range ca. 320 km farther north (Fig. 3D). Since then, numerous additional collections with glabrous and pubescent leaves, and with either racemose or paniculate inflorescences, originating from various localities ranging from Makirovana in the northeast to sites near Tolagnaro in the extreme southeast have likewise been associated with this broad interpretation of *D. maritima* as a widely distributed, polymorphic species.

Recent phylogenomic and population genomic analyses based on more than 600 accessions of Malagasy *Dalbergia* (CRAMERI *et al.* in prep.) have shown that this broad concept of *D. maritima* is polyphyletic. The southeastern material with paniculate inflorescences included in the typical variety by BOSSER & RABEVOHITRA (2002), hereafter referred to as the 'southeastern paniculate material', is not closely related to the *D. maritima* population from the type locality. Instead, it belongs to a strongly divergent lineage (the Chapelieri clade) that also includes *D. chapelieri* Baill. *sensu lato* (*s.l.*), a species complex currently under study that also forms paniculate inflorescences (Figs. 1G, 2E) but whose leaves and leaflets are clearly larger and shaped differently (Figs. 1G–H). By contrast, the material clearly referable to *D. maritima* belongs to a separate lineage (the Maritima clade), which includes two other described rosewood species from eastern Madagascar (CRAMERI *et al.* in prep.): (i) *D. louvelii* R. Vig. *s.l.* (Figs. 1C–D, 2C), a species complex with larger leaflets and flowers than *D. maritima*, which co-occurs with *D. maritima* in east-central Madagascar (Fig. 3D) and potentially extends southwards to 'Mananara' according BOSSER & RABEVOHITRA (2002), which could mean the river that flows past Vangaindrano in southeastern Madagascar, and extends to the north towards the Antongil Bay (e.g. *Service Forestier 9144*); and (ii) *D. occulta* Bosser & R. Rabev., which

is characterized by its distinctly shaped leaflets and large inflorescence bracts (BOSSER & RABEVOHITRA 2005), and occurs in Makira Natural Park and Masoala National Park in northeastern Madagascar (Fig. 3D). A fourth described rosewood species, *D. normandii* Bosser & R. Rabev., forms an early-branching group in the Maritima clade (CRAMERI *et al.* in prep.) and is morphologically clearly distinct from all other species by its long leaves with few and large leaflets, and large flowers (Bosser & Rabevohitra 2002).

At least three further entities also belong to the Maritima clade, all of which are distributed in the northeastern part of the island (CRAMERI *et al.* in prep.). They are often confused with either *Dalbergia maritima* or *D. louvelii* despite being morphologically distinct, but the flowering material currently available is insufficient to describe them. One of these undescribed entities corresponds to material from the SAVA region mentioned above and included in *D. maritima* by Bosser and Rabevohitra (hereafter referred to as 'undescribed SAVA material'), but it is more closely related to *D. occulta* and genetically strongly differentiated from topotypic *D. maritima* (CRAMERI *et al.* in prep.). The phylogenomic data further revealed that the material from east-central Madagascar assigned to the two varieties of *D. maritima* forms a clade, which is more closely related to *D. louvelii sensu stricto* (*s.str.*) from littoral forests in the same part of the island than to the southeastern material with racemose inflorescences included in the typical variety by BOSSER & RABEVOHITRA (2002), hereafter referred to as the 'southeastern racemose material'. Moreover, the southeastern racemose material shows strong genetic differentiation from *D. maritima* from east-central Madagascar (CRAMERI *et al.* in prep.). It is thus clear that (i) the southeastern paniculate material cannot be retained in *D. maritima* because the resulting concept would be polyphyletic, (ii) the taxonomic status of the southeastern racemose material needs clarification, and (iii) the broad delimitation of *D. maritima* adopted by BOSSER & RABEVOHITRA (2002) requires amendment.

In this study, we examine the taxonomic status of the southeastern paniculate material and the southeastern racemose material previously associated with *Dalbergia maritima*. We assess the distinctiveness of these entities using multivariate analyses of morphological features of their leaves, leaflets, inflorescences and flowers, as well as ecological data. Our results demonstrate that these two entities are morphologically and eco-geographically coherent and distinct from one another as well as from *D. maritima* from east-central Madagascar, and we formally describe them as two new species, each of which is illustrated with line drawings and photographs taken in the field. We further provide an emended description for *D. maritima*, which we have re-circumscribed so that it comprises a monophyletic and morphologically coherent species from east-central Madagascar, within which we recognize two subspecies, and we present a conservation

assessment based on IUCN Red List categories and criteria (IUCN 2012) for each of the three species.

Materials and Methods

Plant Material

We inspected *Dalbergia* collections deposited at P, TAN, and ZT (acronyms following THIERS 2021), which together hold most type specimens and recent collections of Malagasy *Dalbergia*, including all of the material studied by Bosser & Rabevohitra. Of these, we selected 57 leafy specimens of *D. chapelieri s.l.*, *D. louvelii s.l.*, and material assigned to *D. maritima* from east-central Madagascar as well as the southeastern paniculate and southeastern racemose material, for measurements of leaf and leaflet characters. We further selected 17 flowering specimens of the same entities for measurements of inflorescence and flower characters. The leafy specimens were chosen to represent the observed morphological variation, and the selected flowering specimens include a large proportion of the available material with flowers of these entities.

Morphological Measurements

We assessed 13 leaf and leaflet characters, along with 11 inflorescence and flower characters (Table 1). We initially investigated fruit characters as well, but excluded them from multivariate analyses due to the limited number of collections with mature fruits, substantial intra-individual variation owing to differences in the number of ovules fertilized, and a limited number of measurable characters. Continuous and discrete characters, which were measured several times on a given collection, were recorded as sample medians. Two or three flowers per collection were softened and rehydrated in a warm soapy solution and dissected under a microscope. Measurements of flower characters were made on digital images of the dissected flower organs using ImageJ (SCHNEIDER *et al.* 2012) version 1.53a.

Ecological Characteristics

We downloaded specimen records from the *Catalogue of the Vascular Plants of Madagascar* (MADAGASCAR CATALOGUE 2021) corresponding to all 334 collections known to us of the entities for which morphological measurements were made (Supplementary Material 1). From these, we selected 257 records for collections made since 2000 that contain precise geo-coordinates obtained using a GPS device or were

derived post facto from precise locality data. This was done to minimize the risk of inaccurate associations between occurrences and vegetation class due to imprecise georeferencing or as a result of deforestation that has occurred in the past decades (see VIEILLEDENT *et al.* 2018a).

We assessed 17 potentially relevant ecological characteristics from available spatial raster or vector data for Madagascar (Table 2). We obtained land surface boundaries from the *GADM* database available through the *raster* package (HIJMANS & VAN ETTEN 2012) version 3.4.5, surface lithology classes from the *SERVIR* database available at http://geoportal.rcmrd.org/data/africa_surface_lithology.zip, vegetation classes from MOAT & SMITH (2007), and bioclimate raster data from the *CHELSA* version 1.2 *Bioclim* database (KARGER *et al.,* 2017). We used *R* (R CORE TEAM 2020) version 4.0.2 and the *elevatr* package (HOLLISTER *et al.* 2020) version 0.3.1, the *terra* package (HIJMANS 2021) version 1.0.10, and the *fasterRaster* package (SMITH 2020) version 0.6.0 to download high-resolution elevation data (3 arc seconds, ca. 90 m resolution) and to perform raster calculations. All rasters were projected to Universal Transverse Mercator (UTM) zone 38S and re-sampled to the resolution of the highest-resolved raster (30 m) when needed. We then extracted the ecological characteristics of the 257 selected collections, resulting in an ecological dataset for multivariate analysis. Commented *R* code documenting the download and extraction of ecological characteristics from occurrence data is available in Supplementary Material 2 and on GitHub (https://github.com/scrameri/DalbergiaTaxonomy).

We produced a combined distribution map based on the 334 downloaded occurrence records, supplemented by nine records of the undescribed SAVA material (Supplementary Material 1). We drew alpha convex hulls around the occurrence points of each entity using the *EOO.computing* function in the *ConR* (DAUBY *et al.* 2017) version 1.3.0 package, with an entity-specific alpha parameter (5 for *D. chapelieri s.l.*, 1 for *D. louvelii s.l.*, 10 for the other entities) and buffer (ca. 3 km for *D. pseudomaritima*, ca. 9 km for the other entities). The map was displayed using the *tmap* (TENNEKES 2018) version 3.2 package, with the estimated forest cover for the year 2017 (VIEILLEDENT *et al.* 2018b) drawn at lower resolution (ca. 700 × 1400 pixels), and including the officially recognized terrestrial protected areas in Madagascar (GOODMAN *et al.* 2018). Because these species are under threat from illegal exploitation, we have systematically refrained from making detailed distribution maps and precise geo-coordinates publicly available, and we have standardized latitude and longitude in the published data table. Specimen records are provided in the *Catalogue of the Vascular Plants of Madagascar* (MADAGASCAR CATALOGUE 2021), but with restricted public access to precise geo-locations.

Multivariate Analyses

Analyses of morphological data were carried out separately for the leaf/leaflet and inflorescence/flower datasets, while ten collections with flowers and mature leaves were represented in both datasets. Two missing values in the leaf/leaflet dataset were assigned using class means. For each dataset we calculated a morphological distance matrix using the *daisy* function in the *cluster* package (MAECHLER *et al.* 2019) version 2.1.0 with variable standardization, Gower's distance (GOWER 1971) as the distance metric, a specification of ordinal variables (indument and leaflet margin) as ratio-scaled variables, and binary nominal variables (leaflet texture and inflorescence type) as asymmetric binary variables. We then subjected the resulting distance matrices to principal coordinate analysis (PCoA) using the *pcoa* function in the *ape* package (PARADIS & SCHLIEP 2018) version 5.4.1, and visualized each of the first two principal coordinates and the corresponding rotation matrices as biplots using the *ggplot2* (WICKHAM 2016) version 3.3.3 and the *ggforce* (PEDERSEN 2020) version 0.3.2 packages, and a scaling factor for variable loadings of 0.8 times the smaller ratio of maximum absolute coordinate and maximum absolute variable loading.

We subjected the scaled and centered matrix of ecological data to principal component analysis (PCA) and visualized the first two axes and the rotation matrix as a biplot as above. Commented *R* code documenting the complete workflow for multivariate analyses and the distribution map are available in Supplementary Material 2 and on GitHub (https://github.com/scrameri/DalbergiaTaxonomy).

TABLE 1: Morphological characters of leaves / leaflets (n = 13) and of inflorescences and flowers (n = 11).

Code	Character	Type
Leaf appearance		
LEN_petiole	Petiole length (base to insertion of first basal leaflet)	Continuous
LEN_leaf	Leaf length (petiole + rachis + terminal leaflet)	Continuous
NB_leaflets	Number of leaflets per leaf	Discrete
Leaf indument	Coded as 1 = glabrous; 2 = scattered; 3 = dense	
IND_rachis	Indument on rachis	Ordinal
IND_petiolule	Indument on petiolule	Ordinal
IND_leaflet_up	Indument on upper leaflet lamina	Ordinal
IND_leaflet_low	Indument on lower leaflet lamina	Ordinal
Leaflet size (ratio)		
LEN_petiolule	Petiolule length	Continuous
LEN_leaflet_dist	Length of distal (towards the leaf apex) leaflets (including petiolule)	Continuous
WID_leaflet_dist	Width of distal leaflets at widest point	Continuous
RATIO_leaflet_dist	Ratio of distal leaflet length to width	Continuous

Leaflet texture and margins		
TEX_leaflet	Leaflet texture (thinly coriaceous or coriaceous)	Nominal
MAR_leaflet	Leaflet margin (1 = flat; 2 = thickened but not revolute; 3 = revolute)	Ordinal
Inflorescence		
TYPE_infl	Inflorescence type (racemose or paniculate)	Nominal
IND_infl_axis	Indument on inflorescence axis (coded as above)	Ordinal
Flower		
LEN_pedicel	Pedicel length	Continuous
LEN_flower	Flower length (calyx base to apex of longest petal)	Continuous
LEN_calyx	Calyx length (base to apex of longest calyx lobe)	Continuous
LEN_calyx_up	Length of upper calyx lobes (free part)	
RATIO_flower	Ratio of flower length to calyx length	Continuous
LEN_standard	Length of standard petal (height without claw)	Continuous
WID_standard	Width of standard petal at widest point	Continuous
RATIO_standard	Ratio of standard petal length to width	Continuous
IND_gyn	Indument on gynoecium (coded as above)	Ordinal

TABLE 2: Ecological characteristics (n = 17).

Code	Characteristic	Resolution (m)	Type
Geography			
Latitude	Latitude	30	Continuous
Longitude	Longitude	30	Continuous
Elevation	Elevation	90	Continuous
Topography			
Slope	Slope	90	Continuous
DIST_Coast	Distance to the nearest coast	30	Continuous
Surface lithology			
LITH_Alluvium	Alluvium deposits (fluvial & other)	90	Nominal
LITH_Metaigneous	Metaigneous basement rock	90	Nominal
LITH_Metasediment	Metasedimentary basement rock	90	Nominal
LITH_Silicic	Silicic basement rock	90	Nominal
LITH_Volcanic	Extrusive volcanic (lavas)	90	Nominal
Vegetation class			
VEGE_HumidDegr	Degraded humid forest	30	Nominal
VEGE_Humid	Humid forest	30	Nominal
VEGE_Littoral	Littoral forest	30	Nominal
Bioclimate			
TEMP_Annual	Mean annual air temperature (bio1)	900	Continuous
Isothermality	Isothermality (bio3)	900	Continuous
TEMP_Seasonality	Temperature seasonality (bio4)	900	Continuous
PRECIP_Annual	Annual precipitation (bio12)	900	Continuous

Results

Morphological measurements (medians per individual) are presented in Supplementary Materials 3 (leaf/leaflet dataset) and 4 (inflorescence/flower dataset). The first principal coordinate of foliar (leaf and leaflet) characters (which accounts for 57.15% of the total

variation) was mainly associated with leaf indument (present in collections of *D. louvelii s.l.* and *D. maritima* var. *pubescens*, Table 3), as well as with coriaceous leaflets with revolute margins, which are typical of both varieties of *D. maritima* from east-central Madagascar, but also occur in *D. chapelieri s.l.* and in *D. louvelii s.l.* (Fig. 3A). The second principal coordinate of foliar characters (which explained 25.65% of the total variation) was mainly associated with leaf, petiole and petiolule length and leaflet size, and separated collections with shorter leaves and smaller leaflets (i.e., the southeastern paniculate material) from those with longer leaves and larger leaflets (i.e., most *D. chapelieri s.l.* and some *D. louvelii s.l.*). Leaf length and leaflet size were found to be highly variable within some entities, notably within *D. chapelieri s.l.* and *D. louvelii s.l.* (Fig. 3A).

The first principal coordinate of inflorescence and flower characters (which explained 59.81% of the total variation) was mainly associated with inflorescence type, pedicel length, and the shape of the standard petal (Fig. 3B). It separated entities with paniculate inflorescences, short pedicels, and elliptic to ovate standard petals (*D. chapelieri s.l.* and the southeastern paniculate material [Fig. 4]) from material with racemose inflorescences, longer pedicels, and obovate to orbicular standard petals (*D. louvelii* type, east-central *D. maritima*, and the southeastern racemose material [Fig. 5]). The second principal coordinate of inflorescence and flower characters (which explained 23.11% of the total variation) was mainly associated with flower size (smallest in east-central *D. maritima*, largest in the type of *D. louvelii*, Table 3) and the ratio of flower length to calyx length (small in *D. chapelieri s.l.* and the southeastern paniculate material, large in east-central *D. maritima*, Fig. 3B).

The extracted ecological data are presented in Supplementary Material 5. The first principal coordinate of ecological characteristics (which explained 30.58% of the total variation) separated collections from areas at higher elevation inland the coast and on steeper slopes, where evergreen humid forests are located (*D. maritima* var. *pubescens* and the southeastern racemose material), from sites in flat areas at low elevation and in proximity to the coast, where remnant stands of littoral forests on sand persist (*D. maritima* var. *maritima* and the southeastern paniculate material, Fig. 3C). The two species complexes (*D. chapelieri s.l.* and *D. louvelii s.l.*) occur in both littoral and low-elevation evergreen humid forests (Fig. 3C). The second principal component of ecological characteristics (which explained 19.84% of the total variation) was mainly associated with geography, annual precipitation (which decreases towards southeastern Madagascar), and temperature seasonality (which increases towards southeastern Madagascar), as well as with various types of surface lithology (Fig. 3C). It therefore separated populations from the northeast and the center-east (*D. louvelii s.l.* and *D. maritima*) from those in the

southeast (i.e., the southeastern racemose and the southeastern paniculate material, Figs. 3C–D).

In summary, the southeastern paniculate material has inflorescences and flowers similar to those of *Dalbergia chapelieri s.l.* but differs by its shorter leaves with distinctly smaller and differently shaped leaflets (Table 3), supporting its recognition as a separate species, which we describe below as *D. pseudomaritima* sp. nov. The southeastern racemose material has an inflorescence structure similar to that of east-central *D. maritima* and *D. louvelii s.l.*, but it differs by its consistently glabrous leaves (vs. with indument in *D. louvelii s.l.* and *D. maritima* var. *pubescens*), its thinly coriaceous and narrowly ovate to narrowly elliptic leaflets without revolute margins, and its larger flowers (vs. coriaceous and ovate to elliptic leaflets with revolute margins, and smaller flowers in *D. maritima*, Table 3). We therefore describe the southeastern racemose material as a second new species, *D. razakamalalae* sp. nov. The recognition of these two new species from southeastern Madagascar and their exclusion from *D. maritima sensu* BOSSER & RABEVOHITRA (2002) results in a significantly narrower, more coherent and monophyletic delimitation of *D. maritima*, which is now restricted to populations from the east-central part of the island (Fig. 3D), prompting us to provide an emended description of this economically significant rosewood species. Within *D. maritima*, we have chosen to recognize the two infraspecific entities as subspecies (rather than varieties) because of their clear distinction with regard to both morphology (glabrous vs. pubescent leaves, inflorescence axes and gynoecium, Fig. 3A–B, Table 3) and ecology (littoral forests vs. low-elevation evergreen humid forests, Fig. 3C), and the absence of any overlap in their geographic ranges (Fig. 3D).

TABLE 3: Morphological comparison between *Dalbergia maritima*, *D. pseudomaritima*, *D. razakamalalae* and related taxa discussed in this article. Diagnostic combinations of characters are shown in bold.

Taxon	Leaf length (incl. terminal leaflet, cm)	Leaflet number (per leaf)	Leaflet length (distal, mm)	Leaflet width (distal, mm)	Leaf indument
D. chapelieri s.l.	(8–)10–18(–26)	7–15(–19)	**22–48(–90)**	**11–25(–40)**	Absent
D. pseudomaritima	(4–)5–8(–10)	(8–)10–17(–21)	7–15(–22)	5–8(–12)	Absent
D. maritima subsp. m*aritima*	6–10(–12)	(8–)11–15(–18)	11–15(–19)	6–9(–11)	Absent
D. maritima subsp. *pubescens*	8–13	(9–)13–21(–27)	12–20(–23)	7–9(–11)	**Present**
D. louvelii s.l.	7–19	9–15(–19)	**(16–)20–45**	**(5–)9–18**	**Present**
D. razakamalalae	7–13(–16)	11–19(–23)	**(12–)15–25(–35)**	**5–10(–14)**	Absent

TABLE 3 (continued):

Taxon	Inflorescence type	Pedicel length (mm)	Flower length (mm)	Standard petal width (mm)	Gyn-oecium indument	Fruit width (mm)
D. chapelieri s.l.	**Paniculate**	1.3–4(–5)	8–11	4.0–4.5	Absent	**15–26**
D. pseudomaritima	**Paniculate**	0.5–2.5	8–11	3.5–4.5	Absent	**15–24**
D. maritima subsp. m*aritima*	Racemose	(2–)3–5.5(–7)	8–9	4.0–5.0	Absent	11–16
D. maritima subsp. *pubescens*	Racemose	1.5–2	8–10	ca. 4.5	**Present**	11–19
D. louvelii s.l.	Racemose	(3.5–)5–8	**12–18**	**ca. 9**	Absent	**15–18**
D. razakamalalae	Racemose	2–4(–6)	**10–14**	**5.5–7.5**	Absent	9–15

Taxonomic Treatment

Dalbergia maritima R. Vig., Notul. Syst. (Paris) 14: 185 (1952), emend. Crameri, Phillipson & N. Wilding. TYPE: MADAGASCAR. Atsinanana [Toamasina]: Forêts côtières de l'Est [forêt côtière de Tampina], s.d. (fl), *Louvel 79* (lectotype, designated by Bosser & Rabevohitra, 2002: 346): P [P00060529!], isolectotypes (fr): P [P00060530!, P00060531!]).

Deciduous tree to ca. 10 m tall, or shrub-like when resprouting after felling, bole to ca. 8 m high, dbh to at least 20 cm; bark becoming fissured with age. **Branches** glabrous (subsp. *maritima*) or shortly villose to tomentose on young growth (subsp. *pubescens*), brown *in vivo* (gray-brown to dark purple *in sicco*) when young, becoming gray, lenticels present. **Leaves** alternate, 6–10(–12) cm long (subsp. *maritima*) or 8–13 cm long (subsp. *pubescens*), with (8–)11–15(–18) alternate leaflets (subsp. *maritima*) or (9–)13–21(–27) alternate leaflets (subsp. *pubescens*), petiole and rachis yellow-green *in vivo*, brown to dark purple *in sicco*, glabrous (subsp. *maritima*) or shortly villose to tomentose (subsp. *pubescens*); petiole (6–)9–12 mm long; stipules not seen; leaflets (7–)9–15(–19) × (4–)5–9(–11) mm (subsp. *maritima*) or (7–)10–20(–23) × (5–)6–9(–11) mm (subsp. *pubescens*), often noticeably smaller toward base; petiolule 1.0–1.5 mm long, yellow-green *in vivo*, dark brown *in sicco*, glabrous (subsp. *maritima*) or shortly villose to tomentose (subsp. *pubescens*); lamina ovate to elliptic, coriaceous, base cuneate and often asymmetric,

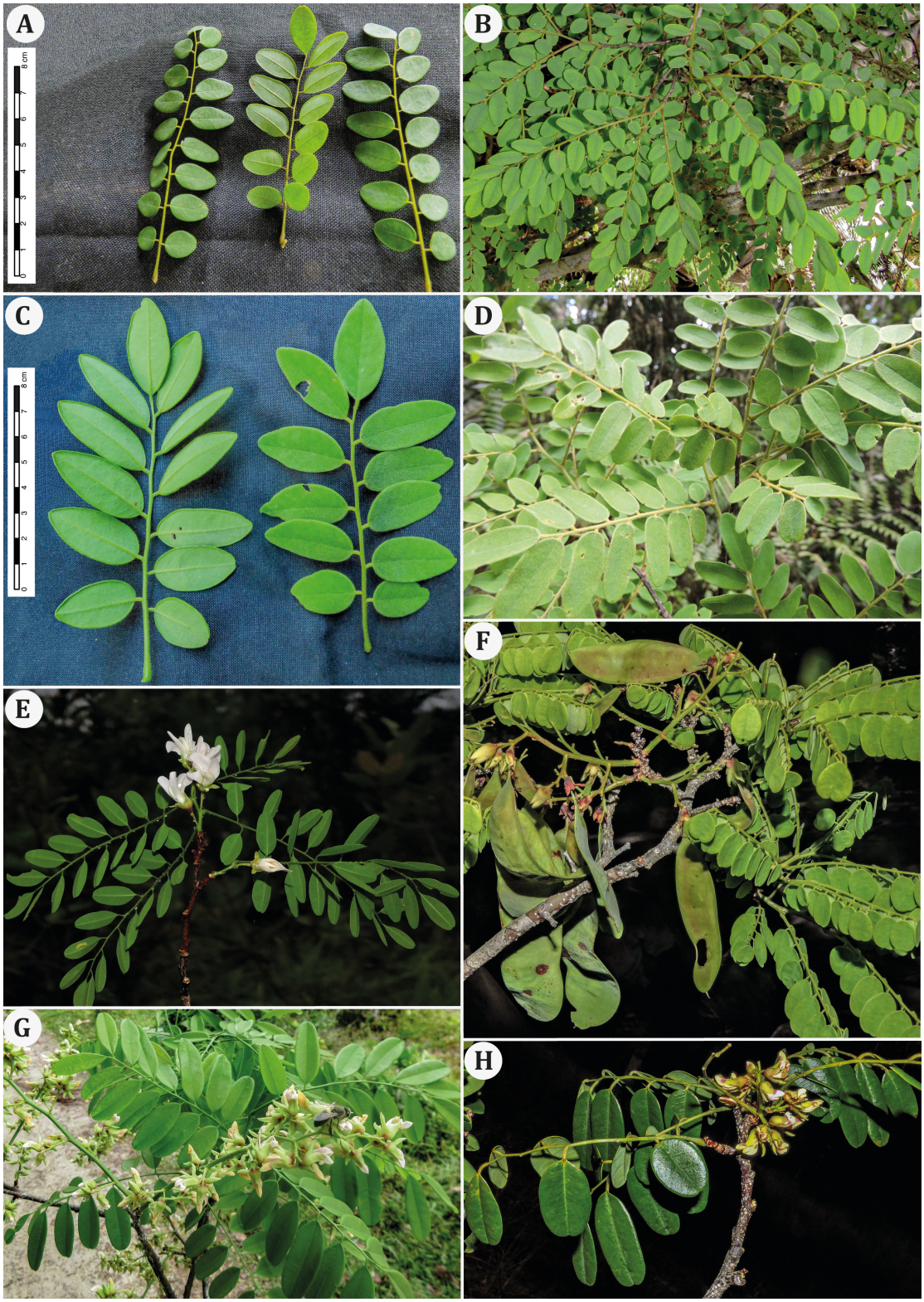

FIGURE 1: Leaves and branches of selected *Dalbergia* taxa from east-central (A–C), northern central-east (D, G), and southeastern Madagascar (E–F, H). **A)** *D. maritima* subsp. *maritima*, Atsinanana Region (*Razakamalala & Rakotovao 8448*). **B)** *D. maritima* subsp. *pubescens*, Atsinanana Region (*Randrianaivo & Sylvain 2928*). **C)** *D. louvelii s.str.*, Atsinanana Region (*Razakamalala & Rakotovao 8432*). **D)** *D. louvelii s.l.*, Analanjirofo Region (*Rakotovao & Bernard 7299*). **E)** *D. razakamalalae*, Anosy Region (*Razakamalala & S. A. Andrianarivelo 8558*). **F)** *D. pseudomaritima*, Anosy Region (*S. A. Andrianarivelo & Razakamalala 63*). **G)** *D. chapelieri s.l.*, Analanjirofo Region (*Lehavana 1109*). **H)** *D. chapelieri s.l.*, Anosy Region (*S. A. Andrianarivelo & Razakamalala 51*). Photos by C. Rakotovao (A, C, D), R. Randrianaivo (B), S. A. Andrianarivelo (E, F, H), A. Lehavana (G).

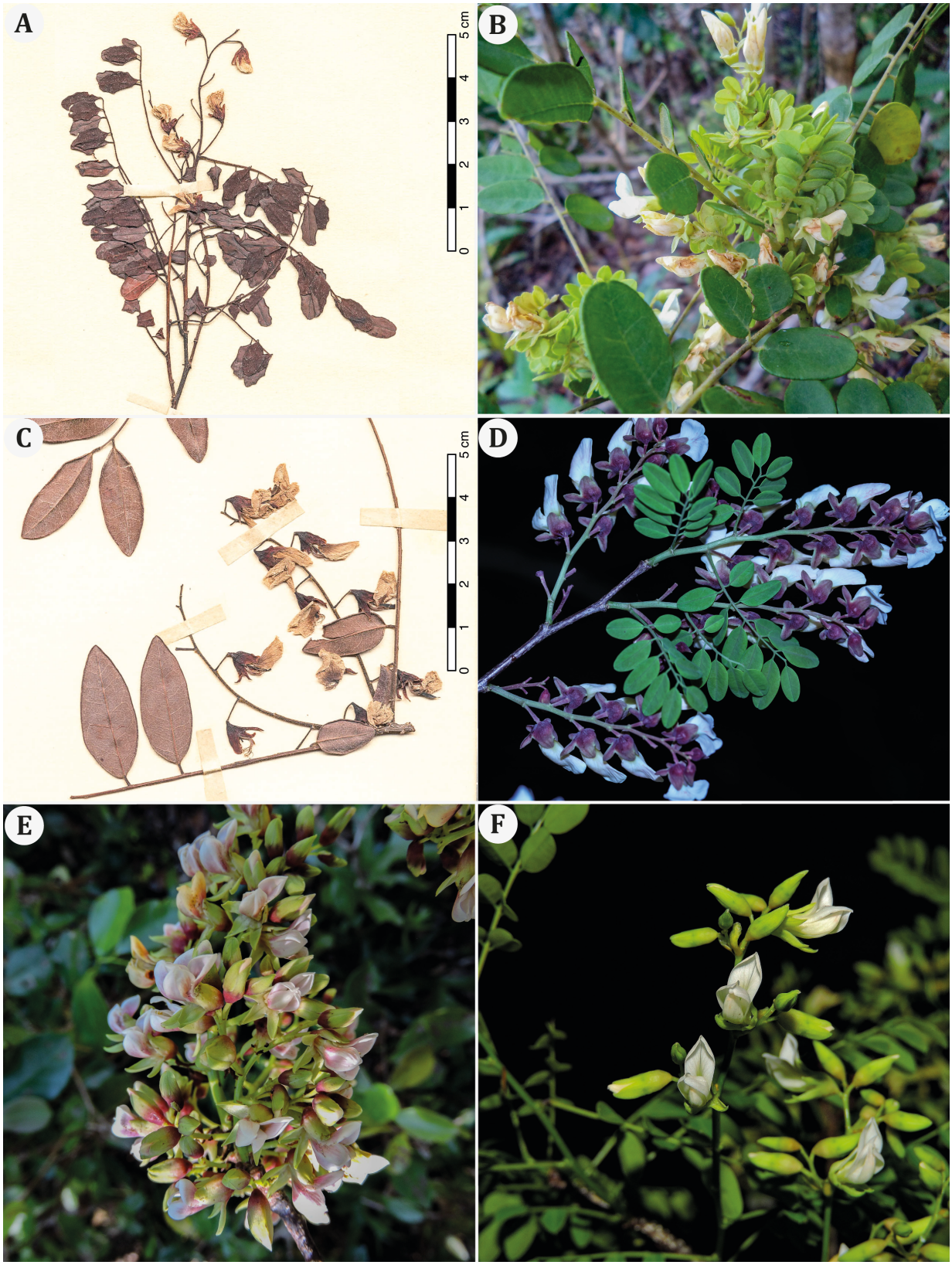

FIGURE 2: Inflorescence structure and flowers of selected *Dalbergia* taxa from east-central (A–C) and southeastern Madagascar (D–F). **A)** *D. maritima* subsp. *maritima*, Atsinanana Region (*Louvel 79*). **B)** *D. maritima* subsp. *pubescens*, Atsinanana Region (*G. Rakotonirina et al. 91*). **C)** *D. louvelii s.str.*, Atsinanana Region (*Louvel 201*). **D)** *D. razakamalalae*, Anosy Region (*Andriamihajarivo et al. 2455*). **E)** *D. chapelieri s.l.*, Atsimo-Atsinanana Region (*N. Rakotonirina & Ravololomanana 1175*). **F)** *D. pseudomaritima*, Anosy Region (*S. A. Andrianarivelo & Razakamalala 65*). Photos by Muséum national d'Histoire naturelle, Paris, France (A, C), F. Rakotoarivony (B), P. Antilahimena (D), N. Rakotonirina (E), S. A. Andrianarivelo (F).

margins revolute *in vivo* and *in sicco*, apex obtuse, sometimes shallowly emarginate, venation brochidodromous, with 5–7 principal lateral veins per side; upper surface matt, mid-green to gray-green *in vivo*, red-brown to dark grayish brown *in sicco*, glabrous (subsp. *maritima*) or pubescent and glabrescent (subsp. *pubescens*), venation inconspicuous (slightly raised *in sicco*), midrib inconspicuous or forming a groove above; lower surface matt, paler than upper *in vivo* and *in sicco*, glabrous (subsp. *maritima*) or pubescent especially along the midrib, becoming puberulous (subsp. *pubescens*), venation forming a loose network of higher-order veins (often paler than matrix *in sicco*) below, midrib prominent. **Inflorescences** racemose, composed of simple racemes with 4–12 alternate flowers each (flowers rarely solitary), often with imparipinnate leafy bracts subtending individual flowers especially near base (thus appearing single-flowered), often pseudo-paniculate with smaller racemes branching off from close to base, 2–5 cm long; axes pale green *in vivo*, dark brown to dark purple *in sicco*, glabrous (subsp. *maritima*) or shortly villose to tomentose (subsp. *pubescens*); anthesis before or concurrent with leaf emergence; peduncle to 9 mm long. **Flowers** often subtended by glabrous (subsp. *maritima*) or pubescent (subsp. *pubescens*), imparipinnate leafy bracts, 12–25 mm long, with 7–13 alternate, ovate to elliptic leaflets, scale-like bracts narrowly ovate, ca. 3.5 × 1.0 mm, caducous; pedicel (1.5–)3–6(–7) mm long, slender, glabrous; bracteoles ca. 2.0 × 0.6 mm, glabrous (subsp. *maritima*), caducous; calyx base to apex of longest petal 8–10 mm long *in sicco*; calyx reddish (subsp. *martitima*, *fide* M. Louvel) or pale yellow-green (subsp. *pubescens*) *in vivo*, purple-brown, darker at base *in sicco*, 4.5–6.0 mm long from base to apex of lower lobe, glabrous (subsp. *maritima*) or with often ciliate lobe margins (subsp. *pubescens*), persistent, 2 upper sepals long-connate, their lobes 1.2–2.0 × 1.3–1.5 mm, apex obtuse, 2 lateral sepals with triangular lobes 1.9–2.6 × 1.0–1.5 mm, lowest sepal with a triangular lobe, margins weakly incurved, apex slightly hooked, 2.1–3.2 × 0.8–1.3 mm; petals glabrous, white at anthesis, becoming cream post anthesis, dark yellow to dark cream *in sicco*; standard petal elliptic to orbicular, claw and lamina forming an obtuse angle, margins incurved forwards when in full flower *in vivo*, apex rounded or notched, 6.0–8.1 × 4.0–5.0 mm, including 1.6–2.6 mm long claw; wing petals 5.4–8.1 × 1.8–2.8 mm, including 1.3–1.8 mm long claw, base distinctly auriculate; keel petals 4.8–7.2 × 1.6–2.7 mm, including 1.0–1.8 mm long claw, base distinctly auriculate; androecium glabrous, monadelphous or diadelphous, 5.8–8.4 mm long; stamens 9–10 or 9+1, free for upper 1.5–3.2 mm; gynoecium 4.0–6.1 mm long, glabrous (subsp. *maritima*) or pubescent (subsp. *pubescens*); stipe ca. 2.0 mm long; ovary 2.4–3.0 mm long, with 3 or 4 ovules; style 1.6–2.4 mm long. **Fruits** (immature) yellow-green *in vivo*, red-brown to purple-brown *in sicco*, with 1–2(–3) seeds, body oblong, 4.5–7.2 × 1.1–1.6(–1.9) cm when single-seeded, up to

8.5 × 1.9 cm when 2-seeded, base cuneate, apex rounded, surface indistinctly net-veined, glabrous; stipe ca. 8 mm long; style rarely persistent. **Seeds** (immature) sub-reniform, flattened, brown, ca. 11 × 6.5 mm. Figs. 1A–B, 2A–B.

Notes—*Dalbergia maritima* was delimited by BOSSER & RABEVOHITRA (2002) to include the populations from southeastern Madagascar recognized here as *D. pseudomaritima* and *D. razakamalalae*, as well as superficially similar collections from the northeastern part of the island (*Service Forestier 2591*, *27751*). However, the populations from the southeast and northeast are genetically distinct and less closely related to *D. maritima* than the latter is to *D. louvelii s.str.* (CRAMERI *et al.* in prep.), with which *D. maritima* co-occurs (Fig. 3D) and from which it is morphologically distinct, as summarized in Fig. 3A–B and Table 3 (but see notes below under *D. maritima* subsp. *pubescens*). The narrower circumscription of *D. maritima* adopted here avoids confusion with the distantly related *D. pseudomaritima* and results in the recognition of monophyletic as well as geographically and morphologically coherent species. The differences in indument on various surfaces (Fig. 3A–B, Table 3) and in habitat (Fig. 3C), and the absence of any overlap in their geographic ranges (Fig. 3D) among the two entities in *D. maritima* support their recognition as subspecies rather than varieties, following the infraspecific taxonomic concepts of CHRISTENSEN (1987). It would not, however, be appropriate to treat them as separate species because they are genetically weakly differentiated and hardly distinguishable based on over 2300 nuclear loci (CRAMERI *et al.* in prep.), and may therefore be inter-fertile.

Dalbergia maritima was first described by R. VIGUIER (ined., 1944) as part of a comprehensive revision of the legumes of Madagascar, but that monumental work was destroyed at the printers in Saint-Lô during a bombardment in June 1944, and it was therefore not effectively published, according to Articles 29.1 & 32.1a of the Shenzhen Code (TURLAND *et al.* 2018). Several years later, H. Humbert validated the names of eleven new *Dalbergia* species described in Viguier's revision, including *D. maritima*, and acknowledging R. Viguier as their posthumous author (VIGUIER 1952).

Conservation Status—*Dalbergia maritima* is known from 29 collection records that represent 23 extant occurrences and 6 occurrences that appear to have been extirpated. Its former Extent of Occurrence (EOO) was at least 2564 km^2 and its former Area of Occupancy (AOO) was at least 80 km^2 (based on a 4 km^2 grid). Its current geographic range has the form of an EOO of 1875 km^2 and an AOO of 60 km^2, and comprises five subpopulations. The species mainly occurs in forest ecosystems (MADAGASCAR CATALOGUE 2021). Forest cover decline between 1953 and 2017 was estimated from the forest cover time series of VIEILLEDENT *et al.* (2018a, 2018b) to be 77% in the altitudinal

range of 0–450 m and within the minimum convex polygon encompassing all known collections of this species. Therefore, *D. maritima* is inferred to have undergone and to be undergoing continuing decline in EOO, AOO, quality of habitat, number of subpopulations, and number of mature individuals. This species occurs at four locations with respect to the most serious plausible threat, which is selective logging for trade in its high-quality heartwood, as inferred from older collections with exploitable diameter and tree stumps observed during recent field work in east-central Madagascar. The occurrences within the protected areas of Betampona, Sahafina, and Vohibola (where a local association provides some level of protection) represent three separate locations. All occurrences outside of protected areas can be inferred to represent a single additional (fourth) location based on the IUCN Red List guidelines (IUCN 2019), because of the large spatial scale at which illegal selective logging (or habitat degradation and loss) can severely reduce the population within a single generation length (at least 30–40 years). Moreover, most known subpopulations of this species can be accessed by road or train, and harvest intensity can be regarded as similar over large spatial scales spanning similarly accessible areas. For these reasons, *Dalbergia maritima* is assigned a preliminary IUCN conservation status of Endangered: EN B1ab(i,ii,iii,iv,v)+2ab(i,ii,iii,iv,v).

Identification Key to the Subspecies of Dalbergia maritima

1. Leaves glabrous on all surfaces; inflorescence axes and gynoecium glabrous; littoral forests on sand, at 0–30 m elevation.....***D. maritima* subsp. *maritima***
1'. Leaves with sparse to dense indumentum on all surfaces, glabrescent on upper surface of leaflets; inflorescence axes shortly villose to tomentose, gynoecium pubescent; low-elevation evergreen humid forest on laterite, at 80–450 m elevation..... .....***D. maritima* subsp. *pubescens***

Dalbergia maritima* subsp. *maritima

Vernacular Names and Uses—Bois de rose (*Rakotovao & Razakamalala 7467, 7474*), Hazomainty (*Rakotovao & Razakamalala 7482*), Volombodipony (*Louvel 79*), Volombodipony à petites feuilles (*Louvel 200*), Volombodipony Isthy (*Louvel 200*), Volombodipony lahy (*Louvel 200*).

The heartwood of this subspecies is burgundy in color (*Rakotovao & Razakamalala 7467*), and with time becomes blackish and similar to ebony (*Louvel 200*). It is considered to be a high-quality rosewood (NORMAND 1988; RICHTER *et al.* 2014).

Habitat, Distribution and Phenology—This taxon is restricted to littoral forests at 0–30 m elevation in east-central Madagascar (Atsinanana region), mainly in the coastal areas between Ambila-Lemaitso and Tampina (Fig. 3D). *Dalbergia maritima* subsp. *maritima* has been collected in full flower from February to March.

Notes—Two specimens with glabrous leaves from the SAVA region (*Service Forestier 2591, Service Forestier (Capuron) 27751*), which were associated with *D. maritima* by Bosser & Rabevohitra (2002), are here excluded from *D. maritima* subsp. *maritima*, since they belong to a different species (CRAMERI *et al.* in prep.) corresponding to the undescribed SAVA material (Fig. 3D).

Additional Specimens Examined—**Madagascar.**—ATSINANANA [Toamasina]: Ambila-Lemaitso, 6 Feb 1951 (fl), *Service Forestier 2860* (P, TAN); same locality, 1 Jan 1952, *Service Forestier 5-R-233* (P); Andranampy forest (Vavony), 2 Jun 2019, *Rakotovao & Razakamalala 7467* (DBEV, MO, P, TAN, ZT); same locality, same date, *Rakotovao & Razakamalala 7468* (DBEV, MO, P, TAN, ZT); same locality, same date (fr), *Rakotovao & Razakamalala 7469* (DBEV, MO, P, TAN, ZT); Tampina, Mar 1924 (fl), *Louvel 200* (P); Vohibola forest (Andranokoditra), 4 Jun 2019, *Rakotovao & Razakamalala 7474* (DBEV, MO, P, TAN, ZT); same locality, same date, *Rakotovao & Razakamalala 7478* (DBEV, MO, P, TAN, ZT); same locality, same date, *Rakotovao & Razakamalala 7479* (DBEV, MO, P, TAN, ZT); same locality, same date, *Rakotovao & Razakamalala 7482* (DBEV, MO, P, TAN, ZT); same locality, same date, *Razakamalala & Rakotovao 8444* (DBEV, MO, P, TAN, ZT); same locality, same date, *Razakamalala & Rakotovao 8448* (DBEV, MO, P, TAN, ZT); same locality, same date, *Razakamalala & Rakotovao 8454* (DBEV, MO, P, TAN, ZT); same locality, 15 Dec 1982 (bud), *Service Forestier 32479* (P, TEF).

Dalbergia maritima **subsp.** ***pubescens*** (Bosser & R. Rabev.) Crameri, Phillipson & N. Wilding, stat. nov. BASIONYM: *Dalbergia maritima* R. Vig. var. *pubescens* Bosser & R. Rabev., Bull. Mus. Natl. Hist. Nat., B, Adansonia Sér. 4, 18(3–4): 208 (1996). TYPE: MADAGASCAR. Atsinanana [Toamasina]: Environs de Foulpointe [Mahavelona], 1985, *Service Forestier 32824* (holotype: P [P00060551!], isotype: TEF [TEF000141]).

Vernacular Names and Uses—Andramena kely ravina (*Bernard & Razakamalala 2247*), Andramena, Hitsika or Volombodinpona (*Service Forestier 18-R-195*).

The heartwood of *Dalbergia maritima* subsp. *pubescens* is used in carpentry, cabinet making, and construction (*Service Forestier 18-R-195*). It is considered to be a high-quality rosewood (*Razakamalala & Bernard 8368*).

Habitat, Distribution and Phenology—This subspecies occurs in low-elevation evergreen humid forests at 80–450 m elevation, and is restricted to east-central Madagascar (Atsinanana Region), recorded from sites between and around the protected areas of Sahafina in the south and Betampona in the north (Fig. 3D). *Dalbergia maritima* subsp. *pubescens* has been collected in full flower in March (*G. Rakotonirina et al. 389*), although immature fruits have been recorded from late January (*Antilahimena 9712, 9720*).

Notes—*Dalbergia maritima* subsp. *pubescens* is distinct from the nominal subspecies on the basis of indument on its leaves, inflorescence axes, and gynoecium, but sterile specimens can potentially be confused with *D. louvelii s.l.* (Figs. 1C–D, 2C), from which it differs in its smaller flowers (as observed in *D. louvelii s.str.*, from littoral forests in east-central Madagascar) and narrower fruits, and in its smaller, more numerous, and differently shaped leaflets, as shown in Figs. 1B, 2B, 3A–B and summarized in Table 3. Moreover, the currently known geographic ranges of these two taxa do not appear to overlap, and the geographically closest similar entity (*D. louvelii s.str.*, which co-occurs with *D. maritima* subsp. *maritima* in littoral forests in east-central Madagascar) occupies a different habitat type (Fig. 3C–D). A single collection of *D. maritima* subsp. *pubescens* (the type, *Service Forestier 32824*) from the surroundings of Mahavelona (Foulpointe), but lacking precise locality details, is suspected to originate from the Analalava protected area, situated ca. 6 km to the southwest of Mahavelona, which would increase its range by ca. 30 km to the north, but no extant occurrences are known from that area, despite intensive recent botanical inventory work at Analalava.

Additional Specimens Examined—**Madagascar.**—ATSINANANA [Toamasina]: Ambodiriana commune, 24 Mar 2017 (fl), *G. Rakotonirina et al. 9*1 (K, MO, P, TAN, UPS); Antetezambaro commune, 12 Oct 2019, *Karatra & Ramanitrinizaka 190* (DBEV, MO, P, TAN, ZT); same locality, 28 Jan 2021 (y.fr), *Antilahimena 9712* (MO, P, TAN); same locality, same date (y.fr), *Antilahimena 9720* (MO, P, TAN); Betampona Special Reserve and surrounding areas, 11 Nov 2016, *Randrianaivo & Sylvain* 2928 (P, TAN, ZT); same locality, 16 Feb 2018 (y.fr), *Randrianaivo 3136* (G, MO, P, TEF, ZT); same locality, 18 Jan 2014, *Razakamalala & Bernard 7704* (BR, G, MO, P, ZT); same locality, 20 Jan 2014, *Bernard & Razakamalala 2247* (BR, G, MO, P, ZT); same locality, 7 Aug 1986 (fr), *Service Forestier 31184* (P, TEF); Masiabarika forest, 17 Dec 1954, *Service Forestier 18-R-195* (P); Sahafina protected area, 16 Apr 2019, *Razakamalala & Bernard 8368* (DBEV, MO, P, TAN, ZT); same locality, 17 Apr 2019, *Bernard & Razakamalala 2734* (DBEV, MO, P, TAN, ZT); Toamasina suburbaine commune, 21 Feb 2018 (y.fr), *G. Rakotonirina et al. 389* (K, MO, P, TAN, UPS).

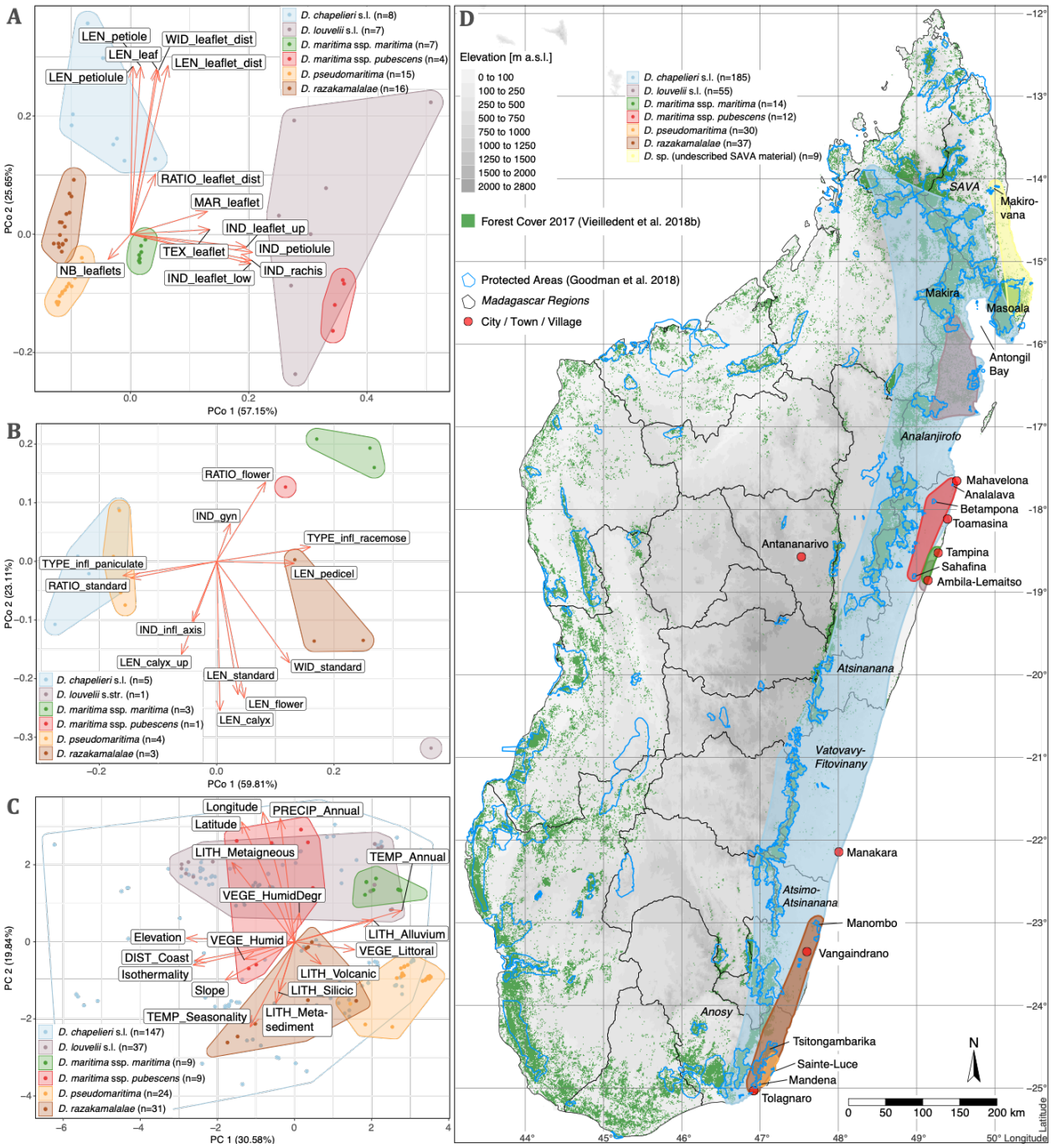

FIGURE 3: Multivariate analysis of morphological and ecological data, and geographic distribution areas of selected *Dalbergia* taxa from eastern Madagascar. **A)** PCoA biplot of 13 leaf and leaflet characters (arrows) measured in 57 collections with mature leaves. **B)** PCoA biplot of 11 inflorescence and flower characters (arrows) measured in 17 collections with flowers. **C)** PCA biplot of 17 ecological characteristics (arrows) for 257 collections made since 2000 with precise geo-coordinates. Shading for *D. chapelieri s.l.* has been omitted for clarity. **D)** Combined distribution map based on buffered alpha hulls drawn around 342 occurrences, including collection records from before 2000, nine records from the undescribed SAVA material, and excluding a doubtfully identified collection (*Service Forestier 37-R-118*). Note that *D. maritima* subsp. *maritima* and *D. louvelii s.str.* co-occur in the Atsinanana Region.

Dalbergia pseudomaritima Crameri, Phillipson & N. Wilding, sp. nov. TYPE: MADAGASCAR. Anosy [Toliara]: Sainte Luce, 13 Feb 2019 (fr), *N. Rakotonirina, R. Razakamalala & R. Bernard 1190* (holotype: P!, isotypes: MO, TAN!, ZT!).

Dalbergia pseudomaritima is similar to *D. chapelieri* Baill. in possessing paniculate inflorescences that appear before or at the same time as the emerging, glabrous leaves, but differs by its shorter leaves [(4–)5–8(–10) cm vs. (8–)10–18(–26) cm long] with distinctly smaller leaflets [(5–)8–14(–22) × (4–)5–8(–12) mm vs. 21–48 × 11–22 mm and sometimes reaching 90 × 40 mm on coppice shoots] that are broadly elliptic to orbicular (vs. elliptic to oblong-elliptic or obovate), resembling in number and size those of *Dalbergia maritima* R. Vig.

Deciduous tree to ca. 12 m tall, or shrub-like when resprouting after felling, bole to ca. 7 m high, dbh to at least 25 cm; bark smooth at first, becoming fissured with age. **Branches** glabrous, orange-brown *in vivo* (dark brown to dark purple *in sicco*) when young, becoming gray, lenticels present. **Leaves** alternate, (4–)5–8(–10) cm long, with (8–)10–17(–21) alternate leaflets, petiole and rachis bright green *in vivo*, purple-brown *in sicco*, glabrous; petiole (6–)8–10(–12) mm long; stipules 4.0–6.5 × 1.0–2.0 mm, obovate, caducous; leaflets (5–)8–14(–22) × (4–)5–8(–12) mm, sometimes noticeably smaller toward base, but often rather uniform; petiolule 0.5–2.0 mm long, yellow-green *in vivo*, dark brown to black *in sicco*, glabrous; lamina broadly elliptic to orbicular, rarely obovate, thinly coriaceous, base broadly cuneate, margins thickened but not revolute *in sicco*, apex shallowly retuse, sometimes mucronulate or rounded, venation brochidodromous, with 5–9 principal lateral veins per side; upper surface matt, yellow-green *in vivo*, olive-green to red-brown *in sicco*, glabrous, venation inconspicuous (slightly raised *in sicco*), midrib inconspicuous or forming a groove; lower surface matt, paler than upper *in vivo* and *in sicco*, glabrous, venation forming a dense network of higher-order veins, contrasting and often darker than matrix *in sicco*, highest-order veins often open-ended, midrib prominent. **Inflorescences** paniculate, composed of numerous lateral and often densely clustered panicles of (2–)6–20 flowers each (sometimes reduced to racemes or solitary flowers), each panicle with a terminal flower, compact, 2–5 cm long, with (2–)4–6 paniculate branches composed of 2–6 flowers each; axes green *in vivo*, dark brown *in sicco*, glabrous or sparsely and minutely ciliate at junctions; anthesis before or concurrent with leaf emergence; peduncle to 8 mm long. **Flowers** subtended by glabrous or minutely ciliate, oblong to obovate bracts, 3.5–6.0 × 1.0–1.5 mm, caducous; pedicel 0.5–2.5 mm long, glabrous; bracteoles cucullate and enclosing flower bud, glabrous or minutely ciliate, caducous; calyx base to apex of longest petal 8–12 mm long *in sicco*; calyx green, reddish

at base *in vivo*, yellow-brown to purple-brown *in sicco*, 5–8 mm long from base to apex of lower lobe, glabrous or sparsely and minutely ciliate, persistent, 2 upper sepals long-connate, their lobes 1.8–2.5 × 1.9–2.5 mm, apex obtuse to subacute, 2 lateral sepals with cymbiform lobes 3.1–4.2 × 1.3–1.9 mm, lowest sepal with a triangular lobe, margins incurved, apex often distinctly hooked, 3.4–4.6 × 1.0–2.5 mm; petals glabrous, white or pinkish-white at anthesis, becoming cream post anthesis, yellow to brown *in sicco*; standard petal ovate to elliptic to obovate, claw and lamina almost perpendicular, margins incurved forwards *in vivo*, apex notched, 8.8–10.0 × 3.7–4.7 mm, including 2.5–3.5 mm long claw; wing petals 7.3–10.3 × 2.0–2.8 mm, including 1.5–2.9 mm long claw, base distinctly auriculate; keel petals 7.4–9.4 × 2.5–3.1 mm, including 1.8–2.7 mm long claw, base distinctly auriculate; androecium glabrous, monadelphous or diadelphous, 9.4–10.6 mm long; stamens 9–10 or 9+1, free for upper 2.7–4.0 mm; gynoecium 6.4–8.2 mm long, glabrous; stipe 3–4 mm long; ovary 3.5–4.5 mm long, with 3–5 ovules; style 1.9–2.5 mm long. **Fruits** (immature) yellow-green becoming red-brown *in vivo*, yellow-brown to red-brown *in sicco*, with 1–3(–4) seeds, body elliptic to oblong, 4.5–6.5 × 1.6–2.3 cm when single-seeded, up to 8.5 × 2.5 cm when 3-seeded, base attenuate, apex rounded or obtuse, surface with reticulate veins, glabrous; stipe 5–10 mm long; style persistent. **Seeds** (immature) sub-reniform, flattened, brown, 8.0–9.0 × 5.0–6.0 mm. Figs. 1F, 2F, 4.

Etymology—The epithet reflects the superficial similarity to and confusion with *Dalbergia maritima*.

Vernacular Names and Uses—Manary (*Ramamonjiarisoa 4*), Manary toloho (*Ramamonjiarisoa 10*), Sambalahy (*Ramison & Ramisy 108*), Tombobitsy (*Razafimandimby et al. 237*).

The heartwood of *Dalbergia pseudomaritima* is orange-brown in color (*S. A. Andrianarivelo & Razakamalala 58*, *Razakamalala & S. A. Andrianarivelo 8566*). Its wood is used as firewood and for charcoal production (R. Randrianaivo, pers. comm.).

Habitat, Distribution and Phenology—*Dalbergia pseudomaritima* occurs in littoral forests on sand and adjacent swamp forests (*Razakamalala et al. 6675*), with one collection from low-elevation evergreen humid forests on sandy lateritic soils near a stream (*Bernard et al. 2654*), at 0–30 m elevation. It is restricted to southeastern Madagascar (Anosy Region), occurring mainly in the protected areas of Mandena and Sainte Luce (Fig. 3D). *Dalbergia pseudomaritima* has been collected in full flower from October to January.

Conservation Status—*Dalbergia pseudomaritima* is known from 42 collection records that represent 29 extant occurrences and 13 occurrences that appear to have been extirpated. Its former Extent of Occurrence (EOO) was at least 275 km^2 and its former Area of Occupancy (AOO) was at least 56 km^2 (based on a 4 km^2 grid), whereas its current

geographic range has the form of an EOO of 252 km^2 and an AOO of 44 km^2, and comprises three subpopulations. The species mainly occurs in forest ecosystems (MADAGASCAR CATALOGUE 2021). Forest cover decline between 1953 and 2017 was estimated from the forest cover time series of VIEILLEDENT *et al.* (2018a, 2018b) to be 35% in the altitudinal range of 0–30 m and within the minimum convex polygon encompassing all known collections of this species. Therefore, *D. pseudomaritima* is inferred to have undergone and to be undergoing continuing decline in EOO, AOO, quality of habitat, number of subpopulations, and number of mature individuals. This species occurs at four locations with respect to the most serious plausible threat, which is habitat degradation or loss due to land clearing and fire for subsistence agriculture. The occurrences within the protected areas of Mandena and Sainte Luce represent two separate locations. Occurrences outside of the Sainte Luce protected area, including sites north of the Ebakika river, represent the third location. The Ampasy forest subpopulation, which is comparatively less accessible, represents the fourth location. For these reasons, *Dalbergia pseudomaritima* is assigned a preliminary IUCN conservation status of Endangered: EN B1ab(i,ii,iii,iv,v)+2ab(i,ii,iii,iv,v).

Notes—Material of *Dalbergia pseudomaritima* has previously been included in or associated with *D. maritima sensu* BOSSER & RABEVOHITRA (2002), mainly owing to their overlapping morphological variation with respect to leaflet size and number, and due to their occurrence in littoral forest. However, *D. pseudomaritima* differs by numerous characters of its leaves, inflorescences, flowers and fruits, as summarized in Fig. 3A–B and Table 3. By contrast, the inflorescence and flower characters of *D. pseudomaritima* are similar to those of the closely related *D. chapelieri s.l.*, with which it shares an often conspicuous reticulate venation with open-ended highest-order veins on the lower leaflet laminae. The currently known geographic ranges of *D. chapelieri s.l.* and *D. pseudomaritima* do not appear to overlap. A single collection of *D. pseudomaritima* is known from a site located outside of (remaining or former) littoral forest habitat (*Bernard et al. 2654*), from the Ampasy forest in Iaboakoho commune, on sandy lateritic soils near a stream. In the same general area, *D. pseudomaritima* might come into contact with neighboring populations from low-elevation evergreen humid forests attributed to the most closely related lineage within *D. chapelieri s.l.* (e.g., *Razakamalala 7739* [P01069112], *Razakamalala 7765* [P01069094], *S. A. Andrianarivelo & Razakamalala 51*, Fig. 1H). However, *D. pseudomaritima* clearly differs from these individuals by its shorter leaves [(4–)5–8(–10) cm vs. (8–)10–13) cm long] with distinctly smaller leaflets [(5–)8–14(–22) × (4–)5–8(–12) mm vs. (17–)23–40(–51) × (8–)10–19(–25) mm] that are broadly elliptic to orbicular (vs. ovate to elliptic) and thinly coriaceous (vs. coriaceous) and have plane (vs.

revolute) margins (Figs. 1F-H, 3A, Table 3), and no individuals with an intermediate genotype have been found (CRAMERI *et al.* in prep.).

***Additional Specimens Examined*—Madagascar.**—ANOSY [Toliara]: Ambanihazo village (Iabakoho commune), 31 Aug 2012, *Ludovic 1570* (TAN); same locality, 25 Nov 2011 (fl), *Razakamalala et al. 6675* (MO, P, TAN); Ampasy forest (Iabakoho commune), 10 Feb 2019, *Bernard et al. 2654* (DBEV, MO, P, TAN, ZT); Mandena protected area and surroundings, 21 Nov 1977, *Ramamonjiarisoa 2* (P); same locality, same date (fl), *Ramamonjiarisoa 4* (P); same locality, same date, *Ramamonjiarisoa 5* (P); same locality, same date, *Ramamonjiarisoa 10* (P); same locality, 12 Jun 1991 (fr), *Zarucchi et al. 7593* (K, MO, P); same locality, 7 Dec 1989 (fl), *Dumetz & McPherson 1139* (K, MO, P); same locality, 16–17 Oct 1989 (bud), *Rabevohitra 2033* (K, MO, P, TEF, WAG); same locality, 7 Apr 2014, *Razakamalala 7783* (MO, P, TAN); same locality, Nov 1978 (fl), *Service Forestier 30547* (P); Mandromodromotra, 6 Dec 2006 (fl), *Ramison & Ramisy 108* (MO, P, TAN); same locality, same date (y.fr), *Ramison & Ramisy 109* (MO, P, TAN); Sainte Luce protected area and surroundings, 22 Nov 2011 (y.fr), *Ratovoson 1713* (MO, P, TAN); same locality, 16 Jan 1990 (y.fr), *McPherson et al. 14804* (MO); same locality, 16 Oct 2008 (fl, y.fr), *Razafimandimby et al. 237* (TEF); same locality, 18 Nov 2004 (fl), *Raharimampionona et al. 1* (MO, P, TEF); same locality, 4 Nov 2003 (fl), *Rabenantoandro et al. 1556* (MO, P, TEF); same locality, 15 Dec 2000 (fl), *Faliniaina et al. 10* (L, MO, P, TEF, WAG); same locality, 18 Dec 1993 (y.fr), *Luckow 4150* (BH, K, MO, TAN, WAG, Z); same locality, 16 Jan 1990 (y.fr), *Dumetz 1195* (K, MO, P); same locality, 26 Apr 1989, *Rabevohitra 1928* (MO, P); same locality, 15–16 Jan 1990 (y.fr), *Rabevohitra 2145* (K, MO, P, TEF); same locality, 17–18 Jan 1990 (fl), *Rabevohitra 2178* (K, MO, P, TEF); same locality, 6 Nov 2019, *Razakamalala & S. A. Andrianarivelo 8566* (DBEV, MO, P, TAN, ZT); same locality, same date (fl), *Razakamalala & S. A. Andrianarivelo 8567* (DBEV, MO, P, TAN, ZT); same locality, same date, *Razakamalala & S. A. Andrianarivelo 8568* (DBEV, MO, P, TAN, ZT); same locality, same date (y.fr), *Razakamalala & S. A. Andrianarivelo 8569* (DBEV, MO, P, TAN, ZT); same locality , same date, *Razakamalala & S. A. Andrianarivelo 8570* (DBEV, MO, P, TAN, ZT); same locality, same date (y.fr), *Razakamalala & S. A. Andrianarivelo 8571* (DBEV, MO, P, TAN, ZT); same locality, 5 Apr 2014, *Razakamalala et al. 7767* (MO, P, TAN); same locality, 20 Oct 2012 (fl), *Razakamalala et al. 7228* (MO, P, TAN); same locality, 17 Oct 2012 (fl, y.fr), *Ramananjanahary et al. 780* (MO, P, TAN); same locality, 20 Oct 2012 (bud, fl, y.fr), *Ramananjanahary et al. 830* (MO, P, TAN); same locality, 29 Mar 1989 (fr), *Gereau et al. 3326* (K, MO, P, WAG); same locality, 7 Nov 2019, *S. A. Andrianarivelo & Razakamalala 58* (DBEV, MO, P, TAN, ZT); same locality, same date (fl), *S. A.*

Andrianarivelo & Razakamalala 60 (DBEV, MO, P, TAN, ZT); same locality, same date (fl, y.fr), *S. A. Andrianarivelo & Razakamalala 63* (DBEV, MO, P, TAN, ZT); same locality, same date, *S. A. Andrianarivelo & Razakamalala 64* (DBEV, MO, P, TAN, ZT); same locality, same date (fl, y.fr), *S. A. Andrianarivelo & Razakamalala 65* (DBEV, MO, P, TAN, ZT).

Dalbergia razakamalalae Crameri, Phillipson & N. Wilding, sp. nov. TYPE: MADAGASCAR. Anosy [Toliara]: Forêt d'Analamahavondjaky (commune de Iaboakoho), 7 Dec 2019 (fl), *T. Andriamihajarivo, N. H. Rakotoarivelo & F. Rakotoarivony 2455* (holotype: P!, isotypes: MO, TAN!, ZT!).

Dalbergia razakamalalae is similar to *D. maritima* R. Vig. in possessing leaves with rather small leaflets and racemose inflorescences, but differs by its consistently glabrous leaves (vs. glabrous or pubescent), larger flowers (10–14 mm vs. 8–10 mm long), and narrowly ovate to narrowly elliptic leaflets (vs. ovate to elliptic) that are thinly coriaceous (vs. coriaceous) and have plane (vs. revolute) margins and frequently an emarginate (vs. obtuse to rounded) apex.

Deciduous tree to ca. 20 m tall, or shrub-like when resprouting after felling, bole to ca. 15 m high, dbh to at least 40 cm; bark smooth at first, becoming fissured with age. **Branches** glabrous, pale brown to purple-brown *in vivo* (dark brown to dark purple *in sicco*) when young, becoming gray-brown, lenticels present. **Leaves** alternate, 7–13(–16) cm long, with 11–19(–23) alternate leaflets, petiole and rachis purplish-green *in vivo*, dark brown to dark purple *in sicco*, glabrous; petiole (9–)12–20(–25) mm long; stipules ca. 3.0 × 1.0 mm, narrowly ovate, caducous; leaflets (8–)13–25(–35) × (4–)5–10(–14) mm, often noticeably smaller toward base or/and apex; petiolule 1.0–2.0 mm long, yellow-green *in vivo*, dark brown to dark purple *in sicco*, glabrous; lamina narrowly ovate to narrowly elliptic, rarely ovate to elliptic, thinly coriaceous, base cuneate, margins not revolute *in sicco*, apex emarginate, rarely obtuse, venation brochidodromous, with 5–7 principal lateral veins per side; upper surface mat, mid-green *in vivo*, dark purple-brown *in sicco*, glabrous, venation inconspicuous (slightly raised *in sicco*), midrib inconspicuous or forming a groove; lower surface paler than upper *in vivo* and *in sicco*, glabrous, venation forming a loose network with higher-order veins (often paler than matrix *in sicco*), midrib prominent. **Inflorescences** racemose, composed of simple racemes with (2–)4–12 alternate flowers each (flowers rarely solitary), often with imparipinnate leafy bracts subtending individual flowers especially near base (thus appearing single-flowered), often pseudo-paniculate with smaller racemes branching off from close to base, 2–5 cm long; axes green

to purple-green especially at apex *in vivo*, dark brown to dark purple *in sicco*, glabrous; anthesis before or concurrent with leaf emergence; peduncle to 6 mm long. **Flowers** often subtended by glabrous, imparipinnate leafy bracts, 22–57 mm long, with 7–13 alternate, narrowly ovate to narrowly elliptic leaflets, scale-like bracts not seen; pedicel 2–4(–6) mm long, slender, glabrous; bracteoles ca. 2.7 × 0.6 mm, glabrous, caducous; calyx base to apex of longest petal 10–14 mm long *in sicco*; calyx bright green to purple and brightly dotted especially at base *in vivo*, purple-brown, darker at base *in sicco*, 7–9 mm long from base to apex of lower lobe, glabrous, persistent, 2 upper sepals long-connate, their lobes 2.3–3.9 × 2.5–2.9 mm, apex obtuse to rounded, 2 lateral sepals with triangular lobes 3.2–4.2 × 1.5–2.2 mm, lowest sepal with a triangular lobe, margins weakly incurved, apex slightly hooked, 3.2–4.2 × 1.4–2.2 mm; petals glabrous, white with often pink or bluish tinged veins at anthesis, dark yellow to dark cream *in sicco*; standard petal broadly obovate to orbicular, claw and lamina forming an obtuse angle, margins slightly incurved backwards when in full flower *in vivo*, apex notched, 9.6–11.5 × 5.8–9.2 mm, including 2.4–3.9 mm long claw; wing petals 7.3–10.8 × 2.2–3.2 mm, including 2.0–2.8 mm long claw, base distinctly auriculate; keel petals 7.3–9.3 × 2.4–2.9 mm, including 2.0–2.9 mm long claw, base distinctly auriculate; androecium glabrous, diadelphous, 6.4–10.3 mm long; stamens 9+1, free for upper 1.7–5.0 mm; gynoecium 7.0–7.7 mm long, glabrous; stipe ca. 3.5 mm long; ovary 4.3–5.5 mm long, with 3–5 ovules; style 1.4–1.8 mm long. **Fruits** (immature) purple-red to carmine *in vivo*, purple-brown *in sicco*, with 1–3 seeds, body oblong or narrowly elliptic, 3.5–5.5 × 0.8–1.5 cm when single-seeded, up to 7.5 × 1.7 cm when 3-seeded, base cuneate, apex rounded or acute, surface indistinctly net-veined, glabrous; stipe ca. 7–10 mm long; style rarely persistent. **Seeds** (immature) sub-reniform, flattened, brown, ca. 6 × 3 mm. Figs. 1E, 2D, 5.

Etymology—*Dalbergia razakamalalae* is named in honor of the botanist Richardson Razakamalala, who has made nearly 9000 high-quality collections over the last two decades, contributing significantly to the knowledge of the flora of Madagascar, and which have included collections of this and many other *Dalbergia* species made while working together with local guides and other members of the Missouri Botanical Garden's research team in Madagascar.

Vernacular Names and Uses—Sambalahimanga (*Andriamihajarivo et al. 2455*), Tombobitsy lahy (*Razakamalala & S. N. Andrianarivelo 8035*), Tongobitsy or Tambobitsy (*Réserves Naturelles 1689*).

The heartwood of *Dalbergia razakamalalae* is beautifully veined and burgundy-colored (*Bernard et al. 2645*, Karatra & *Rakotovao 242*, *Ramanitrinizaka &*

Sandratriniaina 1). It is considered to be a high-quality rosewood (*Humbert 20607*) and is used in cabinet making (*Humbert 20355bis*).

Habitat, Distribution and Phenology—This species occurs in low-elevation evergreen humid forests on lateritic soils, at 20–510 m elevation in southeastern Madagascar, where it has mostly been recorded in and around the Tsitongambarika protected area but extends to forests farther north and more inland (*Karatra & Rakotovao 241, 242* from the Ampotaky forest at Beampingaratry). It has also been recorded in the northern parcel of the Manombo protected area, ca. 160 km to the north (Fig. 3D). *Dalbergia razakamalalae* has been collected in full flower from November to February.

Conservation Status—*Dalbergia razakamalalae* is known from 39 collection records that represent 34 extant occurrences and 5 occurrences that appear to have been extirpated. Its former Extent of Occurrence (EOO) was at least 2737 km^2 and its former Area of Occupancy (AOO) was at least 72 km^2 (based on a 4 km^2 grid). Its current geographic range has the form of an EOO of 2085 km^2 and an AOO of 56 km^2, and comprises three subpopulations. The species mainly occurs in forest ecosystems (MADAGASCAR CATALOGUE 2021). Forest cover decline between 1953 and 2017 was estimated from the forest cover time series of VIEILLEDENT *et al.* (2018a, 2018b) to be 70% in the altitudinal range of 20–650 m and within the minimum convex polygon encompassing all known collections of this species. Therefore, *D. razakamalalae* is inferred to have undergone and to be undergoing continuing decline in EOO, AOO, quality of habitat, number of subpopulations, and number of mature individuals. This species occurs at four locations with respect to the most serious plausible threat, which is selective logging for trade in its high-quality heartwood, as inferred from recent field observations of exploited trees at several sites. The occurrences within in the protected areas of Manombo and Tsitongambarika represent two separate locations. Occurrences outside of the Tsitongambarika protected area represent the third location. The subpopulation from the Ampotaky forest at Beampingaratry, which is situated at higher elevation and appears to be less accessible, represents the fourth location. For these reasons, *Dalbergia razakamalalae* is assigned a preliminary IUCN conservation status of Endangered: EN B1ab(i,ii,iii,iv,v)+2ab(i,ii,iii,iv,v).

Notes—Material of *Dalbergia razakamalalae* has previously been included in or associated with *D. maritima sensu* BOSSER & RABEVOHITRA (2002), mainly owing to their overlapping morphological variation with respect to leaflet size, shape and number, and inflorescence structure. However, *D. razakamalalae* differs in its leaflet texture and margins, and in its larger flowers. Its flowers are similar in size to those of *D. louvelii s.l.*, and it inhabits low-elevation evergreen humid forests like *D. maritima* subsp. *pubescens*,

but unlike these taxa, its leaves are consistently glabrous, as summarized in Figs. 1E, 2D, 3A–B and Table 3. A specimen with both flowers and immature fruits that were collected on different dates (*Réserves Naturelles 1689*) was examined by Bosser & Rabevohitra in 1995, who associated it with both *D. maritima* and *D. louvelii* on account of its small and glabrous leaflets (as in *D. maritima* subsp. *maritima*) and large flowers (as in *D. louvelii*). They suggested that this collection might be a hybrid between these two taxa, without any evidence for the presence of *D. louvelii* in the region, and evidently without realizing that its morphology is consistent with other collections they saw from the same region and habitat type, viz. *Réserves Naturelles 1124* and *Service Forestier 22334*, both included in their broad definition of *D. maritima*, and potentially also *Humbert 20355bis* & *20607*, two sterile collections present in the Paris herbarium at the time. The two collections made in the 1940s (*Humbert 20355bis* and *Réserves Naturelles 1689*) from forests around Manantantely, and the collection from the 1960s from the Ivola forest near Ifarantsa (*Réserves Naturelles 1124*) increase the documented distribution range of *D. razakamalalae* southwards, including to the southern part of the Tsitongambarika protected area, but no extant occurrences are known from that area, despite extensive recent collection efforts, so these populations are presumed to have been extirpated. Likewise, a sterile and poorly preserved collection from the village of Andriana in the Manakara district (*Service Forestier 38-R-118*) may represent *D. razakamalalae* and would increase its range by ca. 110 km to the north, but it probably dates from the 1950s and originates from a site that is not included in a protected area, so this possible subpopulation likewise probably no longer exists.

Additional Specimens Examined—**Madagascar.**—ANOSY [Toliara]: Ampotaky forest (Beampingaratry), 3 Dec 2019, *Karatra & Rakotovao 241* (DBEV, MO, P, TAN, ZT); same locality, 3 Dec 2019, *Karatra & Rakotovao 242* (DBEV, MO, P, TAN, ZT); Ivola forest (near Ifarantsa in the Tolagnaro district), s.d. (y.fr), *Réserves Naturelles 1124* (P); Manampanihy valley (Ampasimena), 18 Mar 1947, *Humbert 20607* (P); Manatantely forest (Tolagnaro distict), 1 Mar 1947, *Humbert 20355bis* (MO, P, TAN); same locality, 30 Nov 1948 (fl, y.fr), *Réserves Naturelles 1689* (P); Tsitongambarika protected area and surroundings (Iabakoho commune), 6 Feb 2019 (fr), *Ramanitrinizaka & Sandratriniaina 1* (DBEV, MO, P, TAN, ZT); same locality, 9 Feb 2019, Ra*manitrinizaka & Sandratriniaina 12* (DBEV, P); same locality, 9 Feb 2019 (fr), *Ramanitrinizaka & Sandratriniaina 13* (DBEV, MO, P, TAN); same locality, same date, *Ramanitrinizaka & Sandratriniaina 18* (DBEV, MO, P); same locality, 12 Feb 2019, *Ramanitrinizaka & Sandratriniaina 25* (DBEV, MO, P); same locality, 7 Feb 2019, *Ramanitrinizaka & Sandratriniaina 57* (DBEV, MO, P); same locality, 16 Feb 2019, *Sandratriniaina & Ramanitrinizaka 23*

(DBEV, MO, P); same locality, same date, *Sandratriniaina & Ramanitrinizaka 26* (DBEV, MO, P); same locality, same date, *Sandratriniaina & Ramanitrinizaka 27* (DBEV, MO, P); same locality, same date, *Sandratriniaina & Ramanitrinizaka 29* (DBEV, MO, P); same locality, 1 Apr 2014, *Razakamalala 7736* (MO, P, TAN); same locality, same date, *Razakamalala 7761* (MO, P, TAN); same locality, same date, *Razakamalala 7762* (MO, P, TAN); same locality, same date, *Razakamalala 7764* (MO, P, TAN); same locality, 12 Feb 2016, *Razakamalala & S. N. Andrianarivelo 8036* (MO, P, TAN, TEF, ZT); same locality, 14 Feb 2016 (fr), *Razakamalala & S. N. Andrianarivelo 8040* (MO, P, TAN, TEF, ZT); same locality, 4 Nov 2019, *S. A. Andrianarivelo & Razakamalala 53* (DBEV, MO, P, TAN, ZT); same locality, Feb 1963 (fl), *Service Forestier (Capuron) 22334* (P, TEF); Tsitongambarika protected area and surroundings (Manantenina commune/Ivohibe-Bemangidy/Antsotso), 11 Feb 2016, *Razakamalala & S. N. Andrianarivelo 8032* (MO, P, TAN, TEF, ZT); same locality, same date, *Razakamalala & S. N. Andrianarivelo 8035* (MO, P, TAN, TEF, ZT); same locality, 1 Nov 2019 (fl), *Razakamalala & S. A. Andrianarivelo 8558* (DBEV, MO, P, TAN, ZT); same locality, 2 Nov 2019, *Razakamalala & S. A. Andrianarivelo 8560* (DBEV, MO, P, TAN, ZT); same locality, 6 Feb 2019 (fr), *Razakamalala et al. 8266* (DBEV, MO, P, TAN, ZT); same locality, same date (fr), *Bernard et al. 2641* (DBEV, MO, P, TAN, ZT); same locality, 9 Feb 2019, *Bernard et al. 2645* (DBEV, MO, P, TAN, ZT); same locality, 11 Feb 2016, *S. N. Andrianarivelo & Razakamalala 255* (MO, P, TAN, TEF, ZT); ATSIMO-ATSINANANA [Fianarantsoa]: Amparihy (Ambitananona - Amparihy - Vangaindrano), 23 Nov 1953 (fl), *Service Forestier 7110* (P, TEF); Manombo Special Reserve, 5 Nov 2019, *Rakotovao & Andriamiarisoa 7522* (DBEV, MO, P, TAN, ZT); same locality, same date, *Rakotovao & Andriamiarisoa 7523* (DBEV, MO, P, TAN, ZT), same locality, same date, *Rakotovao & Andriamiarisoa 7528* (DBEV, MO, P, TAN, ZT), same locality, 28 Jan 2014, *Emeline 23* (MO, P, ZT); same locality, 4 Nov 2019, *Andriamiarisoa & Rakotovao 2424* (DBEV, MO, P, TAN, ZT).

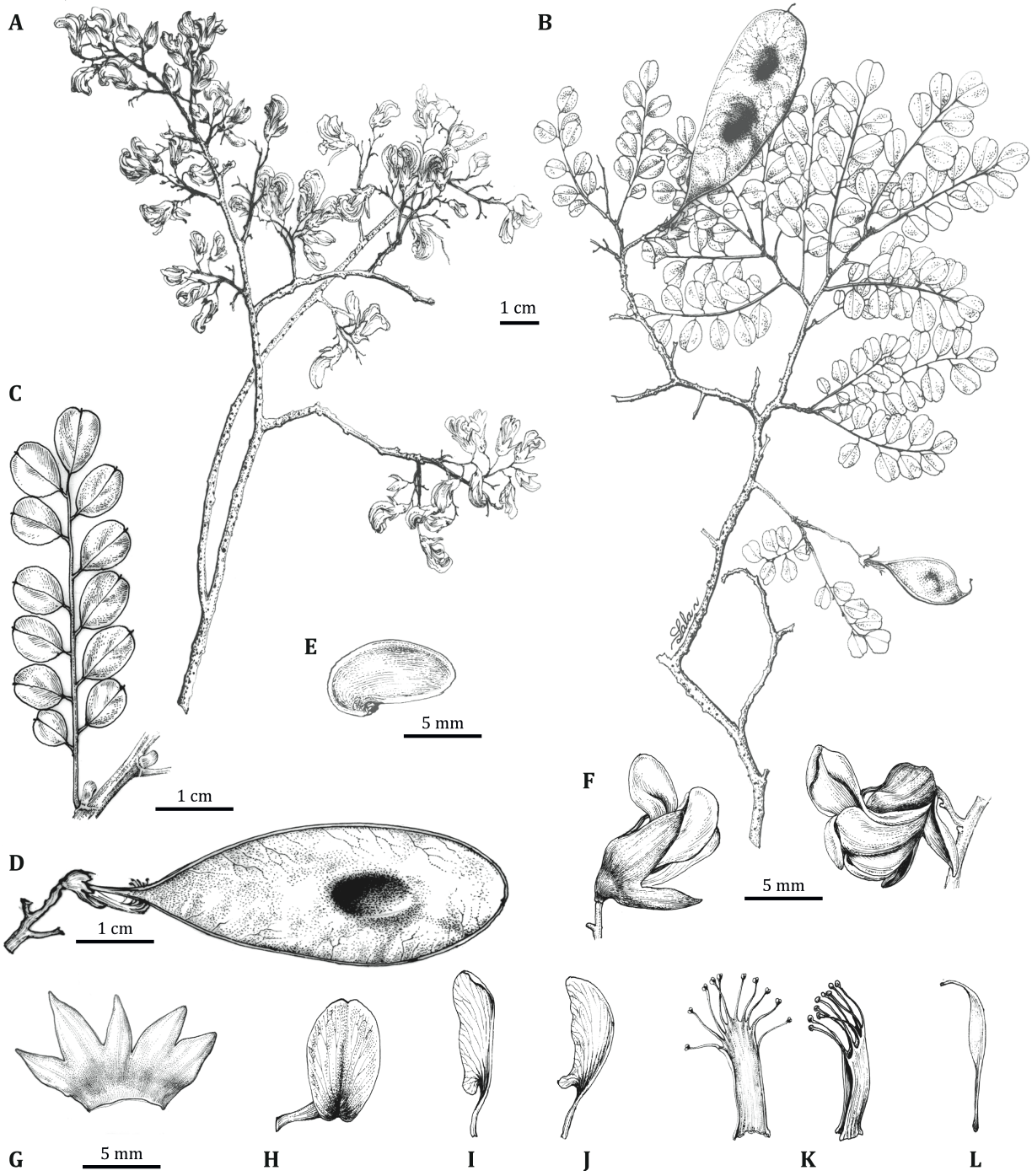

FIGURE 4: *Dalbergia pseudomaritima*. **A)** Flowering branch. **B)** Fruiting branch. **C)** Leaf, top view. **D)** Fruit, single-seeded (immature). **E)** Seed (immature). **F)** Flower, side view (left) and frontal view (right). **G)** Calyx, abaxial surface, split open and flattened, upper lobes on right. **H)** Standard petal, adaxial surface. **I)** Wing petal, adaxial surface. **J)** Keel petal, adaxial surface. **K)** Androecium, adaxial surface, flattened, with ten fused stamens (left) or side view, with nine fused stamens (right). **L)** Gynoecium. Illustration by Roger Lala Andriamiarisoa from *Ramananjanahary et al. 830* (A, F–L), *Rakotonirina et al. 1190* (B, C), and *Ramison & Ramisy 109* (D, E).

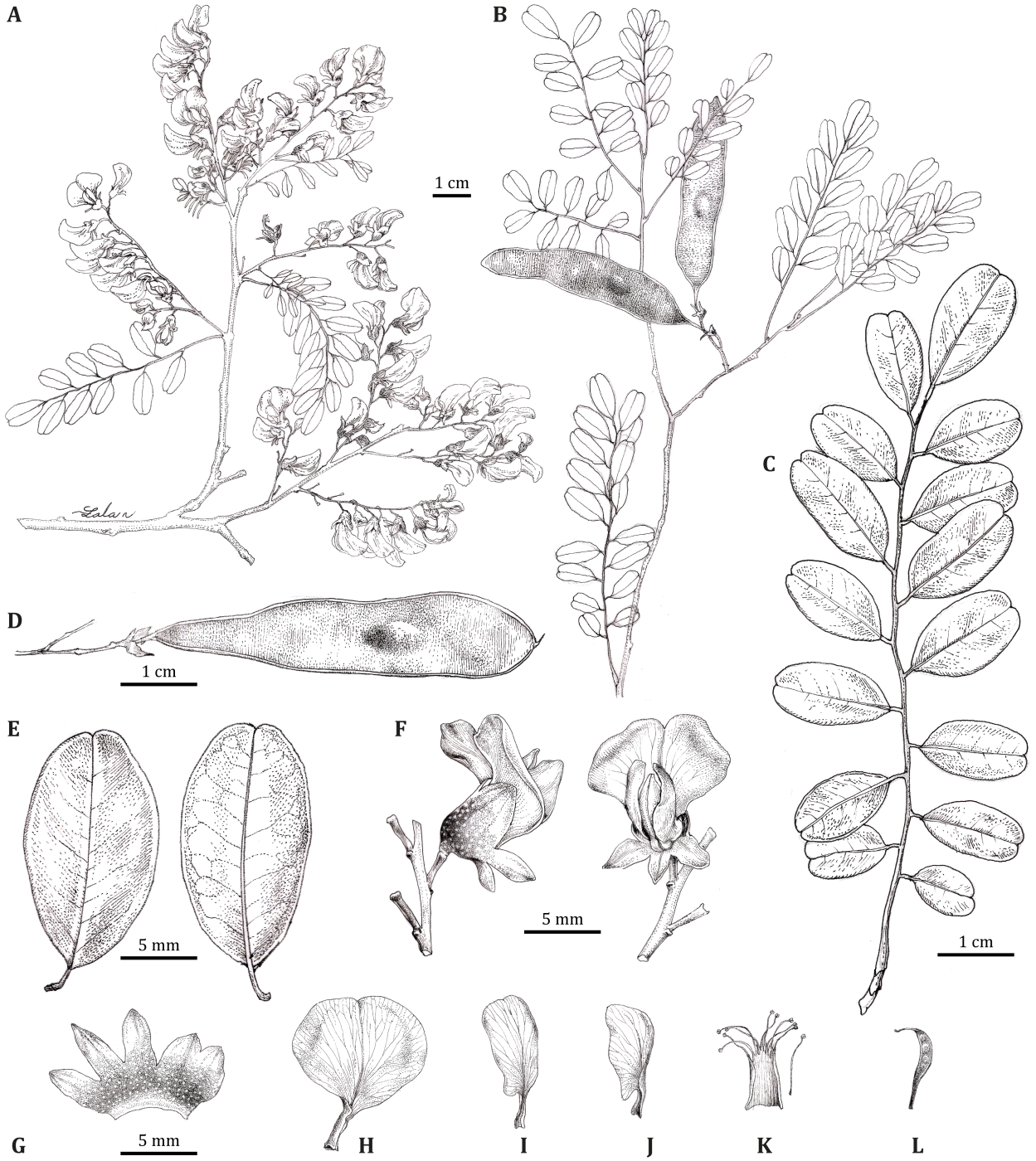

FIGURE 5: *Dalbergia razakamalalae*. **A)** Flowering branch. **B)** Fruiting branch. **C)** Leaf, top view. **D)** Fruit, single-seeded (immature). **E)** Leaflet upper (left) and lower (right) surface. **F)** Flower, side view (left) and frontal view (right). **G)** Calyx, abaxial surface, split open and flattened, upper lobes on right. **H)** Standard petal, adaxial surface. **I)** Wing petal, adaxial surface. **J)** Keel petal, adaxial surface. **K)** Androecium, adaxial surface, with nine or ten fused stamens. **L)** Gynoecium. Illustration by Roger Lala Andriamiarisoa from *Andriamihajarivo et al. 2455* (A, F–L), *Ramanitrinizaka & Sandratriniaina 13* (B, C, E) and *Ramanitrinizaka & Sandratriniaina 1* (D).

Discussion

The morphological and ecological analyses presented here show that the two newly descried species from southeastern Madagascar, *Dalbergia pseudomaritima* and *D. razakamalalae*, form two coherent and distinct morphological clusters (Fig. 3A–B), each associated with a different habitat type (Fig. 3C). Moreover, their geographic ranges are very different from those of *D. maritima* (as re-delimited here) and *D. louvelii s.l.* (Fig. 3D), which differ from one another in several characters of their leaves and flowers, co-occur in east-central Madagascar, and are more closely related to each other than either is to *D. razakamalalae* or *D. pseudomaritima* (CRAMERI *et al.* in prep.). The two subspecies of *D. maritima* can clearly be distinguished from one another based on the presence or absence of indument on the leaves, the inflorescence axis, and the gynoecium (Fig. 3A–B), as well as their apparent allopatric geographic distribution (Fig. 3D), which is associated with different habitat types (Fig. 3C).

In light of the findings presented in this study and the resulting taxonomic changes, in particular the significantly narrowed delimitation of *Dalbergia maritima*, the associated trade names need to be modified accordingly. Precious wood harvested from forests in southeastern Madagascar and traded under the name *D. maritima* was presumably obtained from individuals of *D. razakamalalae*, the only known taxon in the region belonging to the Maritima clade. NORMAND (1988) noted more than three decades ago that rosewood attributed to *D. maritima* was being harvested especially in southeastern Madagascar because individuals of exploitable diameter had already become rare in the region of Tamatave (Toamasina) in east-central Madagascar by the 1920s, and in fact the leaf and leaflet illustrated in NORMAND (1988, p. 91, Figs. 4–5) likely represent a specimen of *D. razakamalalae* (see our Figs. 1E, 2D). Likewise, precious wood exploited from northeastern Madagascar (NORMAND 1988, map on p. 91) and traded as *D. maritima* presumably was actually harvested from individuals of *D. louvelii s.l.*, *D. occulta* (which was only described in 2005), and/or other taxa that occur in the region and still await formal description (Fig. 3D). This taxonomic confusion has important implications regarding the remaining number of large mature individuals of these species and the impact that further harvesting of precious wood would have on them (WAEBER *et al.* 2019), which in turn significantly impacts whether they would meet the requirements for issuing a 'non-detriment finding', as required for international commerce under the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES, Article IV-2a). The present study reveals the importance of conducting targeted field work, detailed taxonomic investigations, and thorough conservation assessments of tropical timber species of high economic value. It also confirms the conservation significance of

Madagascar's remaining low-elevation eastern evergreen humid and littoral forests and their importance as the habitat of a previously underestimated number of comparatively narrowly distributed and threatened rosewood species.

This study, along with several others (CRAMERI *et al.* in prep.; HASSOLD 2015; HASSOLD *et al.* 2016; RAKOTONIRINA *et al.* in prep.; WILDING *et al.* submitted-a, b), is part of an ongoing international effort to develop an improved taxonomy for *Dalbergia* species in Madagascar. The integration of morphological studies and eco-geographic considerations with phylogenomic and population genomic analyses, based on a much larger number of collections now available, has resulted in an increasingly comprehensive understanding of the diversity within this taxonomically complex genus. It would likely not have been possible to resolve the taxonomic confusion regarding *D. maritima* without the significant insights provided by phylogenomic and population genomic analyses. In particular, the results of these analyses led to the important inference that differences in inflorescence structure are associated with two strongly divergent lineages (racemose in the Maritima clade, paniculate in the Chapelieri clade) within which there has been morphological convergence in leaflet size (small leaflets in *D. maritima*, the undescribed SAVA material, and *D. pseudomaritima*), while flower size, the presence or absence of indument, and eco-geography are informative at the species or subspecies level. The integration of morphology, eco-geography, and phylogenomics provides strong support for clarifying species limits and developing an improved taxonomy for the genus, and this approach is now being applied to the other groups of Malagasy *Dalbergia* that present taxonomic issues.

Acknowledgements

We are grateful to the botanists of the Missouri Botanical Garden Madagascar Program (notably to Sandratra Andrianarivelo, Patrice Antilahimena, Roger Bernard, Chris Birkinshaw, Adolphe Lehavana, Jeannie Raharimampionona, Fortunat Rakotoarivony, Charles Rakotovao, Toky Ralainaorina, Richard Randrianaivo, Richardson Razakamalala, and Anselme Tilahimena), Sonja Hassold, Michael Zehnder, Ravo Ramanantsialonina and undergraduate and doctoral students of the University of Antananarivo for botanical inventory work, numerous field collections, providing photographs of living specimens, and fruitful discussions. We also thank the G3D team at the MBG office (notably Sylvie Andriambololonera, Faranirina Lantoarisoa, Nadiah Manjato, and Fano Rajaonary) for the organization of field work and facilitating export permits and the exportation of herbarium vouchers. We would also like to thank the cited herbaria and their curators (notably Odile Poncy at P and Alessia Guggisberg at ZT) for specimen management and access to collections, and for allowing minimal invasive sampling. We further thank George Garnett for helpful comments on an earlier version of this manuscript.

References

Barrett, M. A., J. L. Brown, M. K. Morikawa, J.-N. Labat, and A. D. Yoder. 2010. Conservation. CITES designation for endangered rosewood in Madagascar. *Science* 328:1109–1110.

Bosser, J. and R. Rabevohitra. 1996. Taxa et noms nouveaux dans le genre *Dalbergia* (Papilionaceae) à Madagascar et aux Comores. *Bulletin du Muséum national d'Histoire naturelle, 4ème série – section B, Adansonia* 18:171–212.

Bosser, J. and R. Rabevohitra. 2002. Tribe *Dalbergieae*. Pp. 321–361 in *The Leguminosae of Madagascar*, ed. D. J. Du Puy, J. N. Labat, R. Rabevohitra, J. F. Villiers, J. Bosser, and J. Moat. Kew: Royal Botanical Gardens.

Bosser, J. and R. Rabevohitra. 2005. Espèces nouvelles dans le genre *Dalbergia* (Fabaceae, Papilionoideae) à Madagascar. *Adansonia* 27:209–216.

Cervantes, A., J. Linares, and E. Quintero. 2019. An updated checklist of the Mexican species of *Dalbergia* (Leguminosae) to aid in its conservation efforts. *Revista Mexicana de Biodiversidad* 90:e902528. DOI: 10.22201/ib.20078706e.2019.90.2528.

Christensen, K. l. 1987. Taxonomic revision of the *Pinus mugo* complex and *P. rhaetica* (*P. mugo sylvestris*) (Pinaceae). *Nordic Journal of Botany* 7:383–408.

Crameri, S. 2020. *Phylogenomics, Species Discovery and Integrative Taxonomy in Dalbergia (Fabaceae) Precious Woods from Madagascar*. Doctoral dissertation [Diss. ETH No. 27241]. Zurich: ETH Zurich.

Dauby, G., T. Stévart, V. Droissart, A. Cosiaux, V. Deblauwe, M. Simo-Droissart, M. S. M. Sosef, P. P. Lowry II, G. E. Schatz, R. E. Gereau, and T. L. P. Couvreur. 2017. ConR: An R package to assist large-scale multispecies preliminary conservation assessments using distribution data. *Ecology and Evolution* 7:11292–11303.

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

Dormontt, E. E., M. Boner, B. Braun, G. Breulmann, B. Degen, E. Espinoza, S. Gardner, P. Guillery, J. C. Hermanson, G. Koch, S. L. Lee, M. Kanashiro, A. Rimbawanto, D. Thomas, A. C. Wiedenhoeft, Y. Yin, J. Zahnen, and A. J. Lowe. 2015. Forensic timber identification: it's time to integrate disciplines to combat illegal logging. *Biological Conservation* 191:790–798.

Goodman, S. M., M. J. Raherilalao, and S. Wohlhauser (eds.). 2018. *The terrestrial Protected Areas of Madagascar: Their History, Description, and Biota.* Antananarivo: Association Vahatra.

Gower, J. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27:857–871.

Hassold, S. 2015. *Molecular Identification of Malagasy Dalbergia species (rosewoods) for biodiversity conservation*. Doctoral dissertation [Diss. ETH No. 22999]. Zurich: ETH Zurich.

Hassold, S., P. P. Lowry, M. R. Bauert, A. Razafintsalama, L. Ramamonjisoa, and A. Widmer. 2016. DNA barcoding of Malagasy rosewoods: towards a molecular identification of CITES-listed *Dalbergia* species. *PLoS One* 11: e0157881. DOI: 10.1371/journal.pone.0157881.

Hijmans, R. J. 2021. terra: spatial data analysis. R package version 1.0–10. https://CRAN.R-project.org/package=terra.

Hijmans, R. J. and J. van Etten. 2012. raster: geographic analysis and modeling with raster data. R package version 3.4–5. http://CRAN.R-project.org/package=raster.

Hollister, J. W., T. Shah, A. L. Robitaille, and M. W. Beck. 2020. elevatr: access elevation data from various APIs. R package version 0.3.1. https://CRAN.R-project.org/package=elevatr.

IUCN. 2012. IUCN Red List categories and criteria, version 3.1, ed. 2. Gland, Switzerland and Cambridge, UK: IUCN.

IUCN. 2019. Guidelines for using the IUCN Red List categories and criteria, version 14. Prepared by the Standards and Petitions Committee. Available from http://www.iucnredlist.org/documents/RedListGuidelines.pdf (last accessed January 2021).

Karger, D. N., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, H. P. Linder, and M. Kessler. Data from: Climatologies at high resolution for the earth's land surface areas. Dryad Digital Repository. https://doi.org/10.5061/dryad.kd1d4.

Karger, D. N., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, H. P. Linder, and M. Kessler. 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4:170122. DOI: 10.1038/sdata.2017.122.

Klitgaard, B. B. and M. Lavin. 2005. *Dalbergieae sens. lat.* Pp. 307–335 in *Legumes of the World*, ed. G. P. Lewis, B. D. Schrire, B. A. Mackinder, and M. Lock. Kew: Royal Botanical Gardens.

Lachenaud, O. and L. van der Maesen. 2016. Notes on African *Dalbergia* (Leguminosae–Papilionoideae) with the description of two new species from Atlantic Central Africa. *Symbolae Botanicae Upsalienses* 38:167–194.

Madagascar Catalogue. 2021. *Catalogue of the Plants of Madagascar*. Missouri Botanical Garden, St. Louis, U.S.A. Available from: http://www.tropicos.org/Project/Madagascar (last accessed March 2021].

Maechler, M., P. Rousseew, A. Struyf, M. Hubert, and K. Hornik. 2019. cluster: cluster analysis basics and extensions. R package version 2.1.0. https://CRAN.R-project.org/package=cluster.

Mason, J., M. Parker, L. B. Vary, P. P. Lowry II, S. Hassold, and G. Ruta. 2016. Malagasy precious hardwoods: scientific and technical assessment to meet CITES objectives. Report submitted by the World Resources Institute and the World Bank. https://www.scribd.com/document/318123493/WRI-WB-Malagasy-Precious-Woods-Assessment-1-pdf.

Moat, J. and P. Smith (eds.). 2007. *Atlas of the Vegetation of Madagascar*. Kew: Royal Botanic Gardens.

Normand, D. 1988. A propos des bois de rose de Madagascar. *Revue Bois et Forêts des Tropiques* 217:89–94.

Paradis, E. and K. Schliep. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528.

Pedersen, T. L. 2020. ggforce: accelerating 'ggplot2'. R package version 0.3.2. https://CRAN.R-project.org/package=ggforce.

POWO. 2021. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Available from: http://www.plantsoftheworldonline.org (last accessed March 2021).

Prain, D. 1904. The species of *Dalbergia* of south-eastern Asia. *Annals of the Royal Botanic Garden, Calcutta* 10:1–411.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Rakotonirina, N. H., P. B. Phillipson, S. Crameri, N. Wilding, B. Rakouth, and R. Razakamalala. in prep. Taxonomic studies on Malagasy *Dalbergia* (Fabaceae). IV. *Dalbergia emirnensis* Benth. and closely related species. *Candollea*.

Richter, H. G., K. Gembruch, and G. Koch. 2014. CITESwoodID: descriptions, illustrations, identification, and information retrieval. In English, French, German, and Spanish. Version: 20th August 2019. Available from: https://www.delta-intkey.com (last accessed Mar 2021).

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 9:671–675.

Schuurman, D. and P. P. Lowry II. 2009. The Madagascar rosewood massacre. *Madagascar Conservation and Development* 4:98–102.

Smith, A. B. 2020. fasterRaster: faster raster processing in R using GRASS GIS. R package version 0.6.0. http://www.earthSkySea.org.

Tennekes, M. 2018. tmap: thematic maps in R. *Journal of Statistical Software* 84:1–39.

Thiers, B. 2021. [continuously updated] Index Herbariorum: *A global directory of public herbaria and associated staff*. New York Botanical Garden's Virtual Herbarium. Available from: http://sweetgum.nybg.org/ih/ (last accessed January 2021).

Turland, N. J., J. H. Wiersema, F. R. Barrie, W. Greuter, D. L. Hawksworth, P. S. Herendeen, S. Knapp, W. H. Kusber, D.-Z. Li, K. Marhold, T. W. May, J. McNeill, A. M. Monro, J. Prado, M. J. Price, and G. F. Smith. 2018. *International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth*

International Botanical Congress Shenzhen, China, July 2017. Regnum Vegetabile 159. Glashütten: Koeltz Botanical Books. https://doi.org/10.12705/Code.2018.

UNODC. 2016. *World Wildlife Crime Report: Trafficking in Protected Species*. New York: United Nations.

Vieilledent, G., C. Grinand, F. Rakotomalala, R. Ranaivosoa, J.-R. Rakotoarijaona, T. F. Allnutt, and F. Achard. 2018a. Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. *Biological Conservation* 222:189–197.

Vieilledent, G., C. Grinand, F. A. Rakotomalala, R. Ranaivosoa, J.-R. Rakotoarijaona, T. F. Allnutt, and F. Achard. 2018b. Update to: combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar.

Viguier, R. 1944. Légumineuses de Madagascar. *Archives de Botanique (Mémoires)* 6:685–725. (unpublished manuscript).

Viguier, R. 1952. Leguminosae madagascarienses novae (suite 2). *Notulae Systematicae, Paris* 14:168–187.

Waeber, P. O., D. Schuurman, B. Ramamonjisoa, M. Langrand, C. V. Barber, J. L. Innes, P. P. Lowry, and L. Wilmé. 2019. Uplisting of Malagasy precious woods critical for their survival. *Biological Conservation* 235:89–92.

Wickham, H. 2016. *ggplot2: elegant graphics for data analysis*, ed. 2. New York: Springer.

Wilding, N., P. B. Phillipson, S. Crameri, S. Andriambololonera, R. L. Andriamiarisoa, S. A. F. Andrianarivelo, R. Bernard, N. Rakotonirina, C. Rakotovao, R. Randrianaivo, R. Razakamalala, and P. P. Lowry II. submitted-a. Taxonomic studies on Malagasy *Dalbergia* (Fabaceae). I: Two new species from northern Madagascar, and an emended description for *D. manongarivensis*. *Candollea*.

Wilding, N., P. B. Phillipson, and S. Crameri. submitted-b. Taxonomic studies on Malagasy *Dalbergia* (Fabaceae). II: The reinstatement of *Dalbergia chermezonii* and a new name for *D. mollis*. *Candollea*.

Appendix III

Supplementary Methods

We performed new IUCN Red List assessments for *Dalbergia pseudomaritima*, *D. razakamalalae* and *D. maritima*. The use of IUCN Red List criterion A (IUCN 2012) for *Dalbergia* species from Madagascar currently represents a major challenge, owing to the difficulty of quantifying a population size reduction over three generations based on the insufficient knowledge of relevant parameters, such as generation length, effective species abundance and distribution, habitat preferences or levels of exploitation. Since all three species appear to have relatively small distribution ranges, we decided for criterion B until more comprehensive population-level information and species distribution models become available for the quantification of habitat loss. Specifically, we used the GeoCat online tool (BACHMAN *et al.* 2011) to estimate extent of occurrence and area of occupancy. Red list assessments were informed by the forest cover time series for Madagascar (VIEILLEDENT *et al.* 2018a, 2018b) and a specifically developed R shiny (CHANG *et al.* 2020) application to estimate spatially explicit change in forest cover within a specified species range and altitude, as well as to visualize spatially explicit sampling effort (current version available at https://github.com/scrameri/ConservationAssessments). Figures S1 – S3 are screenshots taken from the shiny application.

IUCN Red List Conservation Assessments

Dalbergia maritima R. Vig. emend. Crameri, Phillipson & N. Wilding

TAXONOMIC NOTES

All *Dalbergia* spp. from Madagascar are currently being subjected to taxonomic review.

Dalbergia maritima R. Vig. has been re-circumscribed to exclude subpopulations from southeast and northeast Madagascar. The species, as previously delimited, is demonstrated to have comprised at least four distinct lineages, which are diagnosable based on inflorescence, flower and leaf characters. The re-circumscribed species is also eco-geographically and genetically distinct from other species based on phylogenomic analysis of target enrichment sequencing data of hundreds of genomic regions (CRAMERI *et al.* in prep.).

IUCN RED LIST CATEGORY AND CRITERIA

Endangered (EN) B1ab(i,ii,iii,iv,v)+2ab(i,ii,iii,iv,v)

JUSTIFICATION

Based on a total of 23 extant and 6 extirpated collection records, the species has an extent of occurrence (EOO) of 1,875 km^2 (formerly 2564 km^2), an area of occupancy (AOO) of 60 km^2 (formerly 80 km^2), and is inferred to comprise five subpopulations (Figure S1). In light of the principal threat facing the species, which is selective logging for trade in its high-quality heartwood, four locations are defined. The species is believed to be undergoing continuing decline in (i) extent of occurrence, (ii) area of occupancy, (iii) area, extent and/or quality of habitat, (iv) number of subpopulations and (v) number of mature individuals, and is therefore assigned an extinction risk of Endangered.

GEOGRAPHIC RANGE INFORMATION

The species is endemic to Madagascar, where extant subpopulations have only been recorded in the Atsinanana region in the central east. The current EOO is 1,875 km^2 using a minimum convex polygon, and the current AOO is 60 km^2 using a 2×2 km grid size. The EOO and AOO were calculated, using GeoCAT (BACHMAN *et al.* 2011), based on 23 georeferenced collections that were inferred to represent extant individuals (Figure S1).

POPULATION INFORMATION

Extant subpopulations

The species comprises five extant subpopulations based on the distribution of the 23 occurrences. Two subpopulations are located in the littoral forests of Vohibola and Andranampy, at a distance of 21 km from each other. Three further subpopulations are located in the protected areas of Sahafina (*Razakamalala 8368*, *Bernard 2734*), in and around the protected area of Betampona (e.g., *Razakamalala 7704*, *Randrianaivo 2928*, *G. Rakotonirina 389*), as well as in the Agnalahely and Menagisa forests, 10 km to the southeast of Betampona (*G. Rakotonirina 91*, *Karatra 190*).

Possibly extinct subpopulations

Further collections from the 1920s (*Louvel 79 & 200* from the Tampina forest) and the 1950s (*Service Forestier 2860 & 5-R-233* from around Ambila-Lemaitso) originate from areas that have seen extensive forest degradation, and were likely collected less than 10 km from the extant subpopulations of Vohibola and Andranampy, respectively. A collection from the 1950s from the Masiabarika forest (*Service Forestier 18-R-195*) could not be precisely georeferenced, but it originates from the canton of Andranobolahy at 200 m elevation, where large areas have been cleared of forest. Another collection from 1985

(*Service Forestier 32824* from "environs de Foulpointe") probably originates from the Analalava protected area, but this suspicion could not be confirmed by any other collection of the species, despite considerable sampling efforts at Analalava.

HABITAT AND ECOLOGY INFORMATION

Habitat

The species occurs in littoral forests on sand (subsp. *maritima*) or low-elevation evergreen humid forests on lateritic soils (subsp. *pubescens*), between 0 — 450 m (MADAGASCAR CATALOGUE 2020).

Phenology

The species has been collected in full flower in the months of February (*Service Forestier 2860*) and March (*Louvel 200*, *G. Rakotonirina 389*).

Life history

Reproductive structures have been observed in individuals at least 4 m tall (*G. Rakotonirina 389*), or with a diameter at breast height (DBH) of at least 1.5 cm (*G. Rakotonirina 91*). The largest documented trees were 10 m tall (*Bernard 2247*) or had a DBH of at least 30 cm (*Service Forestier 18-R-195*). The generation length of this species is not precisely known but estimated to be approximately 30 to 40 years. Logging disproportionally targets large individuals, which likely belong to the most fertile age classes. Coppice shoots have been observed (*Rakotovao 7468*) but it might take up to a decade before reproduction can resume.

THREATS INFORMATION

The most serious plausible threat to *D. maritima* is selective logging for national and often illegal international trade, as it is highly sought after for its dense and burgundy coloured heartwood, which was and probably still is used in carpentry (*Service Forestier 18-R-195*, subsp. *pubescens*), cabinet-making (*Louvel 200*, subsp. *maritima*) and construction (*Service Forestier 31184*, subsp. *pubescens*).

A further serious plausible threat is habitat degradation or loss by land conversion for housing, industrial or agricultural development (see VIEILLEDENT *et al.* (2018a) and references therein), most notably due to burning for pasture and the slash-and-burn practice *tavy*, which is "the traditional and predominant land use practice of eastern Madagascar" (STYGER *et al.* 2007). The species mainly occurs in forest ecosystems (MADAGASCAR

CATALOGUE 2020), and forest cover decline between 1953 and 2017 was estimated from the forest cover time series published in VIEILLEDENT *et al.* (2018a, 2018b) to be 77% in the altitude range of 0 – 450 m and within the minimum convex polygon encompassing all known collections of this species (historic and modern, n = 27, Figure S1). Moreover, subsp. *maritima* has a clear habitat preference for littoral forests, which have experienced significant reductions and are among Madagascar's most threatened forest formations (GANZHORN *et al.* 2001).

Four locations are defined based the principal threat of selective logging. Three locations correspond to the protected areas of Betampona, Sahafina, and Vohibola, and one location corresponds to all known occurrences located outside of protected areas. Given the higher level of protection afforded in protected areas, occurrences inside and outside of protected areas can be considered as different locations. A single location encompassing all subpopulations outside of protected areas can be inferred because of the large spatial scale at which a threatening event (i.e., logging or habitat loss) can severely reduce the population within a single generation length (approximately 30 to 40 years). In addition, most known subpopulations of this species can be accessed by road and/or train, and market trends in rosewood demand, as well as the corresponding changes in harvest intensity, can be regarded as similar over large spatial scales spanning similarly accessible areas.

USE AND TRADE INFORMATION

This species is considered to produce high-quality rosewood or *bois de rose* (R. Razakamalala, pers. comm.) based on its exploitable growth habit and the beautifully veined heartwood of a burgundy colour (*Rakotovao 7467*). The heartwood is used in carpentry (*Service Forestier 18-R-195*, subsp. *pubescens*), cabinet-making (*Louvel 200*, subsp. *maritima*) and construction (*Service Forestier 31184*, subsp. *pubescens*).

Traces of selective logging of this species were documented at Andranampy (*Rakotovao 7468*, subsp. *maritima*). The only available collections of standing mature trees with exploitable DBH (larger than 20 cm) date from 1924 (*Louvel 200*, subsp. *maritima*) and 1954 (*Service Forestier 18-R-195,* subsp. *pubescens*), which suggests that few large individuals of this species remain. Despite large gaps in knowledge about the actual trade volumes, it can be assumed that the species was heavily exploited especially for international trade. This is indicated by the demand for its high-quality heartwood, the high accessibility of source populations near the coast, or transport possibilities along numerous roads and along the Tananarive-Côte Est (TCE) railway line, as well as by the proximity to one of the largest and best-equipped ports of Madagascar (Toamasina), from where large

quantities of rosewood logs were exported in the 1950s (NORMAND 1988) and until after (SCHUURMAN AND LOWRY II 2009). Together with *D. louvelii* and *D. baronii*, the heartwood of *D. maritima* is suspected to represent the major source of *Bois de rose* from the Atsinanana region.

Relevant vernacular or trade names are Andramena kely ravina (*Bernard 2247*), *bois de rose* (*Rakotovao 7474*), Hazomainty (*Rakotovao 7482*), Volombodipony (*Louvel 79*), Volombodipony à petites feuilles or Volombodipony lahy (*Louvel 200*).

CONSERVATION ACTIONS INFORMATION

Afforded Protection

One subpopulation of subsp. *maritima* occurs in the Vohibola forest, where some level of protection is afforded by Razan'ny Vohibola, an association of volunteers from four surrounding villages. Two subpopulations of subsp. *pubescens* occur in the protected areas of Betampona and Sahafina, respectively. This species is listed in CITES Appendix II since 2013, whereby the identification of non-detriment findings is considered to be impossible given the currently large gaps in knowledge about population size, distribution and abundance, and an uplifting of all of Madagascar's precious woods to Appendix I was requested (WAEBER *et al.* 2019). An ongoing international research project aims at documenting the occurrence and morphology of *Dalbergia* spp. from Madagascar, as well as collecting leaf samples for genetic analysis and heartwood samples for their anatomical, spectroscopic and spectrometric characterisation. These collections are used to build a reference collection of *Dalbergia* spp. from Madagascar for forensic timber identification.

Conservation Recommendations

Include the littoral forest of Andranampy (second subpopulation of subsp. *maritima*) in Madagascar's protected areas network.

Research Needed

The extent and impact of logging, habitat degradation, and fire on population size and trends, age structure, extent of occurrence and conservation status should be investigated and regularly monitored. It should further be clarified whether there are extant subpopulations at additional locations or protected areas, such as the littoral forests to the north and south of Ambila-Lemaitso (subsp. *maritima*) or in low-elevation humid forests at Analalava or the Ankeniheny-Zahamena Forest Corridor (subsp. *pubescens*). Research on its ecology, growth rate, life history and the ability to recover and recruit following burning or logging would also be valuable.

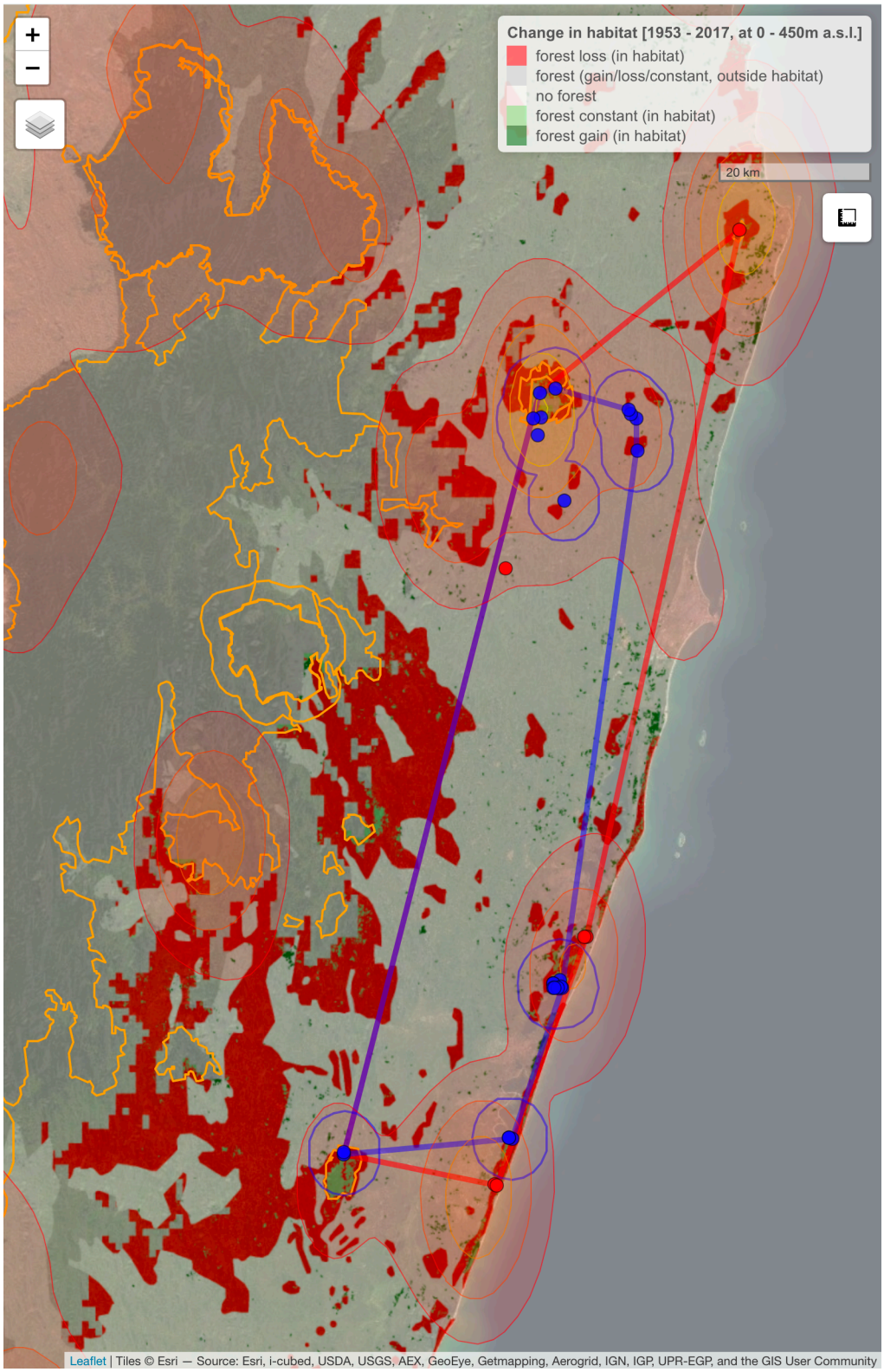

Figure S1: Distribution of *Dalbergia maritima*. Species occurrences are shown as slightly jittered points and withheld geo-coordinates (blue: extant [n = 23], red: not detected since 1985 [n = 6]). The extents of occurrence (minimum convex polygon) are shown as thick lines (blue: extant, red: historic, purple: extant and historic). Thin blue lines denote three extant subpopulations using a radius of 5 km around each occurrence. Orange lines denote protected areas. Shaded areas denote change in forest cover between 1953 and 2017 estimated from VIEILLEDENT *et al.* (2018a, 2018b), within an altitude range of 0 – 450 m (see colour legend), resulting in an estimated loss of 77%. Heat-colored contour lines denote 2D Kernel density estimates of angiosperm collections databased on www.tropicos.org (accessed on Aug 6 2020).

Dalbergia pseudomaritima Crameri, Phillipson & N. Wilding

TAXONOMIC NOTES

All *Dalbergia* spp. from Madagascar are currently being subjected to taxonomic review.

Dalbergia pseudomaritima was previously included in *D. maritima* but was recently recognized as distinct on the basis of its fundamentally different inflorescence structure. This species is also eco-geographically and genetically distinct from *D. maritima* based on phylogenomic analysis of target enrichment sequencing data of hundreds of genomic regions (CRAMERI *et al.* in prep.).

IUCN RED LIST CATEGORY AND CRITERIA

Endangered (EN) B1ab(i,ii,iii,iv,v)+2ab(i,ii,iii,iv,v)

JUSTIFICATION

Based on a total of 29 extant and 13 extirpated collection records, the species has an extent of occurrence (EOO) of 252 km^2 (formerly at least 275 km^2), an area of occupancy (AOO) of 44 km^2 (formerly at least 56 km^2), and is inferred to comprise three subpopulations (Figure S2). In light of the principal threat facing the species, which is habitat degradation or loss, four locations are defined. The species is believed to be undergoing continuing decline in (i) extent of occurrence, (ii) area of occupancy, (iii) area, extent and/or quality of habitat, (iv) number of subpopulations and (v) number of mature individuals, and is therefore assigned an extinction risk of Endangered.

GEOGRAPHIC RANGE INFORMATION

The species is endemic to Madagascar, where extant subpopulations have only been recorded in the Tolagnaro district (Anosy region) in the southeast. The current EOO is 252 km^2 using a minimum convex polygon, and the current AOO is 44 km^2 using a 2×2 km grid size. The EOO and AOO were calculated, using GeoCAT (BACHMAN *et al.* 2011), based on 29 georeferenced collections that were inferred to represent extant occurrences (Figure S2).

POPULATION INFORMATION

Extant subpopulations

The species comprises three extant subpopulations based on the distribution of the 21 occurrences. One is located in the protected area of Mandena, where the species has last been confirmed in 2014 (*Razakamalala 7783*), in the conservation zone M15 (LOWRY *et al.* 2008). Another subpopulation is located approximately 25 km northeast of Mandena, in the littoral forest complex of Sainte Luce (Ambato Atsinanana), where it has last been confirmed in 2019 in the conservation zones S8 and S9 (LOWRY *et al.* 2008). This subpopulation extends to unprotected and degraded coastal forests and marshes surrounding Sainte Luce, including areas north of the Ebakika river (*Ludovic 1570*, *Razakamalala 6675*). A third subpopulation is located on sandy lateritic soils in the low-elevation evergreen humid forest of Ampasy, and is represented by a single collection from 2019 (*Bernard 2654*).

Possibly extinct subpopulations

Nine further collections dating from 1977 to 1991 from the broader Mandena region have likely been made in formerly intact littoral forest parcels that were severely impacted by illegal charcoal producers in the 1990s, who produced charcoal to meet the growing demand in Tolagnaro (P. P. Lowry II, pers. comm.). Two collections from 2006 from former littoral forests around Mandromodromotra (*Ramison* 108 & 109) are also not inferred to represent additional extant subpopulations.

HABITAT AND ECOLOGY INFORMATION

Habitat

The species mainly occurs in littoral forests on sand, but it has also been recorded in adjacent swamp forests and marshes (*Razakamalala 6675*) and low-elevation evergreen humid forests on sandy lateritic soils along streams (*Bernard 2654*), between 0 — 30 m above sea level (MADAGASCAR CATALOGUE 2020).

Phenology

The species has been collected in full flower in the months of October (*Ramananjanahary 830*) to January (*Rabevohitra 2178*). Immature fruits have been observed until February (*N. Rakotonirina 1190*).

Life history

Reproductive structures have been observed in individuals at least 3 m tall (*N. Rakotonirina 1190*) or with a diameter at breast height (DBH) of at least 5 cm (*S. A. Andrianarivelo* 60). The largest documented trees were 12 m tall (*S. A. Andrianarivelo 58*) or had a DBH of 25 cm (*Rabenantoandro 1556*). The generation length of this species is not precisely known but estimated to be approximately 20 to 40 years. Logging disproportionally targets large individuals, which likely belong to the most fertile age classes. Coppice shoots have been observed (*Rabevohitra 2033*) but it might take up to a decade before reproduction can resume.

THREATS INFORMATION

The most serious plausible threat to *D. pseudomaritima* is habitat degradation or loss due to land clearing and fire for subsistence agriculture (see VIEILLEDENT *et al.* (2018b) and references therein), most notably due to burning for pasture and the slash-and-burn practice *tavy*, which is "the traditional and predominant land use practice of eastern Madagascar" (STYGER *et al.* 2007). The species mainly occurs in forest ecosystems (MADAGASCAR CATALOGUE 2020), and forest cover decline between 1953 and 2017 was estimated from the forest cover time series published in VIEILLEDENT *et al.* (2018a, 2018b) to be 35% in the altitude range of 0 – 30 m and within the minimum convex polygon encompassing all known collections of this species (historic and modern, n = 43, Figure S2). This species has a clear habitat preference for littoral forests, which have experienced significant reductions in southeast Madagascar (BOLLEN AND DONATI 2006, LOWRY *et al.* 2008) and are among Madagascar's most threatened forest formations (GANZHORN *et al.* 2001).

A further serious plausible threat is selective logging for firewood and possibly charcoal, or to produce furniture from individuals with an arboreal growth habit (R. Razakamalala, pers. comm.).

Four locations are defined based on the principal threat of habitat degradation or loss. Two locations correspond to the protected areas of Mandena and Sainte Luce. The third location corresponds to occurrences located outside of the protected area of Sainte Luce, including occurrences north of the Ebakika river. The fourth location corresponds to the Ampasy forest subpopulation. Given the higher level of protection afforded in protected areas, occurrences inside and outside of protected areas can be considered as different locations. Moreover, the subpopulation at Ampasy appears to be less accessible by road compared to the subpopulation around Sainte Luce, which merits the recognition of two separate locations. However, the streams might constitute a natural route for transport of logs from remote subpopulations.

USE AND TRADE INFORMATION

This species is used for firewood and possibly charcoal, or to produce furniture from individuals with an arboreal growth habit (R. Razakamalala, pers. comm.).

Traces of selective logging of this species were documented in the Mandena region (*Rabevohitra 2033*). The heartwood of this species is of an orange-brown colour (*S. A. Andrianarivelo 58*, *Razakamalala 8566*) and not considered to be a *bois de rose* (R. Razakamalala, pers. comm.), as indicated by its vernacular name (Manary) and owing to its brownish rather than purplish colour.

Relevant vernacular or trade names are Manary (*Ramamonjiarisoa 4*), Magnaritoloho (*Ratovoson 1713*), Sambalahy (*Ramison 109*), or Tombobitsy (*Razafimandimby 237*).

CONSERVATION ACTIONS INFORMATION

Afforded Protection

The species occurs in the protected areas of Mandena and Sainte Luce. It is listed in CITES Appendix II since 2013, whereby the identification of non-detriment findings is considered to be impossible given the currently large gaps in knowledge about population size, distribution and abundance, and an uplifting of all of Madagascar's precious woods to Appendix I was requested (WAEBER *et al.* 2019). An ongoing international research project aims at documenting the occurrence and morphology of *Dalbergia* spp. from Madagascar, as well as collecting leaf samples for genetic analysis and heartwood samples for their anatomical, spectroscopic and spectrometric characterisation. These collections are used to build a reference collection of *Dalbergia* spp. from Madagascar for forensic timber identification.

Research Needed

The extent and impact of habitat degradation, fire and logging on population size and trends, age structure, extent of occurrence and conservation status should be investigated and regularly monitored. It should further be clarified whether there are extant populations in additional patches of unprotected littoral forest, or forests along streams. Research on its ecology, growth rate, life history and the ability to recover and recruit following burning or logging would also be valuable.

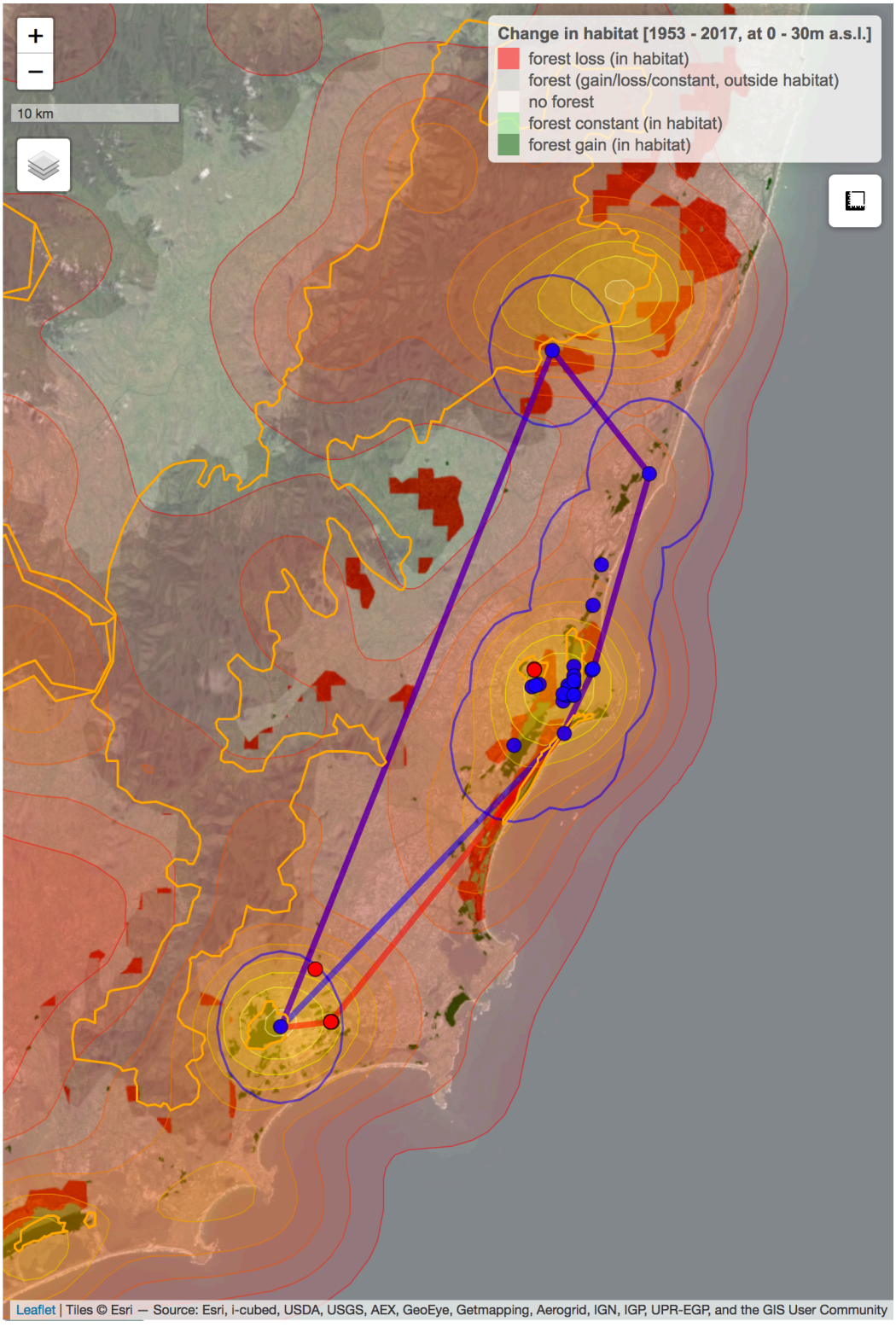

FIGURE S2: Distribution of *Dalbergia pseudomaritima*. Species occurrences are shown as slightly jittered points and withheld geo-coordinates (blue: extant [n = 29], red: not detected since 2006 [n = 13]). The extents of occurrence (minimum convex polygon) are shown as thick lines (blue: extant, red: historic, purple: extant and historic). Thin blue lines denote three extant subpopulations using a radius of 4 km around each occurrence. Orange lines denote protected areas. Shaded areas denote change in forest cover between 1953 and 2017 estimated from VIEILLEDENT *et al.* (2018a, 2018b), within an altitude range of 0 – 30 m (see colour legend), resulting in an estimated loss of 35%. Heat-colored contour lines denote 2D Kernel density estimates of angiosperm collections databased on www.tropicos.org (accessed on Aug 6 2020).

Dalbergia razakamalalae Crameri, Phillipson & N. Wilding

TAXONOMIC NOTES

All *Dalbergia* spp. from Madagascar are currently being subjected to taxonomic review.

Dalbergia razakamalalae was previously included in *D. maritima* but was recently recognized as distinct on the basis of its larger flowers and differences in leaf characters. This species is also eco-geographically and genetically distinct from *D. maritima* based on phylogenomic analysis of target enrichment sequencing data of hundreds of genomic regions (CRAMERI *et al.* in prep.).

IUCN RED LIST CATEGORY AND CRITERIA

Endangered (EN) B1ab(i,ii,iii,iv,v)+2ab(i,ii,iii,iv,v)

JUSTIFICATION

Based on a total of 34 extant and 5 extirpated collection records, the species has an extent of occurrence (EOO) of 2,085 km^2 (formerly 2,737 km^2), an area of occupancy (AOO) of 56 km^2 (formerly 72 km^2), and is inferred to comprise three subpopulations (Figure S3). In light of the principal threat facing the species, which is selective logging for trade in its high-quality heartwood, four locations are defined. The species is believed to be undergoing continuing decline in (i) extent of occurrence, (ii) area of occupancy, (iii) area, extent and/or quality of habitat, (iv) number of subpopulations and (v) number of mature individuals, and is therefore assigned an extinction risk of Endangered.

GEOGRAPHIC RANGE INFORMATION

The species is endemic to Madagascar, where extant subpopulations have only been recorded in the Anosy and Atsimo-Atsinanana regions in the southeast. The current EOO is 2,085 km^2 using a minimum convex polygon, and the current AOO is 56 km^2 using a 2×2 km grid size. The EOO and AOO were calculated, using GeoCAT (BACHMAN *et al.* 2011), based on 34 georeferenced collections that were inferred to represent extant individuals (Figure S3).

POPULATION INFORMATION

Extant subpopulations

The species comprises three extant subpopulations based on the distribution of the 34 occurrences. One is located in and around the northern part of the Tsitongambarika protected area (e.g., F. Ramanitrinizaka 18, R. Razakamalala 8032 & 8040). Another subpopulation is located in the Ampotaky forest (A. Karatra 241 & 242) and separated from the first by the largely forest-free Manampanihy valley. A third subpopulation is located approximately 160 to 170 km to the north of the other two subpopulations, in the northern parcel of the Manombo protected area (e.g., Emeline 23).

Possibly extinct subpopulations

It is currently unclear whether collections that were made in the 1940s in the forests of Manantately (*Humbert 20355bis*, *Réserves Naturelles, Dinard 1689*) and Ivola (*Réserves Naturelles 1124*) represent an additional extant subpopulation. These collections were made near or within the southern part of Tsitongambarika, which received protection only in 2015, and where there has been considerably less sampling effort compared to the northern part. Three further collections from the 1940s and 1950s (*Humbert 20607* from the Manampanihy valley, *Service Forestier de Madagascar 7110* from between Ambitananona and Vangaindrano, and *Service Forestier 38-R-118* from the Manakara-Sud district) originate from areas that have largely been cleared of forest, and are not inferred to represent additional extant subpopulations.

HABITAT AND ECOLOGY INFORMATION

Habitat

The species has been recorded in low-elevation evergreen humid forests on lateritic soils, between 20 — 510 m (MADAGASCAR CATALOGUE 2020).

Phenology

The species has been collected in full flower in the months of November (*Réserves Naturelles, Dinard 1689*) to February (*Service Forestier, Capuron 22334*).

Life history

Reproductive structures have been observed in individuals at least 5 m tall and with a diameter at breast height (DBH) of at least 7 cm (*Razakamalala 8558*). The largest documented trees were 20 m tall (*Andriamihajarivo 2455*) or had a DBH of 40 cm

(*Razakamalala 7736*). The generation length of this species is not precisely known but estimated to be approximately 30 to 40 years. Logging disproportionally targets large individuals, which likely belong to the most fertile age classes. Coppice shoots have been observed (*Ramanitrinizaka 18*, *Emeline 23*) but it might take up to a decade before reproduction can resume.

THREATS INFORMATION

The most serious plausible threat to *D. razakamalalae* is selective logging for national and often illegal international trade, as it is highly sought after for its dense and burgundy coloured heartwood, which was and probably still is used in cabinet-making (*Humbert 20355bis*).

A further serious plausible threat is habitat degradation or loss by land conversion for housing, industrial or agricultural development (see VIEILLEDENT *et al.* (2018a) and references therein), most notably due to burning for pasture and the slash-and-burn practice *tavy*, which is "the traditional and predominant land use practice of eastern Madagascar" (STYGER *et al.* 2007). The species mainly occurs in forest ecosystems (MADAGASCAR CATALOGUE 2020), and forest cover decline between 1953 and 2017 was estimated from the forest cover time series published in VIEILLEDENT *et al.* (2018a, 2018b) to be 70% in the altitude range of 20 – 650 m and within the minimum convex polygon encompassing all known collections of this species (historic and modern, n = 40, Figure S3).

Four locations are defined based on the principal threat of selective logging. Two locations correspond to the protected areas of Manombo and Tsitongambarika. Two further locations correspond to the parts of two subpopulations located outside of protected areas (one in the Ampotaky forest and another outside the limits of Tsitongambarika). Given the higher level of protection afforded in protected areas, occurrences inside and outside of protected areas can be considered as different locations. Moreover, the subpopulation at Ampotaky is situated at higher elevation and appears to be less accessible by road compared to the subpopulation outside Tsitongambarika, which merits its recognition as a separate location. However, the Manampanihy river might constitute a natural route for transport of logs from remote subpopulations (BARRETT *et al.* 2010).

USE AND TRADE INFORMATION

This species is considered to produce high-quality rosewood or *bois de rose* (R. Razakamalala, pers. comm.) based on its exploitable growth habit and the beautifully

veined heartwood of a burgundy colour (*Karatra 242*, *Bernard 2465*). The heartwood was and probably still is used in cabinet making (*Humbert 20355 bis*).

Traces of selective logging of this species were documented in the Tsitongambarika (*Ramanitrinizaka 18*) and Manombo (*Emeline 23*) protected areas. Despite large gaps in knowledge about the actual trade volumes, it can be assumed that this species was heavily exploited especially for international trade. This is indicated by the demand for its high-quality heartwood, its ability to grow to tall trees, the relatively high accessibility of source populations, especially at Manombo, as well as by the proximity of the northern and southern limits of its former extent of occurrence to the ports of Manakara in the north and Tolagnaro in the south. This species has probably also been used locally for a long time, as indicated by reports on the use of *bois de rose* from the Farafangana region to produce coffins for chiefs (NORMAND 1988). Together with *D. baronii*, the heartwood of *D. razakamalalae* is suspected to represent the major source of *bois de rose* from southeast Madagascar.

Relevant vernacular or trade names are *bois de rose* (*Humbert 20607*), Tombobitsy (*Humbert 20607*), Tombobitsy lahy (*Razakamalala 8035*) or Volombodipona (*Service Forestier 38-R-118*).

CONSERVATION ACTIONS INFORMATION

Afforded Protection

This species occurs in the protected areas of Tsitongambarika and Manombo. It is listed in CITES Appendix II since 2013, whereby the identification of non-detriment findings is considered to be impossible given the currently large gaps in knowledge about population size, distribution and abundance, and an uplifting of all of Madagascar's precious woods to Appendix I was requested (WAEBER *et al.* 2019). An ongoing international research project aims at documenting the occurrence and morphology of *Dalbergia* spp. from Madagascar, as well as collecting leaf samples for genetic analysis and heartwood samples for their anatomical, spectroscopic and spectrometric characterisation. These collections are used to build a reference collection of *Dalbergia* spp. from Madagascar for forensic timber identification.

Conservation Recommendations

Include the low-elevation evergreen humid forests near Ampotaky (subpopulation of *Karatra 241 & 242*) in Madagascar's protected areas network, ideally through a forest corridor between the national parks of Andohahela and Midongy du Sud.

Research Needed

The extent and impact of logging, habitat degradation, and fire on population size and trends, age structure, extent of occurrence and conservation status should be investigated and regularly monitored. It should further be clarified whether there are extant subpopulations at additional locations or protected areas with suitable habitat, such as the Andohahela and Midongy du Sud national parks, or the Ambositra-Vondrozo Forest Corridor. Research on its ecology, growth rate, life history and the ability to recover and recruit following burning or logging would also be valuable.

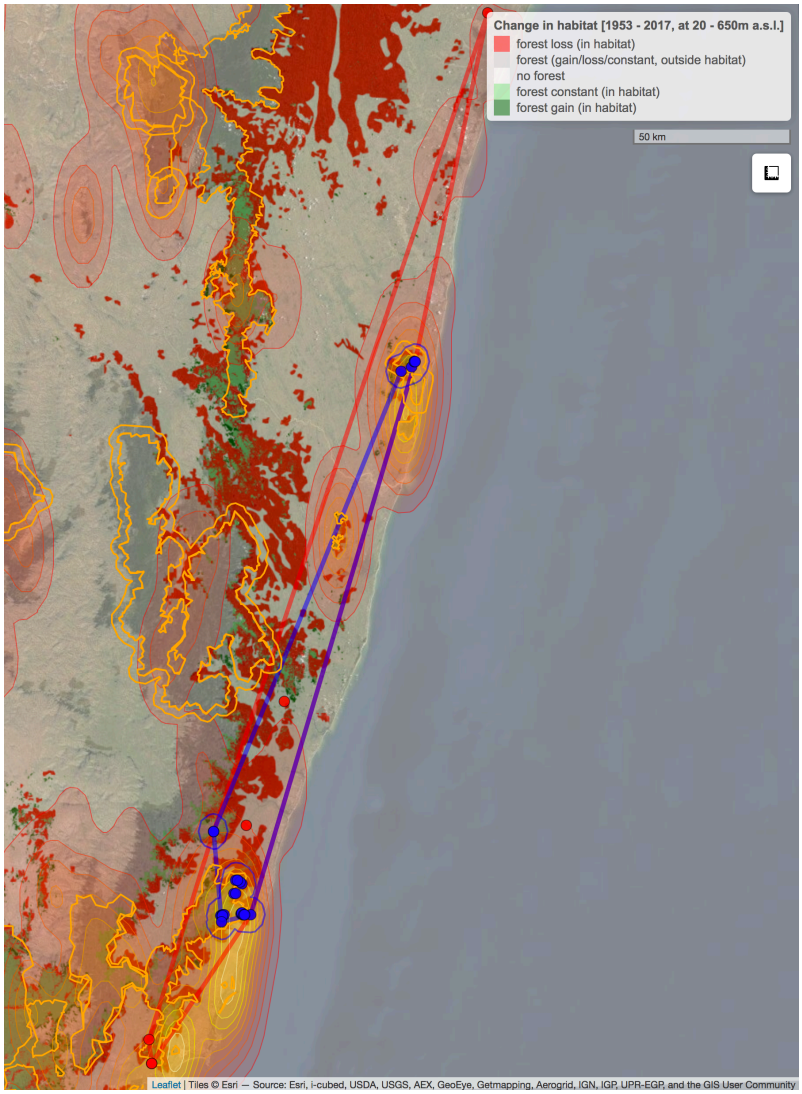

FIGURE S3: Distribution of *Dalbergia razakamalalae*. Species occurrences are shown as slightly jittered points and withheld geo-coordinates (blue: extant [n = 34], red: not detected since the 1950s [n = 5]). The extents of occurrence (minimum convex polygon) are shown as thick lines (blue: extant, red: historic, purple: extant and historic). Thin blue lines denote three extant subpopulations using a radius of 5 km around each occurrence. Orange lines denote protected areas. Shaded areas denote change in forest cover between 1953 and 2017 estimated from VIEILLEDENT *et al.* (2018a, 2018b), within an altitude range of 20 – 650 m (see colour legend), resulting in an estimated loss of 70%. Heat-colored contour lines denote 2D Kernel density estimates of angiosperm collections databased on www.tropicos.org (accessed on Aug 6 2020).

Identification key to species of *Dalbergia* from Madagascar, Supergroup I

The following draft identification key includes all currently (November 2020) known described species and new candidate species of Malagasy *Dalbergia* of Supergroup I (see CRAMERI *et al.* in prep.). Confirmed described species (in ***bold***), confirmed new candidate species (underlined) awaiting description and unconfirmed new candidate species (notably *D.* sp. 14, *D.* sp. 19, *D.* sp. 24 and *D.* sp. 51 within subgroup Maritima) are included, and a selection of these are shown in Figure S4. The unconfirmed new candidate species are included because 1) they often grow to an exploitable diameter (e.g., *B. Rakotonirina 8*, *Hassold 99*) and have frequently been collected as regrowth from felled trees, suggesting that these candidate species are affected by illegal logging, and 2) phylogenomic and population genomic data and analyses (CRAMERI *et al.* in prep.), which are available for all new candidate species except *D. lehavana* ined. (*D.* sp. 54), suggest that these entities represent distinct populations, and possibly distinct species that are merely precluded from formal taxonomic description due to a current lack of fertile material. Therefore, these unconfirmed new candidate species can be considered as evolutionary significant units (ESU) in accordance with RYDER (1986).

1. Leaves 15 — 25(— 32) cm long, with (5 —)7 — 9 leaflets, distal leaflets (4 —)5 — 10(—14) x (1.6 —)2 — 4.5 (— 6) cm, ovate, the apex acuminate; E and NE Madagascar .....***D. normandii*** (Fig. S4F)
1'. Plants without the above combination of characters..... 2.
2. Inflorescence paniculate (racemose in *D. tricolor* var. *breviracemosa*, sessile in *D. brachystachya*), pedicels 1 — 3(— 6) mm long, rarely slender; standard petal often elliptic to oblong or ovate; mature fruits elliptic to oblong, the 1-seeded (1 —)1.5 — 2(— 2.5) cm wide, pericarp yellow-brown to red-brown or purple when dried, net-veined or veins hardly visible; throughout Madagascar, recorded at 0 — 1500 m .... 3.
2'. Inflorescence racemose, often with smaller leaves, which subtend single flowers especially at the base, sometimes pseudo-paniculate with secondary branching close to the base, pedicels (1 —)2 — 6(— 10) mm long, often slender; standard petal often obovate to orbicular; mature fruits oblong or narrowly elliptic, the 1-seeded (0.8 —)1 — 1.5(— 1.8) cm wide, pericarp red-brown when dried, veins not or hardly visible; humid SE, CE and NE Madagascar, recorded at 0 — 800 m ..... 16.
3. Leaves glabrous on all surfaces…… ..... 4.
3'. Leaves loosely to densely pubescent on at least one of these surfaces: petiole and rachis, petiolule, leaflet underside..... 13.

4. Leaflet apex tapering and subacuminate to long-acuminate or acute, finally often obtuse and mucronulate; N, W and SW Madagascar and on the Central Plateau .....5.
4'. Leaflet apex never tapering and acuminate or acute .....11.
5. Inflorescence spicate, dense, 1 — 2 cm long, flowers in clusters of 5 — 8, subtended by large bracts..... *D. brachystachya*
5'. Plants without the above combination of characters .....6.
6. Inflorescence axes finely to densely pubescent; Central Plateau .....7.
6'. Inflorescence axes glabrous.....8.
7. Leaves with 6 — 9 leaflets; leaflets (1.8 —)2.5 — 5 x (0.9 —)1.2 — 3 cm; flowers 9 — 12 mm long; calyx 6 — 7 mm long; C Madagascar.....***D. capuronii***
7'. Leaves with 7 — 11 leaflets; leaflets (2.5 —)4 — 7 x (1.2 —)1.5 — 3 cm; flowers 5 — 6 mm long; calyx 4 — 5 mm long; SC Madagascar..... ***D. erubescens***
8. Leaves with (3 —)5 — 8 leaflets; leaflets 2 — 5 x 1 — 2.5 cm; inflorescences axillary, slender; W and N Madagascar; W and N Madagascar, on limestone 9. ***D. glaberrima***
8'. Leaves with 7 — 13 leaflets; leaflets 2 — 4 x 0.6 — 2.2 cm; inflorescences terminal; C Plateau and W Madagascar; on sandy or ferralitic soils ..... 10. ***D. tricolor***
9. Small tree, leaves with (5 —)6 — 7(— 8) leaflets ; leaflets obovate to narrowly so; calyx with the upper lobes not connate; ovules (3 —)4 — 5 ....*D. glaberrima* subsp. *glaberrima*
9'. Small, spindly treelet, shrubby, with lianescent shoots; leaves with (3)5-6 leaflets; leaflets broadly ovate to orbicular; calyx with the upper lobes connate and forming a single lobe which is emarginate at the apex; ovules 1(— 2) .... *D. glaberrima* subsp. *ankaranensis*
10. Inflorescences paniculate, reaching 20 x 15 cm; calyx 4 — 4.5 mm long, with short, broadly triangular lobes c. 1.5 mm long; W Madagascar and on the Central Plateau.... .....*D. tricolor* var. *tricolor*
10'. Inflorescences racemose or paniculate, short, 2 — 8 cm long; calyx 5.5 — 6 mm long, with narrowly triangular lobes 2 — 2.5 mm long; SC Madagascar ....*D. tricolor* var. *breviracemosa*
11. Calyx burgundy or pink becoming dark purple to black when dried; fruits dark brown to dark purple when dried, 1- to 2-seeded; N, NW, W, SW and S Madagascar and on the Central Plateau.....***D. pervillei s.lat.*** (incl. <u>*D. densicoma*</u> / <u>*D. obtusa*</u> / *D.* sp. 28)
11'. Calyx green with a reddish base becoming yellow-brown with a darker base or dark purple when dried; fruits yellow-brown to red-brown when dried, 1- to 4-seeded; humid SE, E, NE and NW Madagascar.....12.

12. Leaves (12 —)13.5 — 24(— 27) cm long, leaflets elliptic to oblong or obovate, (2.4 —)2.7 — 8.5(— 9.3) x (0.9 —)1.1 — 3(— 4) cm, apex usually rounded; inflorescence axes glabrous or pubescent; humid SE, E, NE and NW Madagascar, recorded at 0 — 1000 m..... ..... ***D. chapelieri s.lat.*** (incl. *D.* sp. 18, *D.* sp. 31, *D.* sp. 38, *D.* sp. 39, *D.* sp. 40)

12'. Leaves (4 —)5 — 8(— 9.5) cm long, leaflets broadly elliptic to orbicular, rarely obovate, (0.6 —)0.8 — 1.4(— 2.2) × (0.4 —)0.5 — 0.9(— 1.2) cm, apex shallowly retuse and sometimes mucronulate or rounded; inflorescence axes glabrous or ciliolate at the junctions; SE Madagascar, recorded at 0 — 100 m .....***D. pseudomaritima***

13. Leaves 15 — 20 cm long, with 9 — 15 leaflets, leaflets (narrowly) ovate to elliptic or oblong, 3 — 7 x 1.5 — 2.3 cm; only known from the Marivorahona massif in NC Madagascar, recorded at 1100 — 1500 m .....***D. aurea***

13'. Plants without the above combination of characters..... 14.

14. Leaves (4 —)5 — 9(— 10.5) cm long, with 5 — 8 leaflets; leaflets obcordate, rarely obovate or oblong-obovate, (1.4 —)2 — 4.1 × 0.8 — 1.8 cm; only known from Daraina .....*D. obcordata* ined. (*D.* sp. 10)

14'. Plants without the above combination of characters..... 15.

15. Leaves with (25 —)29 — 39(— 51) closely spaced leaflets; leaflets elliptic to ovate-elliptic, 0.8 — 1.2(— 1.5) x 0.3 — 0.5(— 0.8) cm, the base subcordate, petiolules 0.5 — 1.5mm long; inflorescence axes densely pubescent; region of Mahajanga and Soalala .....***D. tsiandalana***

15'. Plants without the above combination of characters; N, NW, W, SW, S and extending onto the Central Plateau ***D. pervillei s.lat.*** (incl. *D. densicoma* / *D. obtusa* / *D.* sp. 28)

16. Leaves glabrous on all surfaces ..... 17.

16'. Leaves loosely to densely pubescent on at least one of these surfaces: petiole and rachis, petiolule, leaflet underside..... 21.

17. Leaves (10 —)12 — 20(— 28) cm long, with (4 —)5 — 8(— 11) leaflets and a conspicuously reddish rachis; leaflets ovate to elliptic, (2 —)3.5 — 5(— 8) x (1 —)1.3 — 3.5 cm, the apex slightly emarginate to obtuse; recorded at Pointe à Larrée, at 0 — 100 m.....*D. lehavana* ined. (*D.* sp. 54, Fig. S4E)

17'. Plants without the above combination of characters; NE to SE Madagascar ..... 18.

18. Leaves (7.5 —)8.5 — 12(— 14) cm long, with (7 —)8 — 12(— 14) leaflets; leaflets ovate to elliptic, (1 —)1.4 — 3(— 3.8) x (0.7 —)0.9 — 1.5(— 1.8) cm (smaller when subtending a flower), the apex slightly emarginate to rounded, thinly coriaceous; recorded from W Makira to E Masoala, at 0 — 500 m ..... ***D. occulta*** (Fig. S4B)

18'. Plants without the above combination of characters; NE to SE Madagascar ..... 19.

19. Leaves 6 — 11(— 13) cm long (smaller when subtending a flower), with (8 —)11 — 15(— 20) alternate leaflets; leaflets ovate to elliptic, (0.6 —)0.9 — 1.5(— 2.2) × (0.4 —)0.5 — 0.8(— 1.0) cm, the base often asymmetric, the apex obtuse to rounded, sometimes slightly emarginate, plane with revolute margins, coriaceous; flowers 8 — 10 mm long; CE Madagascar, recorded at 0 — 100 m ...***D. maritima*** subsp. *maritima*

19'. Plants without the above combination of characters; NE to SE Madagascar.....20.

20. Leaves 7 — 13(— 16) cm long (smaller when subtending a flower), with 11 — 19(— 23) alternate leaflets; leaflets narrowly ovate to elliptic or obong, (0.8 —)1.3 — 2.5(— 3.5) × (0.4 —)0.5 — 1.0(— 1.5) cm, the apex emarginate or obtuse to almost rounded, plane without revolute margins, thinly coriaceous; flowers 10 — 14 mm long; SE Madagascar, recorded at 0 — 700 m ..... ***D. razakamalalae***

20'. Leaves (3.3 —)3.6 — 4.5(— 4.7) cm long, with 7 — 11(— 13) leaflets; leaflets (0.5 —)0.7 — 1.2(— 1.5) x 0.5 — 0.9 cm, orbicular to elliptic, the apex emarginate; NE Madagascar, recorded from the eastern part of the Masoala peninsula in the S to Makirovana in the N, at 0 — 400 m .....*D. racemosa* ined. (*D.* sp. 27, Fig. S4A)

21. Leaves (8 —)9 — 12(— 13) cm long, with (17 —)18 — 27(— 29) leaflets, leaflets elliptic to oblong, (0.6 —)1 — 1.5(— 1.9) x (0.4 —)0.5 — 0.7(— 0.8) cm; recorded from Ile Ste Marie and Pointe à Larrée in the S to Ambohitralanana in the N, at 0 — 300 .....*D. hassoldii* ined. (*D.* sp. 24, Fig. S4C)

21'. Plants without the above combination of characters .....22.

22. Leaves (8 —)15 — 19(— 21) cm long, with (13 —)15 — 25(— 27) leaflets; leaflets narrowly ovate to oblong, (1.8 —)2 — 4(— 4.8) x (0.6 —)0.7 — 1.2(— 1.5) cm; NE Madagascar, recorded from around Antalaha to Marojejy and towards Vohémar, at 0 — 750 m .....*D. marojejyensis* ined. (*D.* sp. 14, Fig. S4D)

22'. Plants without the above combination of characters; CE and NE Madagascar, recorded from around Ambila Lemaitso in the S towards Maroantsetra in the N.....23.

23. Leaflets ovate to elliptic, (0.6 —)0.9 — 1.5(— 2.2) × (0.4 —)0.5 — 0.8(— 1.0) cm, the base often asymmetric, the apex emarginate or obtuse to almost rounded; flowers 8 — 10 mm long; CE Madagascar, recorded from Sahafina to around Betampona, at 150 — 450 m .....***D. maritima*** subsp. *pubescens*

23'. Leaflets narrowly ovate to ovate or elliptic, (1.2 —)2 — 4 x 0.7 — 1.5 cm; the apex obtuse or slightly emarginate; flowers 12 — 15(— 18) mm long; CE Madagascar, recorded from around Ambila Lemaitso in the S towards Maroantsetra in the N, at 0 — 750 m .....***D. louvelii* s.*lat.*** (incl. *D.* sp. 19, *D.* sp. 51)

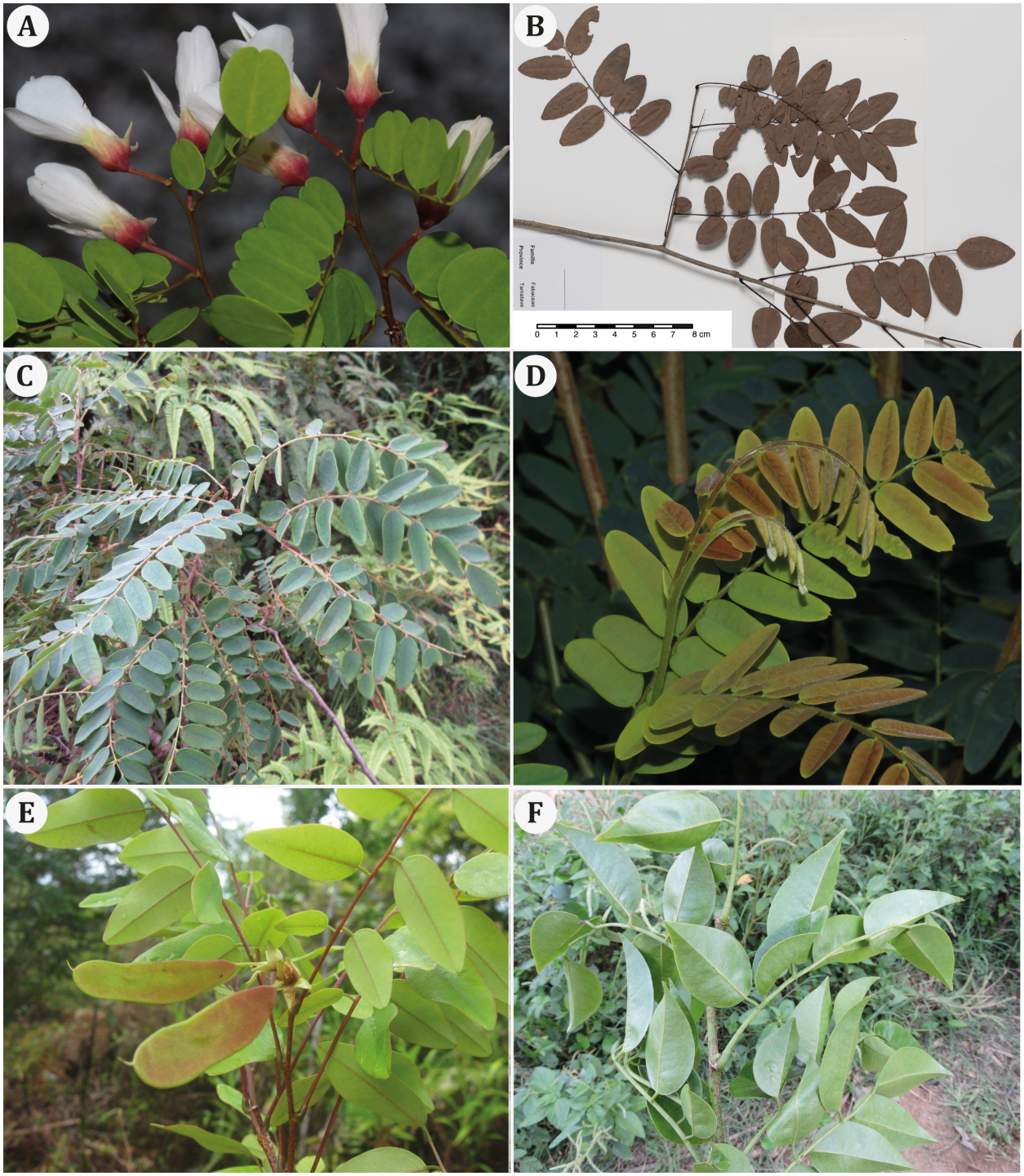

FIGURE S4: Representative photographs of further described *Dalbergia* taxa and new candidate species of subgroup Maritima. **A)** *D.* sp. 27 (*D. racemosa* ined., N. Rakotonirina 1220). **B)** *D. occulta* (S. Hassold 452). **C)** *D.* sp. 24 (*D. hassoldii* ined., A. Lehavana 1132). **D)** *D.* sp. 14 (*D. marojejyensis* ined., P. Phillipson 6650). **E)** *D.* sp. 54 (*D. lehavana* ined., A. Lehavana 902). **F)** *D. normandii* (R. Bernard 2486). — Photos: A & D by P. B. Phillipson; B by S. Crameri; C & E by A. Lehavana; F by R. Bernard.

References

Bachman, S., J. Moat, A. Hill, J. de la Torre, and B. Scott. 2011. Supporting Red List threat assessments with GeoCAT: geospatial conservation assessment tool. ZooKeys **150**:117-126.

Barrett, M. A., J. L. Brown, M. K. Morikawa, J.-N. Labat, and A. D. Yoder. 2010. Conservation. CITES designation for endangered rosewood in Madagascar. Science **328**:1109-1110.

Bollen, A., and G. Donati. 2006. Conservation status of the littoral forest of south-eastern Madagascar: a review. Oryx **40**:57-66.

Chang, W., J. Cheng, J. J. Allaire, and Y. Xie. 2020. shiny: Web Application Framework for R. R package version 1.4.0.2. https://CRAN.R-project.org/package=shiny.

Crameri, S., P. B. Phillipson, N. Wilding, T. Radanielina, Muséum national d'Histoire naturelle (P), P. P. Lowry II, and A. Widmer. in prep. Species discovery, phylogenomics and integrative species delimitation in *Dalbergia* (Fabaceae) precious woods from Madagascar and the Comoros.

Ganzhorn, J. U., P. P. Lowry, G. E. Schatz, and S. Sommer. 2001. The biodiversity of Madagascar: one of the world's hottest hotspots on its way out. Oryx **35**:1-3.

IUCN. 2012. IUCN Red List Categories and Criteria. Version 3.1. Second edition. IUCN Species Survival Commission, Gland & Cambridge.

Lowry II, P. P., F. Randriatafika, and J. Rabenantoandro. 2008. Conservation status of vascular plant species from the QMM / Rio Tinto mining area at Mandena, Tolagnaro (Fort Dauphin) region, southeast Madagascar. Madagascar Conservation and Development **3**:55-63.

Madagascar Catalogue. 2020. Catalogue of the Plants of Madagascar. Missouri Botanical Garden, St. Louis, USA. http://www.tropicos.org/Project/Madagascar on June 6 2020.

Normand, D. 1988. A propos des bois de rose de Madagascar. Revue Bois et Forêts des Tropiques **217**:89-94.

Ryder, O. A. 1986. Species conservation and systematics: the dilemma of subspecies. Trends in Ecology and Evolution 1:9-10.

Schuurman, D., and P. P. Lowry II. 2009. The Madagascar rosewood massacre. Madagascar Conservation and Development **4**:98-102.

Styger, E., H. M. Rakotondramasy, M. J. Pfeffer, E. C. M. Fernandes, and D. M. Bates. 2007. Influence of slash-and-burn farming practices on fallow succession and land degradation in the rainforest region of Madagascar. Agriculture, Ecosystems & Environment **119**:257-269.

Vieilledent, G., C. Grinand, F. Rakotomalala, R. Ranaivosoa, J.-R. Rakotoarijaona, T. F. Allnutt, and F. Achard. 2018a. Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. Biological Conservation **222**:189-197.

Vieilledent, G., C. Grinand, F. A. Rakotomalala, R. Ranaivosoa, J.-R. Rakotoarijaona, T. F. Allnutt, and F. Achard. 2018b. Update to: Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar.

Waeber, P. O., D. Schuurman, B. Ramamonjisoa, M. Langrand, C. V. Barber, J. L. Innes, P. P. Lowry II, and L. Wilmé. 2019. Uplisting of Malagasy precious woods critical for their survival. Biological Conservation **235**:89-92.

General Discussion

During my dissertation I developed and validated molecular and bioinformatic methods supporting population genomic and phylogenomic analyses in a non-model system (chapter 1). I then applied these methods in a basic research project exploring species diversity of *Dalbergia* in Madagascar using an integrative taxonomy approach (chapter 2). Next, I prepared taxonomic descriptions and conservation assessments for three species (chapter 3). In the following I will briefly discuss the most important findings, address some limitations and provide an outlook for future research.

Relevance of findings

We validated target enrichment sequencing as an applicable and highly informative genomic approach for basic research on the biological diversity in an understudied system and region. The unprecedented resolution at various evolutionary time scales, including those relevant for speciation, informed integrative research in the economically relevant and taxonomically complex genus *Dalbergia*. We also demonstrated that sequencing data obtained from high-quality DNA samples can be combined with the analysis of informative herbarium material dating back at least 100 years. This finding is not novel but reaffirms the increasing perception of museum collections as possibly underappreciated sources of available biological information waiting to be exploited (HART *et al.* 2016).

The developed set of target enrichment baits and the assembly of 2,396 reference sequences for target enrichment sequencing facilitate genome-wide analysis of *Dalbergia* from Madagascar and elsewhere. Our findings reveal that the diversity of *Dalbergia* species in Madagascar has been vastly underestimated. This has implications not only for future studies on the ecology and evolution of these species, but also for conservation management and ultimately also for CITES regulations. To become relevant, however, it is essential that the newly identified candidate species be scientifically described such that they are recognized by the international scientific community and in conservation legislation and regulation. The completion of our integrative taxonomic workflow for three threatened species, including the preparation of Red List assessments, represents a first step in this direction and translates basic research findings in conservation prioritisations and potential conservation action.

The well-supported phylogenetic backbone of Malagasy *Dalbergia* species resulted in the identification of eleven highly supported subgroups, one evolutionary grade with some morphological coherence, one group of phylogenetically unplaced species and two isolated lineages. These can now be assessed for potential synapomorphies in leaf, bark, wood, inflorescence, flower, and/or fruit characters, which could facilitate the development of a functioning morphological identification key for the entire genus on Madagascar. In addition, the phylogenetic subgroups and species delimitations also hold the potential to

facilitate the identification of diagnostic wood anatomical, spectroscopic or spectrometric characters for forensic timber identification.

Potential limitations

The present study encompasses the most comprehensive sampling of Malagasy *Dalbergia* to date. Because of the unexpected and staggering diversity encountered, the logistical challenges and restrictions due to accessibility during field work, coupled with the scarcity of encountered individuals with flowers and/or fruits, the number of available high-quality collections was still found to be limited for many (putative) species. This had consequences for the choice of our species discovery approach. Some authors (CARSTENS *et al.* 2013) suggest to approach species discovery, which is often synonymous with population discovery, using unguided clustering techniques such as Structurama (HUELSENBECK *et al.* 2011), Structure (PRITCHARD *et al.* 2000) or TESS (CHEN *et al.* 2007). Use of these approaches would increase the objectivity compared to our multivariate approach based on principal component analysis and neighbour-joining trees. However, clustering approaches are negatively affected by uneven sampling (PUECHMAILLE 2016, MEIRMANS 2019) and deliver numbers of clusters and clustering solutions, which all need to be interpreted and evaluated *post hoc* (EVANNO *et al.* 2005, MEIRMANS 2015, JANES *et al.* 2017). This implies a potential subjectivity in the selection of the most likely number of clusters, and also requires some *a priori* knowledge about population structure to develop a high-quality sampling design. This was not compatible with the little knowledge available before analysis and the difficult conditions during collection and export of collected material, which resulted in a largely blind and opportunistic sampling approach.

The species validation step, which we here approached by integration of morphology and eco-geography without a model-based framework, is increasingly being approached using the multispecies coalescent (MSC) model (YANG AND RANNALA 2014, RANNALA AND YANG 2017). However, our trials of the widely used BPP program (FLOURI *et al.* 2018) required subsampling of the genomic dataset for computational reasons, and revealed a strong tendency of the algorithm to support almost any putative species (or population) as a distinct species validated under the MSC model. We could not reconcile these results with our awareness of the geographically often discontinuous sampling and our knowledge of morphological characters and suspected ecological niches of the different putative species. Therefore, we adopted an integrative species delimitation and candidate species approach similar to the one adopted by VIEITES *et al.* (2009) for integrative species delimitation in Amphibians from Madagascar. Nevertheless, species discovery, delimitation and validation approaches are a highly active field of research (CAMARGO *et al.* 2012, CARSTENS *et al.* 2013, YANG AND RANNALA 2014, SOLIS-LEMUS

et al. 2015, JACKSON *et al.* 2017, SMITH AND CARSTENS 2019, CAMPILLO *et al.* 2020, HAUSDORF & HENNIG 2020) and there are vigorous debates on how best to approach these tasks consistently and objectively (SUKUMARAN AND KNOWLES 2017, LEACHÉ *et al.* 2018, CHAMBERS AND HILLIS 2020). Therefore, these research debates and advances must be considered when designing future studies.

Directions for future studies

The high number of newly discovered putative species, and the opportunities arising from our genome-wide dataset call for continued research. Firstly, the newly discovered species diversity requires further investigation to test and clarify the status of multiple unconfirmed new candidate species, and additional confirmed new candidate species need formal taxonomic descriptions to ensure that taxonomists, other researchers, conservation practitioners and amateurs work on and talk about the same units (PANTE *et al.* 2015). New or updated conservation assessments are needed for more of the here revised species hypotheses. The use of IUCN Red List criterion A (IUCN 2012) for Dalbergia species from Madagascar currently represents a major challenge because of the need to quantify populations size reductions. There is currently insufficient knowledge on multiple relevant parameters to assess population reductions over time, such as generation length, effective species abundance and distribution, habitat preferences and levels of exploitation. Species distribution models could provide additional insights and inform conservation assessments (SYFERT *et al.* 2014), but these also rely on a representative set of collections for model training and validation (GRIMMETT *et al.* 2020), which is currently lacking for many Malagasy Dalbergia species, and calls for continued field collecting efforts.

Secondly, the large amounts of genome-wide data accumulated during this thesis can be leveraged to gain insights into species diversification rates across time and space (STADLER 2011, MANDEL *et al.* 2019), as well as to test hypotheses on species diversification mechanisms in Madagascar (VENCES *et al.* 2009). Such studies would also greatly benefit from (ideally) complete taxon sampling and may help shed light on the processes that have shaped the fascinating biodiversity of Madagascar.

Thirdly, our highly informative sequencing approach could, in combination with a genus-wide and dense taxon sampling, result in a highly supported genus-level phylogeny and contribute to biodiversity assessments of rosewoods worldwide (EGAN & VATANPARAST 2019). This could reveal many biogeographic connections and synapomorphies that would increase our evolutionary understanding of *Dalbergia* on a global scale. In the second chapter, we have demonstrated that integration of available and informative herbarium material with modern collections is possible, an approach that could be extended to include all known *Dalbergia* species.

Lastly, our research was motivated by recent concerted efforts (DORMONTT *et al.* 2015) to reduce the devastating environmental and socio-economic consequences of illegal selective logging of precious timber species (reviewed in DUMBRELL *et al.* 2020). The third thesis chapter serves as an example for how basic research presented in chapters one and two can be applied. However, further applied research questions, for example in relation to forensic timber identification, need to be addressed. We are therefore working towards the development of a cost-effective SNP genotyping apprach to distinguish rapidly and reliably between *Dalbergia* species from Madagascar, and to collaborating with researchers in Madagascar and the US conducting Near-Infrared Spectroscopy (NIRS, RAOBELINA *et al.* 2020) and direct analysis in real time and time-of-flight mass spectrometry (DART TOF-MS, LANCASTER & ESPINOZA 2012) on collected heartwood samples. We anticipate that such an integrative approach will facilitate the development and evaluation of a set of tools for forensic timber identification and contribute to improved control of illegal timber trade and hence to the conservation of *Dalbergia* diversity in Madagascar.

References

Camargo, A., M. Morando, L. J. Avila, and J. Sites, Jack W. 2012. Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the Liolaemus darwinii complex (Squamata: Liolaemidae). Evolution **66**:2834-2849.

Campillo, L. C., A. J. Barley, and R. C. Thomson. 2020. Model-Based Species Delimitation: Are Coalescent Species Reproductively Isolated? Systematic Biology **69**:708-721.

Carstens, B. C., T. A. Pelletier, N. M. Reid, and J. D. Satler. 2013. How to fail at species delimitation. Molecular Ecology **22**:4369-4383.

Chambers, E. A., and D. M. Hillis. 2020. The Multispecies Coalescent Over-Splits Species in the Case of Geographically Widespread Taxa. Systematic Biology **69**:184-193.

Chen, C., E. Durand, F. Forbes, and O. François. 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Molecular Ecology Notes **7**:747-756.

Dormontt, E. E., M. Boner, B. Braun, G. Breulmann, B. Degen, E. Espinoza, S. Gardner, P. Guillery, J. C. Hermanson, G. Koch, S. L. Lee, M. Kanashiro, A. Rimbawanto, D. Thomas, A. C. Wiedenhoeft, Y. Yin, J. Zahnen, and A. J. Lowe. 2015. Forensic timber identification: it's time to integrate disciplines to combat illegal logging. Biological Conservation **191**:790-798.

Dumbrell, A., E. Turner, and T. Fayle. 2020. Tropical Ecosystems in the 21st Century. Academic Press.

Egan, A. N., and M. Vatanparast. 2019. Advances in legume research in the genomics era. Australian Systematic Botany **46**:185-125.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology **14**:2611-2620.

Flouri, T., X. Jiao, B. Rannala, and Z. Yang. 2018. Species Tree Inference with BPP Using Genomic Sequences and the Multispecies Coalescent. Molecular Biology and Evolution **35**:2585-2593.

Grimmett, L., R. Whitsed, and A. Horta. 2020. Presence-only species distribution models are sensitive to sample prevalence: Evaluating models using spatial prediction stability and accuracy metrics. Ecological Modelling **431**:109194-109112.

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

Hausdorf, B., and C. Hennig. 2020. Species delimitation and geography. Molecular Ecology Resources **371**:133-111.

Huelsenbeck, J. P., P. Andolfatto, and E. T. Huelsenbeck. 2011. Structurama: Bayesian inference of population structure. Evolutionary Bioinformatics **7**:55-59.

IUCN. 2012. IUCN Red List Categories and Criteria. Version 3.1. Second edition. Gland, Switzerland and Cambridge, U. K.: IUCN Species Survival Commission. .

Jackson, N. D., B. C. Carstens, A. E. Morales, and B. C. O'Meara. 2017. Species delimitation with gene flow. Systematic Biology **66**:799-812.

Janes, J. K., J. M. Miller, J. R. Dupuis, R. M. Malenfant, J. C. Gorrell, C. I. Cullingham, and R. L. Andrew. 2017. The K=2 conundrum. Molecular Ecology **26**:3594-3602.

Lancaster, C., and E. Espinoza. 2012. Analysis of select Dalbergia and trade timber using direct analysis in real time and time-of-flight mass spectrometry for CITES enforcement. Rapid communications in mass spectrometry **26**:1147-1156.

Leaché, A. D., T. Zhu, B. Rannala, and Z. Yang. 2018. The Spectre of Too Many Species. Systematic Biology **25**:1979-1914.

Mandel, J. R., R. B. Dikow, C. M. Siniscalchi, R. Thapa, L. E. Watson, and V. A. Funk. 2019. A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. Proceedings of the National Academy of Sciences of the United States of America:doi:10.1073/pnas.1903871116.

Meirmans, P. G. 2015. Seven common mistakes in population genetics and how to avoid them. Molecular Ecology **24**:3223-3231.

Meirmans, P. G. 2019. Subsampling reveals that unbalanced sampling affects Structure results in a multi-species dataset. Heredity **122**: 276-287.

Pante, E., C. Schoelinck, and N. Puillandre. 2015. From integrative taxonomy to species description: one step beyond. Systematic Biology **64**:152-160.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945-959.

Puechmaille, S. J. 2016. The program structuredoes not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Molecular Ecology Resources **16**:608-627.

Rannala, B., and Z. Yang. 2017. Efficient Bayesian species tree inference under the multispecies coalescent. Oxford University Press.

Raobelina, A. C., G. Chaix, A. R. Razafimahatratra, and T. Ramananantoandro. 2020. Contribution de la Spectroscopie Proche Infra Rouge et des méthodes chimiométriques de classification non supervisée dans la révision taxonomique des espèces de *Dalbergia* de Madagascar. http://agritrop.cirad.fr/597954.

Smith, M. L., and B. C. Carstens. 2019. Process-based species delimitation leads to identification of more biologically relevant species. Evolution **147**:466-414.

Solis-Lemus, C., L. L. Knowles, and C. Ané. 2015. Bayesian species delimitation combining multiple genes and traits in a unified framework. Evolution **69**:492-507.

Stadler, T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proceedings of the National Academy of Sciences of the United States of America **108**:6187-6192.

Sukumaran, J., and L. L. Knowles. 2017. Multispecies coalescent delimits structure, not species. Proceedings of the National Academy of Sciences of the United States of America:doi:10.1073-pnas.1607921114.

Syfert, M. M., L. Joppa, M. J. Smith, D. A. Coomes, S. P. Bachman, and N. A. Brummitt. 2014. Using species distribution models to inform IUCN Red List assessments. Biological Conservation **177**:174-184.

Vences, M., K. C. Wollenberg, D. R. Vieites, and D. C. Lees. 2009. Madagascar as a model region of species diversification. Trends in Ecology and Evolution **24**:456-465.

Vieites, D. R., K. C. Wollenberg, F. Andreone, J. Köhler, F. Glaw, and M. Vences. 2009. Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. Proceedings of the National Academy of Sciences of the United States of America **106**:8267-8272.

Yang, Z., and B. Rannala. 2014. Unguided species delimitation using DNA sequence data from multiple loci. Molecular Biology and Evolution **31**:3125-3135.

Acknowledgements

This dissertation would not have been possible without the help of many people. First and foremost, I would like to thank my doctoral advisor Alex Widmer for his enduring support, guidance, and the helpful scientific discussions over many years. This work would not have been possible without Alex and the Plant Ecological Genetics (PEG) research group at ETH Zurich, which provided a superb scientific environment. I especially thank Martin Fischer for his insights into population genetics, Karsten Rohweder for the many pleasant lunch breaks, and Ursina Walther, Hirzi Luqman and Artemis Treindl for being friendly and empathic fellow Ph.D. students.

A special thanks goes to Sonja Hassold, with whom I closely collaborated during my master thesis, and who encouraged me to continue and expand on the *Dalbergia* project. Among other things, she introduced me to important people of the tropical timber research community, and she instructed me in many administrative and cultural aspects of such a large and international research project. I also hold very fond memories of common field work during two fascinating field trips to Madagascar.

I am very thankful to Claudia Michel for the enormous amount of challenging laboratory work, which she performed with great precision and efficiency. Niklaus Zemp & Stefan Zoller, and Silvia Kobel from the Genetic Diversity Center (GDC) Zurich were always ready to help with bioinformatics and lab support, respectively. Michal Schorro, Heidi Amstalden and Michael Zehnder greatly helped with specimen digitization and thousands of morphometric measurements.

I am also thankful to my co-examiners. Among many other things, Pete Lowry contributed his extensive knowledge on the Malagasy flora to the interpretation of molecular and morphometric results, and Rolf taught me to focus on the big picture after working out the details.

Pete Phillipson and Nicholas Wilding greatly helped to turn the results of this thesis into further formal descriptions of new Malagasy *Dalbergia* species and many new conservation assessments. Both have become esteemed colleagues over the last years, and I hope that we can continue working closely together in the future.

I would like to thank all people involved in the G3D Madagascar precious woods project for the enormous amount of organisation, field trip planning, field collecting, sample management and dispatch to the various laboratories. A special thanks goes to Richard Randrianaivo, Nivohenintsoa Rakotonirina and Felana Rakoto, with whom I could share an unforgettable field trip to the Ankarafantsika National Park, the Ankarana Special

Reserve, and the extreme north of Madagascar. I further thank Ravo Ramanantsialonina, a Ph.D. student at the University of Antananarivo, for the interesting exchanges during her stay at the PEG lab.

Lastly, I would like to thank my family and friends, who also provided enduring support for many years. I would like to thank my twin brother Michel (Micio), who unfortunately and unexpectedly passed away in the summer of 2019, for being my friend, companion, role model and counterpart for 30 years. I especially thank my parents, Nicole, Ben, Fiona, Remo & Katrin and the kids, who provided much support and balance, especially during the last weeks of the writing process; Ramon and Niklaus, who regularly accompanied me on short walks or other distractions; Pädu and Jael, who brought emergency provisions shortly before the submission deadline; and Nadine, who encouraged me through her expressed appreciation of my work.

Curriculum Vitae

Personal details

Name	Simon Crameri
Date of birth	04.04.1989
Nationality	Swiss

Education

2015 – 2020	Doctoral (Ph.D.) thesis at the Institute of Integrative Biology, Plant Ecological Genetics (PEG) lab, ETH Zurich. Supervisors: Prof. Dr. Alex Widmer, Prof. Dr. Porter P. Lowry II, Prof. Dr. Rolf Holderegger
2016 – 2020	Tutorial assistant in ecological genetics at the Institute of Integrative Biology, ETH Zurich.
2015 – 2017	Diploma of Advanced Studies in Applied Statistics, Department of Mathematics, ETH Zurich.
2011 – 2020	Tutorial assistant in systematic botany at the Institute of Integrative Biology, ETH Zurich.
2012 – 2014	Master of Science ETH in Biology, major in Ecology and Evolution, ETH Zurich.
2009 – 2012	Bachelor of Science ETH in Biology, ETH Zurich.

Peer-reviewed publications

Eggens, F., Jafari, F., Thollesson, M., **Crameri, S.**, Zarre, S., Oxelman, B. (2020) Phylogeny and species delimitation in *Silene* sect. *Arenosae* (Caryophyllaceae): a new section. PhytoKeys **159**:1-34.

Manuscripts in review

Crameri, S., Phillipson, P. B., Rakotonirina, N., Wilding, N., Andriamiarisoa, R. L., Lowry II, P. P., Widmer, A. (in review) Taxonomic studies on Malagasy *Dalbergia* L.f. (Fabaceae). III: Two new species from southeastern Madagascar and an emended descripton of the rosewood species *D. maritima* R. Vig. Systematic Botany.

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 II, P. P. (in review) Taxonomic studies on Malagasy *Dalberiga* L.f. (Fabaceae). I: Two new species from northern Madagascar, and an emended descripton for *D. manongarivensis*. Candollea.

Wilding, N., Phillipson, P. B., **Crameri, S.** (in review) Taxonomic studies on Malagasy *Dalbergia* L.f. (Fabaceae). II: The reinstatement of *D. chermezonii* R. Vig. and a new name for *D. mollis* Bosser & R. Rabev. Candollea.

Contribution to the International Tropical Timber Organization (ITTO)

Hassold, S., **Crameri, S.**, Ravaomanalina, H., Lowry II, P. P., Widmer, A. (2018) Creating a base for rosewood identification. ITTO Tropical Forest Update **27**:4-7.

Contribution to the Global Timber Tracking Network (GTTN)

Schmitz, N. (ed.), H. Beeckman, H., Blanc-Jolivet, C., Boeschoten, L., Braga, J.W.B., Cabezas, J.A., Chaix, G., **Crameri, S.**, Degen, B., Deklerck, V., Dormontt, E., Espinoza, E., Gasson, P., Haag, V., Helmling, S., Horacek, M., Koch, G., Lancaster, C., Lens, F., Lowe, A., Martínez-Jarquín, S., Nowakowska, J.A., Olbrich, A., Paredes-Villanueva, K., Pastore, T.C.M., Ramananantoandro, T., Razafimahatratra, A.R., Ravindran, P., Rees, G., Soares, L.F., Tysklind, N., Vlam, M., Watkinson, C., Wheeler, E., Winkler, R., Wiedenhoeft, A.C., Zemke, V.Th., Zuidema, P. (2020). Overview of current practices in data analysis for wood identification. A guide for the different timber tracking methods. Global Timber Tracking Network, GTTN secretariat, European Forest Institute and Thünen Institute. DOI: 10.13140/RG.2.2.21518.79689.

Conference presentations

- Oral presentation at the Swiss Annual Meeting of Biologists, Zurich, Switzerland (2019). Integration of anchored phylogenomics and coalescent-based species delimitation in taxonomically challenging *Dalbergia* precious woods.
- Oral presentation at the 3rd Annual Meeting in Conservation Genetics, Vienna, Austria (2018). Molecular species diagnosis in CITES-listed *Dalbergia* precious woods.

Invited seminar presentations

- Diversity Turn Seminar, University of Göttingen, Germany (2020). Species diversity of rosewoods (*Dalbergia* spp.) revealed by anchored phylogenomics.
- Precious Woods Consortium Meeting, Antananarivo, Madagascar (2019). Génétique et delimitation des espèces de *Dalbergia* de Madagascar.
- Biodiversity and Systematics Seminars, Conservatoire et Jardin Botaniques (CJB) de Genève, Switzerland (2018). Cryptic species diversity in rosewoods (*Dalbergia* spp.) revealed by anchored phylogenomics.