Ecological genomics and adaptation of rosewoods *Dalbergia cochinchinensis* and *D. oliveri* for conservation and restoration

Tin Hang (Henry) Hung

University College & Department of Plant Sciences

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

March 2022

In the darkest, dreadful, daunting day
The poachers permeate the perpetual peace
The footsteps frighten the fortified forest
The seekers smudge the solacing silence

I hear they whisper *Kranhung*
The name so distant
Their call so near
"*Kat*"

"We kill the *Kranhung* with the chainsaws"
"*Kranhung*" "*Kat*"
"We cut open their brothers and sisters"
"*Kranhung*" "*Kat*" "We see the blood"
"*Kranhung*" "*Kat*" "We smell the blood"
"*Kranhung*" "*Kat*" "We take their blood"
"*Kat*" "Because blood is money, blood is gold, blood is future"
"*Kat*" "We"
(The cold barrel smokes hot.)

In a village not far away
The family sent a father to a grave
For the money of rosewood
Be heard the last lament of the poor

在那遥遠的地方 (*In thousands of miles away*)
From a newborn embraced
In a cradle crafted of *Hongmu*
Be heard the first word of *Ma ma*

The cradle is also the grave

We are nothing but the twigs of the
Tree of life

'The bloodwood', from *Wooden Heart*
Henry Hung (2019)

Table of Contents

Foreword

This thesis is a reminder of scholarship to myself. I complete this thesis amidst a time when a war is starting, a pandemic is roaming, and the sixth mass extinction is happening. The longer these humanitarian and environmental crises endure, the more they seem a hopeless air of neglect. We as scholars must find our places. We must continue to alert the world what problems we are facing, and to protect the freedom and integrity of knowledge so that the society will be guided with the best information. We are in a race for new technologies, medical treatments, and sustainable solutions to bring hope to this world.

And most importantly, we must continue to make truth and science count.

I completed this thesis with the best mentors, collaborators, and friends I could never imagine. I am immensely grateful for my two supervisors, John MacKay and David Boshier, for their unfailing support, patience, and insight. I came to Oxford without any knowledge on rosewoods and very little experience in research. John and David led me into a world of curiosity and opened my eyes to how research could bring kindness to humanity. I am enormously fortunate to have a group of supportive peers, including Barley Rose, Gabriele, Hayley, Heather, Laura, and Rosemary in the MacKay Group, and Edgar, Tom, and many more colleagues at the Plant Sciences. I have always known that I can count of them anytime for their knowledge, help, and comfort. I am hugely benefited from the support of my collaborators in Cambodia, Laos, Thailand, Vietnam, and many more places in the world. They endured the tedious fieldwork and the administrative procedures to provide the samples and knowledge which make this thesis possible. I have learned from them the dedication towards conservation. Finally, I thank Stephen Harris, Simon Hiscock, and Andrew Eckert, for their examining and valuable feedback that have improved this thesis.

On a personal note, I really wish to thank all of my friends but if I name all of them, this thesis will exceed the word limit and probably be rejected. Thank you for always offering me a chat when I am suffering, thank you for waiting and having faith in me when I need a bit of silence and focus, thank you for showering me your love (and many times food and drinks) when I am alone in the lab at the midnight with an existential crisis. Thank you, Eric, for your nearly daily phone calls since I left Hong Kong. It all began as I had a fresher's flu and homesickness when I first arrived here and phoned you in tears – and you replied "I am too worried about you, I will phone you every day to make sure you are okay". I would not be here without my family, to whom I owe so much for nurturing me and teaching me to become a better person.

I want to thank myself for finishing this thesis. There were many moments that seemed impossible and I wanted to quit, the latest one being me diagnosed with major depression when this thesis was even so close to finish. I really wish I could travel back in time, give myself a hug, and tell myself this would all pass. This thesis is the best testament to it. I made it.

I dedicate this thesis to every person around the world who is endeavouring to preserve our nature and humanity and to every silent tree that cannot speak for themselves.

Abstract

Global biodiversity, in particular tropical forests, is decreasing under both environmental change and anthropogenic disturbance. Environmental change alters species' adaptability to their current habitat, leading to loss of fitness and range shift, while anthropogenic disturbance reduces their adaptive capacity. Conserving and restoring threatened species and ecosystems thus become a grand challenge for the 21st century. This thesis studies two threatened rosewood species, *Dalbergia cochinchinensis* and *D. oliveri*, which are illegally exploited for their valuable timber in the Greater Mekong Subregion. They became the world's most trafficked wild product between 2005 and 2014, amounting to ~40% of the total global trade. Conservation efforts grew in the last decade to tackle the range-wide challenge, aiming to improve the species' survival, amplify the production of genetic materials, and designate more conservation units. However, a sustainable supply of genetic materials can meet several challenges that compromise the effectiveness of a restoration programme, namely the genetic bottlenecks, maladaptation, and climate change. While knowledge of adaptation can predict and mitigate these risks, standard study approaches such as common garden experiments have become impractical due to the acute threats of illegal logging in these two species, which are lacking in *a priori* knowledge.

This thesis aims to increase the knowledge of genetic and physiological underpinning of adaptation in the two *Dalbergia* species with relevance to application in conservation and restoration strategies. This thesis presents a rich body of genomic resources such as chromosome-scale genomes and reference transcriptomes, which advance the progress in less-represented angiosperm tree genomes and woody legume genomes and enable studies in genetic diversity. Comparative genomic studies revealed insight into the evolution and

potential adaptive role of of certain gene families in tropical *Dalbergia* species. The landscape genomic study provides a comprehensive scan of adaptive signals and reports significant differences of the adaptive variation between the two species, where *D. cochinchinensis* is driven by temperature variability and *D. oliveri* by precipitation variability. The controlled stress experiment provides a physiological understanding of how the two species regulate their water relations and photosynthetic apparatus to respond to drought differently, where *D. cochinchinensis* has a more anisohydric behaviour than *D. oliveri*. These contrasting patterns of adaptation indicate how the two species may differentiate their niches, while co-occurring in some habitats.

The knowledge of adaptive variation identifies hotspots of local adaptation and vulnerability towards climate change, and thus are expected to help conservation practitioners delineate conservation units, compare provenances for assisted germplasm transfer, and prioritise conservation actions. It also opens new avenues for future research, including combining common garden experiments and genomic approaches to more fully unravel genotype-phenotype-environment relationships.

Chapter 1. General Introduction

Global biodiversity is negatively impacted by the rapidly changing environment and anthropogenic disturbances, such as habitat fragmentation and exploitation[1]. In particular, tropical forests support more than two-thirds of the world's biodiversity[2] but it also experiences the greatest total forest loss and the trend is still increasing[3]. Research has thus been increasingly focused on predicting the nature and magnitude of threats to species, in particular long-lived individuals and populations such as forest trees, such as species range shifts[4], loss of fitness[5], and changes in population genetic composition[6]. At the same time, alongside the declaration of the Decade on Ecosystem Restoration between 2021 and 2030 by the United Nations, forest restoration has been seen as an urgent opportunity to enable biodiversity conservation, climate mitigation and adaptation, local likelihoods, and economic gains[7]. Evidence-based restoration is therefore urgently called for to ensure that rigorous, repeatable, and transparent methods can be applied for biodiversity conservation[8].

This thesis studies two endangered rosewood tree species, *Dalbergia cochinchinensis* and *D. oliveri*, which are considered some of the most trafficked and most priced timber species in the world[9]. This introductory chapter starts by discussing how rosewoods have weathered decades of exploitation and become increasingly popular for restoration and plantation. While an effective restoration programme depends heavily on a sustainable supply of genetic materials, a number of associated genetic risks are also identified, namely the genetic bottleneck, maladaptation, and climate change. Understanding adaptation and its genetic and physiological underpinning to the environment is therefore critical for conserving and restoring forest trees like *Dalbergia*[10].

In the literature review, this chapter discusses how gene-environment studies are conducted in other tree species using common garden experiments traditionally, and their insights and challenges. This chapter then moves on to discuss how ecological genomics can help us understand adaptation and guide conservation decisions under climate change.

Finally, this chapter presents the overarching goal of the thesis, which is to develop and understanding of adaptive diversity in *D. cochinchinensis* and *D. oliveri* by comparing provenances across their species' range and searching for gene-environment associations, in relevance of their conservation and restoration, and how the objectives of the subsequent chapters address the goal.

1.1 Overview of *Dalbergia cochinchinensis* and *D. oliveri*

Dalbergia cochinchinensis Pierre and *D. oliveri* Gamble ex Prain belong to the genus *Dalbergia* L. f. (Fabaceae: Faboideae), which contains around 250 species and is predominantly pantropical[11]. Like many other *Dalbergia* species, the rosewood timber of *D. cochinchinensis* and *D. oliveri* is often targeted in illegal harvesting especially for luxury furniture[12]. The growing demand and the diminishing population supply have drastically increased the economic value of rosewood and exploited their natural populations. *D. cochinchinensis* was once the most sought-after rosewood species globally but it is now virtually commercially extinct, whereas *D. oliveri* has been following a similar trend[13]. They were classified as vulnerable and endangered respectively in the IUCN Red List, and their trade is now strictly regulated under CITES Appendix II since 2017. These two species are endemic to the Greater Mekong Subregion (GMS). *D. cochinchinensis* grows in open deciduous and semi-deciduous forests at elevation up to 400 metres, while *D. oliveri* grows in

a wider range of forest types from deciduous to evergreen and can grow at elevation up to 1,200 metres.

The Greater Mekong Subregion (GMS) in Southeast Asia is defined as consisting of Cambodia, China, Laos, Myanmar, Thailand, and Vietnam. The majority (84%) of its area overlaps with the Indo-Burmese mega-biodiversity hotspot[14]. The complex biogeographical and geological history in the GMS may have contributed to the high species richness and endemism levels found in the area[15]. In particular, the changes in the distribution of land and water in the region have been associated with changes in vegetation types and cover[16]. The topography is highly varied, ranging from lowlands to mountains, with many freshwater bodies including the Mekong River and the Tonle Sap Lake[17]. As a result, the GMS has a very heterogeneous landscape and supports a wide range of habitats. However, GMS also harboured the countries with fastest accelerating loss of tree cover between 2001 and 2014, for example, 14.4% in Cambodia and 6.1% in Vietnam[3].

1.2 Routes to endangered status

The exploitation of *D. cochinchinensis* and *D. oliveri* is an integral part of the worldwide rosewood trade. Between 2005 and 2014, rosewoods had the highest share among total seizures of wildlife products by value (35%)[9] – more than all animal products summed together, which still remained (31.7%) until 2018, when the latest data were available[18]. However, even with growing intentions to protect the species and to enforce national law throughout their ranges, the threats have not stopped and the populations have continued to decline[19]. It was suggested that the huge financial profits offered in the Asian market, particular the rising economy in China (which was the country of destination for 32.3% of rosewood seizure)[18], led to continuous exploitation of the species. It was observed that after

the listing of *D. cochinchinensis* on CITES since 2013, the dwindling wild stocks inflated its value increasingly[20]. *D. cochinchinensis* was reported to be the most trafficked rosewood species with 798 metric tons of seizures aggregated between 2005 and 2015[21]. While still not being listed on CITES until 2017, *D. oliveri* was subsequently targeted for illegal harvesting after *D. cochinchinensis* becoming rarer. The export of *D. oliveri* from Vietnam into China was 193,880 m^3, while that of *D. cochinchinensis* was 2,588 m^3 in 2013[22]. The risk of serial depletion of rosewood remains intense.

The population decline of *Dalbergia* species in the region has become and remains prominent. In Thailand, the number of *D. cochinchinensis* trees dropped from 300,000 in 2005 to 80,000 to 2011[19]. In Vietnam, even within the protected areas, *D. cochinchinensis* density was as low as 1–10 trees per hectare[19], similar to 1.37 per hectare in Cambodia[22]. In Laos, field surveys in 2012 confirmed no mature trees could be founded in the natural populations of *D. cochinchinensis*[20]. The endangered status was similar for *D. oliveri*. In Vietnam, forest stands up to 60 years were no longer characterised with *D. oliveri* and they have nearly disappeared from many forest sites[23,24].

There were very limited activities in commercial plantation of *Dalbergia* in the region, thus wild populations remain the target of exploitation and face increasing risk of extinction. Based on silviculture studies, a periodic annual increment of 1 cm in diameter-at-breast-height could be attained in 20–29 year-old plantations of *D. cochinchinensis*[19]. Natural regeneration in both species was possible but was often found to be poor due to low germination rates and unfavourable weather conditions, with negative impacts on growth rates[22].

1.3 Conservation actions of *Dalbergia* around early 2000s

Table 1.1. Conservation actions of *D. cochinchinensis* and *D. oliveri* by the countries in their species' range around early 2000s (compiled from APFORGEN reports[25,26]).

	Cambodia	**Lao**	**Thailand**	**Vietnam**
D. cochinchinensis	• Threat level 5 (2003) • 50 ha seed source (2002) • 5 kg seed demand supplied from CTSP and local people (2003) • 50 ha *in situ* conservation area • 3 ha production area • 2 genetic conservation stands, 69 ha, 147 trees (2006)	• Priority species (1999) • Category A (Endangered) • 200 ha plantation area in Bolikhamxai Forest (1992) • Used mainly for ex situ conservation and demonstration	• Very high priority species (2003) • Recommended to establish more *ex situ* conservation areas • 5 genetic conservation stands, 56 ha, 38 plus trees (1999) • 43 plus trees (2003–2007) • Research on genetic process and variation and distribution and status is needed	• Priority species (2003) • 1 *ex situ* conservation site • 2 seed sources with 2,600 trees planted
D. oliveri	• Threat level 5 (2003) • 4 seed sources, 71.5 ha, 157 mother trees	• Category B (Vulnerable)	• Very high priority species • 34 ha *ex situ* conservation site • Recommended to establish more *ex situ* conservation areas • 3 genetic conservation stands, 34 ha, 30 plus trees (2002) • 20 plus trees (2003–2007)	• Priority species • Very high commercial value

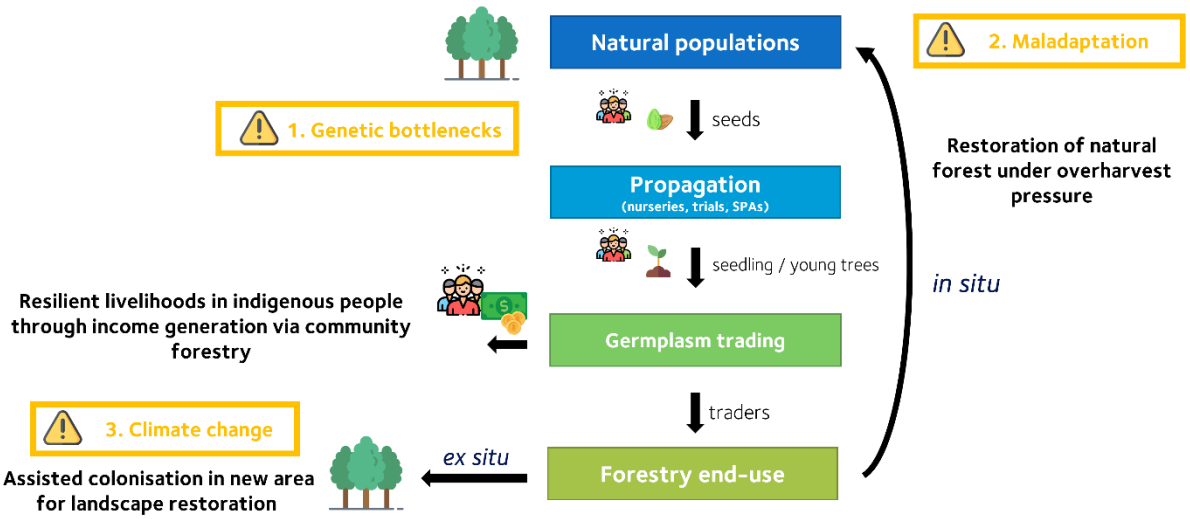
In response to their declining populations, a wide range of actions were initiated around early 2000s to conserve the two threatened *Dalbergia* species (Table 1), for example, *in situ* and *ex situ* conservation stands, seed production areas, plantations, and demonstration

plots. These actions were usually undertaken as part of a larger scale forest genetic resources management and development programmes with other indigenous, threatened tree species in the region, such as teak (*Tectona grandis*), padauk (*Pterocarpus macrocarpus*), and pyinkado (*Xylia xylocarpa*)[27]. By incorporating indigenous villagers, smallholders, and landowners, a mixed-species plantation usually attracts more interests as species with complementary traits can optimise the efficient use of limiting resources towards growth and production[28]. The two *Dalbergia* species were generally ranked high in the priority list of each country and specific land and economic resources were used towards the conservation and research around the two species. However, we note that the scale of conservation and production is still small in scale with usually < 50 plus trees for seed production in one country, and *D. oliveri* has attracted less attention than *D. cochinchinensis* in seed production and plantation initiatives. In short, at the time, the countries lacked capacity to provide a sustainable supply of genetic materials to meet the growing demand and conservation goals.

The range-wide threats to *Dalbergia* should ideally be met with a range-wide response to conserve the genetic diversity and ensure the survival of these populations. Since 2010s, there have been more concerted efforts across the countries to develop novel approaches, and strengthen existing capacity for genetic materials collection, nursery management, and value chain development[26]. For example, activities under the Asia Pacific Forest Genetic Resources Programme (APFORGEN) bring together all of the range countries of *D. cochinchinensis* and *D. oliveri*. In an APFORGEN project launched in 2018, the countries have identified a set of impactful goals towards conservation of the two *Dalbergia* species: including ≥ 50% increase in number of *in situ* or *ex situ* conservation units, ≥ 20% increase in forest-related income, and 10–25% improvement in species' survival and growth[29]. At the same time, the countries identified common challenges in assessing the conservation status and defining conservation units due to lack of knowledge and research[26].

1.4 Ensuring sustainable supply of adaptively diverse genetic materials

Figure 1.1. A schematic diagram of a conventional seed supply chain from natural populations to end use and their associated genetic risks.



A seed supply chain usually consists of four levels: the seed source, the propagation process, the trading network (wholesaler, retailers, and distributors), and the end users. Conservation of *D. cochinchinensis* and *D. oliveri*, together with other prioritised species, in the region heavily involves the local communities. For example, the governmental Cambodia Tree Seed Project leads a National Forest Programme to establish seed sources and community forests within and around local villages. Domestication activities of *D. cochinchinensis* have also been reported in Laos (Phongoudome, 2003). There has been an observed shift in responsibility of managing forest genetic resources from traditional forestry sector to include more community participation, as new policies have recognised property rights and benefit sharing for community-based forest management by releasing permits and tenure, since the start of the millennium (cf. The Forestry Act [UU No. 41/1999], MOF Decree no.31/Kpts-II/2000, Act No. 29/2000). Continued development of small-scale plantations and provision of nurseries at the village level has the potential to scale up the

capacity to multiplicate germplasm and strengthen the supply to meet the demand[25]. Demonstration plots for *ex situ* conservation, progeny trials, and seedling orchards were also established in the region[27].

The end users of these genetic materials, usually NGOs and forest authorities, decide their intended use and management. They may be planted into *in situ* populations of *Dalbergia*, which may also be defined as an *in situ* conservation area, to bring up the population size. They may also be brought into new habitats where *Dalbergia* have not colonised, which may be deforested land or an early-succession site, for forest landscape restoration.

1.5 Potential genetic risks

There are clear risks associated with declining population size and restoration efforts, which will have negative impacts on population genetic processes. The first genetic risk is associated with the genetic bottleneck during the sourcing of genetic materials. The underrepresented genetic diversity is then passed down the seed supply chain for *in situ* and *ex situ* plantations. Loss of genetic diversity reduces the species' ability to both adapt to the environmental gradient of current species' range (adaptive variation) and to evolve under future environmental change (adaptive potential)[30]. The loss is usually exacerbated in endangered species like *Dalbergia*, because the effect of genetic drift and inbreeding depression are more prominent in their small effective population sizes[31]. Unfortunately, the lack of sufficiently diverse tree seed was found to be a widespread problem in forest landscape restoration projects worldwide[32], usually due to the lack of capacity in seed sourcing such as poorly developed guidelines or location accessibility and logistics.

The second genetic risk is associated with maladaptation. Local adaptation is usually constrained by variation in natural selection (environmental heterogeneity) and gene flow, either temporally or spatially[33]. Local adaptation is usually characterised with both a higher average relative fitness of individuals in their native habitats than those from foreign habitats, and also reciprocally in foreign habitats[34]. There has been substantial evidence of local adaptation in plants globally, as supported by 71.0% of studies surveyed in the first quantitative meta-analysis[35]. In the strictest sense of local adaptation defined by Kawecki and Ebert[33] where a reciprocal experiment has to be conducted, there are still nearly half of the studies support the evidence of local adaptation. Therefore, maladaptation may either arise from mismatched habitat suitability or small effective population size of the source populations, which are more prone to deviation of population trait values from local optima (maladaptation through "inaccuracy")[36]. However, the extent and scale of local adaptation lacks global empirical evidence in tree species[37], and the common approach is to put limits on geographic distances but the accuracy is questionable because of the arbitrarity[38].

The third genetic risk is associated with climate change. The negative effect of climate change on long-term sustainability and effectiveness of restoration projects on endangered species is multi-fold: affecting the ecosystem functioning[39], composition and size of populations[40], and subsequently local extirpations and collapsing species' range[41]. Climate change can also pose indirect, collateral threats to conservation management, for example, the increased cost of seed production due to reduced land and water availability from extreme climates, and the risk of disease outbreak or lack of pollinators and dispersers due to change in the associated biodiversity[42]. It is predicted that trees are more vulnerable than more mobile and short-living organisms, as they have longer generation times and thus experience a higher rate of climate change per generation[43]. Therefore, a body of literature has indicated that local seed sources may not be the best choice, and in some cases even underperform

when the restored populations can no longer adapt[44]. More crucially, climate change makes tree reproduction unpredictable and can lead to more but infertile seeds[45], which has a direct impact for seed sourcing in restoration projects. At the same time, climate change, especially elevating temperature, accelerates aging of seeds, promotes oxidative stress, and decrease their viability[46]. Seed storage and germination will therefore become more challenging, especially in areas with less access to long-term refrigeration. In short, not only may climate change negatively impact the survivability of standing populations, it challenges the multiplication and regeneration of genetic materials either naturally or in a restoration project[47].

1.6 Detecting gene-environment association

Field experiments have been the gold standard to detect environmental variation and differential selection that leads to genetic divergence and local adaptation among populations. Common garden experiments have been widely used and can be dated back as early as a century ago[48]. Reciprocal transplant garden experiments emerged as an extension of the common garden experiments to detect local adaptation, or more specifically, the effect of genotype (G), that of environment (E), and their interaction (G × E). Evidence for local adaptation was observed in hundreds of taxa in the United States, in which evidence of local adaptation was inferred from better survival and fitness in 67% and 90% of the reciprocal transplant studies[49]. A recent meta-analysis reviewed 111 common and reciprocal transplant gardens between 1990–2020 and trees were the dominant study type[49]. While these experiments provide valuable insights into adaptive variation relevant to climate change[50], conservation and restoration[51], knowledge gaps still exist for geographically distributed populations of most plant species[52].

Although common and reciprocal transplant experiments may provide valuable insight into the adaptive variation of *Dalbergia*, they have several constraints and limitations. First, the existing provenance and progeny trials in the region only represent a few populations across the entire species range for the two species (Table 2), and this limits the power for capturing the range-wide pattern of adaptation. A larger-scaled provenance trial is thus needed, which may also serve as an *ex situ* conservation site and a seed production source. Second, these trials are expensive and time-consuming to establish and maintain[53]. Especially, when there are large number of provenances, reciprocal transplant experiments can scale up the cost exponentially, while common garden experiments may not have sufficient power to verify local adaptation[54]. The long generation time of *Dalbergia*, like other trees, is likely to exacerbate the challenges and cost of these experiments.

Table 1.2. Existing provenance or progeny trials of *D. cochinchinensis* and *D. oliveri* across their species' range (pers. comm. with forestry authorities).

Species	**Site country**	**Site location**	**Number of provenance**	**Provenance coverage**	**Number of families**	**Planted**	**Status assessed in 2019**
D. cochinchinensis	Thailand	Khao-Yai Field Station	7	East Thailand	85	1988	?
D. cochinchinensis	Vietnam	Chumomray NP & buffer zone	2		100	2003	Good growth and Survival
D. cochinchinensis	Cambodia	Siem Reap	7	Cambodia	100	2016	
D. oliveri	Vietnam	Cat Tien NP	2		100	2003	

The underlying theory of G × E and local adaptation is that selection pressure across a heterogenous environment favours an adaptive variation in allelic frequencies of different

loci, acting along other non-selection (neutral) forces such as genetic drift and geographical barriers to gene flow[55]. The need to account for environmental heterogeneity in genetic studies has opened the doors to the development of a new field – known as "landscape genetics" – in the last ~15 years[56]. Recent technological advances have stimulated the rapid evolution of a new sub-field known as "landscape genomics" (first coined in 2007[57]) to study adaptive variation due to three driving forces: (1) the accelerating advancements in sequencing and bioinformatic tools allow us to quantify genetic variation with higher genomic coverage and decreasing costs, even for non-model organisms[34], (2) the remote sensing technologies enable a finer spatial and temporal resolution of environmental data[58], and (3) the increased computational capacity and more robust statistical models support more power in analysing G × E[59]. Landscape genomics thus breaks methodological barriers of studying the genetic underpinning of adaptive variation and helps to extend the outputs of common garden and reciprocal transplant experiments.

Landscape genomics is particularly attractive for studying forest trees. Trees are particularly good models for understanding evolution and population processes because of their higher level of genomic and phenotypic variation, compared to other shorter-lived plants[60]. Due to their longevity and sessile lifestyle, they are spatially fixed and their adaptation is mainly limited to dispersal, local adaptation, and phenotypic plasticity[61]. The earliest landscape genomic studies in forest trees analysed a relatively limited number of loci (usually single nucleotide polymorphisms, SNP) and utilised simple correlation or regression statistical models, such as F_{ST} outlier tests and generalised linear mixed models (GLMM) (see Soltis *et al.*[62] for review). They usually detected tens of candidate loci as outliers and associated with environmental variables. As the field developed, there was a large and rapid increase in the marker number up to millions, which are a analysed with more sophisticated statistical models (Table 3). Non-model organisms may particularly benefit from reduced

representation sequencing methods such as start codon targeted markers (SCoT), genotyping-by-sequencing (GBS), and restriction site associated sequencing (RAD-seq). The advanced statistical models, many of which are based on Bayesian inference (e.g. BayeScan), or machine learning (e.g. gradient forest), bring in additional benefits, such as accounting for neutral genetic structure[63], correcting for unobserved confounders[64], and predicting the association in unsampled or future environment[65].

Table 1.3. Selected landscape genomic studies of forest trees using adaptive genetic markers between 2007 and 2021 as an extension of the meta-analysis by Balkenhol et al.[62]

Species	**Habitat**	**Markers**	**Genotyping method**	**Analysis approach**	**Main finding**	**Ref**
Cotinus coggygria	China, warm-temperate zone	1,131 loci	SCoT	FIDST2 + BayeScan	27 outlier loci (1.14%), 13 environmentally associated with temperature and precipitation	[66]
Quercus rugosa	Trans-Mexican Volcanic Belt	5,354 SNPs	GBS	(1) LFMM + (2) GF	74 outlier SNPs and 97 SNPs associated with climate variation, mostly temperature and temperature seasonality	[67]
Cornus florida	United States	3,134 SNPs / 1,660 loci	GBS	LFMM	72 outlier SNPs and 29 SNPs significant for predicting location with disease occurrence	[68]
Platycladus orientalis	Northern China	3,911 SNPs	GBS	BayeScan + Pcadapt + Bayenv2	579 (5.24%) outlier SNPs explained by environment (10.7%) and geography (6.3%)	[69]
Pinus taeda	Southern US	2,822,609 SNPs	GBS	TASSEL (GLM) + RDA + Samβada	611 SNPs associated with 56 climate and geographic variables, mainly longitude, max temperature of the warm months, monthly precipitation	[70]

Pterocarya stenoptera	China, warm-temperate and subtropical zones	1,006 loci	SCoT	Arlequin + BayeScan + LFMM	43 loci associated with environment, mainly minimum temperature of the coldest month and solar radiation in June	[71]
Neolitsea sericea	Taiwan, China, Japan, Korea, warm-temperate	62,978 SNPs	RADseq	GDM	85 non-neutral SNPs (79 under purifying selection and 6 under divergent selection), strongly with annual mean temperature and precipitation of driest month	[72]
Pterocarya stenoptera	China, warm-temperate and subtropical zones	894,250 SNPs	SLAF-seq	BayeScan + LFMM + Samβada	5,827 outlier SNPs, and 801 EAL, strongly with temperature, precipitation, and water vapour pressure	[73]
Juniperus monticola	Trans-Mexican Volcanic Belt	2,925 SNPs	ddRAD-seq	Mantel test	Genetic differentiation related to degree of glacial habitat connectivity	[74]
Berberis alpina	Trans-Mexican Volcanic Belt	3,669 SNPs	ddRAD-seq			
Acacia koa	Hawaii	11,001 SNPs	GBS	GF + GDM	Mean annual rainfall is the most important predictor	[75]
Quercus suber	Western Mediterranean	2,583 SNPs	GBS	BayeScan + RDA + LFMM + GF	265 EALs, majority of which associated with temperature variables	[76]
Achyranthes bidentata	China, warm-temperate zone	202 loci	SCoT	RDA + Arlequin + LFMM	23 outlier loci, 18 EALs, related to temperature and precipitation	[77]
Quercus lobata	California, US	11,019 SNPs	GBS	sNMF + LFMM + GF	536 EALs	

Within these studies, the availability of a reference genome is a powerful resource to increase the density of genomic coverage and power of landscape genomics. When a reference genome is available, landscape genomics provides additional insights into the environment-associated loci, such as their functional annotation, evolutionary trajectory, and genomic arrangement. The lack of a reference genome may also compromise the accuracy in

de novo assembly and variant calling of sequence fragments, especially for genotyping methods associated with restriction enzymes[78]. Reference genomes are available now for *Populus*, *Eucalyptus*, *Picea*, *Pinus*, and *Quercus*, among others which also have relevant landscape genomic studies[79].

Despite the very significant progress, there remains a huge gap in landscape genomic studies of tropical and angiosperm trees, and even more so in the threatened species. *Populus* and *Quercus* have been two emerging models for genomic studies in angiosperm trees, and we benefit from an understanding of the genetic architecture of their phenology, disease resistance, and landscape genomics[80]. Tropical forests harbour a higher ecological density and the evolutionary findings than temperate forests, which make up the majority of the landscape genomic studies to date, may not be congruent with the genomic underpinning of adaptation in tropical areas[81]. Some angiosperm trees (like *Populus* and *Dalbergia*) can undergo clonal reproduction effectively, which is shown to have a strong effect on local genetic structure[16,82]. There has been no landscape genomic study in the Greater Mekong Subregion to date.

1.7 Predicting the fate of forest trees in the changing environment

As discussed earlier, climate change is a substantial risk to many tree species and their conservation programmes. Knowledge of the spatial and temporal patterns of adaptive variation are crucial in developing conservation and forest management strategies as it can predict the genetic response of species and populations in the changing environment[83].

The resilience and productivity of forests in the future greatly depend on the match between genotypes and new environmental conditions. If the local environments change more rapidly than the populations can adapt genetically, they may lose genetic diversity and suffer

from local extirpation[84], and the classical preference of local germplasm for local provenance may no longer hold. Deliberate transfer of germplasm along climate gradients may be necessary[85]. Examples of such strategies include assisted migration based on predictive provenancing to facilitate adaptation of the populations under climate change[86], which can be classified into assisted gene flow and assisted colonization.

Assisted gene flow (AGF) refers to the managed translocation of individuals or germplasms among populations within the current range (i.e. *in situ*), as to facilitate more rapid adaptation to the predicted future climate, or to mitigate local maladaptation[84]. AGF may target isolated populations which have restricted gene flow with other populations and low genetic diversity. AGF aims to supply the population with individuals of higher fitness under new conditions than the original residents, thus helping the survival and productivity of the maladapted populations. In the long term, AGF should expand the genetic variation to facilitate evolutionary responses in the changing environment[87] and adaptation[88,89]. However, AGF has received most attention in temperate and boreal forests, because limited data are available on local adaptation of tropical forests[90].

Assisted colonization (AC) refers to the managed translocation of individuals or germplasms outside contemporary species ranges (i.e. *ex situ*). Unlike AGF, AC is distinctively important when the species cannot adapt to keep up with the changing climate[91]. However, many scientists point out the potential ecological risk of AC similar to invasive species, and the uncertainty of tree survival[92,93]. AC can potentially be more suitable in managed populations or production stands, especially for tree species that have high economic value for production.

Landscape genomics has shown potential in guiding *ex situ* decisions for conservation and forest restoration under climate change. The statistical framework of landscape genomics

involves quantifying the correlation of alleles and environmental variables, which can therefore be used to extrapolate future correlation. A few studies have integrated both current and future climate scenarios in their landscape genomics model. Bay et al. defined the metric "genomic vulnerability" as the mismatch between contemporary and future genomic variation using landscape genomic model and predicted the population shift and decline in yellow warblers[94]. Gugger et al.[75] predicted that changes in rainfall patterns among Hawaiian koa populations result in unmatched genotypes and environment in the future. Supple et al.[95] used both landscape genomics studies and common garden experiments to examine genomic and phenotypic variations, and made empirical based predictions for seed sourcing and restoration sites in Australian *Eucalyptus melliodora*. These studies produced spatially explicit predictive models, which are most useful because the science that underpins the predictions can be of little use to frontline practitioners and managers[67]. The findings can guide a wide range of conservation and forestation decisions, including AGF and AC.

1.8 Rationale, objectives and outline of this thesis

Dalbergia cochinchinensis and *D. oliveri* are two threatened species across their entire species' range, where understanding genetic processes and adaptive variation of *Dalbergia* have been identified as a priority research need in national evaluations[27]. Early conservation actions have been limited in scope and could not capture the range-wide diversity that may be needed to conserve these valuable species. While there are emerging efforts among these countries to reduce the pressure and restore the *in situ* populations, ensuring a sustainably supply of adaptively diverse germplasm becomes the central goal of conservation. We highlighted a number of genetic risks which may compromise the effectiveness of the conservation programme and how understanding adaptive variation may

either mitigate or prevent the risks. Common garden and reciprocal transplant experiments, while proven to be useful, are impractical for these two species: there is no existing capacity and it would be too time- and cost-consuming to establish. With the very limited *a priori* knowledge on hand, landscape genomics thus has promising potential to unravel the adaptive pattern of these two species. The knowledge will be translatable to decision-making processes in the conservation and supply chain of genetic materials, such as delineating *in situ* and *ex situ* conservation areas and predictive provenancing of assisted gene flow.

Led by the University of Oxford and funded by Darwin Initiative, forestry authorities and research organisations from Cambodia, Lao PDR, Thailand, Vietnam, and China recently launched a three-year collaborative project (2018 – 2021) to safeguard the genetic resources of three *Dalbergia* rosewood species of high conservation concern (*D. cochinchinensis*, *D. cultrata*, and *D. oliveri*), with the aims to improve capacity for seed collection, seed sourcing, and nursery management. The project will also build the capacity of rural communities for livelihood benefits from the sustainable use of the forest trees. Working with these community forestry programmes opens up a new opportunity to develop and transfer scientific knowledge to conserve both the genetic diversity and adaptability of the forests. This thesis benefits from the infrastructure of the Darwin-funded project for both acquiring samples for research and establishing routes for impact.

The overarching goal of this project is to develop and understanding of adaptive diversity in *D. cochinchinensis* and *D. oliveri* by comparing provenances across their species' range and searching for gene-environment associations.

Objective 1: To develop genetic resources for analyses of genetic diversity

Chapters 2 and 3 present a genomic information base for the genus, including reference genomes, reference transcriptomes, gene predictions, and annotations. The chapters also illustrate the utility of these genomic resources by enhancing our understanding of genome evolution in the legume family.

Objective 2: To understand the physiological underpinning of adaptation to temperature and water availability

Chapter 4 presents a controlled experiment to test and compare the physiological responses of *D. cochinchinensis* and *D. oliveri* under drought and heat stresses, the two major drivers of adaptive variation between the two species and tree mortality worldwide. The chapter provides a functional explanation of the adaptive patterns of the two species, and predicts the effect of extreme climate on the survival of their seedling establishment.

Objective 3: To detect signals of adaptation and understand the genetic basis of adaptation

Chapter 5 presents a landscape genomic study across the entire species' range to determine the influence of spatio-environmental variables on their genomic variation and to identify loci under natural selection. The chapter also extends the genomic model to incorporate future climate predictions and presents a novel web application which guides assisted germplasm transfer.

1.9 References

1. Lindenmayer, D. B. *et al.* Conservation strategies in response to rapid climate change: Australia as a case study. *Biol. Conserv.* **143**, 1587–1593 (2010).
2. Giam, X. Global biodiversity loss from tropical deforestation. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5775–5777 (2017).
3. Hansen, M. C. *et al.* High-resolution global maps of 21st-century forest cover change. *Science.* **342**, 850–853 (2013).
4. Chen, I. C., Hill, J. K., Ohlemüller, R., Roy, D. B. & Thomas, C. D. Rapid range shifts of species associated with high levels of climate warming. *Science.* **333**, 1024–1026 (2011).
5. Brady, S. P. *et al.* Causes of maladaptation. *Evol. Appl.* **12**, 1229–1242 (2019).
6. Andersen, E. J., Ali, S., Byamukama, E., Yen, Y. & Nepal, M. P. Disease resistance mechanisms in plants. *Genes.* **9**, 7 (2018).
7. Chazdon, R. & Brancalion, P. Restoring forests as a means to many ends. *Science.* **364**, 24–25 (2019).
8. Cooke, S. J. *et al.* Evidence-based restoration in the Anthropocene—from acting with purpose to acting for impact. *Restor. Ecol.* **26**, 201–205 (2018).
9. UNODC. *World Wildlife Crime Report: Trafficking in Protected Species*. (United Nations Publication, 2016).
10. Anderson, J. T. & Song, B. H. Plant adaptation to climate change—Where are we? *J. Syst. Evol.* **58**, 533–545 (2020).
11. Vatanparast, M. *et al.* First molecular phylogeny of the pantropical genus *Dalbergia*: Implications for infrageneric circumscription and biogeography. *South African J. Bot.* **89**, 143–149 (2013).
12. Winfield, K., Scott, M. & Graysn, C. Global status of *Dalbergia* and *Pterocarpus*

rosewood producing species in trade. in *Convention on International Trade in Endangered Species 17th Conference of Parties - Johannesburg* (2016).

13. EIA. *The Rosewood Racket: China's billion dollar illegal timber trade and the devastation of Nigeria's forests*. (Environmental Investigation Agency, 2017).
14. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
15. Woodruff, D. S. Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. *Biodivers. Conserv.* **19**, 919–941 (2010).
16. Hartvig, I. *et al.* Population genetic structure of the endemic rosewoods *Dalbergia cochinchinensis* and *D. oliveri* at a regional scale reflects the Indochinese landscape and life-history traits. *Ecol. Evol.* **8**, 530–545 (2018).
17. Theilade, I., Schmidt, L., Chhang, P. & McDonald, J. A. Evergreen swamp forest in Cambodia: floristic composition, ecological characteristics, and conservation status. *Nord. J. Bot.* **29**, 71–80 (2011).
18. UNODC. *World Wildlife Crime Report: Trafficking in Protected Species*. (2020).
19. Convention on International Trade in Endangered Species of Wild Fauna and Flora. CoP16 Prop. 60. https://cites.org/sites/default/files/eng/cop/16/prop/E-CoP16-Prop-60.pdf (2013).
20. EIA. *Routes of Extinction: The corruption and violence destroying Siamese rosewood in the Mekong*. (Environmental Investigation Agency, 2014).
21. United Nations Office on Drugs and Crime. The World Wildlife Seizures (World WISE) database. (2016) doi:10.18356/76DBEAB6-EN.
22. Winfield, K., Scott, M. & Grayson, C. *CITES CoP17 Information Paper - Global status of Dalbergia and Pterocarpus rosewood producing species in trade*. (2016).

23. Millet, J., Pascal, J. P. & Kiet, L. C. Effects of disturbance over 60 years on a lowland forest in southern Vietnam. *J. Trop. For. Sci.* **22**, 237–246 (2010).

24. Millet, J. & Truong, L. H. Assessment of the diversity and distribution of the threatened tree species in a logged forest in Vietnam: *http://dx.doi.org/10.1177/194008291100400108* **4**, 82–96 (2011).

25. Luoma-aho, T., Hong, L. T., Ramanatha Rao, V. & Sim, H. C. *Forest genetic resources conservation and management: Proceedings of the Asia Pacific Forest Genetic Resources Programme (APFORGEN) Inception Workshop*. (2003).

26. APFORGEN. *Conserving Rosewood genetic resources for resilient livelihoods in the Mekong - Project Inception Workshop Report*. (2018).

27. Jalonen, R., Choo, K. Y., Hong, L. T. & Sim, H. C. *Forest genetic resources conservation and management: status in seven South and Southeast Asian countries*. (2009).

28. Liu, C. L. C., Kuchma, O. & Krutovsky, K. V. Mixed-species versus monocultures in plantation forestry: Development, benefits, ecosystem services and perspectives for the future. *Glob. Ecol. Conserv.* **15**, e00419 (2018).

29. Boshier, D. *et al. Conserving Rosewood genetic diversity for resilient livelihoods in the Mekong - Darwin Initiative Application for Grant for Round 24: Stage 2*.

30. Frankham, R. *et al.* Loss of genetic diversity reduces ability to adapt. in *Genetic Management of Fragmented Animal and Plant Populations* (eds. Frankham, R. et al.) (Oxford University Press, 2017). doi:10.1093/OSO/9780198783398.003.0004.

31. Isagi, Y. *et al.* Significant loss of genetic diversity and accumulation of deleterious genetic variation in a critically endangered azalea species, *Rhododendron boninense*, growing on the Bonin Islands. *Plant Species Biol.* **35**, 166–174 (2020).

32. Jalonen, R., Valette, M., Boshier, D., Duminil, J. & Thomas, E. Forest and landscape

restoration severely constrained by a lack of attention to the quantity and quality of tree seed: Insights from a global survey. *Conserv. Lett.* **11**, e12424 (2018).

33. Kawecki, T. J. & Ebert, D. Conceptual issues in local adaptation. *Ecol. Lett.* **7**, 1225–1241 (2004).

34. Savolainen, O., Lascoux, M. & Merilä, J. Ecological genomics of local adaptation. *Nat. Rev. Genet.* **14**, 807–820 (2013).

35. Leimu, R. & Fischer, M. A Meta-Analysis of Local Adaptation in Plants. *PLoS One* **3**, e4010 (2008).

36. Derry, A. M. *et al.* Conservation through the lens of (mal)adaptation: Concepts and meta-analysis. *Evol. Appl.* **12**, 1287 (2019).

37. Boshier, D. *et al.* Is local best? Examining the evidence for local adaptation in trees and its scale. *Environ. Evid.* **4**, 1–10 (2015).

38. Saari, C. & Glisson, W. Survey of Chicago region restoration seed source policies. *Ecol. Restor.* **30**, 162–165 (2012).

39. Verkerk, P. J. *et al.* Climate-Smart Forestry: the missing link. *For. Policy Econ.* **115**, 102164 (2020).

40. Petit-Cailleux, C. *et al.* Tree mortality risks under climate change in Europe: Assessment of silviculture practices and genetic conservation networks. *Front. Ecol. Evol.* **9**, 582 (2021).

41. Lindner, M. *et al.* Climate change and European forests: What do we know, what are the uncertainties, and what are the implications for forest management? *J. Environ. Manage.* **146**, 69–83 (2014).

42. Singh, R. P., Prasad, P. V. V. & Reddy, K. R. Impacts of Changing Climate and Climate Variability on Seed Production and Seed Industry. *Adv. Agron.* **118**, 49–110 (2013).

43. IPCC. *Climate Change 2014: Mitigation of Climate Change Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. (Cambridge University Press, 2014).

44. Wilczek, A. M., Cooper, M. D., Korves, T. M. & Schmitt, J. Lagging adaptation to warming climate in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7906–7913 (2014).

45. Caignard, T. *et al.* Increasing spring temperatures favor oak seed production in temperate areas. *Sci. Reports 2017 71* **7**, 1–8 (2017).

46. Kijowska-Oberc, J., Staszak, A. M. & Ratajczak, E. Climate change affects seed aging? Initiation mechanism and consequences of loss of forest tree seed viability. *Trees.* **35**, 1099–1108 (2021).

47. Walck, J. L., Hidayati, S. N., Dixon, K. W., Thompson, K. & Poschlod, P. Climate change and plant regeneration from seed. *Glob. Chang. Biol.* **17**, 2145–2161 (2011).

48. Turesson, G. The genotypical response of the plant species to the habitat. *Hereditas* **3**, 211–350 (1922).

49. Johnson, L. C. *et al.* Reciprocal transplant gardens as gold standard to detect local adaptation in grassland species: New opportunities moving into the 21st century. *J. Ecol.* **110**, 1–18 (2021).

50. Kokko, H. *et al.* Can evolution supply what ecology demands? *Trends Ecol. Evol.* **32**, 187–197 (2017).

51. Bucharova, A. Assisted migration within species range ignores biotic interactions and lacks evidence. *Restor. Ecol.* **25**, 14–18 (2017).

52. Falk, D., Palmer, M. A. & Zedler, J. A. Foundations of restoration ecology. in *Society for ecological restoration* (Island Press, 2006).

53. Tsuyama, I., Ishizuka, W., Kitamura, K., Taneda, H. & Goto, S. Ten years of

provenance Trials and application of multivariate random forests predicted the most preferable seed source for silviculture of *Abies sachalinensis* in Hokkaido, Japan. *For.* **11**, 1058 (2020).

54. Berend, K., Haynes, K. & MacKenzie, C. M. D. Common garden experiments as a dynamic tool for ecological studies of alpine plants and communities in northeastern North America. *BioOne.* **121**, 174–212 (2019).

55. Richardson, J. L., Urban, M. C., Bolnick, D. I. & Skelly, D. K. Microgeographic adaptation and the spatial scale of evolution. *Trends Ecol. Evol.* **29**, 165–176 (2014).

56. Dyer, R. J. Is there such a thing as landscape genetics? *Mol. Ecol.* **24**, 3518–3528 (2015).

57. Joost, S. *et al.* A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol. Ecol.* **16**, 3955–3969 (2007).

58. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).

59. Lotterhos, K. E. & Whitlock, M. C. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol. Ecol.* **24**, 1031–1046 (2015).

60. Sork, V. L. *et al.* Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genet. Genomes* **9**, 901–911 (2013).

61. Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T. & Curtis-McLane, S. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol. Appl.* **1**, 95–111 (2008).

62. Balkenhol, N. *et al.* Landscape Genomics: Understanding Relationships Between

Environmental Heterogeneity and Genomic Characteristics of Populations. in *Population Genomics* (ed. Rajora, O.) 261–322 (Springer, Cham, 2017).

63. Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M. & Holderegger, R. A practical guide to environmental association analysis in landscape genomics. *Mol. Ecol.* **24**, 4348–4370 (2015).

64. Frichot, E., Schoville, S. D., Bouchard, G. & François, O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol. Biol. Evol.* **30**, 1687–1699 (2013).

65. Ellis, N., Smith, S. J. & Pitcher, C. R. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**, 156–168 (2012).

66. Miao, C.-Y., Li, Y., Yang, J. & Mao, R.-L. Landscape genomics reveal that ecological character determines adaptation: a case study in smoke tree (*Cotinus coggygria* Scop.). *BMC Evol. Biol.* **17**, 202 (2017).

67. Martins, K. *et al.* Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of *Quercus rugosa*. *Evol. Appl.* **11**, 1842–1858 (2018).

68. Pais, A. L., Whetten, R. W. & Xiang, Q. Y. Population structure, landscape genomics, and genetic signatures of adaptation to exotic disease pressure in *Cornus florida* L.—Insights from GWAS and GBS data. *J. Syst. Evol.* **58**, 546–570 (2020).

69. Jia, K. H. *et al.* Landscape genomics predicts climate change-related genetic offset for the widespread *Platycladus orientalis* (Cupressaceae). *Evol. Appl.* **13**, 665–676 (2020).

70. Lu, M., Loopstra, C. A. & Krutovsky, K. V. Detecting the genetic basis of local adaptation in loblolly pine (*Pinus taeda* L.) using whole exome-wide genotyping and an integrative landscape genomics analysis approach. *Ecol. Evol.* **9**, 6798–6809 (2019).

71. Li, J. X. *et al.* Adaptive genetic differentiation in *Pterocarya stenoptera* (Juglandaceae) driven by multiple environmental variables were revealed by landscape genomics. *BMC Plant Biol.* **18**, 1–12 (2018).

72. Cao, Y. N. *et al.* Inferring spatial patterns and drivers of population divergence of *Neolitsea sericea* (Lauraceae), based on molecular phylogeography and landscape genomics. *Mol. Phylogenet. Evol.* **126**, 162–172 (2018).

73. Li, L. F. *et al.* Landscape genomics reveals genetic evidence of local adaptation in a widespread tree, the Chinese wingnut (*Pterocarya stenoptera*). *J. Syst. Evol.* (2020) doi:10.1111/JSE.12699/SUPPINFO.

74. Mastretta-Yanes, A. *et al.* Long-term in situ persistence of biodiversity in tropical sky islands revealed by landscape genomics. *Mol. Ecol.* **27**, 432–448 (2018).

75. Gugger, P. F., Liang, C. T., Sork, V. L., Hodgskiss, P. & Wright, J. W. Applying landscape genomic tools to forest management and restoration of Hawaiian koa (*Acacia koa*) in a changing environment. *Evol. Appl.* **11**, 231–242 (2018).

76. Vanhove, M. *et al.* Using gradient Forest to predict climate response and adaptation in Cork oak. *J. Evol. Biol.* **34**, 910–923 (2021).

77. Yang, J., Li, Y., Miao, C. Y. & Mao, R. L. Landscape genomics analysis of *Achyranthes bidentata* reveal adaptive genetic variations are driven by environmental variations relating to ecological habit. *Popul. Ecol.* **59**, 355–362 (2017).

78. LaCava, M. E. F. *et al.* Accuracy of de novo assembly of DNA sequences from double-digest libraries varies substantially among software. *Mol. Ecol. Resour.* **20**, 360–370 (2020).

79. Borevitz, J. O., Bragg, J. G., Supple, M. A. & Andrew, R. L. Tansley review Genomic variation across landscapes: insights and applications. doi:10.1111/nph.13410.

80. Porth, I. *et al.* Evolutionary quantitative genomics of *Populus trichocarpa*. *PLoS One*

10, e0142864 (2015).

81. Fetter, K. C. *et al.* Landscape genomics of angiosperm trees: From historic roots to discovering new branches of adaptive evolution. in *Comparative and Evolutionary Genomics of Angiosperm Trees* (eds. Groover, A. & Cronk, Q.) vol. 21 303–333 (Springer, Cham, 2017).
82. Mock, K. E., Rowe, C. A., Hooten, M. B., Dewoody, J. & Hipkins, V. D. Clonal dynamics in western North American aspen (*Populus tremuloides*). *Mol. Ecol.* **17**, 4827–4844 (2008).
83. Valladares, F. *et al.* The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. *Ecol. Lett.* **17**, 1351–1364 (2014).
84. Aitken, S. N. & Whitlock, M. C. Assisted gene flow to facilitate local adaptation to climate change. *Annu. Rev. Ecol. Evol. Syst.* **44**, 367–388 (2013).
85. Bozzano, M. *et al. Genetic Considerations in Ecosystem Restoration Using Native Tree Species*. (FAO and Bioversity International, 2014).
86. Chmura, D. J. *et al.* Forest responses to climate change in the northwestern United States: ecophysiological foundations for adaptive management. *For. Ecol. Manage.* **261**, 1121–1142 (2011).
87. Wadgymar, S. M., Cumming, M. N. & Weis, A. E. The success of assisted colonization and assisted gene flow depends on phenology. *Glob. Chang. Biol.* **21**, 3786–3799 (2015).
88. Broadhurst, L. M. *et al.* Seed supply for broadscale restoration: maximizing evolutionary potential. *Evol. Appl.* **1**, 587–597 (2008).
89. Milesi, P. *et al.* Assessing the potential for assisted gene flow using past introduction of Norway spruce in Southern Sweden: Local adaptation and genetic basis of

quantitative traits in trees. *bioRxiv* 481614 (2018) doi:10.1101/481614.

90. Aitken, S. N. & Bemmels, J. B. Time to get moving: assisted gene flow of forest trees. *Evol. Appl.* **9**, 271–290 (2016).

91. Loarie, S. R. *et al.* The velocity of climate change. *Nature* **462**, 1052–1055 (2009).

92. Kreyling, J. *et al.* Assisted Colonization: A question of focal units and recipient localities. *Restor. Ecol.* **19**, 433–440 (2011).

93. Ricciardi, A. & Simberloff, D. Assisted colonization is not a viable conservation strategy. *Trends Ecol. Evol.* **24**, 248–253 (2009).

94. Bay, R. A. *et al.* Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **359**, 83–86 (2018).

95. Supple, M. A. *et al.* Landscape genomic prediction for restoration of a *Eucalyptus* foundation species under climate change. *Elife* **7**, e31835 (2018).

Chapter 2. Reference transcriptomes and comparative analyses of six species in the threatened rosewood genus *Dalbergia*

Tin Hang Hung[1,*], Thea So[2], Syneath Sreng[2], Bansa Thammavong[3], Chaloun Boounithiphonh[3], David H. Boshier[1], John J. MacKay[1,*]

1. Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, United Kingdom
2. Institute of Forest and Wildlife Research and Development, Cambodia
3. Forest Research Center, National Agriculture and Forestry Research Institute, Lao PDR

2.1 Abstract

Dalbergia is a pantropical genus with more than 250 species, many of which are highly threatened due to overexploitation for their rosewood timber, along with general deforestation. Many *Dalbergia* species have received international attention for conservation, but the lack of genomic resources for *Dalbergia* hinder evolutionary studies and conservation applications, which are important for adaptive management. This study produced the first reference transcriptome for six *Dalbergia* species with different geographical origins and predicted ~32 K to 49 K unique genes. We showed the utility of these transcriptomes by phylogenomic analyses with other Fabaceae species, estimating the divergence time of extant *Dalbergia* species to ~14.78 MYA. We detected over-representation in 13 Pfam terms including HSP, ALDH and ubiquitin families in *Dalbergia*. We also compared the gene families of geographically co-occurring *D. cochinchinensis* and *D. oliveri* and observed that more genes underwent positive selection and there were more diverged disease resistance proteins in the more widely distributed *D. oliveri*, consistent with reports that it occupies a wider ecological niche and has higher genetic diversity. We anticipate that the reference transcriptomes will facilitate future population genomics and gene-environment association studies on *Dalbergia*, as well as contributing to the genomic database where plants, particularly threatened ones, are currently underrepresented.

2.2 Introduction

The genus *Dalbergia* L. f. (Fabaceae: Faboideae) contains around 250 species, many of which are globally recognized for their economic value. *Dalbergia* species encompass a high diversity in their life histories and morphologies as trees, shrubs, and woody lianas[1]. They are distributed pantropically across Central and South Americas, Africa, and Asia[2]. More than 50 *Dalbergia* species are documented to have the ability to fix atmospheric nitrogen with possession of aeschynomenoid type root nodules[3]. Many *Dalbergia* species produce valuable heartwood timber known as rosewood, and are incorporated in a wide range of uses including furniture, boats, and musical instruments[4]. They are often targeted in illegal harvesting and traded in local and global markets with little regulation either in Asia (including the Indochina biodiversity hotspot) or Africa (particularly in Madagascar)[5,6]. Due to overexploitation of their timber, population sizes and areas within their native distribution have significantly diminished[7]. The genus *Dalbergia* is declared as threatened worldwide, with many species classified as endangered or vulnerable in the International Union for Conservation of Nature (IUCN) Red List. The whole genus of *Dalbergia* was listed in the Convention on International Trade in Endangered Species (CITES) Appendix I or II in 2017 to regulate the international trade of *Dalbergia* timber.

Studies in the evolutionary history and genetic resources of *Dalbergia* are still scarce. Genetic markers have been developed for a number of *Dalbergia* species and used in studies of evolutionary history and for conservation. The earliest complete report on infrageneric taxonomy of *Dalbergia* was published by Bentham[8], and the first molecular phylogeny recently supported the monophyletic nature of *Dalbergia* genus, grouped in a clade with other genera including *Machaerium, Aeschynomene*, and *Ormocarpum*[1]. In earlier studies the *Dalbergia* clade was assigned to the Dalbergieae tribe with *Adesmia* and *Pterocarpus* clades[9]. Recent studies utilise genetic markers to infer the phylogeography of populations and identify

landscape features which may explain the population structure[10]. A number of DNA-based barcodes have also been developed that may be used in conservation forensics to track illegal trade and verify species identification[11]. These *Dalbergia* studies have mainly analysed loci such as *rbcL*, *matK*[4,12], *trnL*, and *psbA-trnH*[13] at species level, and microsatellites[10,14,15] at the infraspecific level. Although recent advances in high-throughput sequencing have expanded the assembly repertoire of many species, genomic resources for the genus *Dalbergia* remain scarce for such a big genus: namely one *de novo* transcriptome assembly of *D. odorifera*[16] (without a gene annotation report), and ten chloroplast genomes[17–21].

The genomic resource gap potentially hinders the understanding of evolutionary history in *Dalbergia* and the application of genetic tools in conservation. For example, *D. cochinchinensis* and *D. oliveri* are commonly found in the same geographical localities in South Eastern Asia, but they have significantly different neutral genetic structure[10]. Understanding their adaptive differences using genome-wide analyses would help devise potentially different conservation strategies. Due to the lack of a reference genome for any of the *Dalbergia* species, transcriptomes can be a practical starting point to facilitate evolutionary research and conservation applications. High-throughput sequencing technologies for RNA-seq enable gene prediction and annotation for non-model organisms with scarce genomic information[22].

In this study, we develop a resource and knowledge base to facilitate transferability and utility across the genus. We produced the first reference transcriptomes from *de novo* assemblies for six diverse *Dalbergia* species, including *D. cochinchinensis* Pierre, *D. frutescens* (Vell.) Britton, *D. melanoxylon* Guill. & Perr., *D. miscolobium* Benth., *D. oliveri* Gamble ex Prain, and *D. sissoo* Roxb. ex DC. (Table 2.1). For gene annotation, we used *ab-initio* gene prediction based on the structure of open reading frames, features of protein-coding genes, and sequence homology to gene models of closely related species [23]. To

demonstrate the utility of the transcriptomic resources, we conducted phylogenomic, gene enrichment, and selection analyses comparing the *Dalbergia* and other Fabaceae species.

2.3 Methods

2.3.1 Ethics statement

Dalbergia cochinchinensis and *D. oliveri* are vulnerable and threatened respectively. All *Dalbergia* species are listed in the CITES Appendix II, albeit their seeds are exempted according to Annotation #15. The seed collections of *D. cochinchinensis* and *D. oliveri* were made by local government authorities with permissions and licences in place.

2.3.2 Plant materials and sample preparation

Dried seeds of *Dalbergia cochinchinensis*, *D. frutescens*, *D. melanoxylon*, *D. miscolobium*, *D. oliveri*, and *D. sissoo* were obtained from different sources (Table 2.2) and stored at 4°C until seed germination. The seeds were scarified by placing them in 70°C distilled water, which was then left to cool to room temperature for 1 hour, with the seed soaking in the water for 24 hours. The seeds were germinated in 1% agar in a plant growth cabinet MLR-350 (Sanyo, Watford, United Kingdom) at 25°C and photoperiod 12L/12D. Seedlings were transferred to small pots in a soil-perlite 3:1 (v:v) mixture in the same growth cabinet. The plants were watered to pot capacity, with any moulded or diseased plants immediately removed. After plant height reached a minimum of 10 cm, four plants of each species were randomly selected. Two plants were drought-stressed until soil gravimetric water content dropped below 50%, while the other two were watered as usual. Three tissues (foliage, stem, and root) were harvested from each individual and their total RNA extracted (n = 72) with Monarch Total RNA Miniprep Kit (New England BioLabs, United Kingdom). Multiple tissue types and growth conditions increased the diversity of transcripts towards a more-complete transcriptome[24]. The quantity and quality of total RNA from each sample were determined with NanoDrop 2000 (Thermo, Wilmington, United States). RNA integrity was assessed with the RNA 6000 Nano Assay on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, United States) and RNA samples with a minimum RNA integrity number (RIN)

of 7 (for leaf tissues) and 8 (for root and stem tissues) were retained for RNA-Seq. Samples of the same species were pooled to equimolarity.

2.3.3 Library preparation and sequencing

RNA samples (n = 6) were sent to the Oxford Genomics Centre (Oxford, United Kingdom) for library preparation and sequencing. Polyadenylated transcript enrichment was performed with NEBNext Poly(A) mRNA Magnetic Isolation Module (New England BioLabs), and then individual libraries were prepared with NEBNext Ultra II Directional RNA Library Prep Kit (New England BioLabs). Libraries were amplified on a Tetrad (Bio-Rad) using in-house unique dual indexing primers[25]. Individual libraries were normalised and their size profiles were analysed on the 2200 or 4200 TapeStation (Agilent, RNA ScreenTape). The pooled library was diluted to ~10 nM for storage. The 10 nM library was denatured and further diluted prior to loading on the sequencer. Paired-end sequencing was performed on the HiSeq4000 (Illumina, HiSeq3000/4000 PE Cluster Kit and 150 cycle SBS Kit) with a read length of 150 bp. The raw reads were obtained in fastq files after an in-house preliminary quality check.

2.3.4 Data filtering and de novo assembly

Quality of raw reads was examined using FastQC v0.11.8 and visualized in MultiQC v1.7[26]. Scythe v0.994[27] was used to trim the 3'-end adapter contaminants and Sickle v1.33[28] was used to remove the low-quality reads (Phred quality score < 30). Filtered reads were assessed again with FastQC. As no reference genome was available for the genus *Dalbergia*, we assembled the transcriptomes *de novo*, to avoid the bias that may be introduced by using other species in genome-guided assembly[29]. The filtered reads for each species were *de novo* assembled using Trinity v2.8.4[30] with the default parameters. The assembly and subsequent steps were performed on the University of Oxford Advanced Researching Computing

ARCUS-B cluster. The schematic bioinformatic pipeline of the transcriptome assembly is shown in Figure 2.1.

2.3.5 Assembly quality assessment and optimization

As a first quality assessment, we generated the output statistics of the initial individual *de novo* assemblies with Trinity scripts. We then assessed the read content of the transcriptome assembly for each species by mapping the clean reads to the assembly using Bowtie2 v.2.3.4.3[31] with the options "-p 10 -q --no-unal -k 20", as suggested in the Trinity package.

Optimizations were carried out to improve the performance and accuracy of downstream analyses, as *de novo* assembly often produces highly similar transcript sequences such as isoforms or assembly artefacts. First, we reduced the redundancy of transcripts with CD-HIT-EST v4.8.1[32] by removing transcripts with sequence similarity greater than 95%. Then we estimated candidate coding regions within transcript sequences with TransDecoder v5.5.0[33] to identify the single best predicted open reading frames (ORF) that are at least 100 amino acids long (parameter --single_best_only). Each transcript was represented by the longest translated protein sequence and each gene by the longest transcript in the final assembly.

We compared the transcripts in the final assembly against the OrthoDB v10 eudicotyledons database[34] with BUSCO (Benchmarking Universal Single-Copy Orthologs) v3.0[35] to evaluate the assembly completeness. For full-length transcript analysis, we performed BLASTP searches (--evalue 1e-3) on the non-redundant transcripts against the RefSeq protein data of *Arachis ipaensis* (NCBI: GCF_000816755.2 Araip1.1[36]), which represented the closest relative to *Dalbergia* with an available annotated genome[36]. We then calculated the coverage of aligned transcripts based on their BLAST hits with 'analyze_blastPlus_topHit_coverage.pl' script in the Trinity package. We also used

TransRate v1.0.3[37] to obtain the Conditional Reciprocal Best BLAST (CRBB) and coverage metrics of final assemblies using Araip1.1 as a reference.

2.3.6 Structural and functional annotation

We aligned our final assemblies against the SwissProt database[38], Araip1.1[36], and the *Arabidopsis thaliana* database (Araport11)[39] with BLASTP for best hits with an e-value below the threshold 10^{-3}. We then annotated the protein domains with HMMER v3.2.1 (http://hmmer.org) on the Pfam 32.0 database (version September 2018, 17,929 entries)[40]. We also predicted signal peptides using SignalP 5.0[41] and transmembrane domains using TMHMM 2.0[42]. We finally loaded the blast homologies of three databases (SwissProt, Araport11, and Araip1.1) into an SQLite database and generated the annotation report for each species assembly with Trinotate v3.3.1. GO (Gene Ontology), KEGG (Kyoto Encyclopedia of Genes and Genomes), and COG (Clusters of Orthologous Groups) assignments were transferred from SwissProt annotations as a verified source.

2.3.7 Phylogenomic analysis and estimation of divergence time

We ran OrthoFinder v2.2[43] on the six *Dalbergia* transcriptomes in this study and 10 other Fabaceae species (Table 2.3). After the analysis, only single-copy orthologs among taxa were retrieved as they were the most robust for phylogenetic reconstruction with high confidence and concordance[44]. We performed multiple sequence alignment for each set of single-copy orthologs using MAFFT v7[45], and every corresponding coding sequence was retrieved and matched to ortholog alignment with PAL2NAL v14[46]. Coding sequences of all ortholog alignments were concatenated to create a single multiple sequence alignment (https://github.com/nylander/catfasta2phyml).

The nucleotide substitution model was tested on the concatenated alignment with jModelTest 2.1.10[47] for likelihood scores. The alignment was then used to construct a best-fit

(i.e. GTR + Γ + I) maximum likelihood phylogenetic tree using RAxML (Randomized Axelerated Maximum Likelihood) 8.2.12[48] with 100 rapid bootstrapping. The maximum likelihood (ML) tree was used as a starting tree in both the Bayesian phylogenetic analysis and subsequently in the gene family analysis.

We estimated the species divergence time with BEAST (Bayesian evolutionary analysis by sampling trees) v.2.5.2[49] using a calibrated birth-death model with an uncorrelated lognormal relaxed clock (ULRC). The crown age of the tree (Fabaceae) was calibrated to the oldest definitive legume fossil (wood of *Paracacioxylon frenguellii*) at 63.5 million years ago (MYA)[50]. The crown age of Faboideae was calibrated to 56.3 ± 1.05 MYA[51] and that of the Dalbergoid clade (*Nissolia-Dalbergia* split) was calibrated to 50.7 ± 0.8 MYA[52]. The time of the *A. duranensis-A. ipaensis* split was calibrated to 2.88 ± 0.22 MYA[53]. All nodes were calibrated to normal models and their sigma values estimated *a priori*. We ran 15,000,000 iterations with 150,000 burn-ins for the Monte Carlo Markov chain and also ran 15,000 trees with 10% burn-ins to produce the maximum clade credibility tree.

2.3.8 Enrichment analysis and gene family evolution

Acrocarpus fraxinifolius, *Bauhinia tomentosa*, and *Xanthocercis zambesiaca* were excluded from the subsequent Pfam and CAFE (Computational Analysis of gene Family Evolution) analyses as their BUSCO scores were not reported in their original studies, and incomplete transcriptomes could introduce bias to the enrichment and gene family analyses.

Gene annotations of the *Dalbergia* species from the Trinotate pipeline were subject to enrichment analyses. First, the annotated GO terms were extracted and searched against the WEGO (Web Gene Ontology Annotation Plot) 2.0 database[54] (version 1 November 2018) to count the level-2 GO terms for each of the *Dalbergia* species. A chi-square test of independence was conducted to detect under- and over-represented GO terms among the

species and significant terms were presented in chord diagrams (https://github.com/mattflor/chorddiag). Second, the annotated Pfam domains were extracted for each species and under- and over-represented Pfam terms were determined using a two-tailed Fisher's exact test. The mean Pfam domain counts in *Dalbergia* were compared against the background domain counts of the other Fabaceae species. Row-Z scores for each significant Pfam domain were used to construct a heatmap in R version 3.6.3.

We applied CAFE version 3.1[55] based on a Bayesian method to detect gene family contraction/expansion events, where a gene family is defined as the orthogroup clustered in the previous OrthoFinder pipelines. We used the ultrametric tree resulting from the Bayesian phylogenetic analysis to time-calibrate the gene trees. For each orthogroup we computed the family-wide p-value and branch-specific p-value (using the Viterbi method) to test the significance of a contraction/expansion event at a specific branch. As recommended by the software developers, only orthogroups with a family-wide p-value < 0.05 and a branch Viterbi p-value < 0.001 were considered significant. We then used PANTHER version 15.0[56] to detect significant over-/under-represented GO terms ($p < 0.05$ after Benjamini and Hochberg correction) of biological functions in the significantly expanded gene families after CAFE analysis.

2.3.9 Positive selection analysis

Single-copy orthologs of the 6 *Dalbergia* species were extracted using the Orthofinder pipeline. The rooted trees for each set of orthologs obtained from RAxML were used to support the evolutionary relationship of the species, while gene signatures of positive selection along a specific branch were detected by branch-site models in the codeml function of PAML (Phylogenetic Analysis by Maximum Likelihood) 4.9[57]. We set *D. cochinchinensis* and *D. oliveri*, which show overlapping ranges in South Eastern Asia[10], as the foreground phylogeny and other species as the background phylogeny in the branch-site model. We built

the alternative model (i.e. the foreground phylogeny has genes under positive selection) for each ortholog with the codeml setting: model = 2, NSites = 2, fix_kappa = 0, fix_omega = 0, omega = 1; and the null model (i.e. the foreground phylogeny has genes under neutral selection compared to the background phylogeny) with the codeml setting: model = 2, NSsites = 2, fix_kappa = 0, fix_omega = 1 and omega = 1. Sites under positive selection were defined as those with higher nonsynonymous-to-synonymous substitution ratios (d_N/d_S) > 1, as expected under neutral evolution. The two hypothetical models were tested for likelihood ratio using a chi-squared distribution with one degree of freedom, following the Benjamini and Hochberg method to correct for the significance level[58]. We determined the positively selected genes as those with corrected $p < 0.1$[59]. KEGG pathway and module enrichment tests were performed on positively selected genes using enrichKEGG and enrichMKEGG functions in clusterProfiler v3.0.4[60] respectively, with *Arachis ipaensis* set as the reference organism.

2.4 Results

2.4.1 RNA-seq library construction and sequencing

Total RNA was successfully extracted from leaf, stem and root tissues of each of 6 *Dalbergia* species and the RNA integrity numbers (RIN) of the RNA pools were all above 7.0. HiSeq4000 multiplex sequencing yielded between 71 to 180 million paired end reads of 150 bp length for each of the 6 *Dalbergia* species (Table 2.4). After quality filtering and trimming, more than 90% of the reads were retained with quality scores ≥ 30. The raw read data from Illumina sequencing for each species are deposited in the NCBI Sequence Read Archive (SRR: SRR10592611–SRR10592618) under BioProject PRJNA593817.

2.4.2 De novo transcriptome assembly and transcript filtering

The number of transcripts in initial *de novo* assemblies from Trinity ranged between 195,268 and 376,014 (see Table 2.4 for assembly statistics). As the first step of assembly quality assessment, we successfully mapped 86–90% of the raw filtered reads to individual assemblies, where an alignment rate above 80% indicates a good quality assembly[61].

Redundant transcripts were identified by clustering similar transcripts and open reading frame prediction to produce the final assemblies (Figure 2.1), which filtered roughly 65–75% of the transcripts. In the final assemblies, 67,379 to 92,906 transcripts were captured for individual species, and predicted to correspond to 31,678 to 48,591 unique genes. The final assemblies are deposited in the NCBI Transcript Shotgun Archive (GIHP00000000–GIHU00000000).

The BUSCO procedure confirmed that the majority of eudicot core genes were captured in our transcriptomes indicating high completeness of our transcriptome assemblies. Search for the 2,121 orthologs recovered over 90% of complete BUSCOs in all of our assemblies with fewer than 5% of BUSCOs missing (Table 2.5).

We mapped our transcripts to gene models of *Arachis ipaensis*, with near full-length and fragmented transcripts defined as > 70% and < 30% coverage respectively. We found that roughly 80% of the transcripts were near full-length for all transcriptomes, with only 5–8% of fragmented transcripts (Figure 2.2). There was no evidence for mapping bias among the species when comparing the counts of full-length and fragmented transcripts among our transcriptomes ($p > 0.05$, chi-square test of independence). The TransRate analysis returned a high mean percentage of contigs covered by the ORF (> 99.7% for all assemblies) and a rather low coverage on the *A. ipaensis* reference (~ 34.1% for all assemblies) (Table 2.7). However, the reference coverage depends significantly on the evolutionary distance between the assembled and reference species[37].

2.4.3 Structural and functional annotation

We annotated the *Dalbergia* transcriptome assemblies by using multiple sources and methods to provide a complete set of annotations for each species. We separated the annotations for our full transcriptome assemblies, which contained isoforms from alternative splicing as predicted in the Trinity pipeline and the gene set, which only contained the longest isoform representing each gene. The homology search on *Arachis ipaensis*, *Arabidopsis thaliana*, and SwissProt annotated 69.8–88.9% of the transcripts and 63.9–83.1% of the genes, depending on the *Dalbergia* species. We also identified protein domains (as Pfam terms) on 59.8–69.8% of the genes, transmembrane domains on 16.7–20.2% of the genes, and signal peptides on 6.3–7.7% of the genes. GO, KEGG and EggNOG assignments were transferred from SwissProt/UniProtKB annotations. The annotation report for each species assembly is available (https://doi.org/10.1038/s41598-020-74814-2#Sec25), and the annotation statistics for the transcriptomes are shown in Table 2.6.

2.4.4 Phylogenomic analysis and estimation of divergence time

Analysis using Orthofinder assigned 481,614 genes (84.7% of total genes) in our six *Dalbergia* and 10 other Fabaceae transcriptomes into 34,725 orthogroups (Table 2.8). All species present shared 5,493 orthogroups but only 256 orthogroups contained single-copy genes. The *Dalbergia* species shared 13,149 orthogroups (Figure 2.3). A Bayesian phylogenetic tree constructed using these 256 single-copy orthologs, with a total aligned length of 479,064 bp, supported the monophyly of *Dalbergia* species in the present study and showed the expected relationship of *Dalbergia* species with other major Fabaceae groups (Figure 2.4).

Using the multiple fossil calibration nodes in Fabaceae, we estimated the divergence time of extant members of the genus *Dalbergia* to be around 14.78 MYA (95% HPD: 13.74 – 16.02). The divergence times of other branches are shown in Table 2.9.

2.4.5 Enrichment analyses and gene family evolution

GO enrichment analyses revealed significant differences for GO categories of cellular components, biological processes, and molecular functions among *Dalbergia* species (Table 2.10 & Figure 2.5; $p < 0.05$, chi-square test of independence). In most categories, *D. frutescens* and *D. oliveri* had the most GO term counts, whereas *D. miscolobium* and *D. sissoo* had the fewest counts. The pattern of GO term count reflected the number of genes predicted in the assemblies, where *D. frutescens* had the highest number of genes (49,050) and *D. miscolobium* the lowest (32,107).

We conducted enrichment analyses on the Pfam protein domains to determine over- or under-represented specific groups of genes between *Dalbergia* species and other Fabaceae species (Table 2.11 & Figure 2.6; $p < 0.05$, two-tailed Fisher's exact test). While we reported a list of under-represented protein domains in *Dalbergia* species, we were cautious about the completeness of our transcriptome assemblies, owing to the samples only including juvenile

stage vegetative tissues. We focused on the 13 protein domains that were over-represented in our *Dalbergia* study species. These included two heat shock proteins Hsp70 and Hsp90 (PF00012.20 and PF00183.18), ubiquitin-related proteins (PF13881.6, PF11976.8, PF14560.7, and PF00240.23), aldehyde dehydrogenase family (PF00171.22), ribosomal proteins (PF01248.26 and PF00428.19), KOW motif (PF00467.29), elongation factor (PF03143.17), actin (PF00022.19), and leucine rich repeats (PF12799.7).

To detect the local scale of gene family expansion/contraction events in *D. cochinchinensis* and *D. oliveri*, CAFE analysis revealed 10 and 49 orthogroups that significantly expanded respectively (family-wide p-value < 0.05, branch Viterbi p-value < 0.001; Table 2.12). GO enrichment analysis revealed many over-represented terms (BH p < 0.05, two-tailed Fisher's exact test; Table 2.13 in these significantly expanded gene families, including innate immune response (GO:0045087) and defence response (GO:0006952).

2.4.6 Positive selection analysis

A total of 9,054 single-copy orthologs were identified among the 6 *Dalbergia* species using Orthofinder. A branch-site model, based on their *dN/dS*, detected 371 and 439 positively selected genes for *D. cochinchinensis* and *D. oliveri* respectively (BH p < 0.05, chi-square test of independence, Table 2.14). KEGG and GO vocabularies were searched on these positively selected genes for individual species to better summarise their biological annotations. The GO enrichment test showed a significant difference between the two species in 20 level-6 GO terms (Figure 2.7; p < 0.05, chi-square test of independence), with a majority of GO terms attributed to molecular function and related to binding. We detected no KEGG pathway or module showing a differential representation between these two species.

2.5 Discussion

We produced six *Dalbergia* transcriptome assemblies estimated to each contain 32K to 49K unique genes. Assessments of assembly completeness and quality suggested that they are suitable for molecular and evolutionary analyses and afford fair comparisons as presented in this study. Here, we discuss insights gained from data analyses with relevance to growth habit, divergence time and phylogeny, gene families, positive selection, and potential conservation implications.

2.5.1 Transcriptome assembly statistics

Genome size variation has been an important character in the evolution of higher plants, and may be accompanied in some cases by substantial changes in the number of genes[62]. No genome has been published for the genus *Dalbergia*, but previous cytophotometry estimated that *Dalbergia* species have genome sizes ranging from 1.43–1.98 Gb, while *Dalbergia* is an exclusively diploid genus with 2n = 20 chromosomes[63]. Cytophotometry results also indicated a larger DNA content in climber or liana *Dalbergia* species than the congeneric tree species. A similar tree-liana evolution trend has been suggested in other woody angiosperm taxa[64,65]. A meta-study on 6,949 angiosperms also confirmed that lianas generally have a larger DNA content[66]. In our study, *D. frustescens* was the only liana while others were all tree species. *D. frutescens* had the largest number of genes in its transcriptome, and was the most recently evolved, according to the most recent molecular phylogeny[1]. Both previous cytophotometry results and our transcriptome statistics suggest that the climbing character in *Dalbergia* may have derived from non-climbing tree ancestors, accompanied by both a larger genome size and an increased gene number. The expansion of gene families in lianas may underpin adaptations such as stem flexibility and vascular transport by adapted, derived secondary growth and wider vessel elements[67]. However, our study is limited by the number and choice of species, and we believe that

studying more species in this large genus will give better insights into the tree-climber relationship.

2.5.2 Phylogenomics and divergence time estimation

Molecular phylogenies have suggested *Dalbergia* is a monophyletic group placed in the *Dalbergia* clade with its sister taxon *Machaerium*[9,68]. The estimated age of MRCA of *Machaerium copote* and *Dalbergia congestiflora* was 40.4–43.0 MYA[52]. The most recent and comprehensive molecular phylogeny research in *Dalbergia* suggested *D. miscolobium* as the basal group among extant members[1], but species divergence time in *Dalbergia* is unstudied to date. Using transcriptome resources and fossil calibrations from other Fabaceae species, we estimated the time of divergence of extant *Dalbergia* species to be around 14.78 MYA (Miocene-Langhian). Our estimation was slightly out of previously estimated ranges (Vantaparaset *et al.*[1]: 3.8–12.7 MYA and Lavin *et al.*[69]: 7–12.2 MYA) based on single or a few loci. While most other fossil records of extinct members date to the Miocene (†*D. nostratum*: Lower Miocene 15.97–23.03 MYA[70]; †*D. lucida*: Late Miocene 5.33–11.61 MYA[71]), the earliest fossil record of †*D. phleboptera* was found in a Chattien (27.82–23.02 MYA) deposit[72], which would suggest an earlier origin of the *Dalbergia* genus. However, the morphological details of extinct *Dalbergia* species were not well described from fossils and thus their placement within the genus *Dalbergia* could not be confirmed. Therefore, in our study, these *Dalbergia* fossils were not useful in node calibration to determine the actual divergence time of *Dalbergia*. We believe our *Dalbergia* crown age estimation would at least be useful in providing a minimum bound when phylogenomic information of other *Dalbergia* species becomes available.

The colonisation of *D. cochinchinensis* and *D. oliveri* in the Indochina biodiversity hotspot was estimated to occur ~11.68 MYA (Lower Miocene), coinciding with rapid *in-situ*

diversification events and migrations after the Thai-Malay Peninsula split into Indochina and Borneo at ~15 MYA[73], leading to Indochina's diverse biota.

Divergence time for legumes was estimated to be ~80.16 MYA in this study, which falls within the most recent estimate of its marginal age prior (79.37–109.20 MYA)[74]. The difficulty in accurate divergence time estimation is proposed to be due to both whole genome duplication events near the root, intertwining with extinction and speciation events[74].

2.5.3 Comparative analysis of gene families between Dalbergia and other Fabaceae members

Eukaryotes share a large uniform set of conserved orthologs which encode for essential functional domains, such as DNA replication and repair, stress response, and secretion, and are based on the same genomic architeture[75]. The expansion and contraction of core orthologs contribute to eukaryotic diversity and enable individual species adaptation to their environment[76]. New genes may develop and result in the partitioning of gene function (subfunctionalisation) or the acquisition of new function (neofunctionalisation)[77]. For comparative genomic analyses of lineage-specific expansions and contractions, we used Pfam and CAFE analyses. The former tends to cluster protein into larger gene families, while the latter produces a finer clustering[59].

Our Pfam analysis revealed expanded gene families in *Dalbergia* species compared to other Fabaceae members with potential biological relevance to their adaptive significance. For example, HSP70 and HSP90 heat shock proteins are molecular chaperones important for protein folding that enable active response to different stresses in plants such as heat, drought, pH and hypoxia via different signalling transduction pathways[78,79]. The protection against prolonged heat stress and acute heat shock by these chaperones has enabled heat acclimatization in *Arabidopsis thaliana*[80], such as via stomatal control and abscisic acid signalling[81]. The expansion of HSPs in *Dalbergia* species may enhance their tolerance of

higher temperatures across their pan-tropical range. Another significantly expanded protein family in the *Dalbergia* genus is the aldehyde dehydrogenase (ALDH) superfamily. ALDH is highly conserved in many metabolic pathways in higher plants and plays a significant role in aldehyde homoeostasis and redox balance[82], such as in photorespiration and nitrate assimilation[83]. Increase in ALDH activity is shown to correlate with higher energy production, which fosters faster coleoptile elongation and seedling survival[84]. Many plant ALDH genes are also known to respond to a diversity of stresses including dehydration, heavy metals, salinity, and others[85]. Finally, several ubiquitin-related terms are over-represented in the *Dalbergia* genus. The best-characterised functions of ubiquitin proteins (Ub) are regulation of targeted protein degradation and maintenance of protein load in cells, with a role in manipulation of the proteome in response to abiotic stress conditions[86,87]. For example, an Ub was found to regulate the expression of heat shock proteins in *Brassica napus*[88]. In addition, Ubs can control pattern-recognition receptors, which are crucial for plant defence and immunity against pathogens[89].

2.5.4 Evolution of plant defence genes in Dalbergia cochinchinensis and D. oliveri

CAFE analysis was conducted to detect expanded gene families in *D. cochinchinensis* and *D. oliveri* compared to other *Dalbergia* and Fabaceae species. Both species showed a significant expansion in disease resistance proteins (R proteins): 34 R protein families were detected to expand in *D. oliveri* (294 R proteins), while 6 were detected in *D. cochinchinensis* (52 R proteins). GO enrichment of these significantly expanded gene families also confirmed an over-representation of immune response and defence response genes. R proteins are important in response to biotic stresses, as plants are attacked by many pathogenic organisms such as bacterial, fungi, viruses, and nematodes[90]. Pathogens secrete effector proteins during infection and can be recognised by R proteins in gene-for-gene interactions[91]. Due to the

highly specific nature of R proteins on effectors, the R protein family evolves under diversifying selection for rapid acquisition of novel specificity to pathogens[92].

Although *D. cochinchinensis* and *D. oliveri* are commonly found in the same geographical localities in Thailand, Laos, Cambodia and Vietnam, *D. oliveri* has a wider distribution towards Myanmar and occurs in a broader diversity of forest types[10]. The wider niche of *D. oliveri* may encompass a wider array of biotic stresses and diseases and thus explain the more diverged R protein families than in *D. cochinchinensis*.

Our PAML analysis detected 16 and 22 positively selected genes responsible for defence responses (GO: 0006952) in *D. cochinchinensis* and *D. oliveri*, respectively, suggesting an adaptive divergence in the suite of plant defence genes. Positive selection in PAML analysis is detected based on measuring the ratio of non-synonymous to synonymous substitution (*dN/dS*) for all single-copy orthologs, assuming $dN/dS = 1$ in neutral molecular evolution, $dN/dS > 1$ signals positive selection[93]. Most of the positively selected genes do not belong to the R family, but instead, for example, to the leucine-rich repeats (LRR) family, RNA-binding family, NPK1-related protein kinase family, which also are involved in the detection of pathogenic compounds and triggering of plant defence[94].

Positive selection analysis also revealed several GO terms that were different between the two species, with *D. oliveri* having more positively selected genes in every term than *D. cochinchinesis*. Only 28 genes were positively selected in both *D. cochinchinensis* and *D. oliveri*, whereas they each had 343 and 411 positively selected distinct genes respectively. The difference in selection signals may suggest that even though the two species share similar geographical distributions, they are subject to different selective forces and slightly more genes have undergone positive selection in *D. oliveri* evolution. The only population genetic study revealed that *D. oliveri* maintains higher genetic diversity than *D. cochinchinensis* from ancient genetic bottlenecks, potentially related to higher gene flow and dispersal capacity in

D. oliveri[10]. Potential selection differences between the two species will need further studies, such as through landscape genomics, to fully elucidate their gene-environment associations.

2.6 Conclusion

Of the 14,191 vascular plants that have been listed as threatened (Vulnerable, Endangered and Critically Endangered) on the IUCN Red List (version February 2019)[95], 16 (~ 0.1%) have published genomes and only 64 have published transcriptomes as BioProjects on NCBI (~ 0.5%)[96]. Compared to about 1% of threatened animal species with published genomes on NCBI[97], there are disproportionately few genome-wide resources in threatened plants.

The potential application of genomic tools for conservation theory and practice has been clearly highlighted but its use is still limited in real-world initiatives[98]. One of the limitations is, assembling a reference genome involves considerable expertise, costs, and computational resources[99]. Advances in RNA-seq and transcriptomics offer a cost-effective alternative to facilitate diverse genomic applications[100]. Reference transcriptomes enable the development of an array of genotyping methods, such as microsatellites[97], exon capture[101], and SNP discovery with genotyping-by-sequencing[102]. Although targeted capture probes exist for legumes[103], our transcriptomes capture a larger set of single or low-copy homologous genes exclusive to *Dalbergia*. The genome-wide resource allows us to study genetic diversity and understand both its neutral and adaptive components. This will produce insights into the mechanisms driving interactions between the environment and populations, with the potential to inform adaptive management of threatened populations, such as through assisted gene flow, GWAS, and marker-based or genomic selection[97,104].

Dalbergia is highly threatened as a genus globally because of its economic value, with *D. cochinchinensis* and *D. oliveri* respectively characterised as Vulnerable and Engendered in the IUCN Red List. With overexploitation of these two species, timber markets have already shifted to other *Dalbergia* species leading to serial exploitation within the genus[105]. Our reference transcriptomes hugely expand the genomic resource repertoire for

the genus *Dalbergia* and will facilitate transfer of utility through to other *Dalbergia* species. They will also open the potential for future studies of *Dalbergia* species towards their evolution and conservation in a broader context.

2.7 Acknowledgements

We thank the Royal Botanic Gardens, Kew and the World Agroforestry Centre for their generous sharing of seed materials. We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome Trust grant reference 203141/Z/16/Z) for the generation and initial processing of the sequencing data. We acknowledge using the University of Oxford Advanced Research Computing (ARC) facility in this work (http://dx.doi.org/10.5281/zenodo.22558). This work was supported by funding to T.H.H. from both the Biotechnology and Biological Sciences Research Council (BBSRC) [grant number BB/M011224/1] and The Oxford & Cambridge Society of Hong Kong (Scholarship Grant Award 2019), and to J.J.M, D.H.B., T.S., B.T. from the Darwin Initiative (ref. 25-023).

2.8 Data availability statement

The research materials, including raw sequences and assemblies, supporting this publication can be publicly accessed in NCBI under the BioProject PRJNA593817.

2.9 References

1. Vatanparast, M. *et al.* First molecular phylogeny of the pantropical genus *Dalbergia*: Implications for infrageneric circumscription and biogeography. *South African J. Bot.* **89**, 143–149 (2013).

2. Saha, S. *et al.* Ethnomedicinal, phytochemical, and pharmacological profile of the genus *Dalbergia* L. (Fabaceae). *Phytopharmacology* **4**, 291–346 (2013).

3. Sprent, J. I. *Legume Nodulation: A Global Perspective*. (Wiley-Blackwell, 2009).

4. Bhagwat, R. M., Dholakia, B. B., Kadoo, N. Y., Balasundaran, M. & Gupta, V. S. Two new potential barcodes to discriminate *Dalbergia* species. *PLoS One* **10**, 1–18 (2015).

5. EIA. *Routes of Extinction: The corruption and violence destroying Siamese rosewood in the Mekong.* (Environmental Investigation Agency, 2014).

6. EIA. *The Hongmu Challenge: A briefing for the 66th meeting of the CITES Standing Committee, January 2016*. (2016).

7. Winfield, K., Scott, M. & Graysn, C. Global status of *Dalbergia* and *Pterocarpus* rosewood producing species in trade. in *Convention on International Trade in Endangered Species 17th Conference of Parties - Johannesburg* (2016).

8. Bentham, G. Synopsis of Dalbergieae, a tribe of Leguminosae. *J. Proc. Linn. Soc. London. Bot.* **4**, 1–128 (1860).

9. Lavin, M. *et al.* The dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. *Am. J. Bot.* **88**, 503 (2001).

10. Hartvig, I. *et al.* Population genetic structure of the endemic rosewoods *Dalbergia cochinchinensis* and *D. oliveri* at a regional scale reflects the Indochinese landscape and life-history traits. *Ecol. Evol.* **8**, 530–545 (2018).

11. Hartvig, I., Czako, M., Kjær, E. D., Nielsen, L. R. & Theilade, I. The use of DNA barcoding in identification and conservation of rosewood (*Dalbergia* spp.). *PLoS One* **10**, (2015).

12. Wattoo, J. I., Saleem, M. Z., Shahzad, M. S., Arif, A. & Hameed, A. DNA Barcoding: Amplification and sequence analysis of *rbcL* and *matK* genome regions in three divergent plant species. *Adv. Life Sci.* **4**, 03–07 (2016).

13. Phong, D. T., Tang, D. Van, Hien, V. T. T., Ton, N. D. & Van, H. N. Nucleotide diversity of a nuclear and four chloroplast DNA regions in rare tropical wood species of *Dalbergia* in Vietnam: a DNA barcode identifying utility. *Asian J. Appl. Sci.* **02**, 116–125 (2014).

14. Resende, L. C., Ribeiro, R. A. & Lovato, M. B. Diversity and genetic connectivity among populations of a threatened tree (*Dalbergia nigra*) in a recently fragmented landscape of the Brazilian Atlantic Forest. *Genetica* **139**, 1159–1168 (2011).

15. Buzatti, R. S. de O., Ribeiro, R. A., Filho, J. P. de L. & Lovato, M. B. Fine-scale spatial genetic structure of *Dalbergia nigra* (Fabaceae), a threatened and endemic tree of the Brazilian Atlantic Forest. *Genet. Mol. Biol.* **35**, 838–846 (2012).

16. Liu, F.-M. *et al.* De novo transcriptome analysis of *Dalbergia odorifera* and transferability of SSR markers developed from the transcriptome. *Forests* **10**, 98 (2019).

17. Xu, D.-P., Xu, S.-S., Zhang, N.-N., Yang, Z.-J. & Hong, Z. Chloroplast genome of *Dalbergia cochinchinensis* (Fabaceae), a rare and endangered rosewood species in Southeast Asia. *Mitochondrial DNA Part B* **4**, 1144–1145 (2019).

18. Wariss, H. M., Yi, T.-S., Wang, H. & Zhang, R. Characterization of the complete

chloroplast genome of *Dalbergia odorifera* (Leguminosae), a rare and critically endangered legume endemic to China. *Conserv. Genet. Resour.* **10**, 527–530 (2017). doi:10.1007/s12686-017-0866-2

19. Liu, Y., Huang, P., Li, C.-H., Zang, F.-Q. & Zheng, Y.-Q. Characterization of the complete chloroplast genome of *Dalbergia cultrata* (Leguminosae). *Mitochondrial DNA Part B* **4**, 2369–2370 (2019).

20. Deng, C., Xin, G., Zhang, J. & Zhao, D. Characterization of the complete chloroplast genome of *Dalbergia hainanensis* (Leguminosae), a vulnerably endangered legume endemic to China. *Conserv. Genet. Resour.* **1**, (2018).

21. Song, Y., Zhang, Y., Xu, J., Li, W. & Li, M. F. Characterization of the complete chloroplast genome sequence of *Dalbergia* species and its phylogenetic implications. *Sci. Rep.* **9**, 1–10 (2019).

22. Lateef, A., Prabhudas, S. K. & Natarajan, P. RNA sequencing and de novo assembly of *Solanum trilobatum* leaf transcriptome to identify putative transcripts for major metabolic pathways. *Sci. Rep.* **8**, 15375 (2018).

23. Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O. & Grau, J. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics* **19**, 189 (2018).

24. Wang, B., Kumar, V., Olson, A. & Ware, D. Reviving the transcriptome studies: An insight into the emergence of single-molecule transcriptome sequencing. *Front. Genet.* **10**, 384 (2019).

25. Lamble, S. *et al.* Improved workflows for high throughput library preparation using the transposome-based nextera system. *BMC Biotechnol.* **13**, 104 (2013).

26. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

27. Buffalo, V. Scythe - A Bayesian adapter trimmer (version 0.994 BETA) [Software]. (2011). Available at: https://github.com/vsbuffalo/scythe.

28. Joshi, N. A. & Fass, J. N. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. (2011). Available at: https://github.com/najoshi/sickle.

29. Carruthers, M. *et al.* De novo transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species. *BMC Genomics* **19**, 32 (2018).

30. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).

31. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–9 (2012).

32. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

33. Hass, B. J. TransDecoder. (2018). Available at: https://github.com/TransDecoder/TransDecoder.

34. Kriventseva, E. V *et al.* OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **47**, D807–D811 (2019).

35. Waterhouse, R. M. *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543 (2017).

36. Bertioli, D. J. *et al.* The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. *Nat. Genet.* **48**, 438–446 (2016).

37. Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M. & Kelly, S. TransRate: Reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* **26**, 1134–1144 (2016).

38. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

39. Cheng, C.-Y. *et al.* Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* **89**, 789–804 (2017).

40. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).

41. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).

42. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. . Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *J. Mol. Biol.* **305**, 567–580 (2001).

43. D.M., E. & S., K. OrthoFinder2: fast and accurate phylogenomic orthology analysis from gene sequences. *bioRxiv* 466201 (2018). doi:10.1101/466201

44. Guo, L. *et al.* The opium poppy genome and morphinan production. *Science* **362**, 343–347 (2018).

45. Nakamura, T., Yamada, K. D., Tomii, K. & Katoh, K. Parallelization of MAFFT for

large-scale multiple sequence alignments. *Bioinformatics* **34**, 2490–2492 (2018).

46. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609-12 (2006).

47. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and high-performance computing. *Nat. Methods* **9**, 772 (2012).

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

49. Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).

50. Brea, M., Zamuner, A. B., Matheos, S. D., Iglesias, A. & Zucol, A. F. Fossil wood of the Mimosoideae from the early Paleocene of Patagonia, Argentina. *Alcheringa An Australas. J. Palaeontol.* **32**, 427–441 (2008).

51. Hane, J. K. *et al.* A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: insights into plant-microbe interactions and legume evolution. *Plant Biotechnol. J.* **15**, 318–330 (2017).

52. Lavin, M., Herendeen, P. S. & Wojciechowski, M. F. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Syst. Biol* **54**, 575–594 (2005).

53. Moretzsohn, M. C. *et al.* A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. *Ann. Bot.* **111**, 113–26 (2013).

54. Ye, J. *et al.* WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018

update. *Nucleic Acids Res.* **46**, W71 (2018).

55. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).

56. Mi, H., Muruganujan, A. & Thomas, P. D. PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* **41**, (2013).

57. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

58. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple. *J. R. Stat. Soc. Ser. B Methodol.* **57**, (1995).

59. Sun, J. *et al.* Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nat. Ecol. Evol.* **1**, 0121 (2017).

60. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).

61. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).

62. Soltis, D. E., Soltis, P. S., Bennett, M. D. & Leitch, I. J. Evolution of genome size in the angiosperms. *Am. J. Bot.* **90**, 1596–1603 (2003).

63. Hiremath, S. C. & Nagasampige, M. H. Genome size variation and evolution in some species of Dalbergia Linn.f. (Fabaceae). *Caryologia* **57**, 367–372 (2004).

64. Lawrence, G. H. M. *Taxonomy of vascular plants*. (Oxford & IBH Publishing Co.,

1973).

65. Lombello, R. A. & Forni-Martins, E. R. Chromosome studies and evolution in Sapindaceae. *Caryologia* **51**, 89–93 (1998).

66. Sheremet'ev, S. N. & Gamalei, Y. V. Towards angiosperms genome evolution in time. *arXiv* (2013).

67. Carlquist, S. Anatomy of vine and liana stems: a review and synthesis. in *The biology of vines* (eds. Putz, F. E. & Mooney, H. A.) 53–72 (University of Cambridge Press, 1991).

68. Li, Q. *et al.* The phylogenetic analysis of *Dalbergia* (Fabaceae: Papilionaceae) based on different DNA barcodes. *Holzforschung* **71**, 939–949 (2017).

69. Lavin, M. *et al.* Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. in *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**, 1509–1522 (Royal Society, 2004).

70. Kučerová, J. Miocénna flóra z lokalít Kalonda a Mučín. *Acta Geol. Slovaca* **1**, 65–70 (2009).

71. Gao, S.-X. & Zhou, Z.-K. The Megafossil Legumes from China. in *Advances in Legume Systematics: Part 4, The Fossil Record* (eds. Herendeen, P. S. & Dilcher, D. L.) (The Royal Botanic Gardens, Kew, 1992).

72. de Saporta, G. *Dalbergia phleboptera* Saporta. *Muséum national d'Histoire naturelle* (2015). Available at: https://science.mnhn.fr/institution/mnhn/collection/f/item/14084.?lang=en_US.

73. De Bruyn, M. *et al.* Borneo and Indochina are major evolutionary hotspots for

Southeast Asian biodiversity. *Syst. Biol* **63**, 879–901 (2014).

74. Koenen, E. J. M. *et al.* The origin and early evolution of the legumes are a complex paleopolyploid phylogenomic tangle closely associated with the Cretaceous-Paleogene (K-Pg) boundary. *bioRxiv* 577957 (2019). doi:10.1101/577957

75. Lespinet, O., Wolf, Y. I., Koonin, E. V. & Aravind, L. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.* **12**, 1048–1059 (2002).

76. Ming, Y. *et al.* Molecular footprints of inshore aquatic adaptation in Indo-Pacific humpback dolphin (*Sousa chinensis*). *Genomics* (2018). doi:10.1016/j.ygeno.2018.07.015

77. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).

78. Luengo, T. M., Mayer, M. P. & Rüdiger, S. G. D. The Hsp70-Hsp90 chaperone cascade in protein folding. (2019). doi:10.1016/j.tcb.2018.10.004

79. Jacob, P., Hirt, H. & Bendahmane, A. The heat-shock protein/chaperone network and multiple stress resistance. *Plant Biotechnology Journal* **15**, 405–414 (2017).

80. Yamada, K. *et al.* Cytosolic HSP90 regulates the heat shock response that is responsible for heat acclimation in *Arabidopsis thaliana*. *J. Biol. Chem.* **282**, 37794–37804 (2007).

81. Clément, M. *et al.* The cytosolic/nuclear HSC70 and HSP90 molecular chaperones are important for stomatal closure and modulate abscisic acid-dependent physiological responses in arabidopsis. *Plant Physiol.* **156**, 1481–1492 (2011).

82. Hou, Q. & Bartels, D. Comparative study of the aldehyde dehydrogenase (ALDH)

gene superfamily in the glycophyte *Arabidopsis thaliana* and *Eutrema halophytes*. *Annals of Botany* **115**, 465–479 (2015).

83. Missihoun, T. D. & Kotchoni, S. O. Aldehyde dehydrogenases and the hypothesis of a glycolaldehyde shunt pathway of photorespiration. *Plant Signal. Behav.* **13**, e1449544 (2018).

84. Estioko, L. P. *et al.* Differences in responses to flooding by germinating seeds of two contrasting rice cultivars and two species of economically important grass weeds. *AoB Plants* **6**, (2014).

85. Brocker, C. *et al.* Aldehyde dehydrogenase (ALDH) superfamily in plants: Gene nomenclature and comparative genomics. *Planta* **237**, 189–210 (2013).

86. Sharma, B., Joshi, D., Yadav, P. K., Gupta, A. K. & Bhatt, T. K. Role of ubiquitin-mediated degradation system in plant biology. *Frontiers in Plant Science* **7**, (2016).

87. Walters, K. J., Goh, A. M., Wang, Q., Wagner, G. & Howley, P. M. Ubiquitin family proteins and their relationship to the proteasome: A structural perspective. *Biochimica et Biophysica Acta - Molecular Cell Research* **1695**, 73–87 (2004).

88. Liu, Z.-B. *et al.* A novel membrane-bound E3 ubiquitin ligase enhances the thermal resistance in plants. *Plant Biotechnol. J.* **12**, 93–104 (2014).

89. Macho, A. P. & Zipfel, C. Plant PRRs and the activation of innate immune signaling. *Molecular Cell* **54**, 263–272 (2014).

90. Martin, G. B., Bogdanove, A. J. & Sessa, G. Understanding the functions of plant disease resistance proteins. *Annu. Rev. Plant Biol.* **54**, 23–61 (2003).

91. Cohn, J., Sessa, G. & Martin, G. B. Innate immunity in plants. *Current Opinion in Immunology* **13**, 55–62 (2001).

92. Lehmann, P. Structure and evolution of plant disease resistance genes. *Journal of Applied Genetics* **43**, 403–414 (2002).

93. Jeffares, D. C., Tomiczek, B., Sojo, V. & dos Reis, M. A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. in *Parasite Genomics Protocols: Second Edition* 65–90 (Springer Fachmedien, 2014). doi:10.1007/978-1-4939-1438-8_4

94. Andersen, E. J., Ali, S., Byamukama, E., Yen, Y. & Nepal, M. P. Disease resistance mechanisms in plants. *Genes* **9**, (2018).

95. IUCN. The IUCN Red List of Threatened Species. Veresion 2019-2. (2019). Available at: http://www.iucnredlist.org.

96. Federhen, S. The NCBI Taxonomy database. *Nucleic Acids Res.* **40**, (2012).

97. Brandies, P., Peel, E., Hogg, C. J. & Belov, K. The Value of Reference Genomes in the Conservation of Threatened Species. *Genes 2019, Vol. 10, Page 846* **10**, 846 (2019).

98. Supple, M. A. & Shapiro, B. Conservation of biodiversity in the genomics era. *Genome Biol.* **19**, (2018).

99. Fuentes-Pardo, A. P. & Ruzzante, D. E. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Mol. Ecol.* **26**, 5369–5406 (2017).

100. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57–63 (2009).

101. Bragg, J. G., Potter, S., Bi, K. & Moritz, C. Exon capture phylogenomics: efficacy across scales of divergence. *Mol. Ecol. Resour.* **16**, 1059–1068 (2016).

102. İpek, A., İpek, M., Ercişli, S. & Tangu, N. A. Transcriptome-based SNP discovery by GBS and the construction of a genetic map for olive. *Funct. Integr. Genomics* **17**, 493–501 (2017).

103. Vatanparast, M., Powell, A., Doyle, J. J. & Egan, A. N. Targeting legume loci: A comparison of three methods for target enrichment bait design in Leguminosae phylogenomics. *Appl. Plant Sci.* **6**, (2018).

104. Ouborg, N. J. Integrating population genetics and conservation biology in the era of genomics. *Biol. Lett.* **6**, 3–6 (2010).

105. CITES. *Consideration of proposals for amendment of appendices I and II. Convention on International Trade in Endangered Species of Wild Fauna and Flora*. (Convention on International Trade in Endangered Species of Wild Fauna and Flora, 2017).

106. Asian Regional Workshop (Conservation & Sustainable Management of Trees Viet Nam). *Dalbergia cochinchinensis*. The IUCN Red List of Threatened Species. e.T32625A9719096 (1998). doi:10.2305/IUCN.UK.1998.RLTS.T32625A9719096.en

107. Bernal, R., Gradstein, S. . & Celis, M. *Catálogo de plantas y líquenes de Colombia*. (Instituto de Ciencias Naturales, Universidad Nacional de Colombia, Bogotá, 2015).

108. World Conservation Monitoring Centre. *Dalbergia melanoxylon*. The IUCN Red List of Threatened Species 1998. e.T32504A9710439 (1998). doi:10.2305/IUCN.UK.1998.RLTS.T32504A9710439.en

109. ILDIS. International Legume Database and Information Service V10.39. (2011).

110. Nghia, N. H. *Dalbergia oliveri*. The IUCN Red List of Threatened Species 1998. e.T32306A9693932 (1998). doi:10.2305/IUCN.UK.1998.RLTS.T32306A9693932.en

111. Orwa, C., Mutua, A., Kindt, R., Jamnadass, R. & Anthony, S. Agroforestree Database:

a tree reference and selection guide version 4.0. (2009). Available at: http://www.worldagroforestry.org/sites/treedbs/treedatabases.asp.

2.10 Figures and Tables

Figure 2.1. Bioinformatic pipeline of *de novo* transcriptome analysis and gene annotation for the 6 *Dalbergia* species. For the software details, see methods.

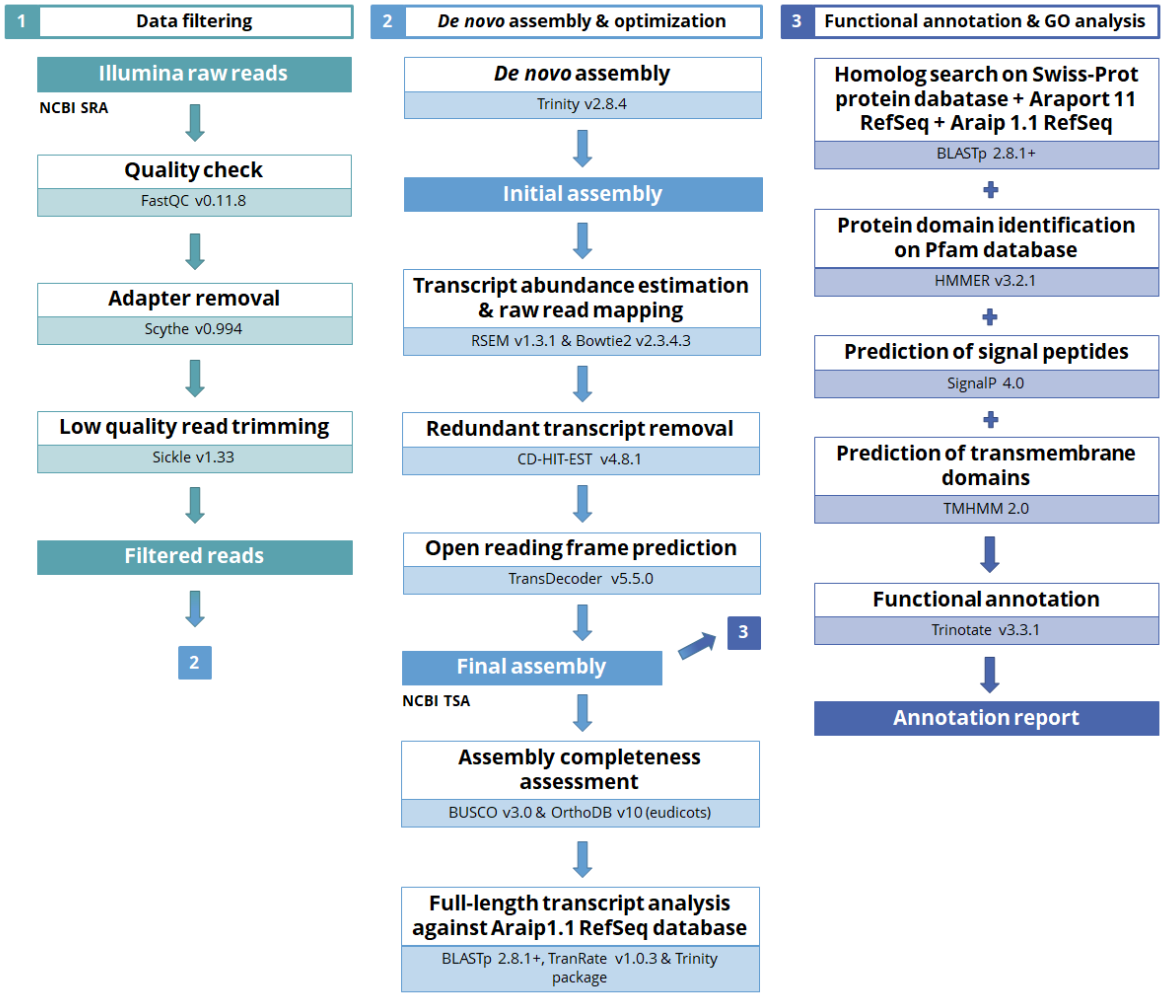


Figure 2.2. Transcripts of the 6 *Dalbergia* species aligned to the *Arachis ipaensis* NCBI RefSeq database at a given coverage.

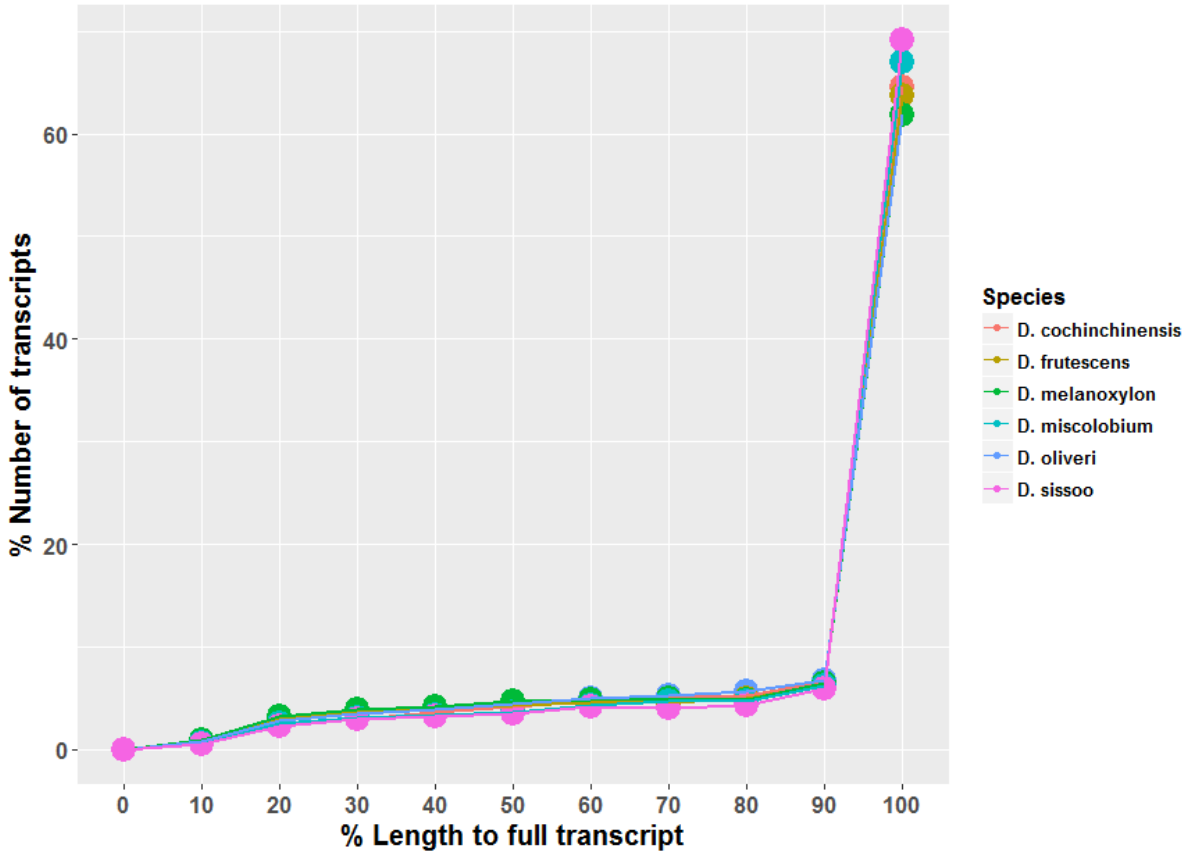


Figure 2.3. Venn diagram showing orthogroups among the 6 *Dalbergia* species visualized with ClusterVenn.

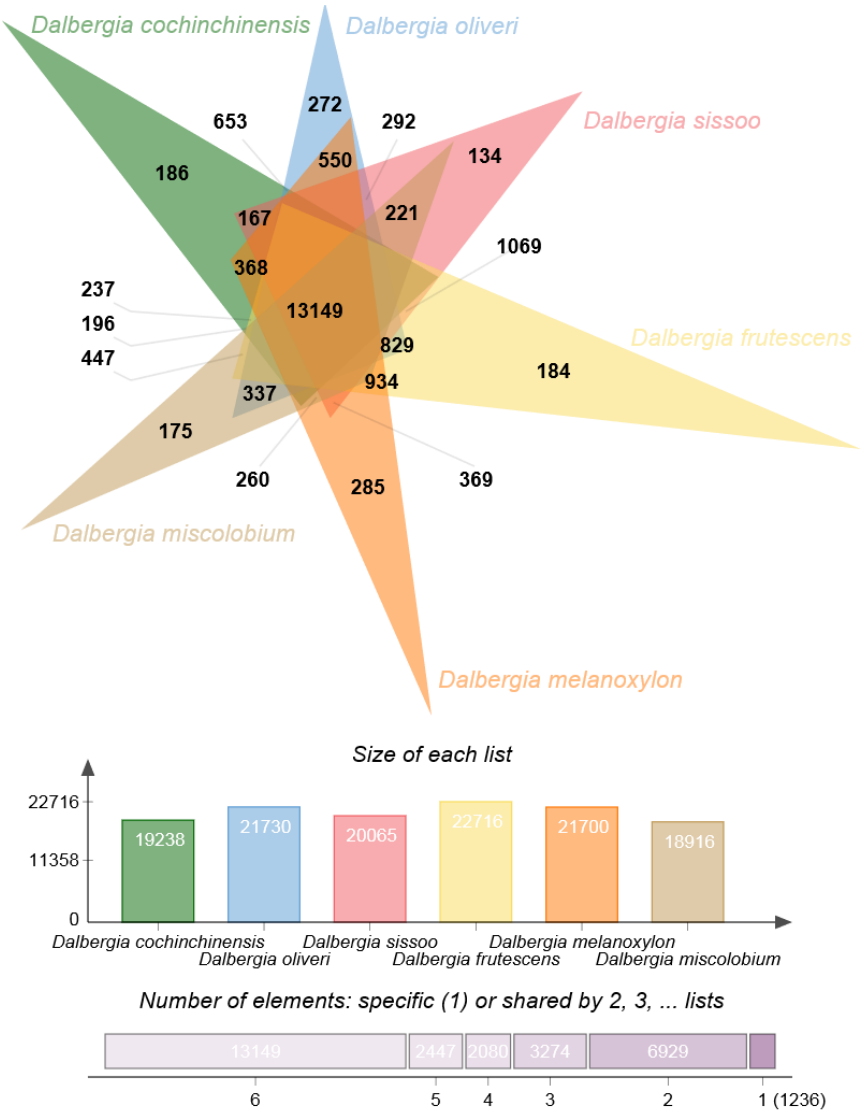


Figure 2.4. Dated phylogeny of 16 Fabaceae species based on Bayesian analysis of a supergene from the 256 single-copy orthologs (479,064 bp) from their transcriptomes. Node bars indicate 95% CI for the estimated divergence time. Numbers on branches indicate posterior probability (1 for all branches).

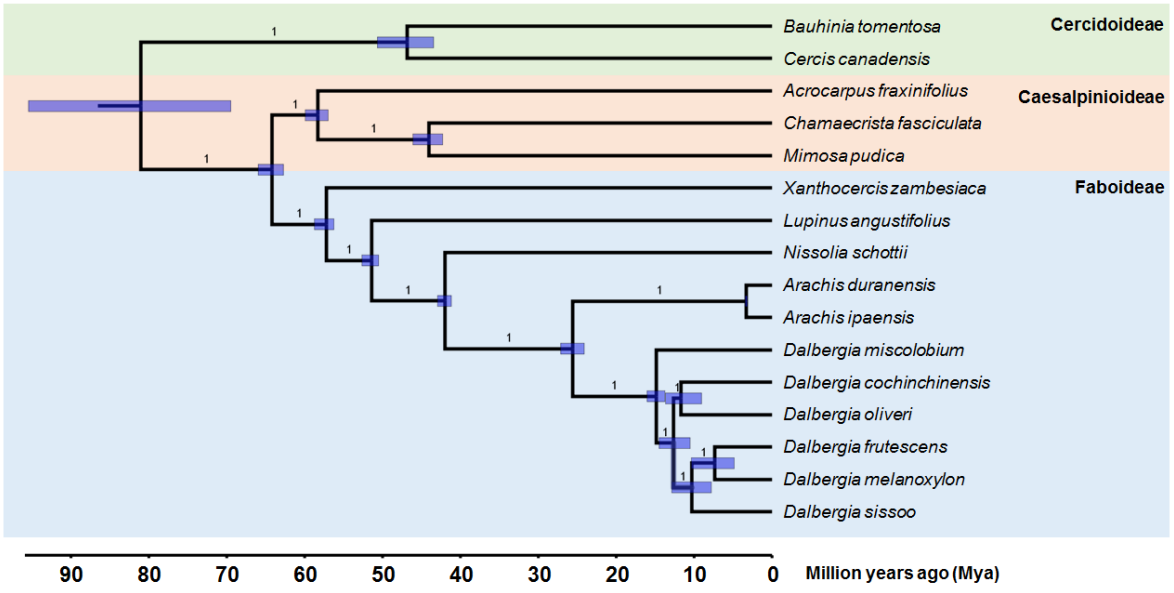
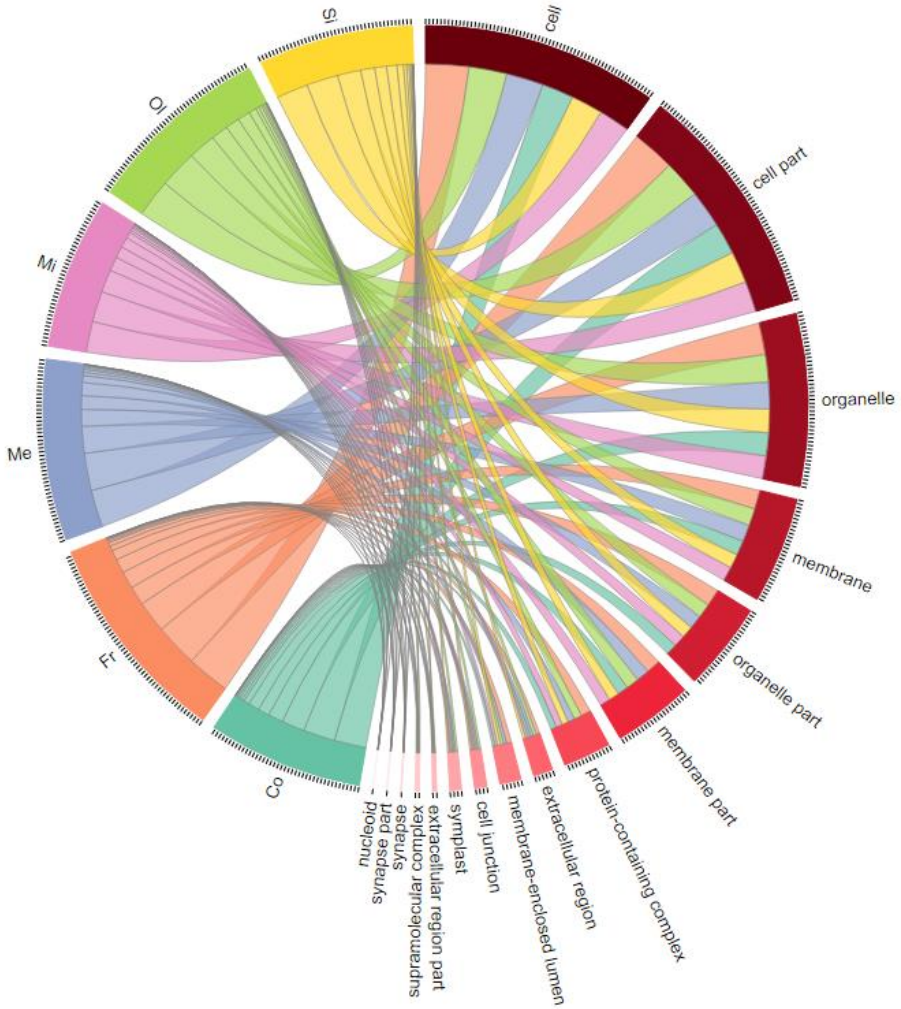
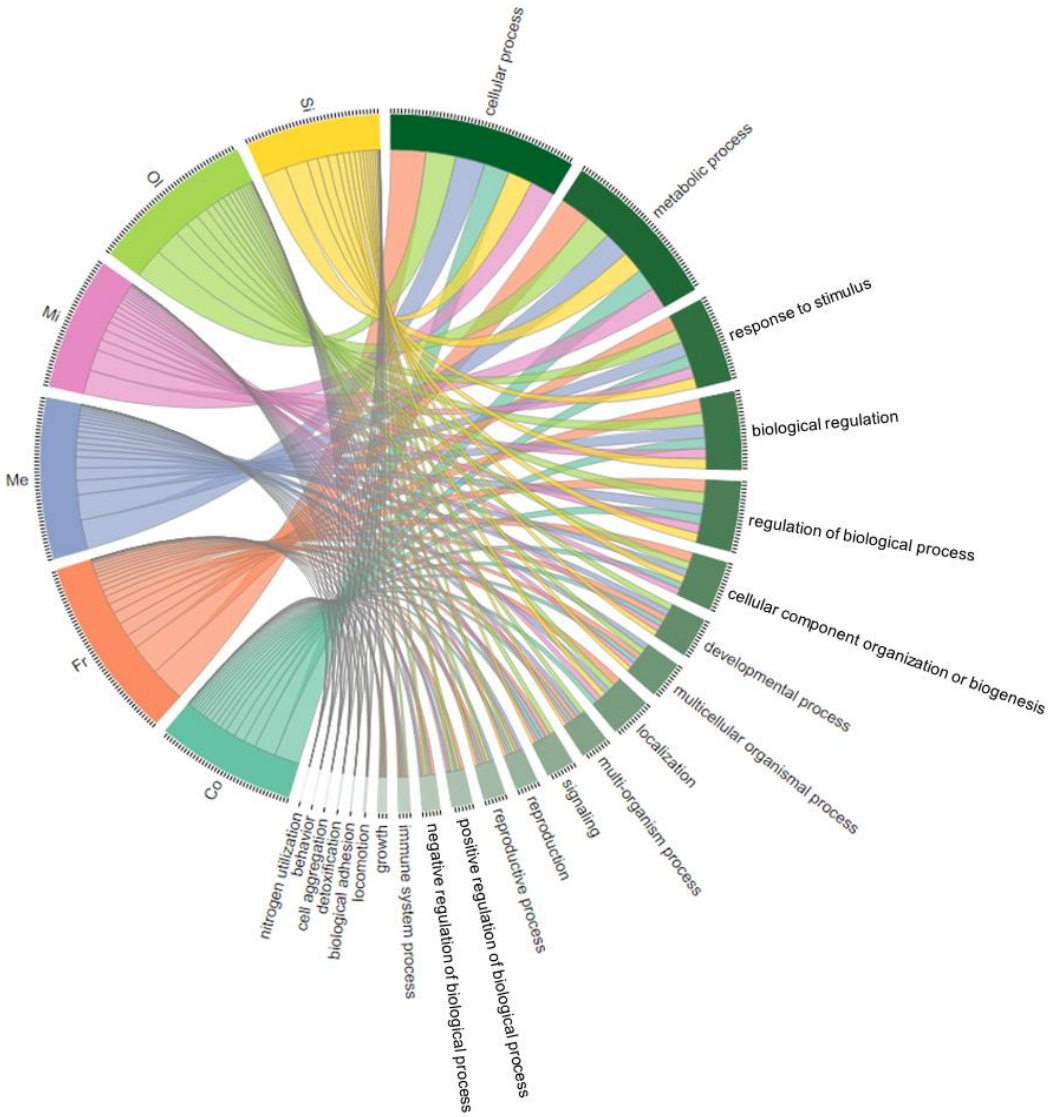


Figure 2.5. Chord diagrams of annotated GO terms (level 2) for (a) cellular components (red); (b) biological processes (green); (c) molecular functions (blue) in the 6 *Dalbergia* species, only showing terms that are significantly different among the species ($p < 0.05$ in chi-square test of independence). Each tick interval is 1,000 counts. The scale is arranged clockwise showing descending GO counts.

(a)



(b)


cellular process
metabolic process
response to stimulus
biological regulation
regulation of biological process
cellular component organization or biogenesis
developmental process
multicellular organismal process
localization
multi-organism process
signaling
reproduction
reproductive process
positive regulation of biological process
negative regulation of biological process
immune system process
growth
locomotion
biological adhesion
detoxification
cell aggregation
behavior
nitrogen utilization
Co
Fr
Me
Mi
Ol
Si

(c)

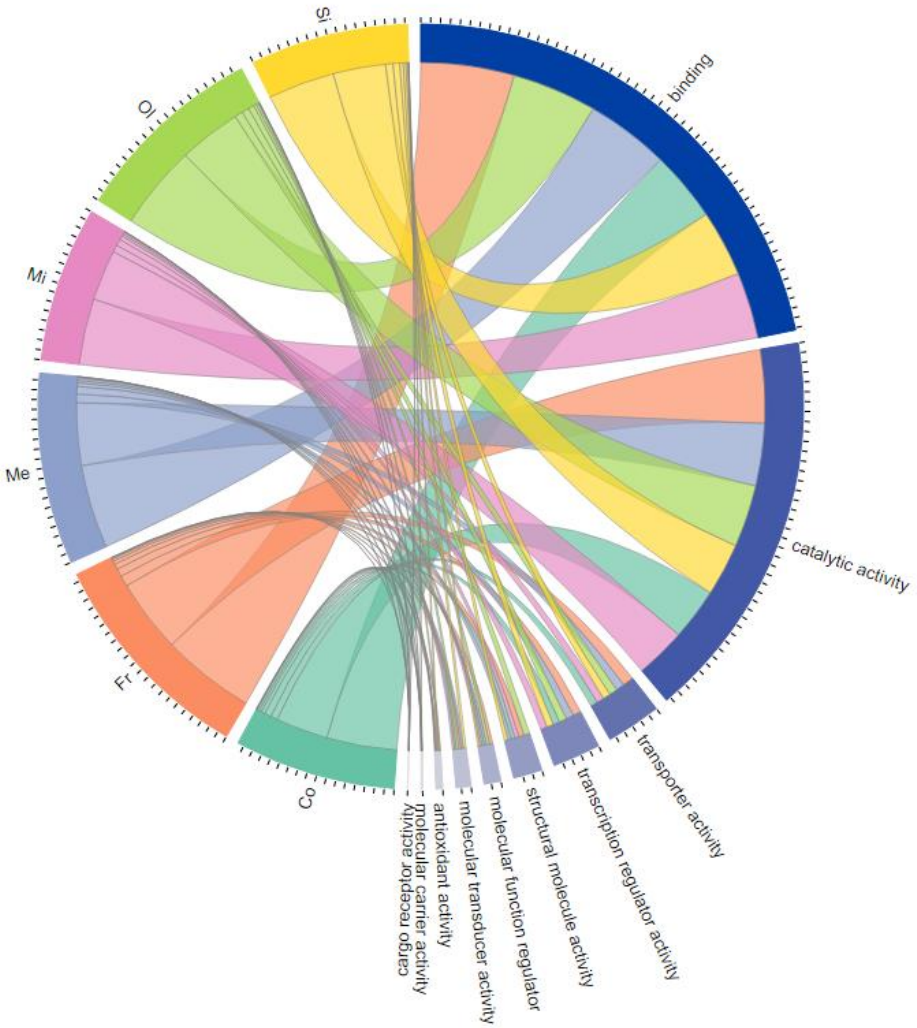

binding
catalytic activity
transporter activity
transcription regulator activity
structural molecule activity
molecular function regulator
molecular transducer activity
antioxidant activity
molecular carrier activity
cargo receptor activity
Co
Ft
Me
Mi
Ol
Si

Figure 2.6. Heatmap of annotated Pfam domains of the 13 Fabaceae species, only showing domains (n = 91) that are significantly contracted (negative) or expanded (positive) in the *Dalbergia* species ($p < 0.05$ in two-tailed Fisher's exact test of independence).

Figure 2.7. Results of GO enrichment analysis on positively selected genes, which are single-copy orthologs, between *D. cochinchinensis* (N = 371, GO annotated n = 299) and *D. oliveri* (N = 439, GO annotated n = 361), only showing terms that are significant ($p < 0.05$ in chi-square test of independence).

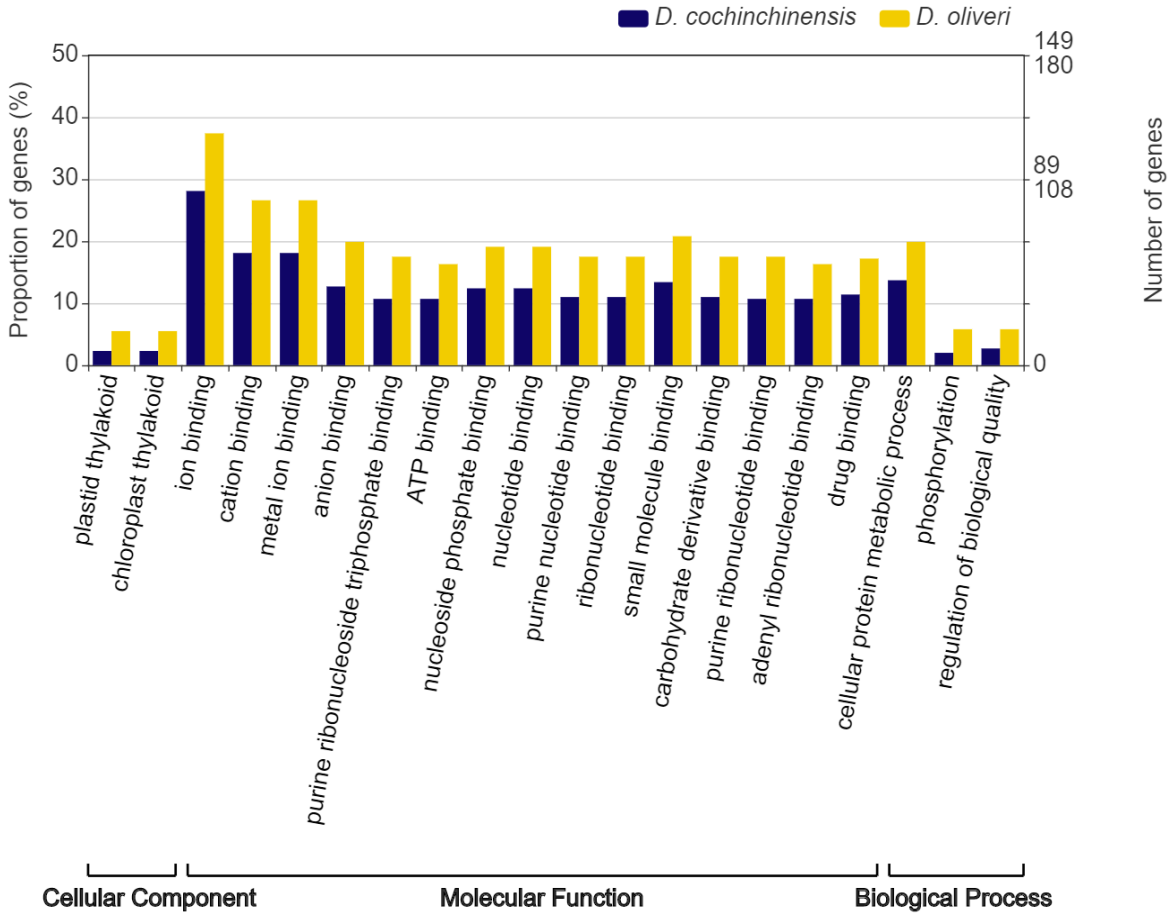


Table 2.1. Basic details and conservation status of the six *Dalbergia* species covered in this study.

Scientific name	Common name	Native occurrence	Habitat	IUCN status	CITES status	References
Dalbergia cochinchinensis	Siamese rosewood	Cambodia, Lao PDR, Thailand, Vietnam	Terrestrial; open semi-deciduous forests	Vulnerable A1cd (1998)	II (2017)	[106]
Dalbergia frutescens	Brazilian tulipwood	Columbia, Amazonia, Andes, Caribbean Plain, Magdalena Valley	Variable, usually as a liana	Unclassified	II (2017)	[107]
Dalbergia melanoxylon	African blackwood	Wide geographical distribution in sub-Saharan countries	Range of woodland habitats	Near threatened (1998)	II (2017)	[108]
Dalbergia miscolobium	Jacaranda-do-cerrado	Brazil, Bolivia	Savannah	Unclassified	II (2017)	[109]
Dalbergia oliveri	Burmese rosewood	Cambodia, Lao PDR, Myanmar, Thailand, Vietnam	Mixed deciduous forests and tropical evergreen	Endangered A1cd (1998)	II (2017)	[110]
Dalbergia sissoo	North Indian rosewood; Shisham	Indian Subcontinent	Deciduous forests	Unclassified	II (2017)	[111]

Table 2.2. Acquisition details for the plant materials of the six *Dalbergia* species covered in this study**.**

Species	**Source**	**Accession**	**Year collected**	**Wild**	**Identifier used by source**
D. cochinchinensis	Forest Research Center, Lao PDR	Bolikhamxay, Khamkend, Laos	2018	Yes	N/A
D. frutescens	Millennium Seed Bank, Royal Botanic Gardens, Kew	Brazil	1992	Yes	98382
D. melanoxylon	World Agroforestry Centre	Kathozweni, Kenya	2011	Yes	05662
D. miscolobium	Royal Botanic Gardens, Kew	Brazil	1994	Yes	102498
D. oliveri	Institute of Forest & Wildlife Research & Development, Cambodia	Cambodia	2018	Yes	N/A
D. sissoo	Commercial retailer (http://www.rarepalmseeds.com)	Unknown	Unknown	Unknown	YRFFDASI

Table 2.3. Transcriptome and annotation resources of the 16 Fabaceae species used in this study.

Species	Abbreviation used in this paper	Source	Assembly method	BUSCO score	Reference
Acrocarpus fraxinifolius	Af	https://ics.hutton.ac.uk/tropiTree/	Trinity	Unreported	[1]
Arachis ipaensis	Ai	NCBI: GCF_000816755.2	Trinity	C: 87.3; F: 5.8; M: 6.9	[2]
Arachis duranensis	Ad	NCBI: GCF_000817695.2	Trinity	C: 88.8; F: 4.7; M: 6.5	[2]
Bauhinia tomentosa	Bt	DOI: 10.5061/dryad.ff1tq	Trinity	Unreported	[3]
Cercis canadensis	Cc	DOI: 10.5524/101044	Trinity	C: 96.2; F: 1.9; M: 1.9	[4]
Chamaecrista fasciculata	Cf	DOI: 10.5524/101045	Trinity	C: 93.2; F: 1.7; M: 5.1	[4]
Dalbergia cochinchinensis	Co	NCBI: GIHU00000000	Trinity	C: 92.2; F: 4.0; M: 2.9	this study
Dalbergia frutescens	Fr	NCBI: GIHP00000000	Trinity	C: 92.1; F: 4.8; M: 3.1	this study
Dalbergia melanoxylon	Me	NCBI: GIHQ00000000	Trinity	C: 92.3; F: 5.1; M: 2.6	this study
Dalbergia miscolobium	Mi	NCBI: GIHR00000000	Trinity	C: 93.1; F: 4.6; M: 2.3	this study
Dalbergia oliveri	Ol	NCBI: GIHS00000000	Trinity	C: 90.9; F. 6.5; M: 2.6	this study
Dalbergia sissoo	Si	NCBI: GIHT00000000	Trinity	C; 94.4; F: 3.3; M: 2.3	this study
Mimosa pudica	Mp	DOI: 10.5524/101049	Trinity	C: 94.3; F: 1.7; M: 4.0	[4]
Lupinus angustifolius	La	NCBI: GCF_001865875.1	Trinity	C: 92.7; F: 4.2; M: 3.1	[5]
Nissolia schottii	Ns	DOI:10.5524/101050	Trinity	C: 95.0; F: 1.5; M: 3.5	[4]
Xanthocercis zambesiaca	Xz	DOI: 10.5061/dryad.ff1tq	Trinity	Unreported	[3]

Table 2.4. Summary of transcriptome assembly statistics of the six *Dalbergia* species.

Feature	***D. cochinchinensis***	***D. frutescens***	***D. melanoxylon***	***D. miscolobium***	***D. oliveri***	***D. sissoo***
Number of paired-end raw reads	168,351,690	71,187,798	74,366,734	91,273,654	181,456,683	73,160,910
Number of paired-end filtered reads	156,116,637 (92.7%)	65,092,217 (91.4%)	67,994,105 (91.4%)	83,178,635 (91.1%)	169,551,748 (93.4%)	67,086,967 (91.7%)
Number of transcripts in initial assembly	277,981	274,663	363,116	208,249	376,014	195,268
Number of genes in initial assembly	161,051	179,085	212,141	123,962	223,289	121,629
Total length of transcripts (bp)	316,346,363	255,266,594	309,909,355	237,557,440	357,336,705	216,910,975
Average transcript length (bp)	1138.01	929.38	853.47	1140.74	950.33	1110.84
N50[1] (bp)	2,159	1,749	1,477	2,074	1,851	2,019
GC (%)	40.25	41.88	41.38	40.60	40.59	41.06
Map representation alignment rate (%)	89.85	87.71	85.69	89.20	87.14	89.02
Number of non-redundant transcripts	224,511	231,281	271,088	174,382	293,334	168,039
Final assembly						
Number of transcripts in final assembly	84,003	84,897	80,484	69,357	92,906	67,379
Total length of transcripts (bp)	81,157,122	75,431,325	70,467,927	68,915,367	83,501,667	67,138,149
Average transcript length (bp)	966.12	888.50	875.55	993.63	898.78	996.43
N50 of transcripts (bp)	1,254	1,152	1,149	1,290	1,179	1,305
GC of transcripts (%)	44.66	46.01	45.24	44.68	44.97	45.00

Number of genes in final assembly	34,655	48,591	43,848	31,678	43,879	32,753
Total length of genes (bp)	33,219,183	41,338,207	37,309,763	31,488,922	37,371,154	32,374,118
Average gene length (bp)	958.57	850.74	850.89	994.03	851.69	988.43
N50 of genes (bp)	1,341	1,145	1,173	1,383	1,182	1,374
GC of genes (%)	45.37	47.43	46.00	45.32	45.97	45.92
BUSCO Score[2] (N = 2,121) (%)	C: 92.2; F: 4.9; M: 2.9	C: 92.1; F: 4.8; M: 3.1	C: 92.3; F: 5.1; M: 2.6	C: 93.1; F: 4.6; M: 2.3	C: 90.9; F: 6.5; M: 2.6	C: 94.4; F: 3.3; M: 2.3

Table 2.5. Summary of transcriptome BUSCO statistics (N = 2,121) of the 6 *Dalbergia* species.

BUSCO statistics	***D. cochinchinensis***	***D. frutescens***	***D. melanoxylon***	***D. miscolobium***	***D. oliveri***	***D. sissoo***
Complete	1,956 (92.2%)	1,953 (92.1%)	1,958 (92.3%)	1,976 (93.1%)	1,927 (90.9%)	2,002 (94.4%)
single-copy	1,180 (55.6%)	1,257 (59.3%)	1,236 (58.3%)	1,254 (59.1%)	1,172 (55.3%)	1,331 (62.8%)
duplicated	776 (36.6%)	696 (32.8%)	722 (34.0%)	722 (34.0%)	755 (35.6%)	671 (31.6%)
Fragmented	104 (4.9%)	102 (4.8%)	109 (5.1%)	97 (4.6%)	137 (6.5%)	69 (3.3%)
Missing	61 (2.9%)	66 (3.1%)	54 (2.6%)	48 (2.3%)	57 (2.6%)	50 (2.3%)

Table 2.6. Transcriptome annotation statistics of the 6 *Dalbergia* species. For the versions of annotation databases, see methods for details. Highest numbers for each row are highlighted in bold.

	D. cochinchinensis	***D. frutescens***	***D. melanoxylon***	***D. miscolobium***	***D. oliveri***	***D. sissoo***
Number of transcripts in final assembly	84,003	84,897	80,484	69,357	**92,906**	67,379
Number of successfully annotated TRANSCRIPTS						
Araip 1.1	74,397 (88.6%)	67,052 (79.0%)	67,164 (83.5%)	**61,653 (88.9%)**	78,245 (84.2%)	58,512 (86.8%)
Araport 11	70,780 (84.3%)	63,438 (74.7%)	62,185 (77.3%)	**58,984 (85.0%)**	73,889 (79.5%)	56,091 (83.2%)
SwissProt	63,175 (75.2%)	61,062 (71.9%)	56,193 (69.8%)	**53,201 (76.7%)**	67,064 (72.2%)	51,051 (75.8%)
GO	61,993 (73.8%)	60,005 (70.7%)	55,022 (68.4%)	**52,043 (75.0%)**	65,740 (70.8%)	50,008 (74.2%)
KEGG	55,538 (66.1%)	52,709 (62.1%)	48,603 (60.4%)	**46,789 (67.5%)**	57,896 (62.3%)	45,190 (67.1%)
EggNOG	52,510 (62.5%)	44,849 (52.8%)	44,802 (55.7%)	**44,221 (63.8%)**	54,059 (58.2%)	41,184 (61.1%)
Pfam	58,589 (69.7%)	56,835 (66.9%)	51,717 (64.3%)	**49,888 (71.9%)**	62,162 (66.9%)	47,842 (71.0%)
TMHMM	**17,486 (20.8%)**	15,424 (18.2%)	14,864 (18.5%)	14,338 (20.7%)	18,359 (19.8%)	13,671 (20.3%)
SignalP	5,603 (6.7%)	5,214 (6.1%)	4,880 (6.1%)	**4,772 (6.9%)**	5,896 (6.3%)	4,643 (6.9%)
Number of genes in final assembly	34,655	**48,591**	43,848	31,678	43,879	32,753
Number of successfully annotated GENES						
Araip 1.1	28,277 (81.6%)	33,452 (68.8%)	33,617 (76.7%)	**26,315 (83.1%)**	32,936 (75.1%)	26,141 (79.8%)
Araport 11	26,388 (76.1%)	31,421 (64.7%)	30,420 (69.4%)	**24,894 (78.6%)**	30,497 (69.5%)	24,837 (75.8%)
SwissProt	24,175 (69.8%)	32,281 (66.4%)	28,022 (63.9%)	**22,920 (72.4%)**	28,658 (65.3%)	23,396 (71.4%)
GO	23,686 (68.4%)	31,733 (65.3%)	27,471 (62.7%)	**22,427 (70.8%)**	28,116 (64.1%)	22,926 (70.0%)
KEGG	20,603 (59.5%)	27,102 (55.8%)	23,609 (53.8%)	19,606 (61.9%)	23,810 (54.3%)	**20,297 (62.0%)**
EggNOG	19,163 (55.3%)	20,886 (43.0%)	21,121 (48.2%)	**18,204 (57.5%)**	21,470 (48.9%)	17,635 (53.8%)
Pfam	23,134 (66.8%)	30,561 (62.9%)	26,161 (59.7%)	**22,077 (69.7%)**	27,332 (62.3%)	22,544 (68.8%)
TMHMM	**7,609 (22.0%)**	8,120 (16.7%)	7,748 (17.7%)	6,417 (20.3%)	8,006 (18.3%)	6,447 (19.7%)
SignalP	2,607 (7.5%)	3,060 (6.3%)	2,763 (6.3%)	**2,453 (7.7%)**	2,874 (6.6%)	2,401 (7.3%)

Table 2.7. TransRate reference-based metrics of the 6 *Dalbergia* transcriptomes mapped on *Arachis ipaensis*. CRBB means Conditional Reciprocal Best BLAST. RBH means reciprocal best hit.

Feature	***D. cochinchinensis***	***D. frutescens***	***D. melanoxylon***	***D. miscolobium***	***D. oliveri***	***D. sissoo***
Number of sequences	34,655	48,591	43,848	31,678	43,879	32,753
Number of contigs with an ORF	23,440	30,717	26,763	22,084	26,737	22,887
Mean % of the contig covered by ORF	99.78	99.80	99.79	99.79	99.79	99.78
CRBB hit	14,963	14,820	15,668	14,778	15,395	14,365
Number of contigs with CRBB	14,963 (43.18%)	14,820 (30.50%)	15,668 (35.73%)	14,778 (46.65%)	15,395 (35.09%)	14,365 (43.86%)
RBH per reference (%)	35.76	35.42	37.45	35.32	36.80	34.33
Number of references with CRBB	13,465 (32.18%)	13,405 (32.04%)	13,649 (32.62%)	13,419 (32.07%)	13,618 (32.55%)	13,196 (31.54%)
Reference coverage (%)	34.38	34.21	34.47	34.43	34.42	34.10

Table 2.8. (a) Basic statistics of the Orthofinder results of the 16 Fabaceae species and (b) the number of shared orthogroups among species.

(a)

Number of genes	568,554
Number of genes in orthogroups	481,614 (84.7%)
Number of orthogroups	34,725
Number of species-specific orthogroups	925
Number of genes in species-specific orthogroups	3,465 (0.6%)
Number of orthogroups with all species present	5,493
Number of single-copy orthogroups	256

(b)

	Ad	Af	Ai	Bt	Cc	Cf	Co	Fr	Me	Mi	Ol	Si	La	Mp	Ns	Xz
Ad	16,595	10,149	15,648	10,595	12,256	11,498	12,346	12,337	12,501	12,221	12,441	12,266	12,013	10,540	12,416	10,806
Af	10,149	13,646	10,253	10,781	10,374	9,791	10,946	11,281	11,440	10,991	11,232	11,005	10,253	9,101	10,245	10,970
Ai	15,648	10,253	16,834	10,687	12,283	11,528	12,429	12,431	12,614	12,311	12,529	12,319	12,034	10,595	12,431	10,892
Bt	10,595	10,781	10,687	13,812	11,051	10,261	11,461	11,598	11,843	11,445	11,668	11,371	10,750	9,520	10,811	11,423
Cc	12,256	10,374	12,283	11,051	14,960	12,393	12,848	12,832	12,973	12,711	12,938	12,710	12,589	11,261	13,112	11,095
Cf	11,498	9,791	11,528	10,261	12,393	13,822	12,032	12,009	12,138	11,931	12,110	11,922	11,892	10,813	12,301	10,372
Co	12,346	10,946	12,429	11,461	12,848	12,032	18,209	15,388	15,739	14,696	16,060	14,769	12,475	11,052	13,022	11,754
Fr	12,337	11,281	12,431	11,598	12,832	12,009	15,388	21,650	17,347	15,693	17,047	16,873	12,464	11,003	12,978	11,925
Me	12,501	11,440	12,614	11,843	12,973	12,138	15,739	17,347	20,719	15,388	16,678	16,007	12,568	11,128	13,128	12,224
Mi	12,221	10,991	12,311	11,445	12,711	11,931	14,696	15,693	15,388	17,963	15,467	14,862	12,397	10,925	12,875	11,762
Ol	12,441	11,232	12,529	11,668	12,938	12,110	16,060	17,047	16,678	15,467	20,592	15,650	12,547	11,132	13,158	11,949

Si	12,266	11,005	12,319	11,371	12,710	11,922	14,769	16,873	16,007	14,862	15,650	19,090	12,394	10,939	12,882	11,681
La	12,013	10,253	12,034	10,750	12,589	11,892	12,475	12,464	12,568	12,397	12,547	12,394	13,882	10,826	12,668	10,902
Mp	10,540	9,101	10,595	9,520	11,261	10,813	11,052	11,003	11,128	10,925	11,132	10,939	10,826	12,882	11,254	9,592
Ns	12,416	10,245	12,431	10,811	13,112	12,301	13,022	12,978	13,128	12,875	13,158	12,882	12,668	11,254	15,005	11,024
Xz	10,806	10,970	10,892	11,423	11,095	10,372	11,754	11,925	12,224	11,762	11,949	11,681	10,902	9,592	11,024	14,319

Table 2.9. Estimation of node ages in the phylogeny of 16 Fabaceae species based on Bayesian analysis of a supergene from the 256 single-copy orthologs (479,064 bp) from their transcriptomes.

Node	**Median of node age**	**95% HPD of node age**
Bauhinia tomentosa–Acrocarpus fraxinifolius	80.1610	69.4302 – 95.4551
Bauhinia tomentosa–Cercis canadensis	46.8064	43.4005 – 50.6522
Acrocarpus fraxinifolius–Chamaecrista fasciculata	58.2181	56.9537 – 59.8703
Chamaecrista fasciculata–Mimosa pudica	43.8327	42.2482 – 46.0581
Acrocarpus fraxinifolius–Xanthocercis zambesiaca	64.0946	62.6783 – 65.9590
Xanthocercis zambesiaca–Lupinus augustifolius	57.1341	56.2045 – 58.7156
Lupinus augustifolius–Nissolia schottii	51.3041	50.4486 – 52.5991
Nissolia schottii–Arachis duranensis	41.8895	41.1156 – 42.8862
Arachis duranensis–Arachis ipaensis	3.2624	3.1225 – 3.4102
Arachis duranensis–Dalbergia miscolobium	25.5414	24.0776 – 27.1138
Dalbergia miscolobium–Dalbergia cochinchinensis	14.7808	13.7365 – 16.0182
Dalbergia cochinchinensis–Dalbergia oliveri	11.6900	9.6624 – 13.6183
Dalbergia cochinchinensis–Dalbergia frutescens	12.7026	10.4926 – 14.4774
Dalbergia frutescens–Dalbergia sissoo	10.0446	7.7557 – 12.8381
Dalbergia frutescens–Dalbergia melanoxylon	7.0789	4.8187 – 10.2959

Table 2.10. Counts of annotated GO terms (level 2) in 6 *Dalbergia* species, only showing terms that are significant different among the species (p-value < 0.05 in chi-square test of independence).

	Co	Fr	Me	Mi	Ol	Si	p-value
Cellular components							
cell	18918	25665	22120	18346	22471	18756	0.00E+00
cell junction	1056	1071	1166	986	1266	934	3.53E-13
cell part	18887	25610	22084	18305	22427	18715	0.00E+00
extracellular region	1587	2179	1650	1567	1819	1544	2.20E-35
extracellular region part	370	392	308	345	415	302	2.30E-05
membrane	8770	11238	10013	8320	10102	8345	2.88E-155
membrane-enclosed lumen	1495	2825	1876	1544	1859	1752	2.01E-131
membrane part	6488	8413	7749	6105	7521	6151	8.28E-138
nucleoid	47	91	76	66	61	79	4.67E-03
organelle	13715	18843	15907	13489	16265	13898	5.45E-302
organelle part	6824	10183	8093	6886	8407	7025	3.76E-229
protein-containing complex	3441	5836	4096	3677	4551	3818	1.52E-191
supramolecular complex	363	416	344	369	437	343	1.20E-03
symplast	991	981	1120	936	1166	878	1.34E-11
synapse	82	121	79	85	130	68	3.05E-06
synapse part	67	96	65	68	103	57	3.22E-04
Molecular functions							
antioxidant activity	199	290	218	201	247	234	1.27E-04
binding	16104	21436	19157	15169	19352	15228	0.00E+00
cargo receptor activity	3	19	11	6	10	5	2.44E-03
catalytic activity	11947	16786	14322	11596	14272	12024	0.00E+00
molecular carrier activity	38	85	57	42	48	55	9.03E-05
molecular function regulator	526	763	566	519	634	421	4.79E-24
molecular transducer activity	471	527	547	438	512	562	5.90E-04
structural molecule activity	843	1233	836	908	1301	778	1.90E-53
transcription regulator activity	1552	1839	1745	1487	1683	1512	7.30E-12
transporter activity	1608	2477	1954	1651	1863	1710	1.71E-57
Biological processes							

behavior	26	45	25	31	46	16	4.40E-04
biological adhesion	76	121	84	78	100	66	4.07E-04
biological regulation	6865	8847	7968	6534	8031	6527	1.11E-130
cell aggregation	34	31	10	46	52	21	5.27E-07
cellular component organization or biogenesis	4070	6294	4680	4021	4912	4280	6.98E-164
cellular process	15997	22257	18818	15510	19098	15839	0.00E+00
detoxification	58	149	100	83	82	85	1.06E-09
developmental process	3630	4014	4035	3567	4074	3411	9.97E-22
growth	778	925	881	730	878	721	8.57E-09
immune system process	1108	1217	1379	849	1394	760	1.88E-66
localization	3349	5211	3926	3409	4058	3535	2.10E-131
locomotion	173	233	156	217	254	169	1.86E-07
metabolic process	13803	19567	16249	13673	16395	14048	0.00E+00
multi-organism process	2627	2901	2935	2303	3053	2180	2.15E-50
multicellular organismal process	3462	3571	3866	3352	3822	3138	3.68E-22
negative regulation of biological process	1462	2104	1747	1478	1715	1519	8.36E-37
nitrogen utilization	15	58	25	16	24	14	1.36E-10
positive regulation of biological process	1745	2403	1977	1713	2057	1756	2.33E-37
regulation of biological process	6179	7853	7137	5833	7183	5816	6.84E-113
reproduction	2037	2244	2276	1955	2234	1849	8.84E-15
reproductive process	2014	2191	2254	1929	2201	1811	7.93E-15
response to stimulus	7706	9228	8974	6972	8989	6820	1.18E-154
signaling	2553	3053	2987	2219	3069	2086	3.99E-76

Table 2.11. Row Z scores of annotated Pfam domains (N = 17,929) of the 13 Fabaceae species, only showing domains (n = 91) that are significantly contracted (negative) or expanded (positive) in the *Dalbergia* species ($p < 0.05$ in two-tailed Fisher's exact test of independence).

Table 2.11 is too big to be included in the main text. It can be accessed at https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-020-74814-2/MediaObjects/41598_2020_74814_MOESM8_ESM.xlsx.

Table 2.12. Results of CAFE analysis on gene expansion/contraction events in (a) *D. cochinchinensis* and (b) *D. oliveri*, only showing 10 and 49 orthogroups that have significantly expanded or contracted (family $p < 0.05$ and Viterbi $p < 0.001$).

Table 2.12 is too big to be included in the main text. It can be accessed at (a) https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-020-74814-2/MediaObjects/41598_2020_74814_MOESM9_ESM.xlsx and (b) https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-020-74814-2/MediaObjects/41598_2020_74814_MOESM10_ESM.xlsx.

Table 2.13. GO enrichment for biological processes of significantly expanded gene families in CAFE analysis in (a) *D. cochinchinensis* and (b) *D. oliveri*, only showing GO terms that are significantly over- or under-presented ($p < 0.05$ in two-tailed Fisher's exact test of independence).

Table 2.13 is too big to be included in the main text. It can be accessed at as Supplementary Table 10 at https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-020-74814-2/MediaObjects/41598_2020_74814_MOESM1_ESM.docx.

Table 2.14. Results of PAML analysis on positive selection in (a) *D. cochinchinensis* and (b) *D. oliveri*, detecting 371 and 439 positively selected genes respectively, out of 9,054 single-copy orthologues tested (BH $p < 0.05$ in chi-square test of independence).

Table 2.14 is too big to be included in the main text. It can be accessed at (a) https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-020-74814-2/MediaObjects/41598_2020_74814_MOESM11_ESM.xlsx and (b) https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-020-74814-2/MediaObjects/41598_2020_74814_MOESM12_ESM.xlsx.

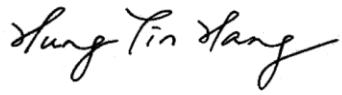
Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Reference transcriptomes and comparative analyses of six species in the threatened rosewood genus *Dalbergia*
Publication Status	☑Published □ Accepted for Publication □Submitted for Publication □Unpublished and unsubmitted work written in a manuscript style
Publication Details	Hung, T. H.*, So, T., Sreng, S., Thammavong, B., Boounithiphonh, C., Boshier, D. H. & MacKay, J. J.* (2020). Reference transcriptomes and comparative analyses of six species in the threatened rosewood genus *Dalbergia*. *Scientific Reports, 10,* 17749. http://doi.org/10.1038/s41598-020-74814-2

Student Confirmation

Student Name:	Tin Hang Hung	
Contribution to the Paper	Designed the study, conducted the sample preparation, conceived and conducted the bioinformatic analyses, drafted and revised the manuscript	
Signature 	Date	8 MAR 2022

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: John MacKay, Wood Professor of Forest Science		
Supervisor comments		
Signature 	Date	8 MAR 2022

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 3. The first chromosome-level genome assembly of Siamese rosewood *Dalbergia cochinchinensis* and comparative genomics of resistance genes

Tin Hang Hung[1,*], David H. Boshier[1], John J. MacKay[1,*]

1. Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, United Kingdom

3.1 Abstract

Dalbergia cochinchinensis produces valuable rosewood, which is the world's most trafficked wild product. The escalating demand exploits the natural populations and the species was classified as Vulnerable in the IUCN Red List in 1998 and regulated under CITES Appendix II since 2017. Despite the growing interest of incorporating genomic tools into conservation of threatened species like *D. cochinchinensis*, genomic resources remain scarce and disproportionate for endangered plants and in particular *Dalbergia* species. We presented a complete, chromosome-level assembly of *D. cochinchinensis* using a combination of ~260X long-read sequencing using PacBio and Oxford Nanopore and ~88.52X Hi-C proximity ligation. We produced 27,952 unique gene models, the majority of which were functionally annotated and are an improved version of a previous study. Protein domain prediction and motif discovery predicted the identity and arrangement of 166 resistance genes (R genes) in the genome of *D. cochinchinensis*. Systematic analysis in 44 species from Fabaceae and closely related families further revealed a large repertoire of R genes ranging from 32 to 2,763, the majority of which CNLs outnumbered and were positively correlated with TNLs. We also found that the number of R genes was significantly higher in perennial plants, which could imply the evolutionary significance of a larger repertoire of specific resistance responses in longer lifespans. Our study offers a reference-grade genome of *D. cochinchinensis*, which has important potential for use, such as discovering adaptive variation and marker-assisted selection for adaptive traits, and also studying legume evolution through R genes.

3.2 Introduction

Dalbergia cochinchinensis Pierre ($2n = 2x = 20$) is one of 250 species in the pantropical genus *Dalbergia* L. f., many of which produce extremely valuable rosewood timber. Rosewoods has been the most trafficked wild product since 2015 and *D. cochinchinensis* is the top trafficked rosewood species[1]. The growing demand for *D. cochinchinensis* is driving its exploitation to commercial extinction. It was classified as vulnerable in the IUCN Red List in 1998, with international trade regulated since 2017 under CITES Appendix II[2]. The species is endemic to Cambodia, Laos, Thailand, and Vietnam within the Indo-Chinese biodiversity hotspot. *D. cochinchinensis* is an intermediate pioneering species that fixes nitrogen, has fast juvenile growth, and drought tolerance[2], and thus is suitable for use in restoring degraded forests and deforested sites. Its conservation could produce ecological and economic benefits both locally and at regionally, and would be facilitated by an improved understanding of genomic properties.

The foundation of conservation genetics theory is based on the premise that small, isolated populations may be genetically threatened, where random genetic drift and inbreeding significantly decrease the genetic variation as a function of their effective sizes[3]. By supplementing conservation biology with genome-wide information, there are two dimensions of advancement: firstly, the increased density of markers across the genome would provide more informative parameters including genetic diversity, demographic history, and inbreeding coefficients[4]; secondly, the wider repertoire of genetic markers would provide more insight into evolutionary mechanisms and adaptations[5]. A reference genome will enable analysis, including characterisation of functional genes, higher-resolution species delineation, association mapping and adaptation, genetic rescue, and genome editing[5]. These in turn will help to address important conservation and forest restoration questions such as genetic

monitoring of introduced and relocated populations, predicting population viability, disease resistance, synthetic alternatives, and de-extinction[6,7].

Despite growing interest and the potential of genomic tools for conservation biology, genomic resources for endangered species remain critically scarce, most of which are non-model organisms[8]. Genomic resources are limited with disproportionate coverage across taxa. Of the 14,191 vascular plants that are listed as either Vulnerable, Endangered, and Critically Endangered in the IUCN Red List, only 0.1% have their genomes published, far fewer than the 1% reported for animals[9].

Genomic resources remain limited for the *Dalbergia* genus which consists of more than 250 species[10]. As of January 2022, the only available genomes are for *D. hupeana*[11] and *D. odorifera*[12], and seven transcriptomes[9,13]. In addition, previous studies in *Dalbergia* mainly focused on a limited number of loci and microsatellites[14–20], which potentially hinders a broader understanding of their evolutionary and ecological processes and the application of genetic tools to conservation and forest restoration.

Filling in the resource and knowledge gaps in *Dalbergia* will ultimately contribute to the understanding of legume evolution. The legume family (Fabaceae) evolved about 60 million years ago and is the third-largest angiosperm family after Orchidaceae and Asteraceae[21]. It contains many ecologically and economically important species. Most legumes host bacteria symbionts in their root nodules and thus are able to fix atmospheric nitrogen and improve soil fertility[22]. Many legumes are major agricultural crops, for example, soybean, peanut, and chickpea, which together contribute to more than 20% of global primary crop production[23]. However, their production has potentially been limited due to a wide range of plant diseases such as rusts, mildews, leaf blights, and bacterial wilts[24]. Most legume crops that have reference genomes are herbaceous; therefore, analysis of species with a perennial

and woody life habit such as *Dalbergia* will open the doors to comparative studies of legumes with a wider range of ecological niches.

Understanding the evolution of disease defence mechanisms enables assisted breeding in legumes, and could inform both conservation and breeding of endangered species like *D. cochinchinensis*. In particular, plants have evolved highly diverse resistance gene (R gene) families, which produce pathogen-resistance proteins (R proteins). R proteins contain highly conserved domains and motifs, and are involved in effector-triggered immunity for pathogen recognition[25]. Plant R proteins are characterised by an N-terminal effector domain, a NB-ARC domain (nucleotide-binding adaptor shared by APAF-1 [apoptotic protease-activating factor 1], R proteins, and CED-4 [the *Caenorhabditis elegans* homolog]) and have a series of LRRs (leucine-rich repeat) making up the sensor domain[26]. The N-terminal effector domain largely defines the R proteins, such as the TIR (toll interleukin-1 receptor) domain for the TNL (TIR-NB-ARC-LRR) subfamily and the coiled-coil (CC) structure for the CNL (CC-NB-ARC-LRR) subfamily[27]. Comprehensive genomic studies on R genes have focused on *Arabidopsis*[28] and crop species[29–31], but rarely on complete taxonomic groups. The only previous study on legumes identified rapid evolution of R genes under purifying selection in seven species[32].

This study presents a complete, chromosome-level assembly of *D. cochinchinensis* developed by using a combination of PacBio and Oxford Nanopore Technologies sequencing and Hi-C scaffolding. We produced the gene models with both *ab initio* prediction, a previously published transcriptome, and protein homology evidence with *Arabidopsis thaliana* and *Arachis ipaensis*. We then conducted genome-wide identification of R genes in *D. cochinchinensis* to understand their genomic architecture. We further predicted R genes in 44 genomes from Fabaceae and closely related families to conduct evolutionary analyses.

3.3 Methods

3.3.1 Plant materials and sample preparation

Dried seeds of *Dalbergia cochinchinensis* were collected from the Bolikhamxay provenance in Khamkend, Lao PDR by the Forest Research Center, Lao PDR in 2018 and stored at 4°C until germination. We scarified the seeds by placing them in 70°C distilled water, which was then left to cool to room temperature. After soaking the seeds in the water overnight, they were germinated in 1% agar in a plant growth cabinet MLR-350 (Sanyo, Watford, United Kingdom) at 25°C and photoperiod 12L/12D. Germinants were transferred to 3-litre pots in a soil-perlite 3:1 (v:v) substrate in a greenhouse at 30°C and 16L/8D photoperiod. The plants were regularly watered to pot capacity and fertilised once a week using N-P-K 20:20:20 fertiliser (Chempak, Suffolk, United Kingdom), and any moulded or diseased plants immediately removed. Leaf tissue was harvested from one selected 9-month-old individual and ground in liquid nitrogen with a mortar and pestle.

High-molecular-weight genomic DNA was extracted with Carlson lysis buffer (100 mM Tris-HCl, pH 9.5, 2% CTAB, 1.4 M NaCl, 1% PEG 8000, 20 mM EDTA) followed by purification using the QIAGEN Genomic-tip 500/G. The quantity and quality of genomic DNA were determined with NanoDrop 2000 (Thermo, Wilmington, United States) and Qubit 1 (ThermoFisher Scientific, United Kingdom). DNA integrity was preliminary assessed with a 0.4% agarose gel against a NEB Quick-Load® 1 kb Extend DNA Ladder. DNA samples passed the quality check only when a single band could be mapped near a lambda DNA band (~ 48.5 kb).

3.3.2 Nanopore library preparation and sequencing

Prior to library preparation, 9 µg of extracted DNA was size-selected using the Short Read Eliminator Kit v2.0 (Circulomics, Baltimore, United Kingdom). 400 ng of size-selected

DNA was used in each library preparation using the Rapid Sequencing Kit (SQK-RAD004) (Oxford Nanopore Technologies, United Kingdom). The library was sequenced on two R9.4 flow cells on a MinION platform (127.0.0.1). For each flow cell, the total sequencing time was ~ 72 hours. Every 24 hours, the sequencing was paused and the flow cell washed with the Flow Cell Wash Kit (EXP-WSH003). A new library was prepared each time and loaded on the same flow cell to resume the sequencing. Real-time basecalling was performed in MinKNOW release 19.10.1. Raw reads with a Phred score lower than 8 were filtered.

3.3.3 PacBio library preparation and sequencing

The Genomics & Cell Characterization Core Facility at the University of Oregon for PacBio prepared the DNA library and carried out the sequencing. Throughout the sample preparation, the quality of DNA was assessed using Fragment Analyzer 1.2.0.11 (Agilent, United States). 20 µg of unsheared genomic DNA was used for library preparation using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, United States). The library was size selected using the BluePippin system (Sage Science, United States) at 45 kb and then sequenced on a single SMRT 8M cell on a Sequel II System (2.0 chemistry) using the Continuous Long-Read Sequencing (CLR) mode with a movie time of 30 hours.

3.3.4 Hi-C proximity ligation library preparation and sequencing

We harvested 0.5 g of fresh leaf from the same reference individual and immediately cross-linked the finely chopped tissue in 1% formaldehyde for 20 minutes. The cross-linking was then quenched with glycine (125 mM). The cross-linked samples were then ground in liquid nitrogen with a mortar and pestle and shipped to Phase Genomics (Seattle, USA) for library preparation and sequencing. The Hi-C library was prepared with the restriction enzyme DpnII, proximity-ligated, and reverse-crosslinked using Proximo Hi-C Kit (Plant)

v2.0 (Phase Genomics, Seattle, USA). The library was sequenced on a HiSeq4000 for ~300 M 150-bp paired-end sequencing.

3.3.5 Assembly pipeline

The raw reads were obtained in BAM format from PacBio platform and converted to FASTA format using samtools 1.9[33]. PacBio and Nanopore raw reads shorter than 500 bp were filtered. We tried different assemblers for the *de novo* draft genome: Raven v0.0.7[34], Flye 2.7-b1585[35], wtdbg2[36], Canu 2.0[37]. Among all assemblers, Canu-assembly had the highest contiguity and retained halotigs which were useful for genome phasing. BUSCO[38] was used to estimate the redundancy of the draft genome. GenomeScope[39] was used to estimate the genome profile. We used purge_haplotigs v1.1.1[40] to collapse the redundant contigs by separating the primary assembly and haplotigs.

The Hi-C reads were mapped to the draft genome assembly using hicstuff 2.3.2[41] to generate the contact matrix, which was then used to scaffold and polish the assembly using instaGRAAL 0.1.2[42] with default options. The contact matrices before and after scaffolding were visualised as heatmaps of contact frequencies using hicstuff.

3.3.6 De novo repeat library

A *de novo* repeat library was constructed using RepeatModeler 2.0.1[43], which incorporated RECON 1.08[44], RepeatScout 1.0.6[45], and TRF 4.0.9[46] for identification and classification of repeat families. We then used RepeatMasker 4.1.1[47] to mask low complex or simple repeats only ("-noint"). A *de novo* library of long terminal repeat (LTR) retrotransposons was constructed on the simple-repeat-masked genome using LTRharvest[48] and annotated with the GyDB database and profile HMMs using LTRdigest[49] module in the genometools 1.6.1 pipeline. Predicted LTR elements with no protein domain hits were removed from the library. We applied the RepeatClassifier module in RepeatModeler to

format both repeat libraries. We merged the libraries together and clustered the sequences that were ≥ 80% identical by CD-HIT-EST 4.8.1[50] ("-aS 80 -c 0.8 -g 1 -G 0 -A 80") to produce the final repeat library.

3.3.7 Genome annotation

Filtered mRNA-sequencing data (50.5 Gbp) for *D. cochinchinensis* from a previous project[9] were aligned against the genome assembly using STAR v2.7.6[51]. The alignment was used to assemble the transcripts using the genome-guided mode of Trinity v2.11.0[52]. Protein sequences were obtained from *Arabidopsis thaliana* (Araport11)[53] and *Arachis ipaensis* (Araip1.1)[54]. After soft-masking the genome with the *de novo* repeat library using RepeatMasker (Dfam libraries 3.2), the transcript and protein evidences were used to produce gene models using MAKER 3.01.03[55]. The MAKER pipeline was iteratively run for two more rounds to produce the final gene models. In between each run of MAKER, the gene models were used to train the *ab initio* gene predictors SNAP (version 2006-07-28)[56] and AUGUSTUS 3.3.3[57] which were used in the MAKER pipeline. tRNA genes were predicted with tRNAscan-SE 1.3.1[58].

3.3.8 Nanopore methylation calling

Oxford Nanopore sequencing is capable of detecting base modifications. We applied Nanopolish 0.13.2[59] onto the raw Nanopore reads to detect 5-methylcytosine (5mC) in CpG sites. We filtered the methylation calls based on the log-likelihood ratio and calculated the frequency of methylation across various genomic regions.

3.3.9 Identification and evolution of R genes (NBS-LRR)

Whole genomes and their gene models of 44 species from the Fabaceae family and closely related families, including Betulaceae, Cucurbitaceae, Euphorbiaceae, Juglandaceae,

Phrymaceae, Rosaceae, Salicaceae, and Vitaceae were obtained from publicly available sources (Table 3.1).

We identified the *R*-genes in the legume species by filtering the gene models with Pfam entries NB-ARC (PF00931), TIR (PF01582), RPW8 (PF05659) and LRR 1 to 6, 8, 9, and LRRNT (PF00560, PF07723, PF07725, PF12799, PF13306, PF13516, PF13855, PF14580, and PF01462 respectively)[25] using HMMER 3.1b2[60] and the Pfam 34.0 database (March 2021)[61] with a e-value threshold of 10. We used the ncoils program to predict coiled-coil secondary structure elements in the putative *R*-genes. We used NLR-Annotator[62] to detect putative *R*-genes, while motif discovery and scanning were performed on MEME Suite 5.3.3 using the default parameters[63]. Putative *R*-genes were characterised by the presence of motif 1 (P-loop)[64], at least three consecutive motifs from the series 1, 6, 4, 5, 10, 3, 12, or 2 (NB-ARC domain), and at least one motif of 9, 11, or 19 (LRR-domain), and classified into CNL, NL, TCNL and TNL classes, based on the 20 sequence motifs characteristic of *R*-genes[29].

3.3.10 Sequence alignment, clustering, and phylogenetic analyses

We extracted the NB-ARC domains of the putative *R*-genes for phylogenetic analyses. To optimise for computational efficiency, sequences with similarity higher than 60% were clustered using CD-HIT. Multiple sequence alignment of the NB-ARC domains was performed on MAFFT 7.575[65] using the E-INS-i algorithm (--genafpair --maxiterate 1000). The nucleotide substitution model was tested on the multiple sequence alignment with ModelTest-NG 0.1.6[66] for likelihood scores. The best model (JTT) was used to construct the maximum likelihood phylogeny using RAXML-NG v. 1.0.2[67].

3.4 Results

3.4.1 Genome assembly

PacBio and Oxford Nanopore whole-genome sequencing of *Dalbergia cochinchinensis* produced 140 Gb and 25 Gb of long reads respectively, with a N50 of 9 Kb and 48.5 Kb respectively. The raw reads approximately had a 260-fold coverage of the *Dalbergia* haploid genome. K-mer analysis using GenomeScope revealed two distinct strong peaks (Figure 3.1), implying that the genome of *D. cochinchinensis* was highly heterozygous (1.26%). K-mer analysis also predicted that the genome size was 449.89 Mb, which was likely an underestimate because of the high heterozygosity. Among all tested assemblers, Canu produced the most contiguous, diploid-aware assembly of 1.35 Gb with 6,443 contigs, achieving a contig N50 of 1.35 Mb and the longest contig being 33.2 Mb at chromosome arm-length (Table 3.2). Purging the diploid assembly with an alignment value of 70% successfully reduced the redundancy of the genome, where BUSCO analysis showed that the duplicated genes went down from 80.3% to 4.0%, without overpurging or collapsing the genome size (Figure 3.2 and Table 3.3).

Hi-C proximity ligation yielded ~366 M read pairs totalling 54.97 Gb (~88.52-fold coverage of the haploid genome), which were used to scaffold the genome assembly. Of these, 25.37% were considered high quality and 6.82% were 10 Kbp apart. In total, 5.21% of the read pairs were informative for genome scaffolding (Table 3.4). instaGRAAL mapped 98.3% of the contig-level assembly into 10 pseudomolecules (Figure 3.3), which were considered the putative chromosomes. After polishing and gap-closing, the final assembly (Dacoc 1.2) of 621 Mb contained 512 scaffolds, with a scaffold N50 of 60.0 Mb (Figure 3.4).

To assess the completeness of our final assembly, we compared it to 2,326 core genes in the eudicotyledon lineage using BUSCO and obtained 98.0% of complete genes (4.3% duplicated), 0.3% fragmented, and only 1.7% missing. We also confirmed that the vast

majority (92.1%) of RNA-seq reads of leaves, stems, and roots from a previous sequencing experiment were uniquely mapped to the final assembly (Table 3.5).

3.4.2 Genome annotation

RepeatModeler and LTRharvest predicted 2,119 and 6,215 repeat families respectively. The two repeat libraries were merged after removing the redundant families. 64.80% of the genome was masked with the final repeat libraries. LTR elements were the dominant type of repeats (46.63%) found in the genome, such as Ty1/Copia (15.25%) and Gypsy/DIRS1 (30.51%), while 14.59% could not be identified (Table 3.6).

Iterative running of the MAKER pipeline finally produced 40,222 gene models, which were supported by transcript and protein evidence and *ab initio* prediction. These included 27,852 models predicted to be unique genes, with an average length of 4,284.20 bp. 98.3% of the gene models had an AED score below 0.5, which was considered a good gene model (Figure 3.5). 96.2% of the gene models were complete BUSCOs (92.1% single-copy), 1.3% fragmented, and only 2.5% missing.

The gene models were functionally annotated using different sources. We found a homologue match with the protein models of *Arachis ipaensis*, *Arabidopsis thaliana*, and SwissProt in 25,802 (92.6%), 24,820 (89.1%), and 21,468 (77.08%) gene models respectively. We obtained a target hit for at least one PFAM domain in 24,107 (86.55%) gene models. We predicted at least one transmembrane helix and a signal peptide (Sec/SPI) in 6,820 (24.49%) and 2,628 (9.44%) gene models respectively.

3.4.3 Genome-wide identification of R genes in D. cochinchinensis

Protein domain prediction and motif discovery based on Pfam and Meme predicted 166 putative R genes in the genome of *D. cochinchinensis*, with an average coding sequence

length of 2,466.99 bp. The R genes had different motif combinations made up by C (coiled coil), T (TIR), N (NR-ABC), and L (leucine-rich repeat). CNL, TNL, and TCNL were regarded putatively as full R genes while others were partial. 70 were complete CNLs and 42 were complete TNLs. Phylogenetic analysis confirmed that CNLs and TNLs formed distinct clades, except that one predicted CNL (Dacoc13882-RA_nlr_1) was more closely related to TNLs than other CNLs (Figure 3.6). The genomic landscape of CNLs and TNLs showed that they were distributed in all 10 assembled chromosomes (Figure 3.7), however unevenly ($P <$ 0.02648, Pearson's Chi-squared test).

3.4.4 Evolution of R genes

Protein domain prediction and motif discovery based on Pfam and Meme predicted 16,540 R genes by analysing the gene models of 44 genomes in the Fabaceae and eight other families of flowering plants (Table 3.7). The number of putative R genes ranged from 32 in *Cucurbita maxima* to 2,763 in *Rosa chinensis.* Average length of the coding sequences of all R genes identified was 2622.12 bp.

Motif discovery revealed the conventional domain organisation in angiosperms. N-terminus started with a pre-nucleotide-biding domain (motives 17, 16 in CNL; motives 18, 15, 13 in TNL), a conserved central NB-ARC (motives 1, 6, 4, 5, 10, 3, 12, 2), a linker region (motives 8, 7), and a highly polymorphic LRR region (motives 9, 11, 19). We observed that motif 11 (in the LRR domain) has the most distinct variation among the R proteins and did not cluster with any other motifs (Figure 3.8). The median number of motif 11 ranged from 2 in *Vigna radiata* to 12 in *Lupinus angustifolius*, with the extreme case of 46 in a putative CNL gene identified in *Arachis duranensis* (Figure 3.9). On the other hand, N-terminal domains, including the motives 17 and 16 in CNL and motifs 18, 15, 13 in TNL, were highly conserved and had high intercorrelation ($R > 0.8$, $P < 0.05$, Spearman's rank correlation)

(Figure 3.10). Phylogenetic analysis also confirmed that CNL and TNL formed distinct clades (Figure 3.11).

We found that the frequency of CNLs was positively correlated with that of TNLs, both considering all genomes (R = 0.86, $P = 5.4e^{-13} < 0.05$, Spearman's rank correlation) (Figure 3.12a) and only legume genomes (R = 0.83, $P = 5.5e^{-7} < 0.05$, Spearman's rank correlation) (Figure 3.12b).

We compared the frequency of R genes between annual and perennial plants, and found that there was a significant difference between the two. Considering all genomes, perennial plants had a higher frequency of complete R genes (Figure 3.13a). The median frequency of CNLs was significantly higher in perennial plants (174) than annual plants (73.5) (P = 0.031, Wilcoxon signed rank test). The median frequency of TNLs was also significantly higher in perennial plants (90) than annual plants (24.5) (P = 0.0065, Wilcoxon signed rank test). The trend held for most of the partial R genes: 22 CNs in perennial and 8.5 in annual (P = 0.0069), 12 Ns in perennial and 4 in annual (P = 0.0052), 27 NLs in perennial and 5.5 in annual (P = 0.0012), and 14 TNs in perennial and 4 in annual (P = 0.062). The same trend was observed when only considering legume genomes, but was not statistically significant, potentially due to the smaller sample size (Figure 3.13b): the median frequency of CNLs was higher in perennial plants (122) than annual plants (85.5) (P = 0.32), and that of TNLs was higher in perennial plants (80.5) than annual plants (42) (P = 0.078). We observed no significant difference in any class of R genes between Fabaceae and non-Fabaceae species ($P > 0.05$) (Figure 3.14).

3.5 Discussion

3.5.1 The first genome assembly and improved gene models

We developed a chromosome-level genome of *Dalbergia cochinchinensis* which provides a genomic resource for the study of the species and other legumes. The K-mer analysis, BUSCO analysis, and high mapping rate of transcriptome sequences indicated the completeness of our genome assembly. We obtained 27,852 high-confidence unique gene models with both *ab initio* prediction and protein and transcriptome evidence. Our study substantially improved the gene models from our previous dataset based on transcriptome assemblies[9], where the average transcript length increased from 966.12 to 1826.50 bp and the complete BUSCO score increased from 92.2 to 96.2. More genes were also functionally annotated, including their protein homology and prediction of presence of transmembrane helices and signal peptides.

The number and quality of plant genome assemblies have been drastically increased, owing to the advance of sequencing and bioinformatic technology over the past decades. As of January 2021, 798 land plant genome assemblies are publicly available and contig N50 increased from 99.5 Kbp to 3.40 Mbp between 2010 and 2020[68]. In particular, long-read sequencing (e.g. PacBio, Oxford Nanopore) helps produce more contiguous contigs while optical mapping (e.g. BioNano) and chromosomal conformation capture (e.g. 3C, Hi-C) help reach the chromosome-level structure[42]. However, available plant genomes still have a major over-representation of economically important and model species with small diploid genomes and there is a paucity of wild species. This imbalance is even more extreme for endangered plants, where only ~0.1% of threatened (Vulnerable, Endangered and Critically Endangered on the IUCN Red List) vascular plants had published genome on NCBI as of 2020[9], constraining the application of genomic tools to species conservation.

3.5.2 R genes in the genome of D. cochinchinensis

We have characterised the complete set of 112 CNL and TNL genes in this genome assembly of *D. cochinchinensis* (Dacoc_1.2), which represent ~0.40% of all gene models. CNL genes outnumber TNL genes by nearly two to one in the *D. cochinchinensis* genome. Our phylogenetic tree based on the NB-ARC domain also confirmed that the CNL clade has longer branch lengths, consistent with the previous observations that CNL genes have accumulated a greater degree of evolutionary change and represent a more ancient group than TNL genes[69]. However, different evolutionary patterns of relative frequency between CNL and TNL genes have been observed when considering land plants. For example, similar to *D. cochinchinensis*, CNL sequences are found to be more numerous in rice[70], potato[29], grapevine, and poplar[71], ranging from approximately 1.7× to 3.8× relative to TNLs. On the other hand, TNL sequences are more numerous in *Arabidopsis*[28] and field mustard[72]. In the latter cases, TNL genes may have undergone a recent amplification and outnumbers CNL genes in specific lineages[28]. TNLs are not found in the order Lamiales, early-diverged dicots (*Aquilegia coerulea*) and all monocots[73]. However, they are abundant among conifers, which may suggest a parallel evolved mechanism to trigger immunity[25].

The presence of multiple collinear R genes, usually defined as within regions up to 1.5 Mb, on chromosomes 2, 5, 7, and 10 of the *D. cochinchinensis* genome is a frequently observed phenomenon in plant genomes. Clustering of R genes can emerge from tandem gene duplication, gene conversion, and unequal crossing-over during recombination[74], and may further lead to new resistance specificities[75]. Clustering of gene sequences can enable co-transcribed gene expression, for example, *PigmR* and *PigmS* reside in the same epigenetically regulated cluster and function antagonistically in broad-spectrum resistance to blast fungus[76].

3.5.3 Molecular evolution and R genes in legumes

R genes comprise one of the largest gene families in plants and contribute to substantial genetic variation[70]. They have an ancient origin and have been identified in the ancestors of early land plants. In particular, a novel gene family has been identified in moss (*Physcomitrella patens*) with sequence homology to TNL genes. We identified a high abundance of both CNL and TNL genes in legumes.

The most recent study dates the origin of legumes to ~66 MYA (million years ago) with in the Maastrichtian or Early Paleocene, which is closely associated with the Cretaceous–Paleogene (K–Pg) boundary (KPB)[77]. It also provided evidence for at least three WGD events in the early evolution of legumes, which could explain the higher abundance of multiple-copy orthologs than single-copy orthologs in legumes[32]. The most recent study on legumes found a range of 227–952 R genes in 7 legume species, while our study identified 36–889 R genes across 24 legume species, which was made possible with the availability of new legume genomes[32]. The number of R genes did not correlate with genome size, which may be explained by the rapid gain-and-loss events in R genes during evolution of specific lineages.

We report a consistent trend of higher frequency and a positive correlation when considering the number of CNL genes compared to TNL genes in both legumes and all 44 species we surveyed. This contrasts with previous studies suggesting the distinct expansion of R proteins with TIR domain in dicots when compared to monocots[78], and also an additional expansion in legumes conferring diversified resistance functions[32]. Although R genes are generally responsible for specific effector recognition of pathogens to activate immune responses, R genes are hypothesised to be part of various immune signalling pathways[27]. CNLs may signal through the non-race specific disease resistance 1 gene (NDR1), whose loss of function results in susceptibility to bacterial and fungal pathogens with a repertoire of

effectors[79]. TNLs signal through the enhanced disease susceptibility 1 gene (EDS1), which is recruited with its interacting partners, phytoalexin deficient 4 (PAD4) and senescence-associated carboxylesterase 101 (SAG101) to elicit immune response[80,81].

We also note the high variability of length in the leucine-rich repeats (LRR) among legume species in contrast to the relatively conserved central NB-ARC domain. The LRR domains are shown to be the most polymorphic part of R proteins[82], and are variable in length with a tandem repeat of 10–40 short LRRs[69]. They have a consensus sequence of LxxLxLxxNxL (where L is a leucine, x any residue, and N an asparagine, threonine, serine, or cysteine), which is critical for the structural arrangement of the domain[83]. They are regulatory in nature and physically associated with the NB-ARC domain, as demonstrated in studies of Bs2, resistance to *Pseudomonas syringae* 5 (RPS5), and Rx[84–86]. In particular, a study reported that the deletion of only the first four LRRs would result in auto-activation of RPS5[87]. Therefore, the variability of length of LRR repeats contribute to diversification and regulation of immune response to specific pathogens, with a high level of diversifying selection[88].

The higher abundance of R genes in perennial plants imply their evolutionary significance. Plants need to adapt to the changing environments over their lifespan, and in particular, to defend against invading pathogens *in situ*. Long-lived perennials, such as trees are exposed to a rapidly evolving pathogen environment. Our results are consistent with a previous literature review on 13 species that perennial plants possess a larger and more diverse repertoire of R genes[89]. Favoured, but parallel, expansion of R gene families is also observed across multiple long-lived tree species, suggesting that immunity marks an essential contribution to longevity[90]. Defence capacity can be further amplified by alternative splicing and post-translational modification, on top of the existing gene diversity, at transcript and protein level[91–93].

3.5.4 Implications of the Dalbergia reference genomes for conservation

Conserving genomic variation is an important target for conservation across the tree of life and could be facilitated by reference genomes. Maintaining genetic variation and maximising effective population size (N_e) remains one of the primary goals of conservation[94]. Below we discuss briefly three potential avenues to incorporate reference genomes in plant conservation and restoration.

First, reference genomes are powerful tools for species delimitation and our understanding of the speciation process. Lack of a robust species taxonomy is a key barrier to conservation assessment, as species delimitation is critical for assessing distribution, population abundance, and threats to target species[95]. For example, a previous barcoding study confirmed that *D. bariensis*, *D. mammosa*, and *D. dognaiensis* were all synonyms of the same species *D. oliveri*[17]. At the same time, discriminating species with risk of illegal harvesting or trafficking can help authorities with legal enforcement such as for CITES or national conservation laws. Reference genomes allow discovery of new polymorphic markers, which in turn provides higher statistical power for species delimitation.

Second, they enable the study of adaptive variation. Conservation theory of population persistence focuses on the loss of potentially adaptive alleles and subsequent reduced productivity and fitness under novel selection pressure brought by rapid environmental change[96]. While traditional conservation genetic studies usually focus on neutral molecular data such as microsatellites or a limited number of loci[97], new evidence has suggested that neutral genetic diversity does not necessarily predict adaptive potential of populations[98]. Reference genomes help understand functional genetic variation which is relevant for future climate change, especially through identifying the significant environmental factors and associated risks. This helps guide the delineation of conservation units and assisted migration to facilitate adaptation of endangered populations.

Third, they can assist the discovery of molecular markers that inform adaptive traits including growth, known as marker assisted selection (MAS). Development of genetic resources and artificial propagation have been central parts of many conservation programmes, especially in plants, to multiply genetic materials for restoration and reintroduction[99]. Mathematical models show that if the between-individual variation is controlled by equalizing family size and the within-individual variation minimized by MAS, N_e of diploid species can in theory be maximised until infinity [94]. An example of a favourable trait to be selected is disease resistance, as epidemics of insects and diseases, both introduced and native, pose substantial ecological damage and economic loss[100,101]. Previous research has identified important markers for genotype screening of resistance in specific species, for example, *Ruv2* confers rust resistance in cowpea[102].

3.6 Conclusion

The potential application of genomic tools for conservation practice has been clearly highlighted but is still limited in practice, constrained by the disproportionally scarce genomic resources for endangered species. Our genome of *D. cochinchinensis* provides a valuable reference for the endangered genus and more widely the Fabaceae family, which has been mainly studied for its crop species. The present study particularly enhanced our understanding of the evolution of resistance genes and its implications. We believe that this paper will stimulate more intraspecific studies on *D. cochinchinensis* and comparative genomic studies on Fabaceae of relevance for their evolution and conservation.

3.7 Acknowledgements

This work was supported by funding to T.H.H. from the Biotechnology and Biological Sciences Research Council (BBSRC) [grant number BB/M011224/1]; to T.H.H., J.J.M. from the Google Cloud Academic Grant; and to D.H.B., J.J.M from the UK Darwin Initiative (ref. 25-023).

3.8 References

1. UNODC. *World Wildlife Crime Report: Trafficking in Protected Species*. (United Nations Publication, 2016).
2. Hung, T. H. *et al.* Physiological responses of rosewoods *Dalbergia cochinchinensis* and *D. oliveri* under drought and heat stresses. *Ecol. Evol.* (2020) doi:10.1002/ece3.6744.
3. Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. (Kuke) & Hedrick, P. W. Conservation genetics in transition to conservation genomics. *Trends Genet.* **26**, 177–187 (2010).
4. Benestan, L. M. *et al.* Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Mol. Ecol.* **25**, 2967–2977 (2016).
5. Supple, M. A. & Shapiro, B. Conservation of biodiversity in the genomics era. *Genome Biol.* **19**, (2018).
6. Allendorf, F. W., Hohenlohe, P. A. & Luikart, G. Genomics and the future of conservation genetics. *Nat. Rev. Genet.* **11**, 697–709 (2010).
7. Desalle, R. & Amato, G. Conservation Genetics, Precision Conservation, and De-extinction. *Hastings Cent. Rep.* **47**, S18–S23 (2017).
8. Barbosa, S. *et al.* Integrative approaches to guide conservation decisions: Using genomics to define conservation units and functional corridors. *Mol. Ecol.* **27**, 3452–3465 (2018).
9. Hung, T. H. *et al.* Reference transcriptomes and comparative analyses of six species in the threatened rosewood genus *Dalbergia*. *Sci. Rep.* **10**, 17749 (2020).
10. Vatanparast, M. *et al.* First molecular phylogeny of the pantropical genus *Dalbergia*: Implications for infrageneric circumscription and biogeography. *South African J. Bot.* **89**, 143–149 (2013).

11. Li, C., Zheng, Y., Liu, Y., Lin, F. & Huang, P. Development of genomic SSR for the subtropical hardwood tree *Dalbergia hupeana* and assessment of their transferability to other related species. *For. 2021, Vol. 12, Page 804* **12**, 804 (2021).

12. Hong, Z. *et al.* The chromosome-level draft genome of *Dalbergia odorifera*. *Gigascience* **9**, 1–8 (2020).

13. Liu, F.-M. *et al.* De novo transcriptome analysis of *Dalbergia odorifera* and transferability of SSR markers developed from the transcriptome. *Forests* **10**, 98 (2019).

14. Bhagwat, R. M., Dholakia, B. B., Kadoo, N. Y., Balasundaran, M. & Gupta, V. S. Two new potential barcodes to discriminate *Dalbergia* species. *PLoS One* **10**, 1–18 (2015).

15. Wattoo, J. I., Saleem, M. Z., Shahzad, M. S., Arif, A. & Hameed, A. DNA Barcoding: Amplification and sequence analysis of *rbcL* and *matK* genome regions in three divergent plant species. *Adv. Life Sci.* **4**, 03–07 (2016).

16. Phong, D. T., Tang, D. Van, Hien, V. T. T., Ton, N. D. & Van, H. N. Nucleotide diversity of a nuclear and four chloroplast DNA regions in rare tropical wood species of *Dalbergia* in Vietnam: a DNA narcode identifying utility. *Asian J. Appl. Sci.* **02**, 116–125 (2014).

17. Hartvig, I., Czako, M., Kjær, E. D., Nielsen, L. R. & Theilade, I. The use of DNA barcoding in identification and conservation of rosewood (*Dalbergia* spp.). *PLoS One* **10**, (2015).

18. Hartvig, I. *et al.* Population genetic structure of the endemic rosewoods *Dalbergia cochinchinensis* and *D. oliveri* at a regional scale reflects the Indochinese landscape and life-history traits. *Ecol. Evol.* **8**, 530–545 (2018).

19. RIBEIRO, R. A. *et al.* Development of polymorphic microsatellite markers for *Dalbergia nigra* (Papilionoideae), an endangered tree from the Brazilian Atlantic

Forest. *Mol. Ecol. Resour.* **9**, 203–206 (2009).

20. Resende, L. C., Ribeiro, R. A. & Lovato, M. B. Diversity and genetic connectivity among populations of a threatened tree (*Dalbergia nigra*) in a recently fragmented landscape of the Brazilian Atlantic Forest. *Genetica* **139**, 1159–1168 (2011).
21. O'Rourke, J. A., Bolon, Y. T., Bucciarelli, B. & Vance, C. P. Legume genomics: understanding biology through DNA and RNA sequencing. *Ann. Bot.* **113**, 1107 (2014).
22. Wang, Q., Liu, J. & Zhu, H. Genetic and molecular mechanisms underlying symbiotic specificity in legume-rhizobium interactions. *Front. Plant Sci.* **9**, 313 (2018).
23. Schaefer, H. *et al.* Systematics, biogeography, and character evolution of the legume tribe Fabeae with special focus on the middle-Atlantic island lineages. *BMC Evol. Biol.* **12**, 1–19 (2012).
24. Kankanala, P., Nandety, R. S. & Mysore, K. S. Genomics of plant disease resistance in legumes. *Front. Plant Sci.* **10**, 1345 (2019).
25. Van Ghelder, C. *et al.* The large repertoire of conifer NLR resistance genes includes drought responsive and highly diversified RNLs. *Sci. Rep.* **9**, 1–13 (2019).
26. Urbach, J. M. & Ausubel, F. M. The NBS-LRR architectures of plant R-proteins and metazoan NLRs evolved in independent events. *PNAS* **114**, 1063–1068 (2017).
27. van Wersch, S., Tian, L., Hoy, R. & Li, X. Plant NLRs: The whistleblowers of plant immunity. *Plant Commun.* **1**, 100016 (2020).
28. Meyers, B. C., Kozik, A., Griego, A., Kuang, H. & Michelmore, R. W. Genome-Wide analysis of NBS-LRR–encoding genes in *Arabidopsis*. *Plant Cell* **15**, 809 (2003).
29. Jupe, F. *et al.* Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics 2012 131* **13**, 1–14 (2012).
30. Gebhardt, C. & Valkonen, J. P. T. Organization of genes controlling disease resistance

in the potato genome. *Annu. Rev. Phytopathol.* **39**, 79–102 (2001).

31. Luo, S. *et al.* Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes in the grass family. *Plant Physiol.* **159**, 197–210 (2012).
32. Zheng, F. *et al.* Molecular phylogeny and dynamic evolution of disease resistance genes in the legume family. *BMC Genomics* **17**, 1–13 (2016).
33. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, 1–4 (2021).
34. Vaser, R. & Šikić, M. Time- and memory-efficient genome assembly with Raven. *Nat. Comput. Sci. 2021 15* **1**, 332–336 (2021).
35. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol. 2019 375* **37**, 540–546 (2019).
36. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods 2019 172* **17**, 155–158 (2019).
37. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, gr.215087.116 (2017).
38. Manni, M., Berkeley, M. R., Seppey, M., Sim~ Ao, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
39. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1–10 (2020).
40. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 1–10 (2018).

41. Matthey-Doret, C. *et al.* koszullab/hicstuff: Use miniconda layer for docker and improved P(s) normalisation. (2020) doi:10.5281/ZENODO.4066363.

42. Baudry, L. *et al.* InstaGRAAL: Chromosome-level quality scaffolding of genomes using a proximity ligation-based scaffolder. *Genome Biol.* **21**, 1–22 (2020).

43. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 9451–9457 (2020).

44. Bao, Z. & Eddy, S. R. Automated de novo identification of repeat sequence families in sequenced henomes. *Genome Res.* **12**, 1269 (2002).

45. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21** Suppl 1, (2005).

46. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).

47. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.* **25**, 4.10.1-4.10.14 (2009).

48. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 1–14 (2008).

49. Steinbiss, S., Willhoeft, U., Gremme, G. & Kurtz, S. Fine-grained annotation and classification of de novo predicted LTR retrotransposons. *Nucleic Acids Res.* **37**, 7002–7013 (2009).

50. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

51. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15 (2013).

52. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the

Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).

53. Cheng, C.-Y. *et al.* Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* **89**, 789–804 (2017).

54. Bertioli, D. J. *et al.* The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. *Nat. Genet.* **48**, 438–446 (2016).

55. Holt, C. & Yandell, M. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).

56. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 1–9 (2004).

57. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).

58. Chan, P. P. & Lowe, T. M. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* **1962**, 1 (2019).

59. Simpson, J. T. *et al.* Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods 2017 144* **14**, 407–410 (2017).

60. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**, e121–e121 (2013).

61. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).

62. Steuernagel, B. *et al.* The NLR-Annotator Tool Enables Annotation of the Intracellular Immune Receptor Repertoire. *Plant Physiol.* **183**, 468–482 (2020).

63. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic*

Acids Res. **43**, W39 (2015).

64. BC, M. *et al.* Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* **20**, 317–332 (1999).

65. Nakamura, T., Yamada, K. D., Tomii, K. & Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**, 2490–2492 (2018).

66. Darriba, Di. *et al.* ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Mol. Biol. Evol.* **37**, 291–294 (2020).

67. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).

68. Marks, R. A., Hotaling, S., Frandsen, P. B. & VanBuren, R. Representation and participation across 20 years of plant genome sequencing. *Nat. Plants 2021 712* **7**, 1571–1578 (2021).

69. Cannon, S. B. *et al.* Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. *J. Mol. Evol.* **54**, 548–562 (2002).

70. Yang, S. *et al.* Genome-wide investigation on the genetic variations of rice disease resistance genes. *Plant Mol. Biol.* **62**, 181–193 (2006).

71. Yang, S., Zhang, X., Yue, J. X., Tian, D. & Chen, J. Q. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genomics* **280**, 187–198 (2008).

72. Mun, J. H., Yu, H. J., Park, S. & Park, B. S. Genome-wide identification of NBS-encoding resistance genes in *Brassica rapa*. *Mol. Genet. Genomics* **282**, 617–631 (2009).

73. Collier, S. M., Hamel, L. P. & Moffett, P. Cell death mediated by the N-Terminal

domains of a unique and highly conserved class of NB-LRR Protein. *http://dx.doi.org/10.1094/MPMI-03-11-0050* **24**, 918–931 (2011).

74. Field, B. & Osbourn, A. Order in the playground. *http://dx.doi.org/10.4161/mge.19348* **2**, 46–50 (2012).
75. Hulbert, S. H., Webb, C. A., Smith, S. M. & Sun, Q. Resistance gene complexes: evolution and utilization. *Annu. Rev. Phytopathol.* **39**, 285–312 (2001).
76. Zhai, K. *et al.* RRM transcription factors interact with NLRs and regulate broad-spectrum blast resistance in rice. *Mol. Cell* **74**, 996-1009.e7 (2019).
77. Koenen, E. J. M. *et al.* The origin of the legumes is a complex paleopolyploid phylogenomic tangle closely associated with the Cretaceous–Paleogene (K–Pg) mass extinction event. *Syst. Biol.* **70**, 508–526 (2021).
78. Zhang, R., Murat, F., Pont, C., Langin, T. & Salse, J. Paleo-evolutionary plasticity of plant disease resistance genes. *BMC Genomics* **15**, (2014).
79. Century, K. S., Holub, E. B. & Staskawicz, B. J. NDR1, a locus of Arabidopsis thaliana that is required for disease resistance to both a bacterial and a fungal pathogen. *Proc. Natl. Acad. Sci.* **92**, 6597–6601 (1995).
80. Wiermer, M., Feys, B. J. & Parker, J. E. Plant immunity: the EDS1 regulatory node. *Curr. Opin. Plant Biol.* **8**, 383–389 (2005).
81. Zhu, S. *et al.* SAG101 forms a ternary complex with EDS1 and PAD4 and is required for resistance signaling against turnip crinkle virus. *PLOS Pathog.* **7**, e1002318 (2011).
82. Qi, D. & Innes, R. W. Recent advances in plant NLR structure, function, localization, and signaling. *Front. Immunol.* **4**, 348 (2013).
83. Stange, C., Matus, J. T., Domínguez, C., Perez-Acle, T. & Arce-Johnson, P. The N-homologue LRR domain adopts a folding which explains the TMV-Cg-induced HR-like response in sensitive tobacco plants. *J. Mol. Graph. Model.* **26**, 850–860 (2008).

84. Moffett, P., Farnham, G., Peart, J. & Baulcombe, D. C. Interaction between domains of a plant NBS–LRR protein in disease resistance-related cell death. *EMBO J.* **21**, 4511–4519 (2002).

85. Leister, R. T. *et al.* Molecular genetic evidence for the role of SGT1 in the intramolecular complementation of Bs2 protein activity in *Nicotiana benthamiana*. *Plant Cell* **17**, 1268–1278 (2005).

86. Ade, J., DeYoung, B. J., Golstein, C. & Innes, R. W. Indirect activation of a plant nucleotide binding site–leucine-rich repeat protein by a bacterial protease. *Proc. Natl. Acad. Sci.* **104**, 2531–2536 (2007).

87. Qi, D., de Young, B. J. & Innes, R. W. Structure-function analysis of the coiled-coil and leucine-rich repeat domains of the RPS5 disease resistance protein. *Plant Physiol.* **158**, 1819–1832 (2012).

88. Wang, C. I. A. *et al.* Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *Plant Cell* **19**, 2898–2912 (2007).

89. Tobias, P. A. & Guest, D. I. Tree immunity: growing old without antibodies. *Trends Plant Sci.* **19**, 367–370 (2014).

90. Plomion, C. *et al.* Oak genome reveals facets of long lifespan. *Nat. Plants* **4**, 440–452 (2018).

91. Gutierrez, J. R. *et al.* Prf immune complexes of tomato are oligomeric and contain multiple Pto-like kinases that diversify effector recognition. *Plant J.* **61**, 507–518 (2010).

92. Ashikawa, I. *et al.* Two adjacent nucleotide-binding site–leucine-rich repeat class genes are required to confer pikm-specific rice blast resistance. *Genetics* **180**, 2267–2276 (2008).

93. Ghosh, J. *et al.* Sp185/333: A novel family of genes and proteins involved in the purple sea urchin immune response. *Dev. Comp. Immunol.* **34**, 235–245 (2010).

94. Wang, J. & Hill, W. G. Marker-assisted selection to increase effective population size by reducing Mendelian segregation variance. *Genetics* **154**, 475 (2000).

95. Vogel Ely, C., Bordignon, S. A. de L., Trevisan, R. & Boldrini, I. I. Implications of poor taxonomy in conservation. *J. Nat. Conserv.* **36**, 10–13 (2017).

96. Jump, A. S., Marchant, R. & Peñuelas, J. Environmental change and the option value of genetic diversity. *Trends Plant Sci.* **14**, 51–58 (2009).

97. Kirk, H. & Freeland, J. R. Applications and implications of neutral versus non-neutral markers in molecular ecology. *Int. J. Mol. Sci.* **12**, 3966 (2011).

98. Teixeira, J. C. & Huber, C. D. The inflated significance of neutral genetic diversity in conservation genetics. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

99. Jones, T. A. Conservation biology and plant breeding: special considerations for the development of native plant materials for use in restoration. *Ecol. Restor.* **27**, 8–11 (2009).

100. Wingfield, M. J., Slippers, B., Wingfield, B. D. & Barnes, I. The unified framework for biological invasions: a forest fungal pathogen perspective. *Biol. Invasions 2017 1911* **19**, 3201–3214 (2017).

101. Brockerhoff, E. G. & Liebhold, A. M. Ecology of forest insect invasions. *Biol. Invasions* **19**, 3141–3159 (2017).

102. Wu, X. *et al.* Fine mapping Ruv2, a new rust resistance gene in cowpea (*Vigna unguiculata*), to a 193-kb region enriched with NBS-type genes. *Theor. Appl. Genet.* **131**, 2709–2718 (2018).

3.9 Figures and Tables

Figure 3.1. K-mer analysis and GenomeScope profile of raw reads from PacBio and Oxford Nanopore sequencing.

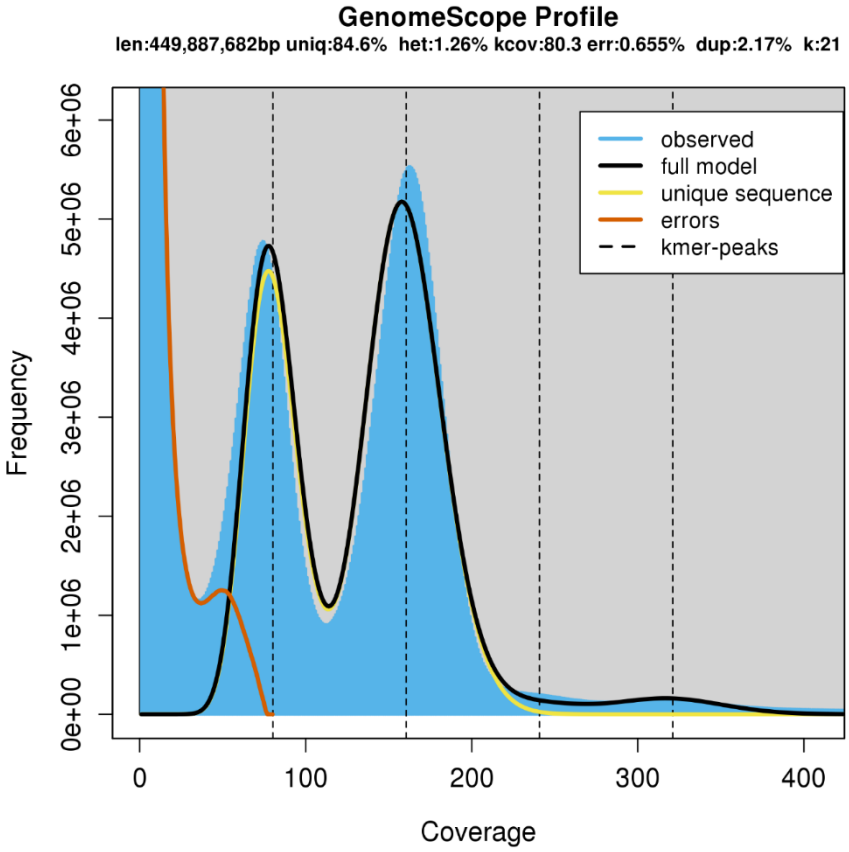


Figure 3.2. Coverage histogram of cleaned reads mapped on the unpurged genome assembly.

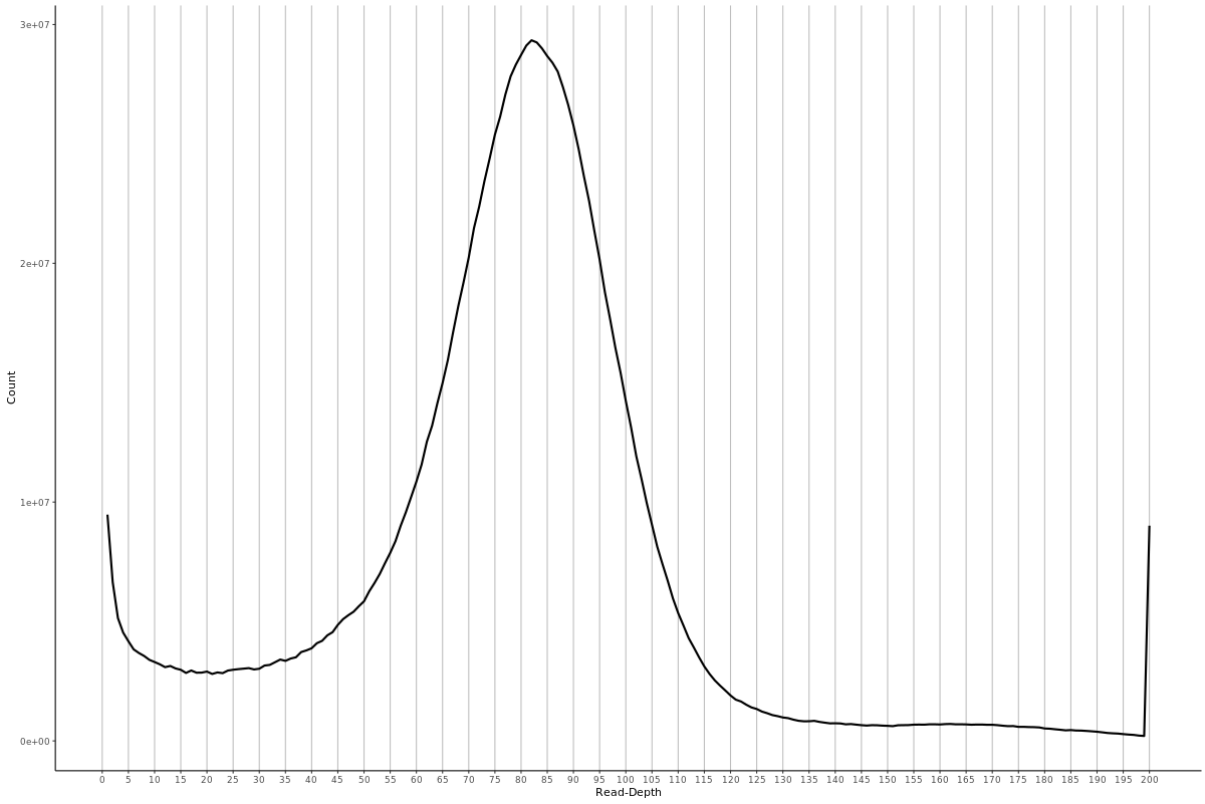
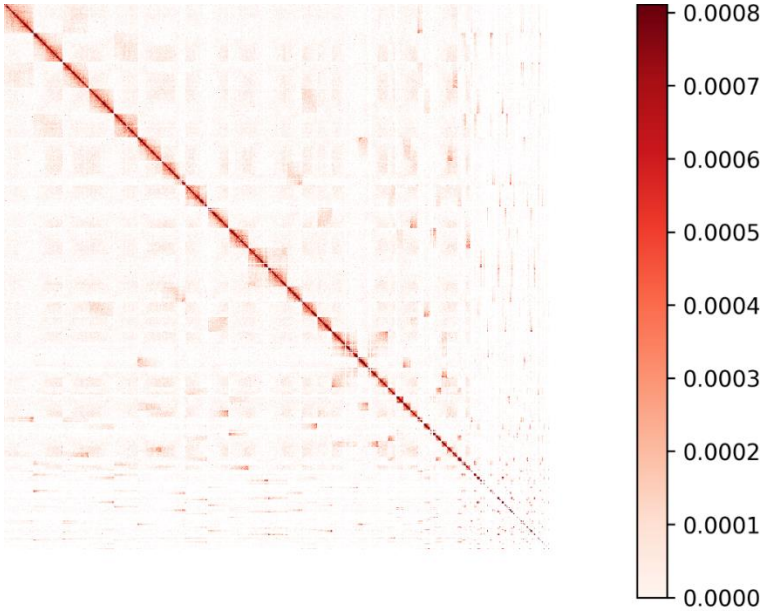


Figure 3.3. High-resolution contact probability map of **(a)** draft genome assembly after purging and **(b)** final genome assembly after scaffolding, revealing the 10 pseudomolecules (putative chromosomes) of the *D. cochinchinensis* genome at 100 Kbp resolution.

(a)



(b)

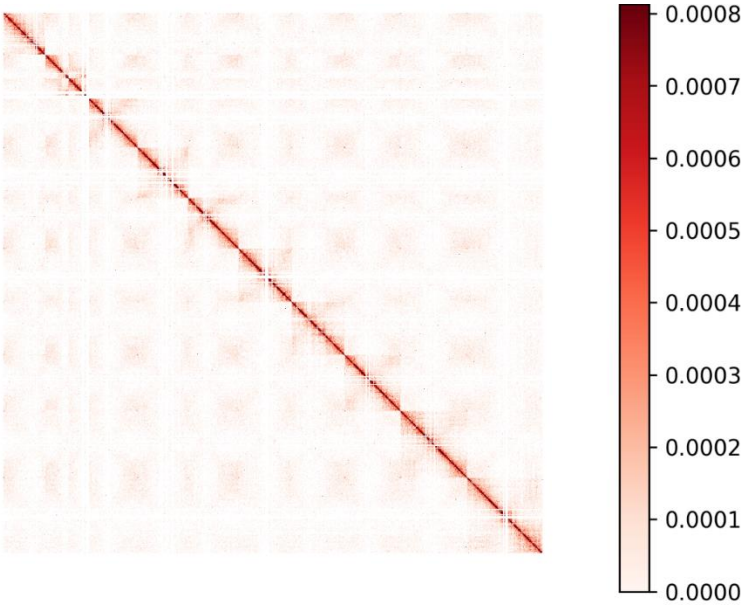


Figure 3.4. Genomic landscape of the 10 assembled chromosomes (pseudomolecules) of *D. cochinchinensis*, showing (a) tick marks every 1 Mb, (b) gene density, (c) repeat density, (d) 5-mC density, and (e) interchromosomal syntenic arrangement. (b) – (e) are calculated in 1-Mb sliding window.

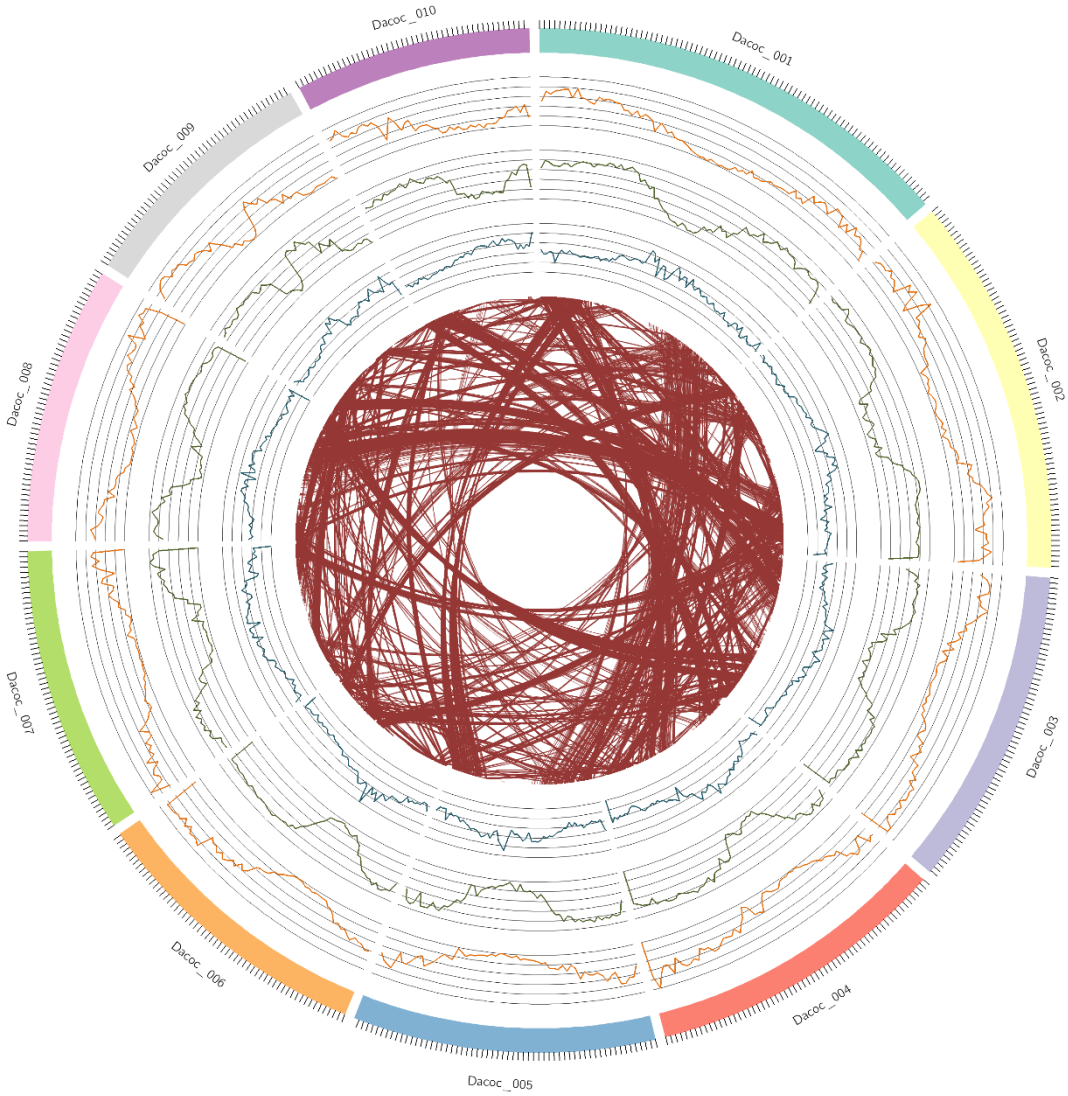


Figure 3.5. Cumulative frequency of gene models of Dacoc_1.2 against annotation edit distance (AED). A gene model with a AED below 0.5 is considered good.

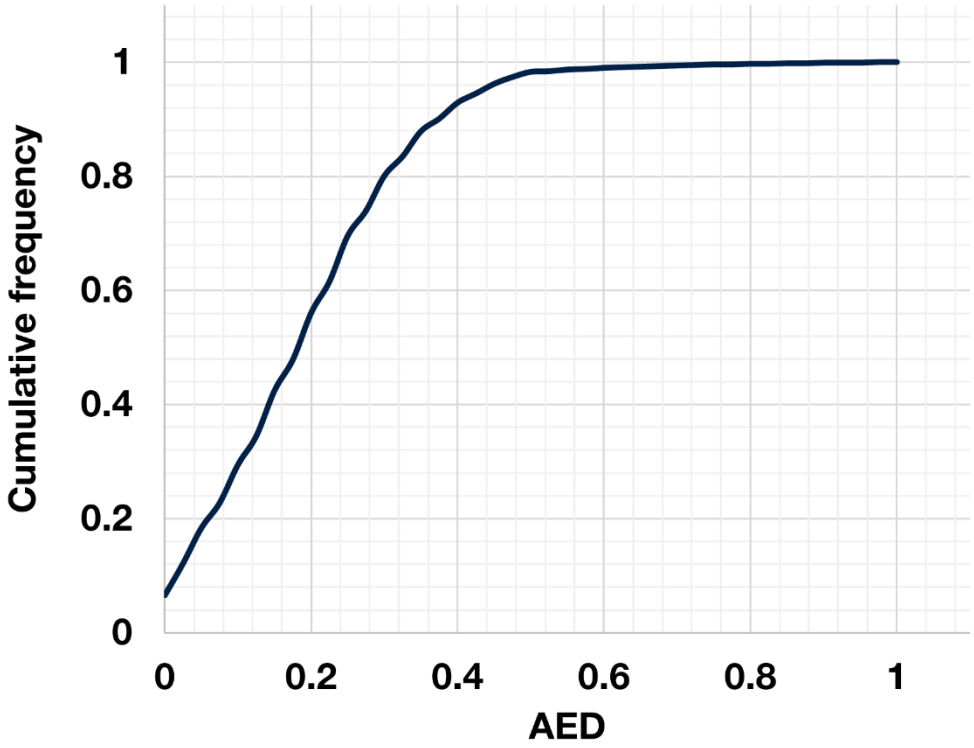


Figure 3.6. Maximum-likelihood tree of complete CNL proteins and TNL proteins identified in the genome of *D. cochinchinensis* using the JTT+G4 substitution matrix with 1,000 Felsenstein bootstrapping. Red tip labels denote CNL proteins and blue denote TNL proteins.

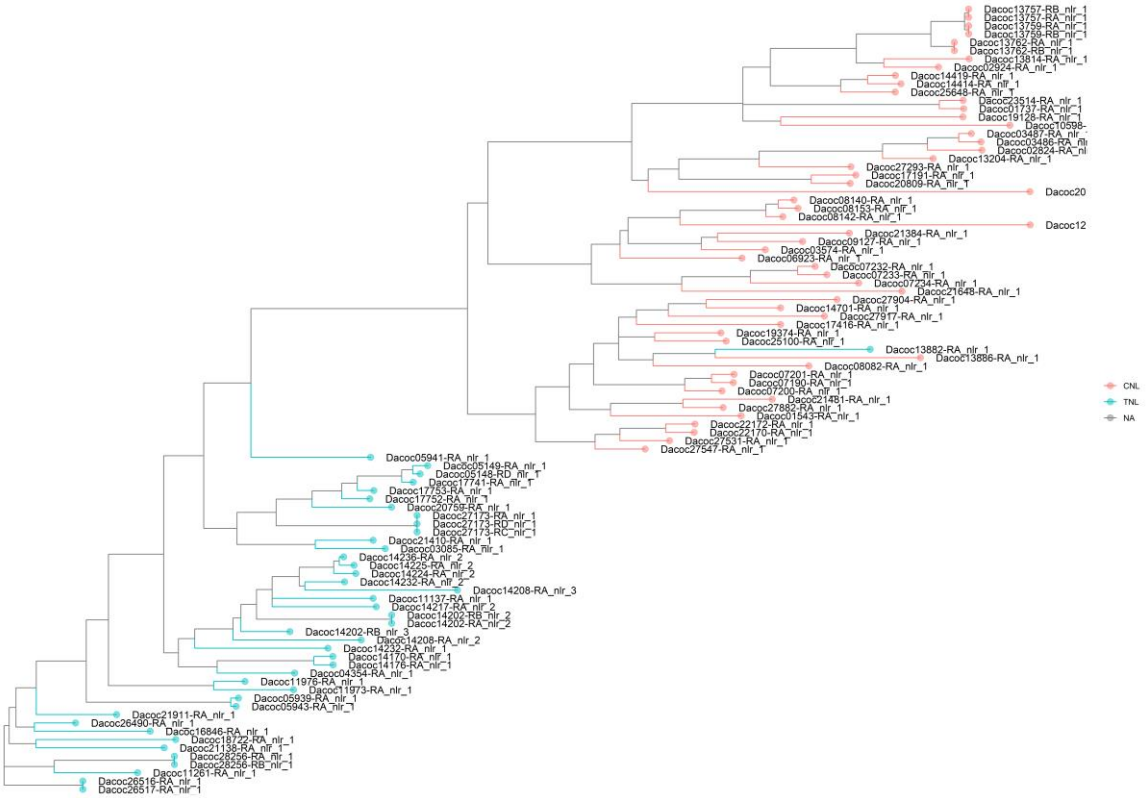


Figure 3.7. Genomic landscape of the complete CNL (red) and TNL (green) genes on the 10 assembled chromosomes (pseudomolecules) of *D. cochinchinensis*, showing tick marks every 1 Mb.

Figure 3.8. Heatmap of scaled number of motifs 1–20 (rows) among all predicted R genes (columns) with hierarchical clustering analysis. Motif 11 had the most dominant abundance and did not cluster with any other motifs.

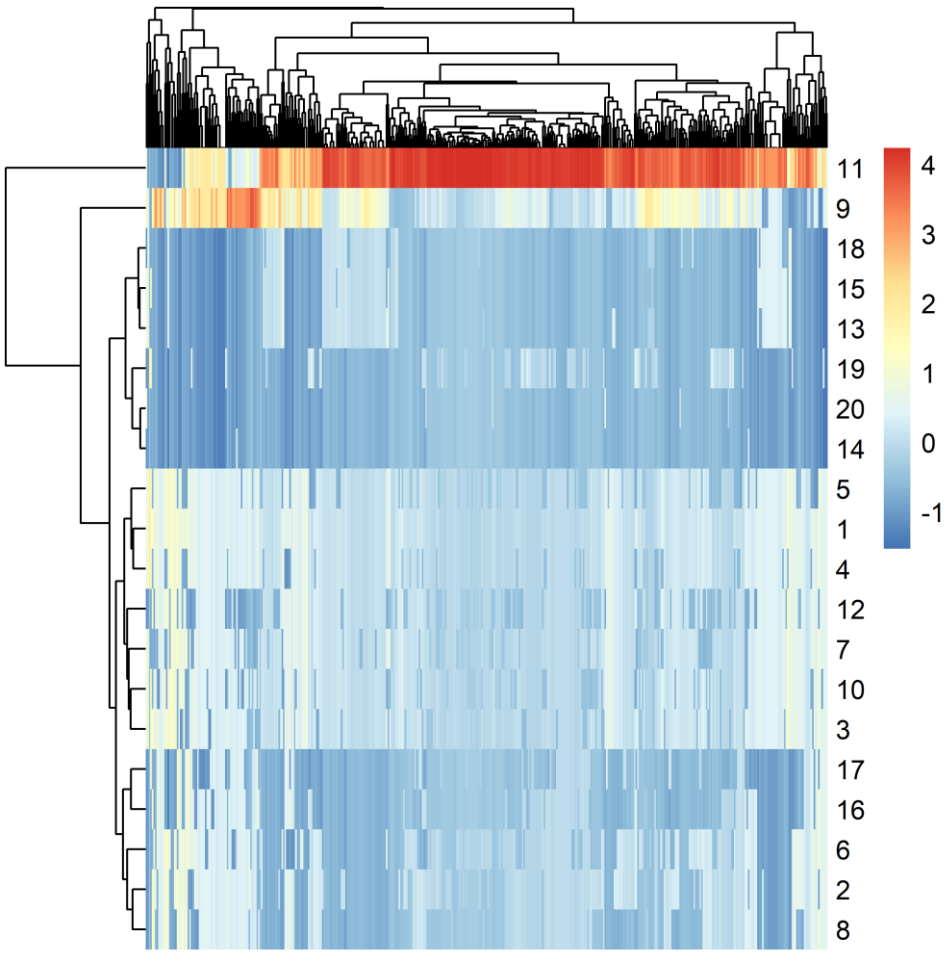


Figure 3.9. Boxplot of number of motif 11 in the leucine-rich repeat (LRR) domain among legume species.

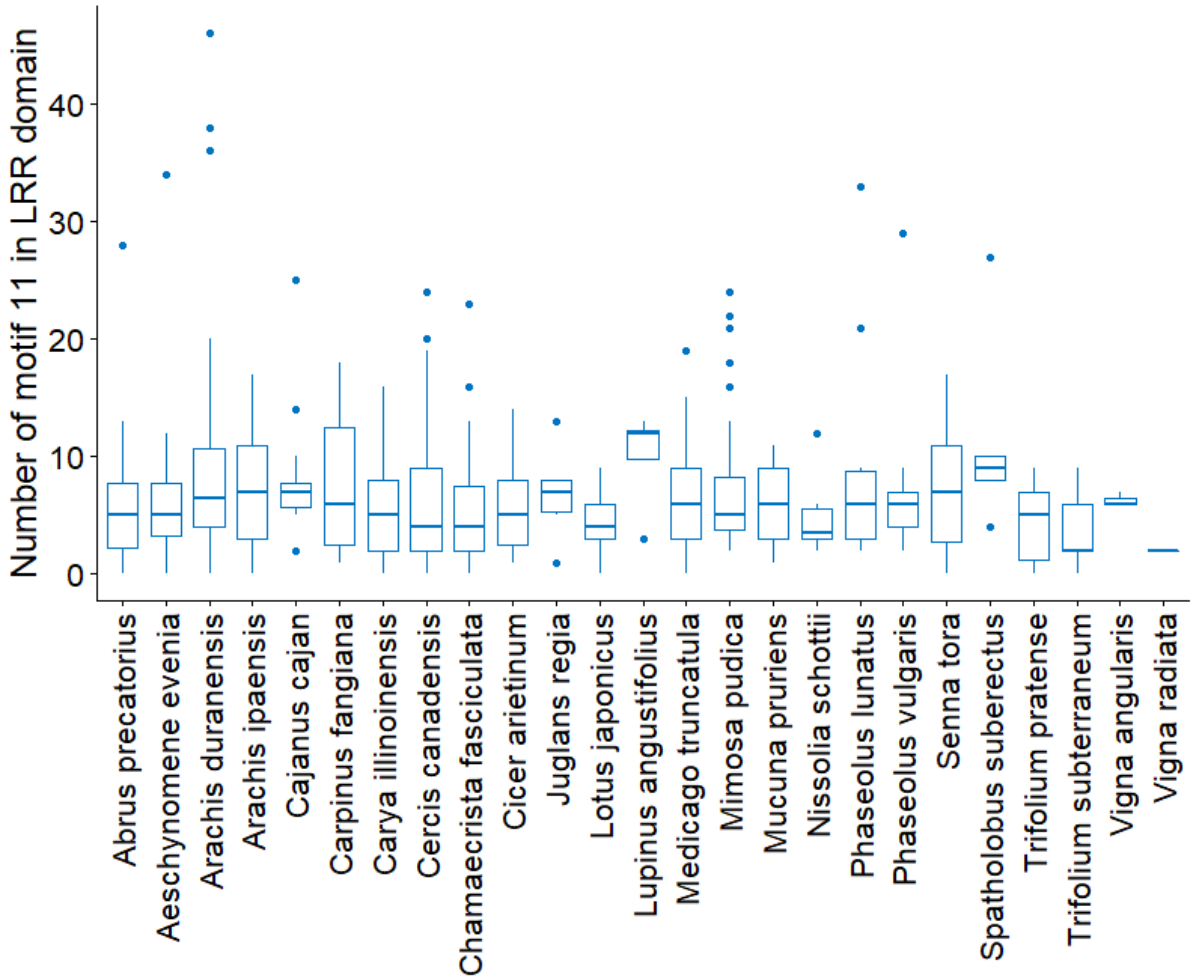


Figure 3.10. Pairwise correlation of frequency of motifs 1–20, only showing those with significant correlation ($P < 0.05$, Spearman's rank correlation). Motif 1 has a frequency distribution with zero standard deviation and thus no correlation value with any other motif.

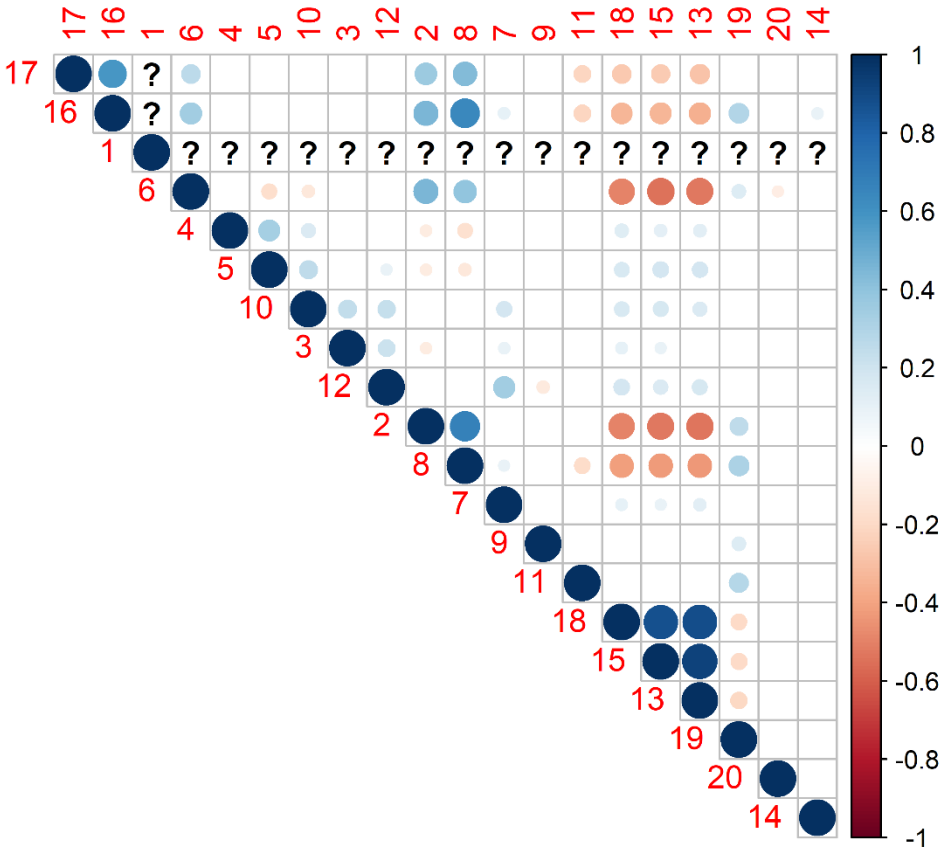


Figure 3.11. Maximum-likelihood phylogenetic tree of R genes across 44 genomes with 1,000 Felsenstein bootstrapping and their motif compositions.

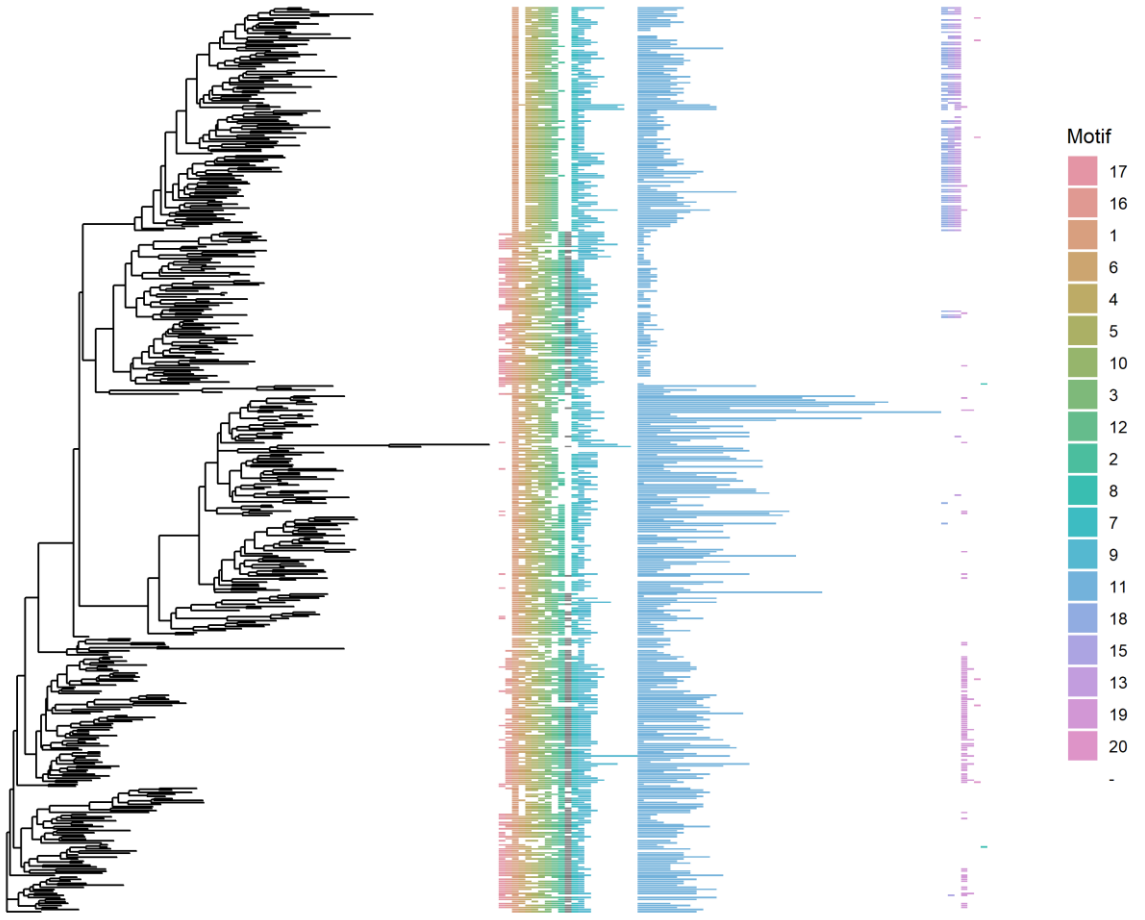


Figure 3.12. Number of TNL proteins against number of CNL proteins among **(a)** all 44 genomes and **(b)** only the legume genomes. The coefficient of determination (*R*), significance (*p*) and the regression equation (*y*) are provided.

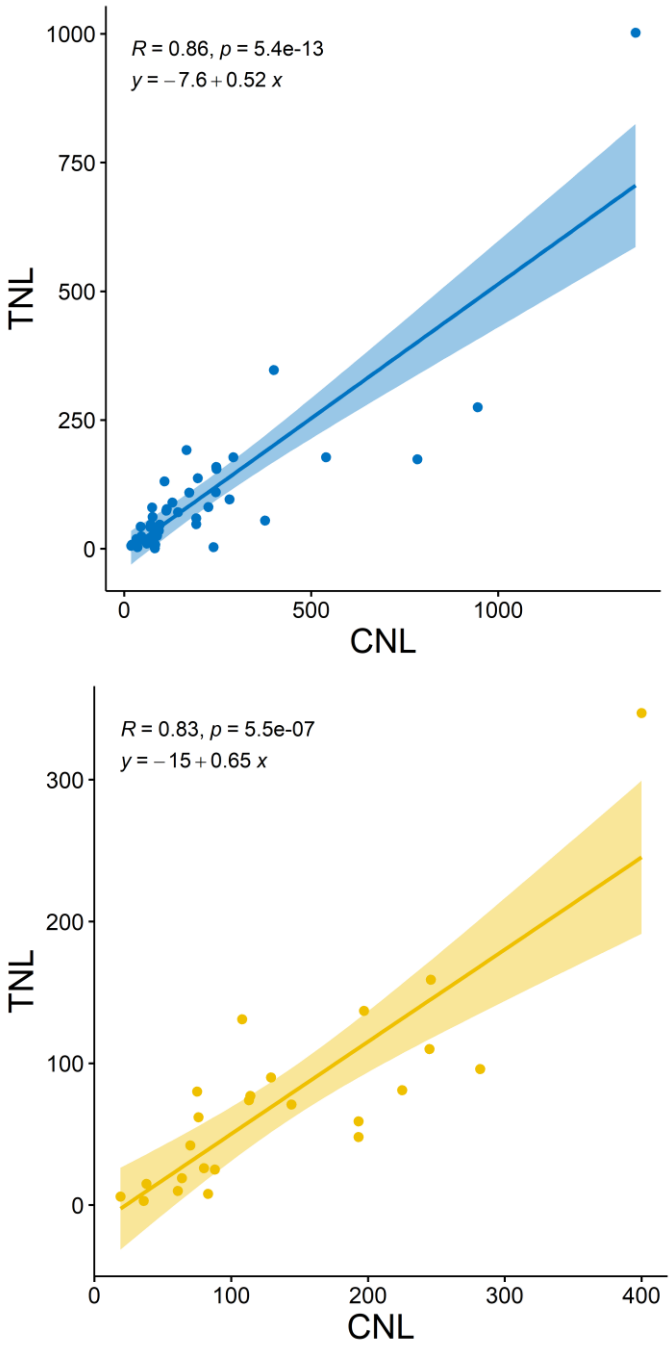
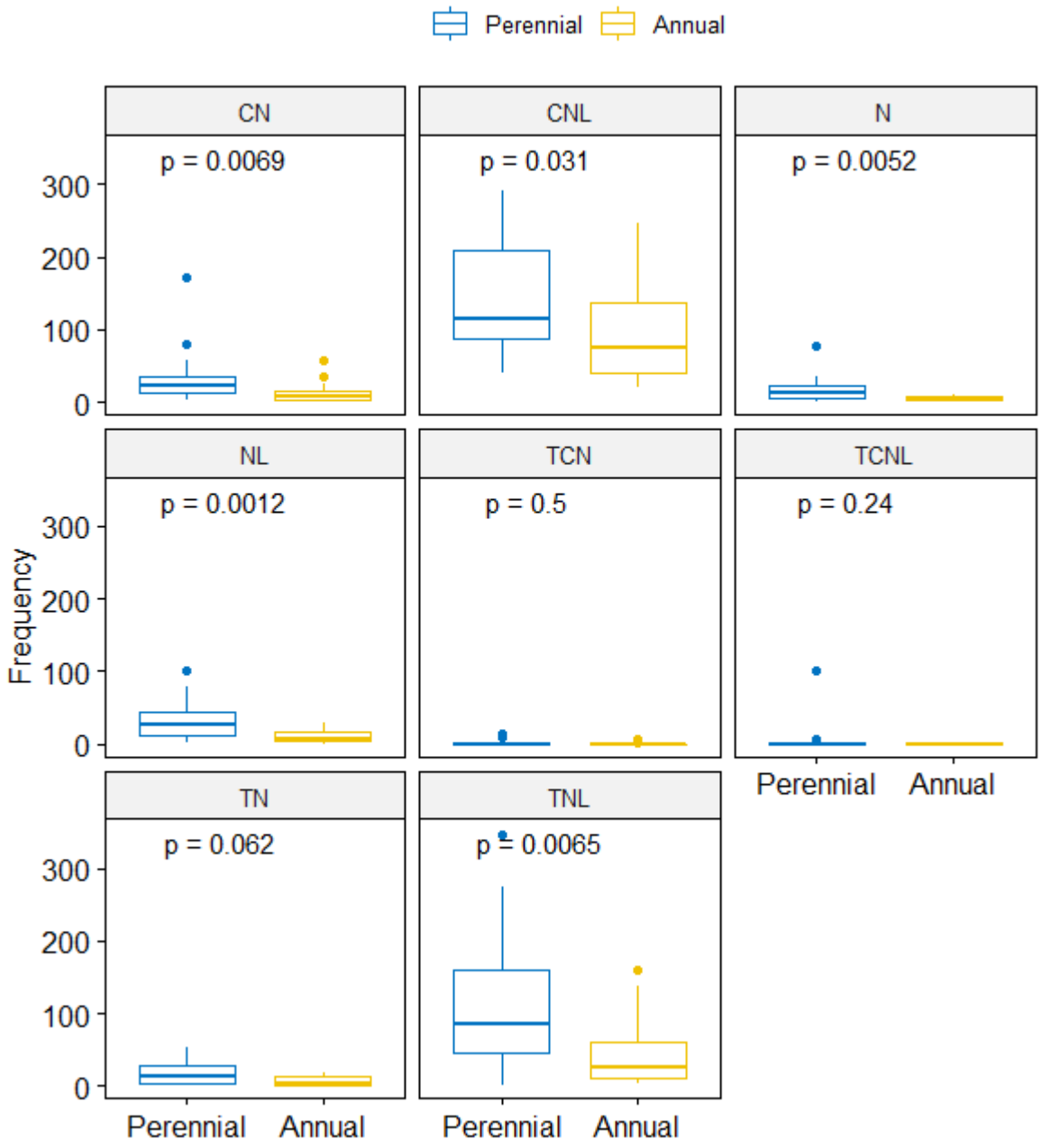


Figure 3.13. Frequency of R genes of different classes between annual and perennial plants for **(a)** all genomes (N = 44) and **(b)** Fabaceae genomes only (N = 24). Pairwise P-values are obtained with Wilcoxon signed rank test.

(a)



(b)

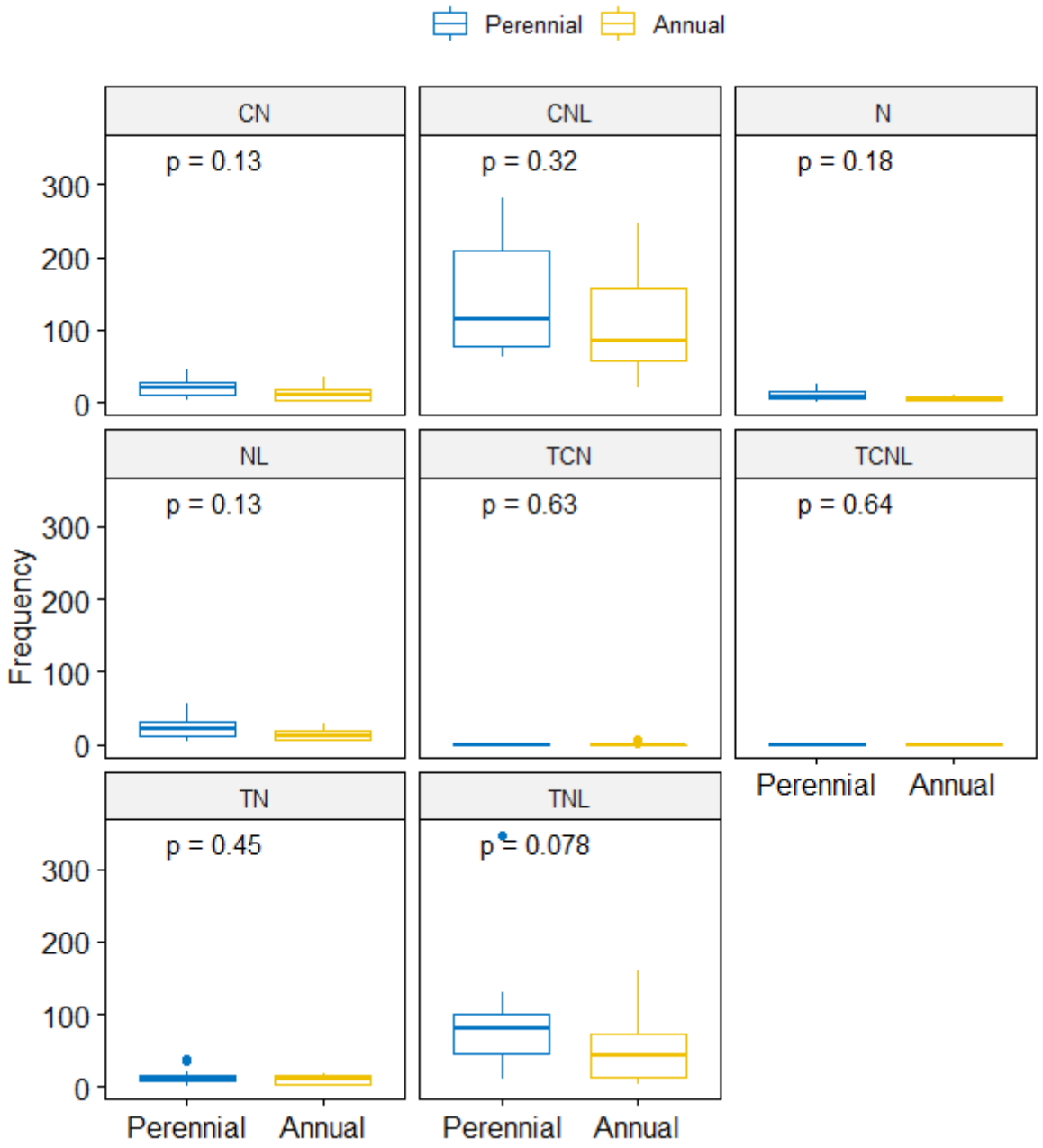

Perennial
Annual
CN
p = 0.13
CNL
p = 0.32
N
p = 0.18
NL
p = 0.13
TCN
p = 0.63
TCNL
p = 0.64
TN
p = 0.45
TNL
p = 0.078
Frequency
0
100
200
300
Perennial
Annual

Figure 3.14. Frequency of R genes of different classes between Fabaceae and non-Fabaceae species. Pairwise P-values are obtained with Wilcoxon signed rank test.

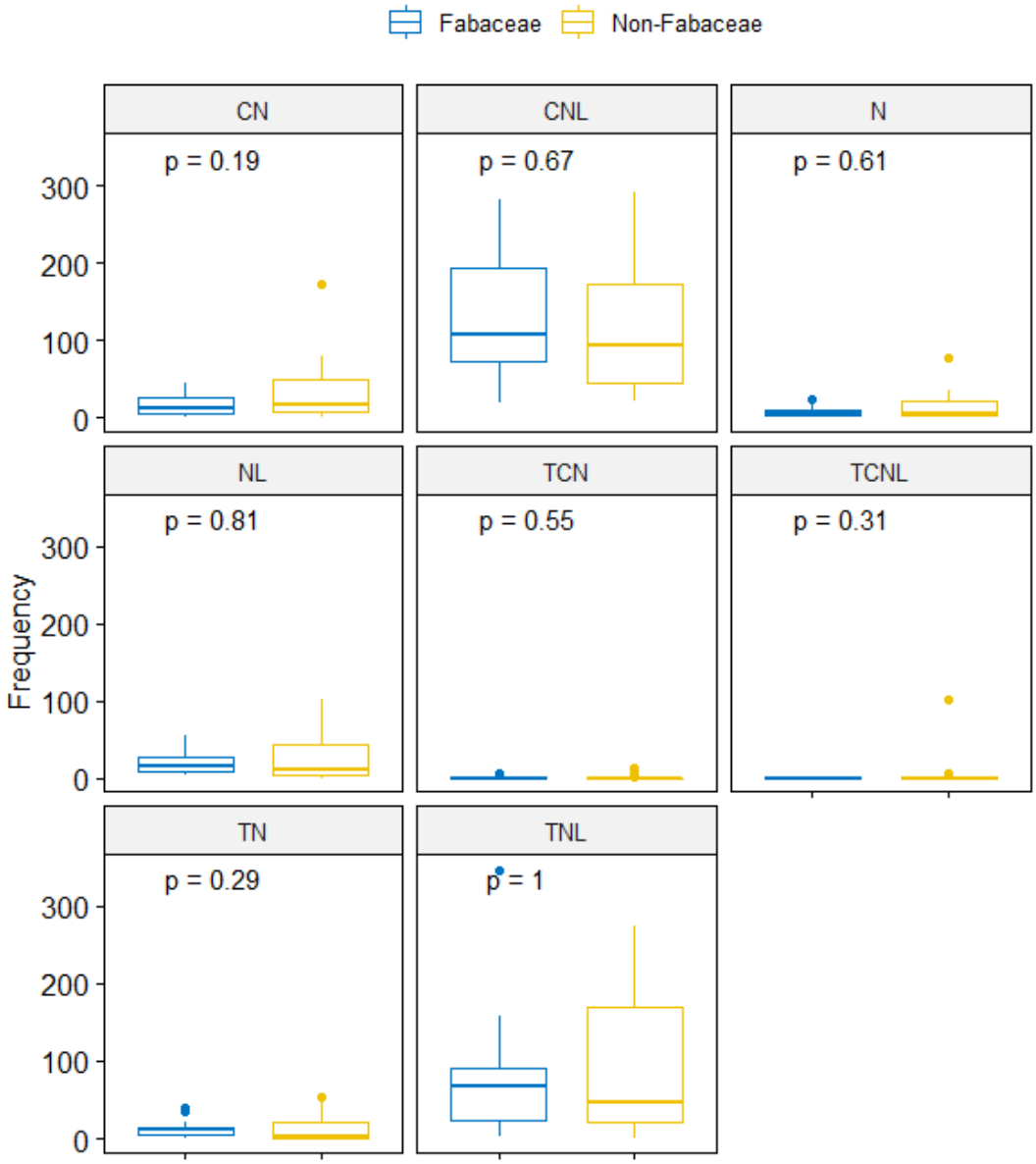


Table 3.1. Details of whole genomes and their gene models of 44 species from the Fabaceae family and closely related families used for comparative genomic analysis in this study.

Species	**Genome**	**Source**	**Life_History**	**Family**
Abrus precatorius	Abrus_2018	NCBI	Perennial	Fabaceae
Aeschynomene evenia	CIAT22838.gnm1.ann1.ZM3R	Legume_Federation	Perennial	Fabaceae
Arachis duranensis	aradu.V14167.gnm1.SWBf	Legume_Federation	Annual	Fabaceae
Arachis ipaensis	araip.K30076.gnm1.bXJ8	Legume_Federation	Annual	Fabaceae
Benincasa hispida	ASM972705v1	NCBI	Annual	Cucurbitaceae
Cajanus cajan	cajca.ICPL87119.gnm1.SBGP	Legume_Federation	Perennial	Fabaceae
Carpinus fangiana	ASM693729v1	NCBI	Perennial	Betulaceae
Carya illinoinensis	C.illinoinensisPawnee_v1	NCBI	Perennial	Juglandaceae
Cercis canadensis	cerca.ISC453364.gnm1.B05Z	Legume_Federation	Perennial	Fabaceae
Chamaecrista fasciculata	chafa.MN87.gnm1.JXFN	Legume_Federation	Annual	Fabaceae
Cicer arietinum	cicar.ICC4958.gnm2.bg5m	Legume_Federation	Annual	Fabaceae
Cucumis melo	ASM31304v1	NCBI	Annual	Cucurbitaceae
Cucumis sativus	Cucumber_9930_V3	NCBI	Annual	Cucurbitaceae
Cucurbita maxima	Cmax_1.0	NCBI	Annual	Cucurbitaceae
Dalbergia cochinchinensis	Dacoc_1.2	this_paper	Perennial	Fabaceae
Erythranthe guttata	Mimgu1_0	NCBI	Annual	Phrymaceae
Hevea brasiliensis	ASM165405v1	NCBI	Perennial	Euphorbiaceae
Jatropha curcas	RJC1_Hi-C	NCBI	Perennial	Euphorbiaceae
Juglans regia	Walnut_2.0	NCBI	Perennial	Euphorbiaceae
Lotus japonicus	lotja.MG20.gnm3.QPGB	Legume_Federation	Perennial	Fabaceae
Lupinus albus	lupal.Amiga.gnm1.F4NR	Legume_Federation	Annual	Fabaceae
Lupinus angustifolius	lupan.Tanjil.gnm1.Qq0N	Legume_Federation	Annual	Fabaceae
Malus baccata	Malus_baccata_v1.0	NCBI	Perennial	Rosaceae
Manihot esculenta	M.esculenta_v8	NCBI	Perennial	Euphorbiaceae
Medicago truncatula	medtr.A17_HM341.gnm4.2GZ9	Legume_Federation	Perennial	Fabaceae
Mimosa pudica		GigaDB	Perennial	Fabaceae
Momordica charantia	ASM199503v1	NCBI	Annual	Cucurbitaceae
Mucuna pruriens	ASM337056v1	NCBI	Annual	Fabaceae
Nissolia schottii		GigaDB	Perennial	Fabaceae
Phaseolus lunatus	phalu.G27455.gnm1.7NXX	Legume_Federation	Annual	Fabaceae
Phaseolus vulgaris	phavu.G19833.gnm1.zBnF	Legume_Federation	Perennial	Fabaceae
Pisum sativa	pissa.Cameor.gnm1.P4FG	Legume_Federation	Annual	Fabaceae
Populus trichocarpa	Pop_tri_v3	NCBI	Perennial	Salicaceae
Prunus avium	PAV_r1.0	NCBI	Perennial	Rosaceae
Ricinus communis	JCVI_RCG_1.1		Perennial	Euphorbiaceae
Rosa chinensis	RchiOBHm-V2		Perennial	Rosaceae
Salix dunnii	SDv1.1		Perennial	Salicaceae
Senna tora	ASM1485142v1	NCBI	Annual	Fabaceae

Spatholobus suberectus	ASM432916v1	NCBI	Perennial	Fabaceae
Trifolium pratense	tripr.MilvusB.gnm2.gNmT	Legume_Federation	Perennial	Fabaceae
Trifolium subterraneum	TSUd_r1.1	NCBI	Annual	Fabaceae
Vigna angularis	vigan.Gyeongwon.gnm3.JyYC	Legume_Federation	Annual	Fabaceae
Vigna radiata	vigra.VC1973A.gnm6.3nL8	Legume_Federation	Annual	Fabaceae
Vitis riparia	EGFV_Vit.rip_1.0	NCBI	Perennial	Vitaceae

Table 3.2. Assembly statistics of draft genomes of *D. cochinchinensis* using different assemblers.

Assembler	**CPU Hour**	**Genome size**	**Contigs**	**Contig N50**	**Longest contig**
Flye 2.7	~ 3,000	814145473	9951	470066	7001933
wtdbg2	714	1137325019	24354	88235	2952015
Raven 1.3.0	~ 5,096	1478339536	4184	403082	2115674
Canu 2.0	**> 500,000**	**1353086645**	**6443**	**1345847**	**33225722**

Table 3.3. Assembly statistics of draft genomes with haplotig-purging using different thresholds of alignment values. Bold one has the highest percentage of single-copy BUSCOs.

Alignment value	Genome size	N50	BUSCO score
unpurged	1.35 Gb	1.35 Mb	C:97.4%[S:17.1%,D:80.3%],F:0.6%,M:2.0%
a60	584 Mb	22.0 Mb	C:96.1%[S:92.2%,D:3.9%],F:0.9%,M:3.0%
a70	**621 Mb**	**21.4 Mb**	**C:96.8%[S:92.8%,D:4.0%],F:0.9%,M:2.3%**
a75	654 Mb	17.0 Mb	C:96.9%[S:92.5%,D:4.4%],F:0.9%,M:2.2%
a80	689 Mb	16.9 Mb	C:97.1%[S:91.8%,D:5.3%],F:0.9%,M:2.0%
a85	724 Mb	16.5 Mb	C:97.3%[S:91.4%,D:5.9%],F:0.8%,M:1.9%
a90	779 Mb	13.9 Mb	C:97.2%[S:89.9%,D:7.3%],F:0.9%,M:1.9%
a95	854 Mb	10.4 Mb	C:97.2%[S:86.6%,D:10.6%],F:0.8%,M:2.0%
a97	904 Mb	8.59 Mb	C:97.1%[S:82.2%,D:14.9%],F:0.7%,M:2.2%

Table 3.4. Statistics of Hi-C library prepared with the restriction enzyme DpnII and proximity-ligated.

Total read pairs (RPs)	366,490,675
RP length	150 bp
High quality (HQ) RPs	25.37%
RPs > 10 Kb apart	6.82%
RPs > 10 Kb apart (CTGs > 10 Kb)	10.37%
Intercontig RPs	33.22%
Intercontig hig-quality (HQ) RPs	6.31%
Same strand RPs	7.22%
Split reads	13.53%
Duplicate reads (extrapolated)	8.72%
Genome scaffolding sufficiency	
Same-strand high-quality* (HQ) read pairs (RPs)	5.03%
Informative RPs**	5.21%

Table 3.5. Mapping statistics of RNA-seq raw reads from a previous experiment using pooled samples from leaves, stems, and roots of *D. cochinchinensis* on the Dacoc_1.2 genome.

Number of input reads	132,696,766
Average input read length	299
Unique reads	
Number of uniquely mapped reads	122,215,722 (92.1%)
Average mapped length	296.69
Number of splices GT/AG GC/AG AT/AC Non-canonical	96,847,851 95,280,260 1,148,136 44,025 375,430
Mismatch rate per base (%)	2.87
Deletion rate per base (%)	0.03
Deletion average length	2.87
Insertion rate per base (%)	0.03
Insertion average length	2.50
Multi-mapping reads	
Number of reads mapped to multiple loci	4,875,679 (3.67%)
Number of reads mapped to too many loci	230,932 (0.17%)
Unmapped reads	
Number of reads with too many mismatches	0 (0.00%)
Number of reads too short	4,954,750 (3.73%)
Number of other reads	419,683 (0.32%)
Chimeric reads	
Number of chimeric reads	0 (0.00%)

Table 3.6. Composition and percentage of the repeat families in the final genome of *D. cochinchinensis* identified with RepeatModeler and LTRharvest.

	Number of elements	Total length (bp)	Percentage of sequence
Dacoc_1.2		621009970	
Bases masked		402405433	64.80
Retroelements	405732	293464285	47.26
SINEs	0	0	0.00
Penelope	0	0	0.00
LINEs	9518	3856482	0.62
CRE/SLACS	0	0	0.00
L2/CR1/Rex	0	0	0.00
R1/LOA/Jockey	0	0	0.00
R2/R4/NeSL	0	0	0.00
RTE/Bov-B	3633	803221	0.13
L1/CIN4	5885	3053261	0.49
LTR elements	396214	289607803	46.63
BEL/Pao	0	0	0.00
Ty1/Copia	209546	94712214	15.25
Gypsy/DIRS1	176990	189459322	30.51
Retroviral	0	0	0.00
DNA transposons	11462	6635025	1.07
hobo-Activator	1830	871791	0.14
Tc1-IS630-Pogo	0	0	0.00
En-Spm	0	0	0.00
MuDR-IS905	0	0	0.00
PiggyBac	0	0	0.00
Tourist/Harbinger	3272	1401349	0.23
Other (Mirage, P-element, Transib)	0	0	0.00

Rolling-circles	3613	2108774	0.34
Unclassified	359305	90585676	14.59
Total interspersed repeats		390684986	62.91
Simple repeats	131568	8474809	1.36
Low complexity	23616	1136864	0.18

Table 3.7. Number of different classes of R genes in 44 species identified with protein domain prediction and motif discovery based on Pfam and Meme. The R genes had different combination of motifs made up by C (coiled coil), T (TIR), N (NR-ABC), and L (leucine-rich repeat).

Species	CN	CNL	N	NL	TCN	TCNL	TN	TNL	UN	Total
Abrus precatorius	26	225	6	3	2	1	7	81	0	351
Aeschynomene evenia	4	80	1	14	0	0	2	26	0	127
Arachis duranensis	22	197	8	24	6	1	19	137	1	415
Arachis ipaensis	26	246	6	30	4	2	14	159	0	487
Benincasa hispida	9	71	3	3	0	0	4	47	0	137
Cajanus cajan	12	114	12	15	0	0	9	77	0	239
Carpinus fangiana	2	39	6	7	0	0	0	0	1	55
Carya illinoinensis	16	247	35	79	0	1	54	155	0	587
Cercis canadensis	27	282	13	31	0	2	14	96	0	465
Chamaecrista fasciculata	8	113	8	18	0	0	18	74	0	239
Cicer arietinum	14	64	3	10	0	0	3	19	0	113
Cucumis melo	5	44	4	4	0	0	2	43	0	102
Cucumis sativus	2	46	3	2	0	0	3	24	0	80
Cucurbita maxima	1	21	2	0	0	0	0	8	0	32
Dalbergia cochinchinensis	17	70	6	27	0	0	4	42	0	166
Erythranthe guttata	56	239	2	2	0	1	0	3	0	303
Hevea brasiliensis	80	784	11	52	0	1	14	174	0	1116
Jatropha curcas	18	95	15	9	0	2	2	47	0	188
Juglans regia	27	167	22	45	0	0	16	192	0	469
Lotus japonicus	36	75	25	37	1	0	39	80	0	293
Lupinus albus	1	19	3	3	0	0	4	6	0	36
Lupinus angustifolius	1	36	2	3	0	0	1	3	0	46
Malus baccata	19	174	12	43	0	1	31	109	0	389
Manihot esculenta	33	377	1	29	2	0	1	55	0	498
Medicago truncatula	30	400	18	56	1	1	36	347	0	889

Mimosa pudica	11	108	11	28	1	0	21	131	0	311
Momordica charantia	12	33	6	5	1	0	2	19	0	78
Mucuna pruriens	16	76	5	17	0	1	18	62	0	195
Nissolia schottii	5	129	4	12	1	0	9	90	0	250
Phaseolus lunatus	12	144	11	19	0	1	14	71	0	272
Phaseolus vulgaris	3	245	2	12	0	0	12	110	0	384
Pisum sativum	14	82	2	4	0	0	0	1	1	104
Populus trichocarpa	56	540	24	11	8	5	28	178	1	851
Prunus avium	58	292	20	26	0	4	34	178	0	612
Ricinus communis	8	93	2	14	0	0	1	35	0	153
Rosa chinensis	44	1367	78	101	14	101	53	1002	3	2763
Salix dunnii	12	97	5	1	0	0	0	0	0	115
Senna tora	8	83	8	5	0	0	0	8	0	112
Spatholobus suberectus	22	61	6	5	0	0	9	10	0	113
Trifolium pratense	44	193	25	33	0	1	14	48	0	358
Trifolium subterraneum	36	193	11	14	1	2	14	59	0	330
Vigna angularis	1	88	3	6	0	0	8	25	0	131
Vigna radiata	2	38	4	8	0	0	5	15	0	72
Vitis riparia	172	945	35	66	0	7	14	275	0	1514

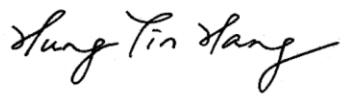
Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Phenotypic responses of rosewoods *Dalbergia cochinchinensis* and *D. oliveri* under isolated and combined stresses of heat and drought
Publication Status	☑Published □ Accepted for Publication □Submitted for Publication □Unpublished and unsubmitted work written in a manuscript style
Publication Details	Hung, T. H.*, Gooda, R., Rizzuto, G., So, T., Thammavong, B., Tran, H. T., Jalonen, R., Boshier, D. H. & MacKay, J. J.* (2020). Physiological responses of rosewoods *Dalbergia cochinchinensis* and *D. oliveri* under drought and heat stresses. *Ecology and Evolution, 10*(19), 10872–10885. http://doi.org/10.1002/ece3.6744

Student Confirmation

Student Name:	Tin Hang Hung		
Contribution to the Paper	Designed the study, conducted the greenhouse experiment, ran the statistical analyses, drafted and revised the manuscript		
Signature 		Date	8 MAR 2022

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title:	John MacKay, Wood Professor of Forest Science		
Supervisor comments			
Signature 		Date	8 MAR 2022

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 4. Physiological responses of rosewoods *Dalbergia cochinchinensis* and *D. oliveri* under drought and heat stresses

Tin Hang Hung[1,^,*], Rosemary Gooda[1,^], Gabriele Rizzuto[1], Thea So[2], Bansa Thammavong[3], Hoa Thi Tran[4], Riina Jalonen[5], David H. Boshier[1], John J. MacKay[1,*]

1. Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, United Kingdom
2. Institute of Forest and Wildlife Research and Development, Cambodia
3. National Agriculture and Forestry Research Institute, Forestry Research Center, Vientiane, Lao PDR
4. Forest Genetics and Conservation, Center for Biodiversity and Biosafety, Institute of Agricultural Genetics, Vietnam Academy of Agricultural Sciences, Pham Van Dong Road, Hanoi, Vietnam
5. Bioversity International, Malaysia Office, c/o TNCPI, University Putra Malaysia, off Lebuh Silikon, 43400 UPM Serdang, Malaysia

4.1 Abstract

Dalbergia cochinchinensis and *D. oliveri* are classified as vulnerable and endangered respectively in the IUCN Red List and under continued threat from deforestation and illegal harvesting for rosewood. Despite emerging efforts to conserve and restore these species, little is known of their responses to drought and heat stress, which are expected to increase in the Greater Mekong Subregion where the species co-occur and are endemic. In this study of isolated and combined drought and heat effects, we found that *D. oliveri* had an earlier stomatal closure and more constant midday water potential in response to increasing drought level, suggesting that *D. oliveri* is relatively isohydric while *D. cochinchinensis* is relatively anisohydric. Heat shock and drought had synergistic effects on stomatal closure. Our results indicate contrasting relationships in water relations, photosynthetic pigment levels and total soluble sugars. An increase in chlorophyll a was observed in *D. cochinchinensis* during drought and a concomitant increase in carotenoid content likely afforded protection against photo-oxidation. These physiological changes correlated with higher total soluble sugars in *D. cochinchinensis*. By contrast, *D. oliveri* avoided drought by reducing chlorophyll content and compromising productivity. Anisohydry and drought tolerance in *D. cochinchinensis* are adaptations which fit well with its ecological niche as a pioneering species with faster growth in young trees. We believe this understanding of the stress responses of both species will be crucial to their effective regeneration and conservation in degraded habitats and in the face of climate change.

4.2 Introduction

The pantropical genus *Dalbergia* L. f. (Fabaceae: Faboideae) contains around 250 species[1], many of which produce valuable heartwood timber known as rosewood[2], which is used to manufacture luxury furniture, boats, and musical instruments[3]. Growing demand and diminishing supply have drastically increased the economic value of rosewood, resulting in much illegal harvesting and poorly regulated exploitation of natural populations. Among these are *Dalbergia cochinchinensis* Pierre and *D. oliveri* Gamble ex Prain (Figure 4.1), both of which are endemic to Cambodia, Laos, Thailand, and Vietnam within the Greater Mekong Subregion (GMS). *D. cochinchinensis* was once the most sought-after rosewood species globally but it is now virtually commercially extinct, as is *D. oliveri*[4]. They were classified as vulnerable and endangered respectively in the IUCN Red List in 1998, with international trade strictly regulated since 2017 under CITES Appendix II.

Responses to their declining populations have seen new initiatives in the Greater Mekong Subregion to conserve and restore both species, such as through improving capacity for seed collection, germplasm supply, and propagation[5–7]. In addition, some *Dalbergia* species are deemed suitable candidates for use in forest landscape restoration in the region[8], where intense land conversion has degraded sites[9]. Successful incorporation of valuable and endangered species like *Dalbergia* in forest restoration programmes can achieve both conservation and restoration goals[10,11]. A sustainable supply of diverse germplasm is essential for the success of such projects[12]. Recently the importance of investigating the response to climatic stresses of germplasm used in forest restoration and other tree planting has been recognised[13], since many forest restoration projects around the world have neglected the adaptability of seeds in the planting sites[14]. For reforestation to succeed, it is crucial to know about the ranges of abiotic conditions that a species can tolerate, allowing their planting in appropriate regions. There is particular concern for the effect of climate change within the

Greater Mekong Subregion with evidence pointing to increased temperature, variability in runoff, and prolonged agricultural droughts[15]. Simultaneously, water availability may decrease in the region due to increasing urbanisation and changes in river infrastructure such as upstream dams in China[16]. However, we have little understanding of the physiological adaptations of *D. cochinchinensis* and *D. oliveri*, which hinders informed decision-making in both conservation and forest restoration initiatives.

Imminent effects of climate change will, in many areas, result in rapid increases in temperatures, drought and extreme weather[17]. The sessile habit and longevity of tree species mean they must have sufficient phenotypic plasticity to tolerate the wide range of environmental conditions that occur during their lifespans[18]. Similarly trees are potentially more vulnerable than other plants under changing climate scenarios as they have long generation times, with the persistence of a particular tree species dependent on adaptive capacity to stress, plasticity, and migration potential[19]. Extreme abiotic conditions caused by climate change will potentially be detrimental to primary productivity, ecological functions, and associated biodiversity of these forests[20]. Vulnerability of forests to tree die-off due to climate change has become a focus of forest sustainability[21], with extensive tree climate-induced mortality well-documented worldwide[22–24]. In particular, seedling recruitment and survival are considered critical bottlenecks in tree life history and have an important role in shifts in species' range under changing climate[25]. Migration to adapted ecological niches and acclimation in an existing range are important mechanisms for tree survival in a changing climate[26].

Although knowledge of the ability of tree species to tolerate environmental stresses is essential for our understanding of how trees will respond to the effects of climate change[18], we lack data on the stress tolerance of specific taxa[27]. Water availability varies spatially and temporally, and has also been found to be an important determinant of functional traits in

trees[28]. Trees have evolved strategies to balance hydraulic conductance and resource allocation in response to water deficit. One of these strategies is establishing barriers for evaporation to achieve homeostasis of tissue water status[18]. Other strategies exist to maintain their metabolism at a lower water potential[29] through osmotic adjustment and protection from photo-oxidation[30]. However, prolonged drought stress that exceeds the drought resistance threshold can lead to mortality associated with hydraulic failure, carbon starvation and the demography of biotic agents[23].

Temperature also plays a major role in determining the distribution of tree species, as it significantly regulates tree growth and development. Trees can usually utilise transpiration to allow water evaporation for temperature regulation[31] but excessive heat stress is an important driver of tree die-off in natural ecosystems[32]. Studies have shown that heat stress impairs photosystems, stimulates photorespiration, and encourages production of volatile compounds[33,34]. Extreme heat reduces tree vigour, fecundity and growth, significantly affecting survival[35].

Different stresses often occur simultaneously in the field but there are few reports on interactions between such stresses in trees[27]. It has been proposed that drought and heat stresses were intrinsically linked and produce positive feedbacks to intensify their effects[36], but new evidence indicates that combinations of stresses can invoke responses that are distinct from those of the individual stresses[37]. For example, the combination of heat and drought stress could reduce the negative effect of drought stress by preserving pre-dawn water potential and malondialdehyde[38]. These interactions and responses are species-specific, and are expected to be complex and difficult to predict, as response mechanisms vary between stresses[39].

The objective of the present study is to develop an understanding of the physiological responses of 3-month-old *D. cochinchinensis* and *D. oliveri* seedlings under controlled

conditions of heat and drought stress. First, we compare and characterise the hydraulic responses in the two species when exposed to drought. Second, we determine the effects of isolated and combined stresses of heat and drought on hydraulic, leaf, and photosynthetic traits. We discuss the findings in relation to life-history traits and ecological niches of these species and formulate implications for their conservation and use in restoration.

4.3 Methods

4.3.1 Plant materials

Dried seeds of *D. cochinchinensis* and *D. oliveri* were provided by the Forest Research Center, Lao PDR and the Institute of Forest & Wildlife Research & Development, Cambodia respectively in 2018. We scarified the seeds by placing them in 70°C distilled water, which was then left to cool to room temperature overnight and germinated them on 1% agar in a plant growth cabinet MLR-350 (Sanyo, Watford, United Kingdom) at 25°C and photoperiod 12L/12D. Germinants were transferred to 1L-pots in a soil-perlite 3:1 (v:v) mixture in a greenhouse set to 30°C, 80% RH, and 12L/12D. 128 plants were kept in trays of 8 pots and randomised with equal numbers of individuals of *D. cochinchinensis* and *D. oliveri*. Plants were watered to maintain at substrate capacity and fertilised once a week using N-P-K 20:20:20 fertiliser (Chempak, Suffolk, United Kingdom).

4.3.2 Experimental design

After 3 months of growth in the greenhouse, the experiment began on the 12th June 2019 (Day 0), on which all plants were watered. The design included a total of 4 treatments: drought treatment (D), well-watered control (W), heat shock treatment (H) or non-heat shock control (N) (Figure 4.2). Each of the trays of 8 plants was randomly assigned as either D or W, while each plant of the trays was then randomly assigned to H or N. This resulted in each individual being assigned to one of four treatment combinations: drought and heat shock (DH), well-watered and heat shock (WH), drought and non-heat shock (DN), and well-watered and non-heat shock (WN). The well-watered plants were watered every other day to maintain them at substrate capacity and all water was withheld from the drought plants. Each individual plant was randomly assigned to one of the three sampling points: 5, 9 and 13 days from the beginning of treatments. 5 biological replicates were present for each sampling group.

For the heat shock treatment, we placed half of the individuals to be sampled at a given time point in a plant growth cabinet MLR-350 (Sanyo, Watford, United Kingdom) at 38°C for 4 hours from 08:00 to 12:00 before sampling. The samples were made up of homogeneous leaves that were snap-frozen in liquid nitrogen and kept at –80°C for subsequent analyses. Fully expanded branches were also cut at the nodes and kept in cooler bags for water potential, mass and area measurements.

4.3.3 Water relation measurements

Soil water content (SWC) was measured daily using a ML3 ThetaProbe Soil Moisture Sensor (Delta-T Devices Ltd. Cambridge, England) by placing the sensor probe in the soil at the centre of the pot. Stomatal conductance was determined daily using an SC-1 leaf porometer (METER Group, Inc., Washington, USA) to measure vapour flux through the stomata on fully expanded leaves for 30 seconds. Midday water potential (Ψ_{MD}) was measured by using a scholander-type pressure chamber SKPM140 (Skye Instruments Ltd. Llandrindod Wells, Wales, UK) on a fully expanded branch. Pressure was increased inside the chamber with compressed nitrogen, until moisture appeared on the cut end and the reading recorded.

4.3.4 Leaf dry matter content and specific leaf area

An office scanner was used to obtain electronic images of leaves on a fully expanded branch from each plant. The program ImageJ 1.52s[40] was then used to measure the total leaf area of a fully expanded branch for each individual. These leaves were then dried in an oven at 65°C until the weight reading remained constant for two consecutive days.

Leaf dry matter content (LDMC) and specific leaf area (SLA) were deduced using the following equations[41]:

$$\text{Leaf dry matter content} = \frac{\text{Oven-dried mass of leaf (mg)}}{\text{Fresh mass of leaf (mg)}} \quad \text{(eqn. 1)}$$

$$\text{Specific leaf area } (\text{cm}^2\ \text{mg}^{-1}) = \frac{\text{Area of leaf } (\text{cm}^2)}{\text{Oven-dried mass of leaf (mg)}} \quad \text{(eqn. 2)}$$

4.3.5 Pigment quantification

Photosynthetic pigments were extracted using cold acetone-50mM Tris buffer pH 7.8 (80:20 v:v) following Sims & Gamon's protocols[42]. Absorbances of the extract were read at 470, 537, 647, and 663 nm using Helios Gamma UV-Vis Spectrophotometer (Thermo Fisher Scientific, United States). Concentrations of anthocyanin (*Ac*), chlorophyll a (Chl_a), b (Chl_b), and carotenoids were determined using Sims & Gamon's formulae:

$$Ac\ (\mu\text{mol ml}^{-1}) = 0.08173A_{537} - 0.00697A_{647} - 0.002228A_{663} \quad \text{(eqn. 3)}$$

$$Chl_a\ (\mu\text{mol ml}^{-1}) = 0.01373A_{663} - 0.000897A_{537} - 0.003046A_{647} \quad \text{(eqn. 4)}$$

$$Chl_b\ (\mu\text{mol ml}^{-1}) = 0.02405A_{647} - 0.004305A_{537} - 0.005507A_{663} \quad \text{(eqn. 5)}$$

$$\text{Carotenoids } (\mu\text{mol ml}^{-1}) = \left(A_{470} - \left(17.1 \times \left(Chl_a + Chl_b\right) - 9.479Ac\right)\right) \times 119.26 \quad \text{(eqn. 6)}$$

4.3.6 Total soluble sugars quantification

Total soluble sugars (TSS) were extracted from 50 mg of frozen leaves in 80% (v/v) ethanol for 1 hour at 80°C. TSS concentration was determined by Osaki's anthrone method[43], in which the extract was mixed with anthrone and sulphuric acid. After heating the mixture at 100°C for exactly 10 min, absorbance was read at 625 nm using Helios Gamma UV-Vis Spectrophotometer (Thermo Fisher Scientific, United States). The concentration was determined according to a D-glucose standard curve.

4.3.7 Statistical analysis

The experiment used a split-plot design: the drought treatment was randomised at the tray level (main-plot), while heat shock was randomised at the individual plant level (split-

plot). We log-transformed the stomatal conductance, SLA, anthocyanin content, Chl_b, and carotenoids content to correct for normality.

To compare and characterise the hydraulic responses in the two species over time, we applied linear-plateau regression of g_s against SWC in the drought treatment groups for both species. Best-fit estimates of the intercept (c), the slope (m), and the critical value of SWC (x') were predicted using nonlinear least squares method[44] with the following equation:

$$g_s = \begin{cases} mx + c & \text{, if } x < x' \\ mx' + c & \text{, if } x \geq x' \end{cases} \quad \text{(eqn. 7)}$$

To test the effects of isolated and combined stresses on the hydraulic, leaf, and photosynthetic traits, we performed main-plot analyses on data subsets which averaged the response variables, including g_s, Ψ_{MD}, LDMC, SLA, Ac, Chl_a, Chl_b, carotenoids, and TSS, over the combinations of block, species, and two treatments, and performed split-plot analyses on the full dataset while controlling for block. The fixed effects of the two stresses and their interaction were tested with the following equation:

$$\begin{aligned} Y = \mu + drought_a + block_i + \eta_{ai} + species_b + day_c + heatshock_d + \\ \left(drought * heatshock * day * species\right)_{abcd} + \varepsilon_{abcdi} \end{aligned} \quad \text{(eqn. 8)}$$

(Y: response variable; μ: mean; $drought_a$: fixed effect of drought treatment; $block_i$: fixed effect of block; η_i: whole-plot error; $species_b$: fixed effect of species; day_c: fixed effect of number of days since water withholding; $heatshock_d$: fixed effect of heat shock treatment; $(drought*heatshock*day*species)_{abcd}$: interaction between explanatory variables; ε_{abcdi}: split-plot error)

We used R 3.6.2 to perform the randomisation for experimental design, statistical analysis and data visualisation. An analysis of variance (ANOVA) table was computed for the statistical model of each response variable.

4.4 Results

The plants were divided into drought (D) and well-watered control (W) treatment groups following a split-plot experimental design. Physiological traits were measured after 5, 9 and 13 days of treatment. On the sampling days only, we also applied a heat shock (H) treatment to half of the individuals in each sampling group with the other not exposed to heat (N). Hence, there were four treatment groups: DH, DN, WH, WN (Figure 4.2).

4.4.1 Stomatal conductance

We found a statistically significant difference in slopes of stomatal conductance (g_s) (p = 0.0134; Table 4.1) between the drought and well-watered (control) treatments but not in intercepts (p = 0.2308; Table 4.2). This observation indicated that water relations were similar among the two groups at the start of the experiment and that g_s decreased with the drought treatment in both species.

We employed a linear-plateau nonlinear least squares fit to model the relationship between g_s and soil water content (SWC) and searched for a critical value of SWC that affects the stomatal aperture (Figure 4.3). We found that the critical SWC of *D. cochinchinensis* (32.65%, p = 4.65e-11) was lower than *D. oliveri* (47.23%, p < 2e-16), suggesting that *D. cochinchinensis* closed its stomata later than *D. oliveri* in response to decreasing SWC.

When analysing all four treatment groups on sampling days, we obtained the same significant decrease in g_s in drought treatment (p = 2.401e-5; Table 4.3 and Figure 4.4a). We found that *D. oliveri* had a lower stomatal conductance than *D. cochinchinensis* (p = 3.366e-06). In addition, we found a significant decrease in heat shock treatment (p = 2.936e-15) and a significant drought × heat interaction in g_s (p = 0.000666). The interaction effect was the strongest on day 13: The g_s of combined stress was lower (27.54) than either drought (50.48)

or heat (61.8) alone in *D. cochinchinensis*. In *D. oliveri*, the g_s of combined stress was also lower (24.56) than either drought (37.46) or heat (34.66) alone.

4.4.2 Midday water potential

We observed a significant decrease in midday water potential (Ψ_{MD}) in the drought treatment, ($p = 0.03955$; Table 4.4 and Figure 4.4b). A significant interaction was found between the drought treatment and species, ($p = 0.004137$), indicating differential responses and vulnerabilities, with *D. cochinchinensis* displaying a faster decrease in MWP over time than *D. oliveri*. The Ψ_{MD} of the *D. cochinchinensis* drought treatment group dropped drastically from –7.27 to –13.84 bar between day 5 and 13, while that for *D. oliveri* remained relatively stable from –11.75 to –13.78 bar.

4.4.3 Leaf dry matter content and specific leaf area

We found no significant effect of the drought and heat treatments on leaf dry matter content (LDMC) and specific leaf area (SLA) (Table 4.5 and Figure 4.5a). However, SLA was significantly different between species ($p = 1.061e-12$; Table 4.6 and Figure 4.5b), with *D. cochinchinensis* having a lower SLA (0.30 cm^2/mg) than *D. oliveri* (0.48 cm^2/mg).

4.4.4 Pigments

We observed no significant effect of either drought or heat shock stresses on anthocyanin content (Table 4.7 and Figure 4.6a). Chlorophyll a content (Chl_a) was significantly affected by drought treatment ($p = 0.02927$; Table 4.8 and Figure 4.6b), species ($p = 0.039284$), and the interaction between both stresses and species ($p = 0.034303$). We found opposite trends of change in Chl_a content between species, such that combined stresses increased Chl_a from 1.15 to 1.43 μmol/g FW in *D. cochinchinensis* and decreased Chl_a from 0.876 to 0.568 μmol/g FW in *D. oliveri*. The drought treatment seemed to amplify the difference between species, resulting in the highest Chl_a in *D. cochinchinensis* and lowest in

D. oliveri. We observed no significant effect of drought and heat shock stresses on chlorophyll b content (Chl_b) (Table 4.9 and Figure 4.6c).

Carotenoid content was significantly affected by the drought treatment (p = 0.03966; Table 4.10 and Figure 4.6d) and the combination of drought, species, and days (p = 0.003470). An opposite trend between species was observed in carotenoids similar to Chl_a, where carotenoid content increased over time following drought from 11.27 to 11.54 mmol/g FW in *D. cochinchinensis* but decreased from 14.27 to 8.63 mmol/g in *D. oliveri*.

4.4.5 Total soluble sugars

We found that the total soluble sugars content (TSS) changed significantly with the drought treatment (p = 0.03993; Table 4.11 and Figure 4.7) and increased slightly from 22.66 to 27.14 mg/g FW in *D. cochinchinensis*, but decreased slightly from 20.34 to 18.58 mg/g FW in *D. oliveri*. *D. cochinchinensis* maintained a higher TSS than *D. oliveri* (p = 2.586e-6).

4.5 Discussion

We report short term physiological changes in seedlings of *D. cochinchinensis* and *D. oliveri* in response to isolated and combined stress treatments of drought and heat. Most of the responses are found to be induced by drought only but the two stresses have an additive effect on reducing stomatal conductance. These stress responses included several physiological factors and may highlight metabolic vulnerabilities, which follow from adaptations to environmental conditions. Our observations indicate changes in water relations, photosynthetic pigments, and soluble sugars. Our results suggest possible trade-offs between hydraulic integrity and gas exchange associated with photosynthesis during drought, or between hydraulic integrity and temperature regulation during combined heat and drought stress. We discuss our findings as potential indicators of adaptation to different ecological niches for *D. cochinchinensis* and *D. oliveri* and the implications for their conservation and restoration.

4.5.1 Water relations

The observed differences in stomatal conductance (g_s) and water potential (Ψ_{MD}) over time reflected contrasting hydraulic responses in the two species: (1) stomatal closure occurred at a higher soil water content (or lower water deficit) in *D. oliveri* leaves than in *D. cochinchinensis,* and (2) Ψ_{MD} was relatively stable in *D. oliveri* until the last day of the drought treatment but decreased rapidly in *D. cochinchinensis.* The ability to maintain Ψ_{MD} and regression of g_s in response to increasing water deficit are common differentiators of iso-anisohydric behaviours[45]: *D. oliveri* fit more closely with an isohydric response, characterized by earlier stomatal closure and thus maintaining a more constant Ψ_{MD}[46]. By contrast, *D. cochinchinensis* was more anisohydric, maintaining stomatal aperture until experiencing severe drought, when Ψ_{MD} decreases drastically[47]. However, recent research suggests that species naturally present a continuum of iso-anisohydry, and very few species could be categorised in

a strict dichotomy of iso-anisohydry[48,49]. Iso-anisohydry is closely related to a plant's life history traits and survival strategy. Isohydry implies a more conservative water-balance management strategy to prevent loss of water via transpiration[50], which can protect plants from hydraulic failures and xylem embolism. Anisohydry maintains higher carbon assimilation at mild drought conditions and thus achieves an overall higher productivity[51]. A previous study among species of Bornean rainforest found that anisohydry is more prevalent and reduces the risk of drought-induced hydraulic failure[52]. However, satellite data suggested the contrary, i.e. that isohydry is more dominant in wet areas[53]. These conflicting results reflect an important role that environment plays in shaping water relations and call for investigation of more aspects of plant hydraulic functions[45]. More studies on specific systems and species are also needed to improve our understanding of species' responses to drought in tropical forests under progressive climate change.

The heat shock applied in this study resulted in rapid closure of stomata in both well-watered and drought stressed plants of both species, suggesting that drought and heat-shock stress were additive in reducing g_s. Transpiration via the stomatal aperture is a mechanism of heat regulation[54]. Previous studies have shown that many plants achieve heat loss via transpiration, with stomatal conductance either maintained at normal levels[38] or increased[31] to prevent leaf temperatures from reaching harmful levels. By contrast, our findings are similar to those reported in olive trees[55], in which g_s decreased with heat shock, and this could be a mechanism for plants adapted to low water availability to reduce the risk of xylem embolism. The level of heat stress tested here produced a conservative response in both *D. oliveri* and *D. cochinchinensis*. Stomatal responses may vary at different levels and durations of heat.

4.5.2 Leaf traits

We found no significant effect of drought or heat shock on leaf dry matter content (LDMC) or specific leaf area (SLA). LDMC and SLA are two leaf traits related to resource use and trade-offs. High SLA or low LDMC imply rapid assimilation for growth and production, whereas low SLA or high LDMC imply efficient conservation of nutrients within structural and well-protected tissues[56]. The temporal limitation in this study may not reflect effects in SLA as leaf development potentially operates over a longer timeframe. Previous studies mainly suggested that drought stress caused decreased SLA to reduce the surface area for transpiration and thus prevent water loss[57,58].

We found a higher SLA in *D. oliveri* than *D. cochinchinensis*, however *D. oliveri* was more isohydric in terms of hydraulic traits. This observation seems to contradict the above claim that SLA decreases plastically in response to drought. However, there is little understanding of variation of SLA across species. Poorter et al.[59] in their meta-analysis suggested that SLA varies strongly with many factors including light, temperature, nutrient, and functional groups. An increase in SLA could be a compensating mechanism for reduced carbon allocation to the leaves during stress[60]. In response to heat stress, increase in SLA could provide a mechanism for temperature regulation by increasing surface area[61]. These features suggest that isohydric *D. oliveri* may suffer an impaired carbon assimilation under drought stress and compensate through associated leaf traits.

4.5.3 Photosynthetic pigments and carbon assimilation

We found that levels of chlorophyll a and carotenoids increased in *D. cochinchinensis* in response to drought and decreased in *D. oliveri*. Chlorophyll content is one of the most commonly used parameters to measure the severity of drought stress[62] as drought stress is thought to damage photosynthetic apparatus and diminish chlorophyll content[63]. By

maintaining lower chlorophyll content under severe drought conditions, plants are protected from photo-oxidative damage by inhibiting photosynthesis and avoiding excess light excitation energy[30,64]. As such, the decreasing trend of chlorophyll a in *D. oliveri* implies correspondence to drought avoidance as may be expected with an isohydric behaviour. On the contrary, some studies have suggested that chlorophyll content is positively associated with drought tolerance and recovery[65], and that the potential photo-oxidative damage can be remediated by increasing levels of carotenoids, which play a central role in the assembly of the light-harvesting complex in the photosystem, regulate photomorphogenesis and provide photoprotection[66]. Carotenoids are also precursors for two plant hormones, strigolactones and abscisic acid (ABA) from carotenoid cleavage[67], while ABA has a dominant role in regulation of stomatal conductance in response to drought stress[68]. Therefore, the contrasting trends of photosynthetic pigments further support the conclusion that *D. oliveri* avoids drought by reducing carbon assimilation and maintaining water potential, while *D. cochinchinensis* tolerates drought and maintains its productivity.

The higher levels of photosynthetic pigments associated with these increases in *D. cochinchinensis* may explain its higher total soluble sugars (TSS). Higher TSS may also provide an energy buffer to sustain metabolic activity during drought stress[69]. The lower TSS in *D. oliveri* potentially implies a carbon limitation, which is commonly associated with isohydric species and could lead to carbon starvation in the long-term[23]. Our finding that TSS increased in response to drought stress may suggest a mechanism of osmoregulation[70], in which osmotically active soluble sugars can help maintain water potential and hydraulic integrity[71].

In our study, we found little response specifically to heat shock in either species, thus potentially suggesting that these trees could have a higher heat threshold than 38°C for 4 hours. We did not observe any synergistic or antagonistic effect of both drought and heat stresses as

in previous studies[37–39], except g_s. There is an emerging interest in studying plant responses to a suite of stress factors, as different stresses may lead different signaling pathways to interact and conflict to produce novel physiological responses[37]..

4.5.4 Relating life-history traits with potential ecological niches

D. cochinchinensis and *D. oliveri* are endemic to and geographically co-occur in Cambodia, Laos, Thailand, and Vietnam. *D. cochinchinensis* is described as an intermediate pioneering species, characteristic of faster growth rate in young age[72]. Its anisohydric behaviour, described here as maximising carbon assimilation at the risk of hydraulic failure during drought, may be associated with its higher productivity for more efficient colonisation in early ecological succession. This fits well with an earlier formulation by Smith and Huston[73] that light-demanding pioneers would optimise growth and outcompete shade-tolerant species in drier habitats. Other studies also support the theory that pioneers run the risk of drought-induced cavitation in order to maintain higher hydraulic efficiency and meet the water demand for photosynthesis and fast growth during drought[74,75]. Such a trade-off between short-term growth and long-term survival aligns with the spectrum from pioneers to shade-tolerant species[76].

On the other hand, *D. oliveri* demonstrates a wider ecological amplitude within the deciduous forest, from relatively rich, deep soils to poor, shallow soils[77]. Aerts et al.[77] suggested that *D. oliveri* can grow on shallow, eroded regosols where drought stress is persistent during the dry season. We report a more isohydric behaviour in *D. oliveri* and hypothesise that it is more conservative in order to allow survival under low water availability of the region's deciduous forests.

The controlled nature of the experiment in this study allowed studying the effects of multiple stressors and assessing their interactions[78]. Although the controlled experiment does

not replicate the environmental and growth conditions in the field, its focus on seedlings shows close relevance to a crucial regeneration stage when mortality and selection pressures are high[79,80]. We believe our results need further validation with field experiments and observations. At the same time, our results inform design of such field studies, for example by indicating the need to study multiple stressors and the divergent response strategies of individual species.

The contrasting hydraulic strategies of *D. cochinchinensis* and *D. oliveri* may reflect adaptation to their current ecological niche, but such dynamics may change drastically with changing climate and affect their survival and regeneration. The extent to which these water-carbon trade-offs affect growth and survival depend on species-specific vulnerability to and capacity for recovery from hydraulic failure and carbon starvation. We recommend that research is needed to study responses and recovery to the extreme stresses at different life stages and under probable climate scenarios within their ranges.

4.6 Conclusion

To date, our understanding of physiology and ecological function in *D. cochinchinensis* and *D. oliveri* is poorly developed compared to that of their mating and genetic structure[81,82]. This knowledge gap hinders informed decision-making for the conservation and restoration of these valuable species. We present the first study on hydraulic traits and carbon assimilation of *D. cochinchinensis* and *D. oliveri* in their seedling stages, and characterise their iso-anisohydric behaviour in response to drought and heat stresses. We suggest potential thresholds and interactions of drought and heat stresses, which will open the opportunity for further studies to gain a better understanding of their physiology and stress response. Such an approach would also be beneficial to studying stress responses in other *Dalbergia* species to similarly inform their conservation and use.

Dalbergia species regenerate naturally when circumstances permit, but are also suited to use for planting in restoring degraded forests and restoring deforested sites[83]. Seedling establishment is a crucial life stage of trees and an understanding of their stress responses will contribute to appropriate site selection and effective long-term regeneration[84,85], especially with regard to restoration efforts and climate change.

4.7 Acknowledgements

This work was supported by funding to T.H.H. from the Biotechnology and Biological Sciences Research Council (BBSRC) [grant number BB/M011224/1], to T.S., B.T., H.T.T, R.J., D.H.B., J.J.M from the UK Darwin Initiative (ref. 25-023), and to J.J.M from the Oxfordshire Local Enterprise Partnership.

4.8 Data accessibility statement

The research materials supporting this publication can be publicly accessed in Dryad (doi.org/10.5061/dryad.v6wwpzgt1).

4.9 References

1. Vatanparast, M. *et al.* First molecular phylogeny of the pantropical genus *Dalbergia*: Implications for infrageneric circumscription and biogeography. *South African J. Bot.* **89**, 143–149 (2013).

2. Winfield, K., Scott, M. & Graysn, C. Global status of *Dalbergia* and *Pterocarpus* rosewood producing species in trade. in *Convention on International Trade in Endangered Species 17th Conference of Parties - Johannesburg* (2016).

3. Bhagwat, R. M., Dholakia, B. B., Kadoo, N. Y., Balasundaran, M. & Gupta, V. S. Two new potential barcodes to discriminate *Dalbergia* species. *PLoS One* **10**, 1–18 (2015).

4. EIA. *The Rosewood Racket: China's billion dollar illegal timber trade and the devastation of Nigeria's forests*. (2017).

5. CDRI. Community forestry for sustainable forest management and livelihhods: a case study of Osoam Community Forest users group. *Cambodia Dev. Rev.* **18**, (2014).

6. Luoma-aho, T., Hong, L. T., Ramanatha Rao, V. & Sim, H. C. *Forest genetic resources conservation and management: Proceedings of the Asia Pacific Forest Genetic Resources Programme (APFORGEN) Inception Workshop*. (2003).

7. APFORGEN. *Conserving Rosewood genetic resources for resilient livelihoods in the Mekong: Project inception workshop report*. (2018).

8. Aerts, R. *et al.* Site requirements of the endangered rosewood *Dalbergia oliveri* in a tropical deciduous forest in northern Thailand. *For. Ecol. Manage.* **259**, 117–123 (2009).

9. Tanaka, N. *et al.* A review of evapotranspiration estimates from tropical forests in Thailand and adjacent regions. *Agric. For. Meteorol.* **148**, 807–819 (2008).

10. Kaewkrom, P., Gajaseni, J., Jordan, C. F. & Gajaseni, N. Floristic regeneration in five types of teak plantations in Thailand. *For. Ecol. Manage.* **210**, 351–361 (2005).

11. Sakai, A. *et al.* Effect of species and spacing of fast-growing nurse trees on growth of an indigenous tree, *Hopea odorata* Roxb., in northeast Thailand. *For. Ecol. Manage.* **257**, 644–652 (2009).

12. Maningo, E. V. & Thea, S. *Regional project for prootion of forest rehabilitation in Cambodia and Vietnam through demonstration models and improvement of seed supply system: lesson learned*.

13. FAO. Climate-smart crop production practices and technologies. in *Climate Smart Agriculture Sourcebook*.

14. Jalonen, R., Valette, M., Boshier, D., Duminil, J. & Thomas, E. Forest and landscape restoration severely constrained by a lack of attention to the quantity and quality of tree seed: Insights from a global survey. *Conserv. Lett.* **11**, e12424 (2018).

15. MRC. *State of the basin report 2010*. (2010).

16. Hughes, A. C. Understanding the drivers of Southeast Asian biodiversity loss. *Ecosphere* **8**, e01624 (2017).

17. Li, Y. *et al.* Local cooling and warming effects of forests based on satellite observations. *Nat. Commun.* **6**, 1–8 (2015).

18. Estravis-Barcala, M. *et al.* Molecular bases of responses to abiotic stress in trees. *J. Exp. Bot.* erz532 (2019) doi:10.1093/jxb/erz532.

19. Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T. & Curtis-McLane, S. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol. Appl.* **1**, 95–111 (2008).

20. SITCH, S. *et al.* Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs). *Glob. Chang. Biol.* **14**, 2015–2039 (2008).

21. Allen, C. D., Breshears, D. D. & McDowell, N. G. On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene. *Ecosphere* **6**, 1–55 (2015).

22. Allen, C. D. *et al.* A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *For. Ecol. Manage.* **259**, 660–684 (2010).

23. McDowell, N. *et al.* Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? *New Phytol.* **178**, 719–739 (2008).

24. Pollastrini, M., Puletti, N., Selvi, F., Iacopetti, G. & Bussotti, F. Widespread crown defoliation after a drought and heat wave in the forests of tuscany (Central Italy) and their recovery—A case study from summer 2017. *Front. For. Glob. Chang.* **2**, 74 (2019).

25. Canham, C. D. & Murphy, L. The demography of tree species response to climate: seedling recruitment and survival. *Ecosphere* **7**, e01424 (2016).

26. Brodribb, T. J., Powers, J., Cochard, H. & Choat, B. Hanging by a thread? Forests and drought. *Science* vol. 368 261–266 (2020).

27. Chaves, M. M. *et al.* How plants cope with water stress in the field? Photosynthesis and growth. *Ann. Bot.* **89**, 907–916 (2002).

28. Terra, M. de C. N. S. *et al.* Water availability drives gradients of tree diversity,

structure and functional traits in the Atlantic–Cerrado–Caatinga transition, Brazil. *J. Plant Ecol.* **11**, 803–814 (2018).

29. Polle, A., Chen, S. L., Eckert, C. & Harfouche, A. Engineering drought resistance in forest trees. *Frontiers in Plant Science* vol. 9 1875 (2019).

30. Pintó-Marijuan, M. & Munné-Bosch, S. Photo-oxidative stress markers as a measure of abiotic stress-induced leaf senescence: advantages and limitations. *J. Exp. Bot.* **65**, 3845–3857 (2014).

31. Urban, J., Ingwers, M., McGuire, M. A. & Teskey, R. O. Stomatal conductance increases with rising temperature. *Plant Signal. Behav.* **12**, (2017).

32. McDowell, N. *et al.* Drivers and mechanisms of tree mortality in moist tropical forests. *New Phytologist* vol. 219 851–869 (2018).

33. Rizhsky, L., Liang, H. & Mittler, R. The combined effect of drought stress and heat shock on gene expression in Tobacco. *Plant Physiol.* **130**, 1143–1151 (2002).

34. Li, X. *et al.* Comparative physiological and proteomic analyses of poplar (*Populus yunnanensis*) plantlets exposed to high temperature and drought. *PLoS One* **9**, e107605 (2014).

35. TESKEY, R. *et al.* Responses of tree species to heat waves and extreme heat events. *Plant. Cell Environ.* **38**, 1699–1712 (2015).

36. Stéfanon, M., Drobinski, P., D'Andrea, F., Lebeaupin-Brossier, C. & Bastin, S. Soil moisture-temperature feedbacks at meso-scale during summer heat waves over Western Europe. *Clim. Dyn.* **42**, 1309–1324 (2014).

37. Zandalinas, S. I., Mittler, R., Balfagón, D., Arbona, V. & Gómez-Cadenas, A. Plant adaptations to the combination of drought and high temperatures. *Physiologia*

Plantarum vol. 162 2–12 (2018).

38. Correia, B. *et al.* Combined drought and heat activates protective responses in *Eucalyptus globulus* that are not activated when subjected to drought or heat stress alone. *Front. Plant Sci.* **9**, 819 (2018).

39. Bansal, S., Hallsby, G., Löfvenius, M. O. & Nilsson, M.-C. Synergistic, Additive and antagonistic impacts of drought and herbivory on *Pinus Sylvestris*: leaf, tissue and whole-plant responses and recovery. *Tree Physiol.* **33**, (2013).

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

41. Pérez-Harguindeguy, N. *et al.* New handbook for standardised measurement of plant functional traits worldwide. *Aust. J. Bot.* **61**, 167–234 (2013).

42. Sims, D. A. & Gamon, J. A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.* **81**, 337–354 (2002).

43. Osaki, M., Shinano, T. & Tadano, T. Redistribution of carbon and nitrogen compounds from the shoot to the harvesting organs during maturation in field crops. *Soil Sci. Plant Nutr.* **37**, 117–128 (1991).

44. Bates, D. M. & Watts, D. G. *Nonlinear Regression Analysis and Its Applications*. (Wiley, 1988).

45. Hochberg, U., Rockwell, F. E., Holbrook, N. M. & Cochard, H. Iso/Anisohydry: A plant-environment interaction rather than a simple hydraulic trait. *Trends Plant Sci.* **23**, 112–120 (2018).

46. Lavoie-Lamoureux, A., Sacco, D., Risse, P.-A. & Lovisolo, C. Factors influencing

stomatal conductance in response to water availability in grapevine: a meta-analysis. *Physiol. Plant.* **159**, 468–482 (2017).

47. Bergel-Landefeldt, U. *Das Wasserhaushalt der Alpenpflanzen (Bibliotheca Botanica Vol. 115)*. (Schweizerbartsche Verlagsbuchhandlung, 1936).

48. Martínez-Vilalta, J., Poyatos, R., Aguadé, D., Retana, J. & Mencuccini, M. A new look at water transport regulation in plants. *New Phytol.* **204**, 105–115 (2014).

49. Klein, T. The variability of stomatal sensitivity to leaf water potential across tree species indicates a continuum between isohydric and anisohydric behaviours. *Funct. Ecol.* **28**, 1313–1320 (2014).

50. Sade, N., Gebremedhin, A. & Moshelion, M. Risk-taking plants: anisohydric behavior as a stress-resistance trait. *Plant signaling & behavior* vol. 7 767–770 (2012).

51. Urli, M. *et al.* Xylem embolism threshold for catastrophic hydraulic failure in angiosperm trees. *Tree Physiol.* **33**, 672–683 (2013).

52. Kumagai, T. & Porporato, A. Strategies of a Bornean tropical rainforest water use as a function of rainfall regime: Isohydric or anisohydric? *Plant, Cell Environ.* **35**, 61–71 (2012).

53. Li, Y. *et al.* Estimating Global Ecosystem Isohydry/Anisohydry Using Active and Passive Microwave Satellite Data. *J. Geophys. Res. Biogeosciences* **122**, 3306–3321 (2017).

54. Zhou, H. H., Chen, Y. N., Li, W. H. & Chen, Y. P. Photosynthesis of *Populus euphratica* in relation to groundwater depths and high temperature in arid environment, northwest China. *Photosynthetica* **48**, 257–268 (2010).

55. Haworth, M. *et al.* The impact of heat stress and water deficit on the photosynthetic

and stomatal physiology of olive (*Olea europaea* L.)—A case study of the 2017 heat wave. *Plants* **7**, (2018).

56. Vitra, A. *et al.* Responses of plant leaf economic and hydraulic traits mediate the effects of early- and late-season drought on grassland productivity. *AoB Plants* **11**, (2019).

57. Liu, F. & Stützel, H. Biomass partitioning, specific leaf area, and water use efficiency of vegetable amaranth (*Amaranthus* spp.) in response to drought stress. *Sci. Hortic. (Amsterdam).* **102**, 15–27 (2004).

58. Pandey, P., Ramegowda, V. & Senthil-Kumar, M. Shared and unique responses of plants to multiple individual stresses and stress combinations: Physiological and molecular mechanisms. *Frontiers in Plant Science* vol. 6 (2015).

59. Poorter, H., Niinemets, Ü., Poorter, L., Wright, I. J. & Villar, R. Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis. *New Phytol.* **182**, 565–588 (2009).

60. Aspelmeier, S. & Leuschner, C. Genotypic variation in drought response of silver birch (*Betula pendula* Roth): Leaf and root morphology and carbon partitioning. *Trees - Struct. Funct.* **20**, 42–52 (2006).

61. Pandey, P., Ramegowda, V. & Senthil-Kumar, M. Shared and unique responses of plants to multiple individual stresses and stress combinations: physiological and molecular mechanisms. *Front. Plant Sci.* **6**, 723 (2015).

62. Ying, Y. Q. *et al.* Physiological response to drought stress in *Camptotheca acuminata* seedlings from two provenances. *Front. Plant Sci.* **6**, 1–8 (2015).

63. Fu, J. & Huang, B. Involvement of antioxidants and lipid peroxidation in the

adaptation of two cool-season grasses to localized drought stress. *Environ. Exp. Bot.* **45**, 105–114 (2001).

64. Aranjuelo, I., Molero, G., Erice, G., Avice, J. C. & Nogué, S. Plant physiology and proteomics reveals the leaf response to drought in alfalfa (*Medicago sativa* L.). *J. Exp. Bot.* **62**, 111–123 (2011).

65. Chen, D. *et al.* Genotypic variation in growth and physiological response to drought stress and re-watering reveals the critical role of recovery in drought adaptation in maize seedlings. *Front. Plant Sci.* **6**, 1241 (2016).

66. Munné-Bosch, S. & Alegre, L. Changes in carotenoids, tocopherols and diterpenes during drought and recovery, and the biological significance of chlorophyll loss in *Rosmarinus officinalis* plants. *Planta* **210**, 925–931 (2000).

67. Nisar, N., Li, L., Lu, S., Khin, N. C. & Pogson, B. J. Carotenoid metabolism in plants. *Molecular Plant* vol. 8 68–82 (2015).

68. Pirasteh-Anosheh, H., Saed-Moucheshi, A., Pakniyat, H. & Pessarakli, M. Stomatal responses to drought stress. in *Water Stress and Crop Plants: A Sustainable Approach* vols 1–2 24–40 (wiley, 2016).

69. Duan, H. *et al.* Drought responses of two gymnosperm species with contrasting stomatal regulation strategies under elevated [CO2] and temperature. *Tree Physiol.* **35**, 756–770 (2015).

70. Granda, E. & Camarero, J. J. Drought reduces growth and stimulates sugar accumulation: new evidence of environmentally driven non-structural carbohydrate use. *Tree Physiol.* **37**, 997–1000 (2017).

71. O'Brien, M. J., Leuzinger, S., Philipson, C. D., Tay, J. & Hector, A. Drought survival

of tropical tree seedlings enhanced by non-structural carbohydrate levels. *Nat. Clim. Chang.* **4**, 710–714 (2014).

72. So, N. V. The potential of local tree species to accelerate natural forest succession on marginal grasslands in Southern Vietnam. in *Proceedings of the Workshop on Forst Restoration for Wildlife Restoration for Wildlife Conservation* (eds. Elliot, S., Kerby, J. & Blakesly, K. et al.) (International Tropical Timber Organization and The Forest Restoration Unit, Chiang Mai University).

73. Smith, T. & Huston, M. A theory of the spatial and temporal dynamics of plant communities. *Vegetatio* **83**, 49–69 (1989).

74. Markesteijn, L., Poorter, L., Paz, H., Sack, L. & Bongers, F. Ecological differentiation in xylem cavitation resistance is associated with stem and leaf structural traits. *Plant, Cell Environ.* **34**, 137–148 (2011).

75. Markesteijn, L., Poorter, L., Bongers, F., Paz, H. & Sack, L. Hydraulics and life history of tropical dry forest tree species: coordination of species' drought and shade tolerance. *New Phytol.* **191**, 480–495 (2011).

76. L, P. & F, B. Leaf traits are good predictors of plant performance across 53 rain forest species. *Ecology* **87**, (2006).

77. Aerts, R. *et al.* Site requirements of the endangered rosewood *Dalbergia oliveri* in a tropical deciduous forest in northern Thailand. *For. Ecol. Manage.* **259**, 117–123 (2009).

78. Seebacher, F. & Franklin, C. E. Determining environmental causes of biological effects: The need for a mechanistic physiological dimension in conservation biology. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 1607–1614 (2012).

79. Muscarella, R. *et al.* Life-history trade-offs during the seed-to-seedling transition in a subtropical wet forest community. *J. Ecol.* **101**, 171–182 (2013).

80. Lan Qie *et al.* Drought cuts back regeneration in logged tropical forests. *Environ. Res. Lett.* **045012**, (2019).

81. Hartvig, I. *et al.* Population genetic structure of the endemic rosewoods *Dalbergia cochinchinensis* and *D. oliveri* at a regional scale reflects the Indochinese landscape and life-history traits. *Ecol. Evol.* **8**, 530–545 (2018).

82. Hartvig, I. *et al.* Conservation genetics of the critically endangered Siamese rosewood (*Dalbergia cochinchinensis*): recommendations for management and sustainable use. *Conserv. Genet.* 1–16 (2020) doi:10.1007/s10592-020-01279-1.

83. Koonkhunthod, N., Sakurai, K. & Tanaka, S. Composition and diversity of woody regeneration in a 37-year-old teak (*Tectona grandis* L.) plantation in Northern Thailand. *For. Ecol. Manage.* **247**, 246–254 (2007).

84. Fontaine, M. *et al.* Elevation and exposition rather than soil types determine communities and site suitability in Mediterranean mountain forests of southern Anatolia, Turkey. *For. Ecol. Manage.* **247**, 18–25 (2007).

85. Abiyu, A., Teketay, D., Glatzel, G. & Gratzer, G. Seed production, seed dispersal and seedling establishment of two afromontane tree species in and around a church forest: implications for forest restoration. *For. Ecosyst.* **3**, 1–10 (2016).

4.10 Figures and Tables

Figure 4.1. Photo of (**a)** *Dalbergia cochinchinensis* and **(b)** *D. oliveri* in a greenhouse (not taken during this study).

Figure 4.2. Conceptual diagram of the experimental design in this study.

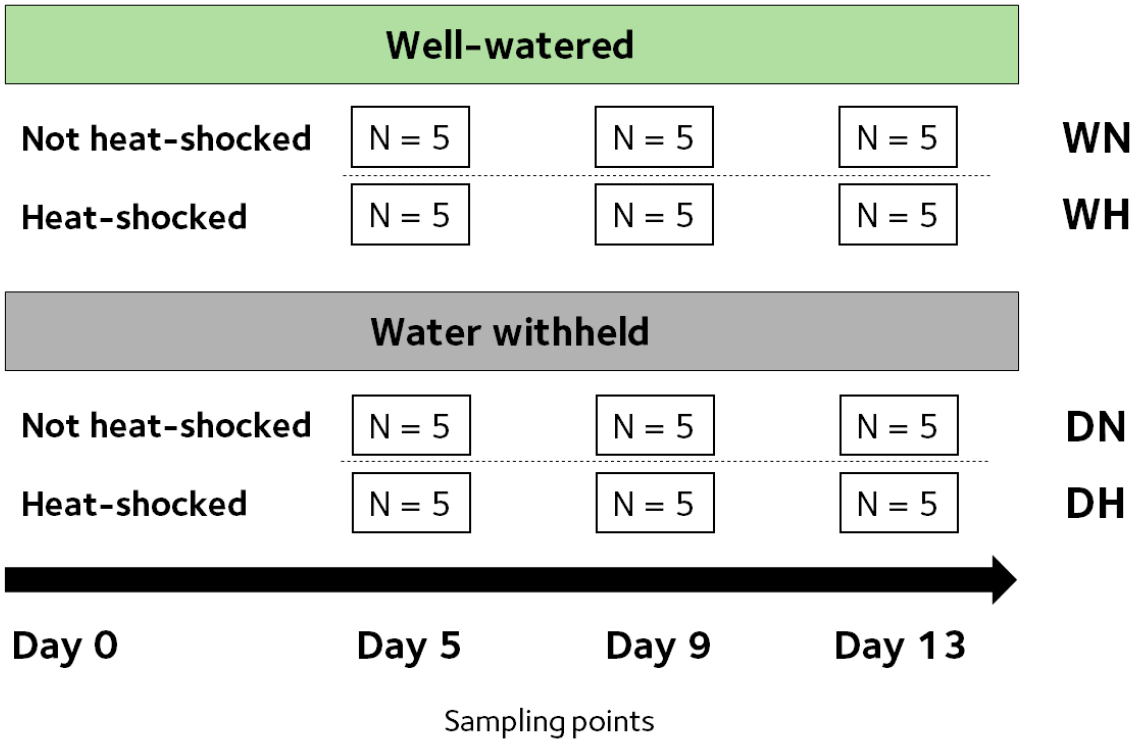


Figure 4.3. Linear-plateau regression of stomatal conductance (g_s) against soil water content (SWC) of *Dalbergia cochinchinensis* (red) and *D. oliveri* (blue). Goodness-of-fit was tested using Nagelkerke method. Values in the brackets represent 95% confidence interval.

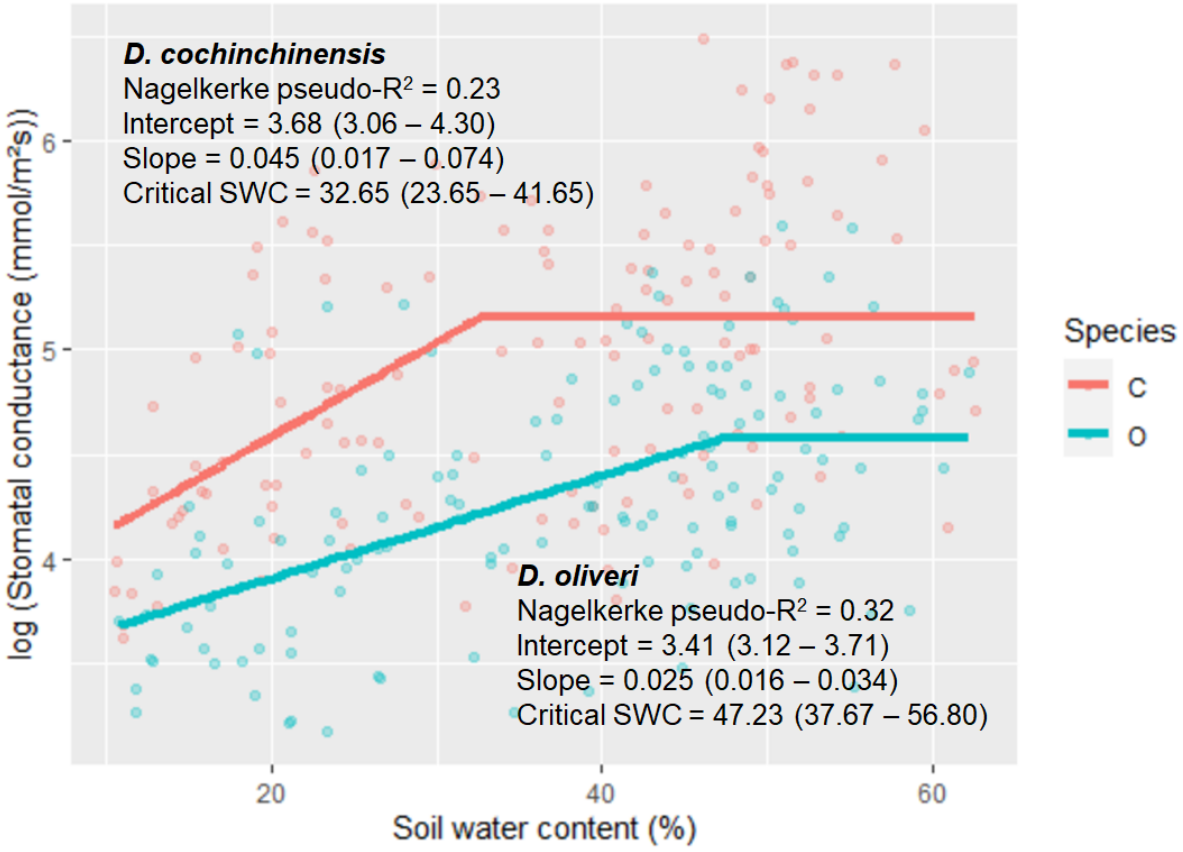
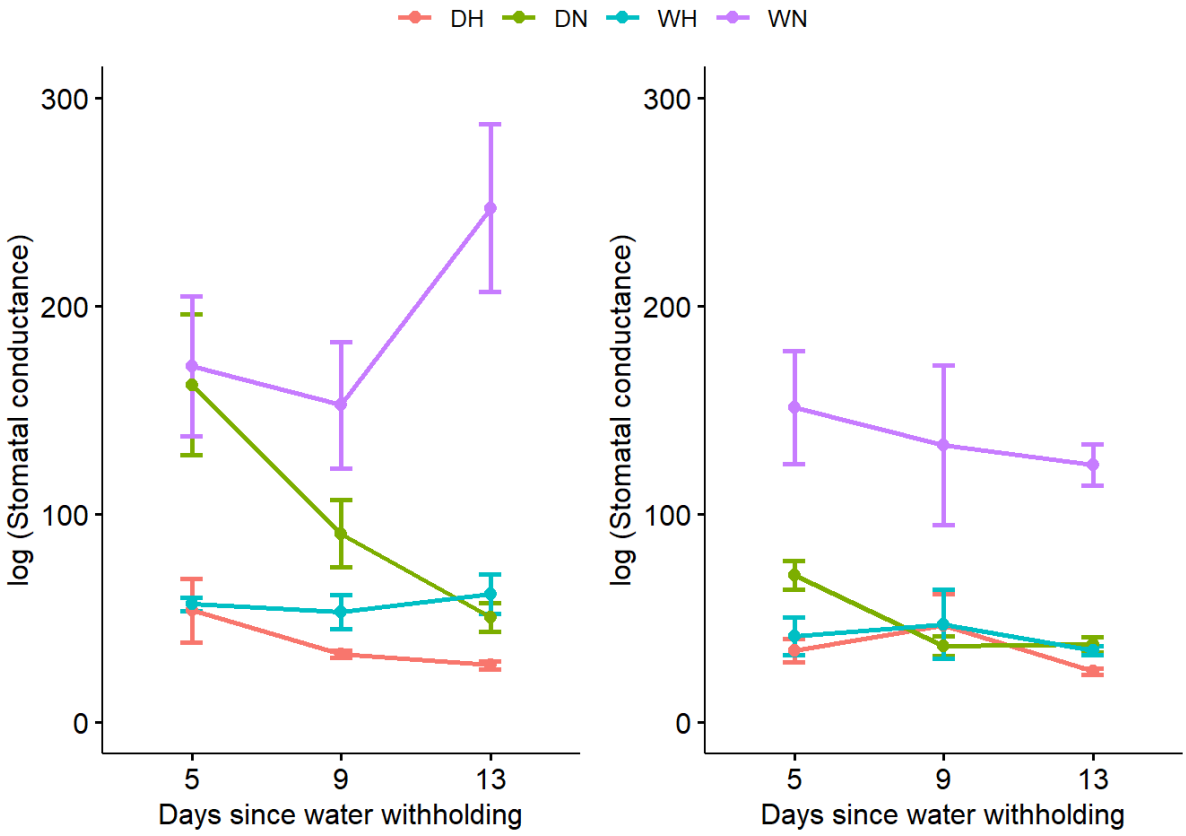


Figure 4.4. **(a)** Stomatal conductance (g_s) and **(b)** midday branch water potential (Ψ_{MD}) of *Dalbergia cochinchinensis* (left) and *D. oliveri* (right) at 5, 9, and 13 days from the beginning of drought treatment. Treatment groups were: DH (drought and heat shock), DN (drought and non-heat shock), WH (well-watered and heat shock), and WN (well-watered and non-heat shock).

(a)



(b)

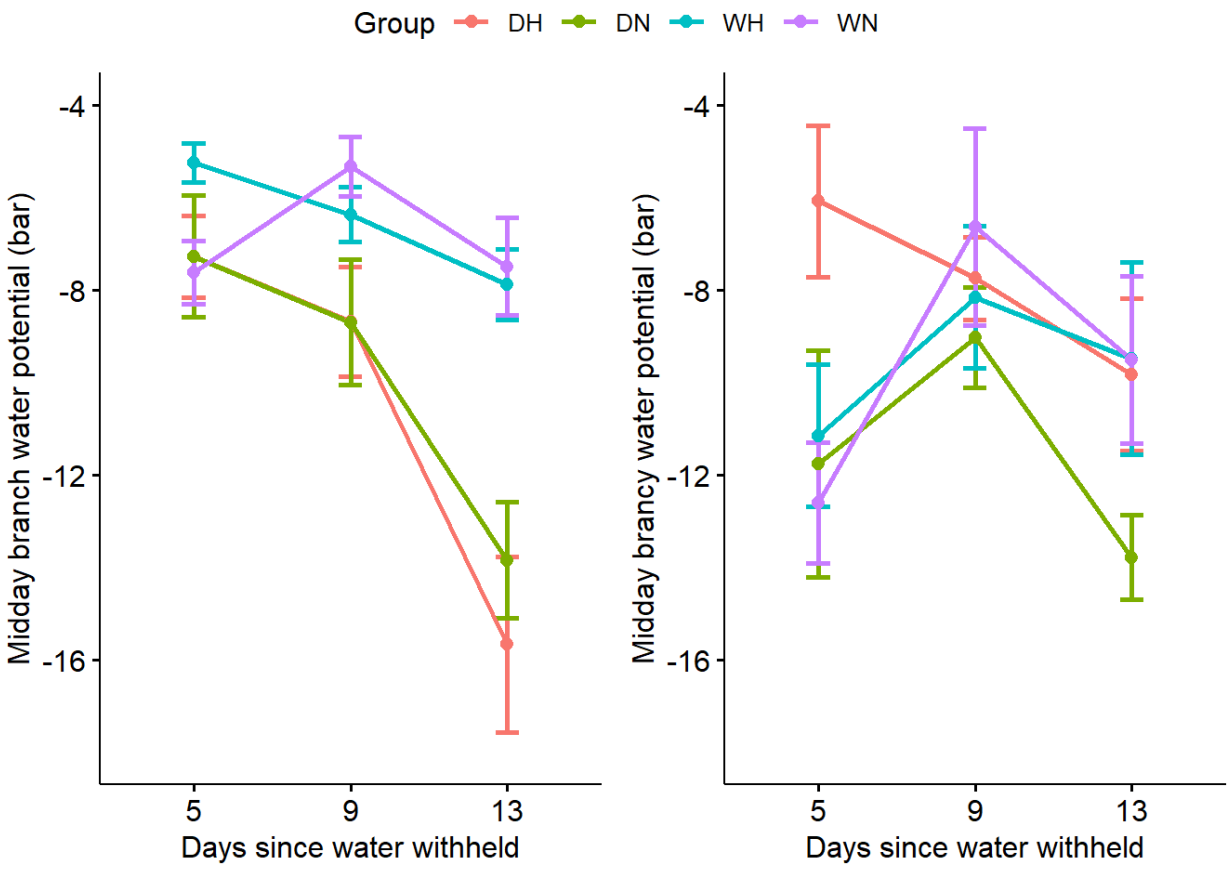
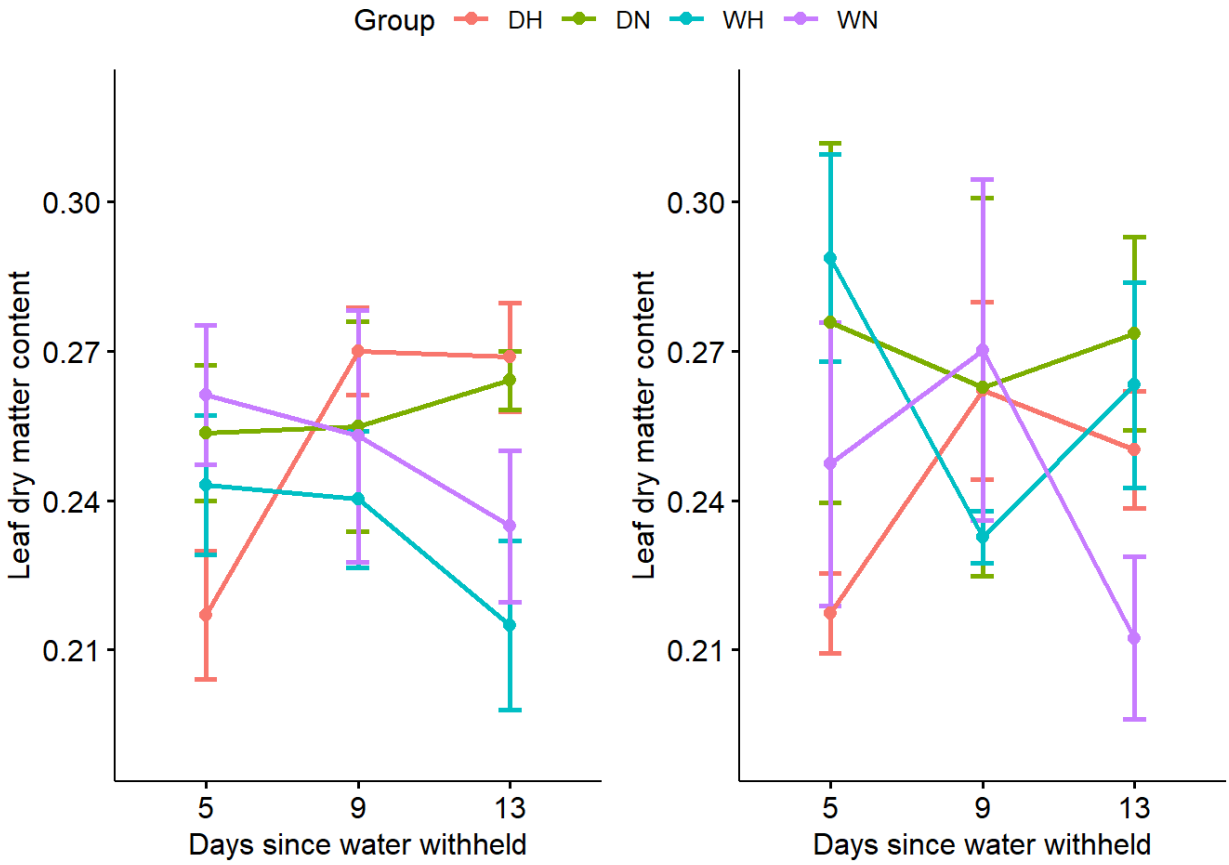

Group
DH
DN
WH
WN
Midday branch water potential (bar)
-4
-8
-12
-16
5
9
13
Days since water withheld
Midday brancy water potential (bar)
-4
-8
-12
-16
5
9
13
Days since water withheld

Figure 4.5. **(a)** Leaf dry matter content (LDMC) and **(b)** specific leaf area (SLA) of *Dalbergia cochinchinensis* (left) and *D. oliveri* (right) at 5, 9, and 13 days from the beginning of drought treatment. Treatment groups were: DH (drought and heat shock), DN (drought and non-heat shock), WH (well-watered and heat shock), and WN (well-watered and non-heat shock).

(a)



(b)

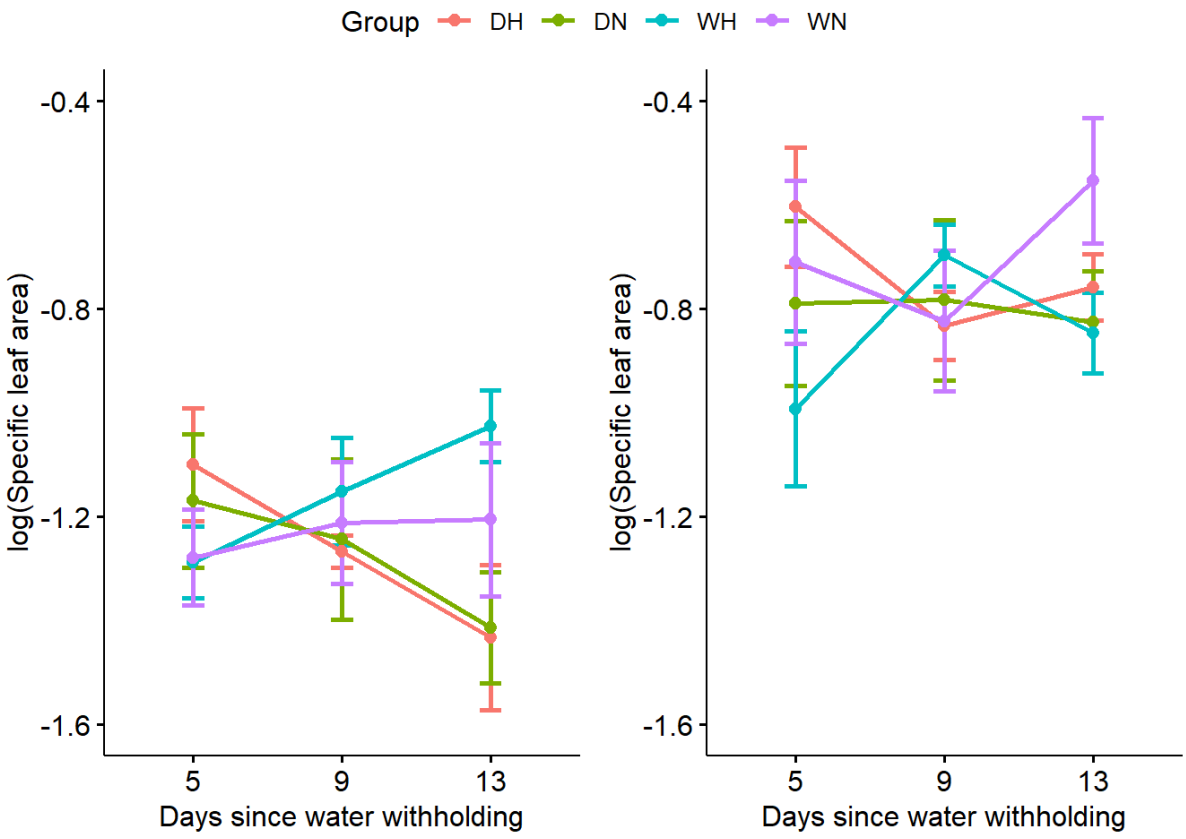
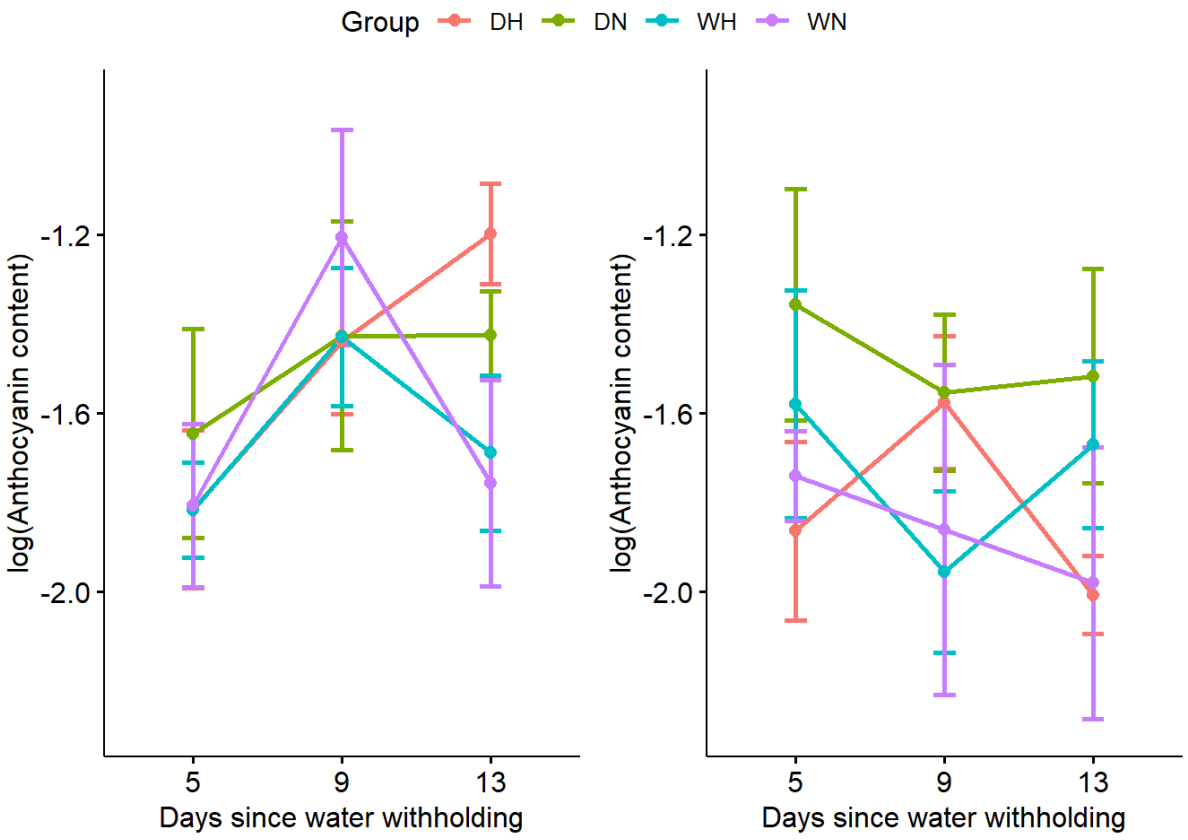
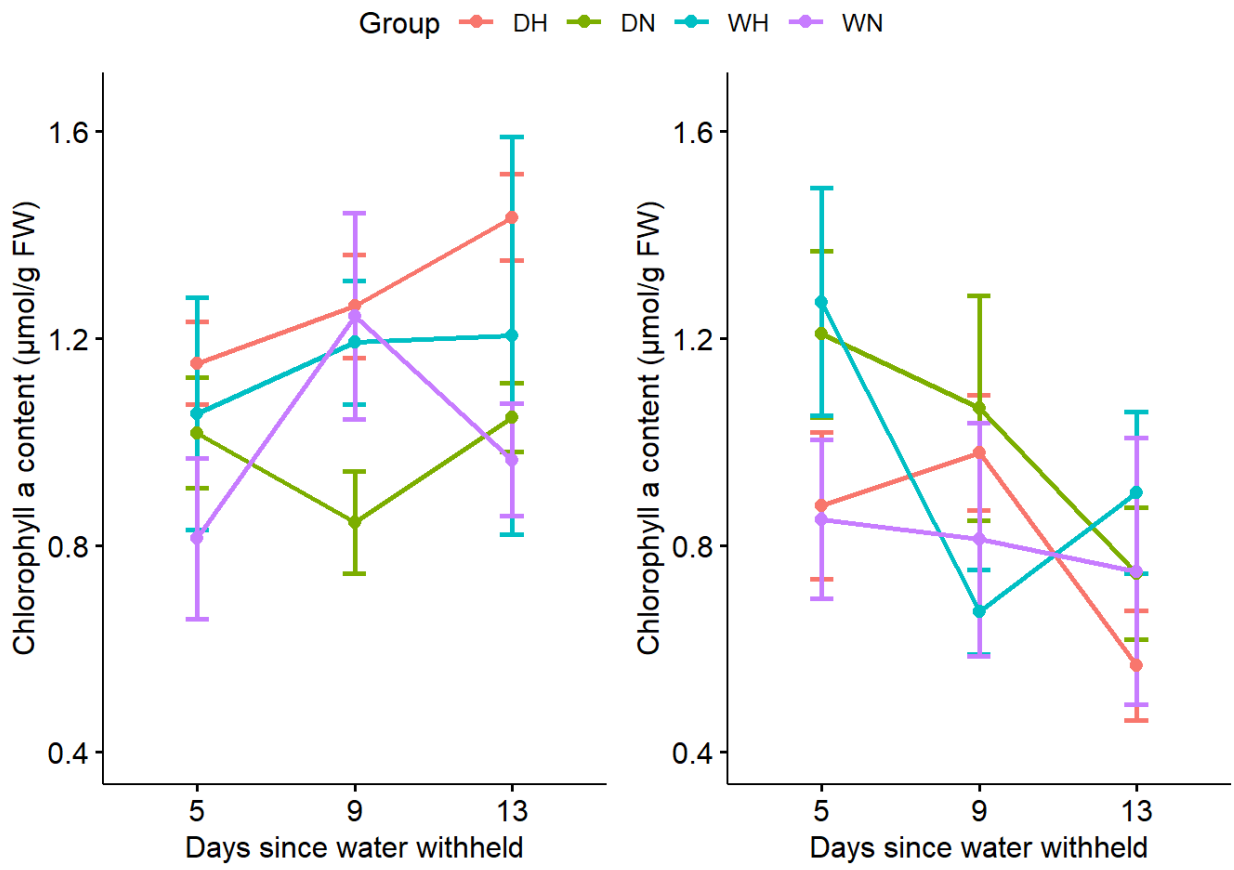

Group
DH
DN
WH
WN
log(Specific leaf area)
-0.4
-0.8
-1.2
-1.6
5
9
13
Days since water withholding

Figure 4.6. **(a)** Anthocyanin content, **(b)** chlorophyll a content (Chl_a), **(c)** chlorophyll b content (Chl_b), and **(d)** carotenoid content of *Dalbergia cochinchinensis* (left) and *D. oliveri* (right) at 5, 9, and 13 days from the beginning of drought treatment. Treatment groups were: DH (drought and heat shock), DN (drought and non-heat shock), WH (well-watered and heat shock), and WN (well-watered and non-heat shock).

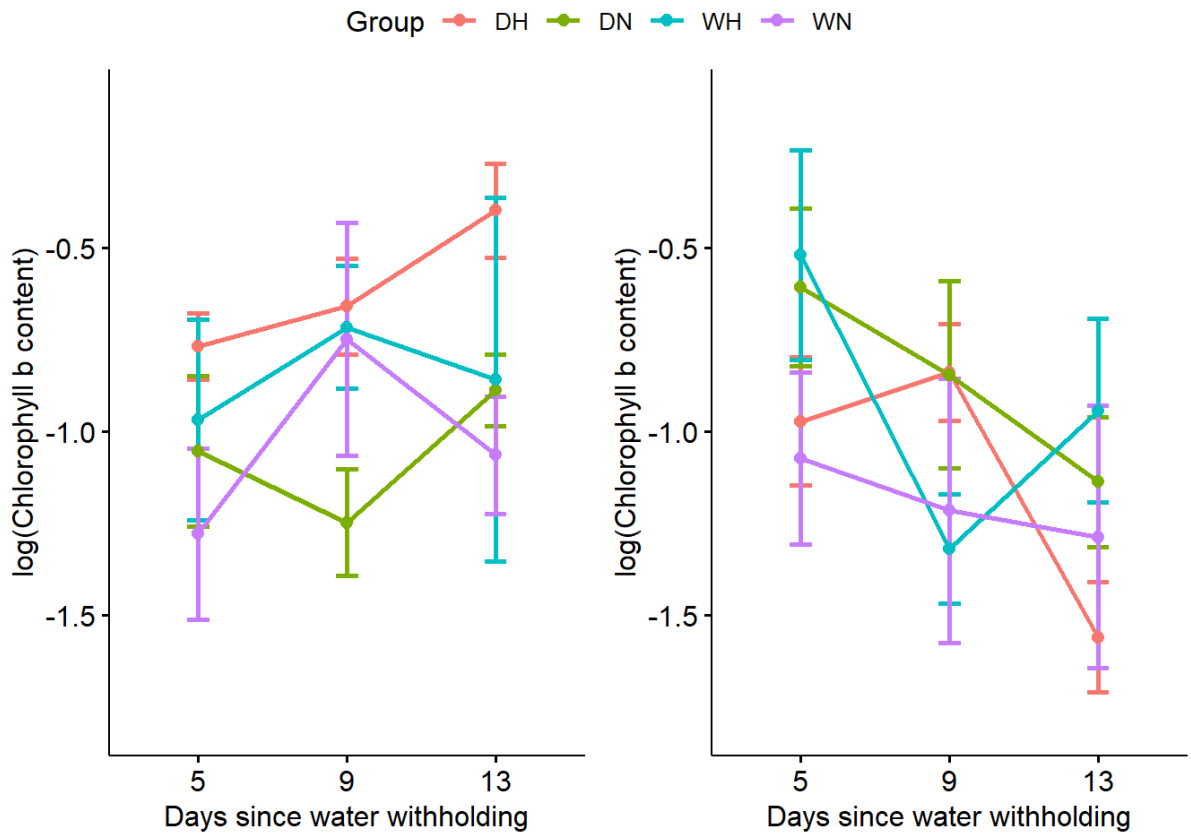
(a)



(b)


Group
DH
DN
WH
WN
Chlorophyll a content (µmol/g FW)
1.6
1.2
0.8
0.4
5
9
13
Days since water withheld
Chlorophyll a content (µmol/g FW)
1.6
1.2
0.8
0.4
5
9
13
Days since water withheld

(c)


Group
DH
DN
WH
WN
log(Chlorophyll b content)
-0.5
-1.0
-1.5
5
9
13
Days since water withholding
log(Chlorophyll b content)
-0.5
-1.0
-1.5
5
9
13
Days since water withholding

(d)

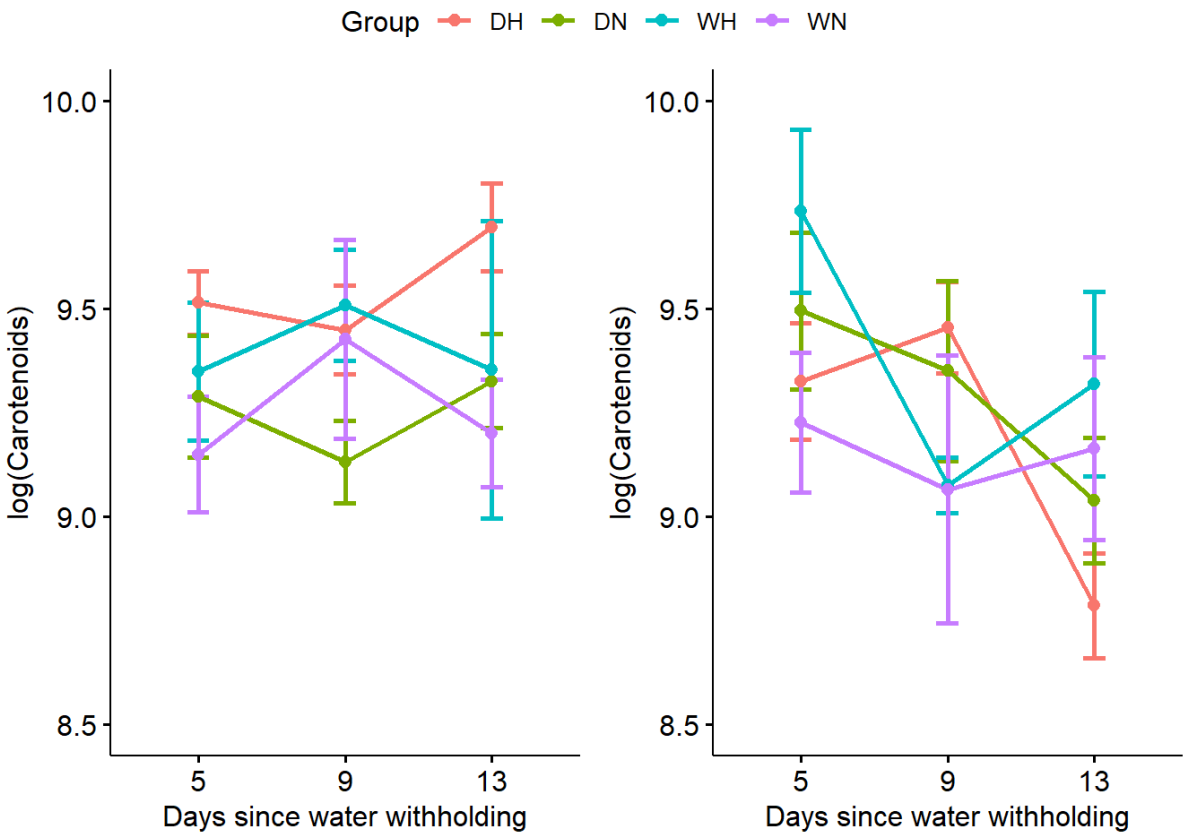

Group
DH
DN
WH
WN
10.0
9.5
9.0
8.5
log(Carotenoids)
5
9
13
Days since water withholding
10.0
9.5
9.0
8.5
log(Carotenoids)
5
9
13
Days since water withholding

Figure 4.7. Total soluble sugars content (TSS) of *Dalbergia cochinchinensis* (left) and *D. oliveri* (right) at 5, 9, and 13 days from the beginning of drought treatment. Treatment groups were: DH (drought and heat shock), DN (drought and non-heat shock), WH (well-watered and heat shock), and WN (well-watered and non-heat shock).

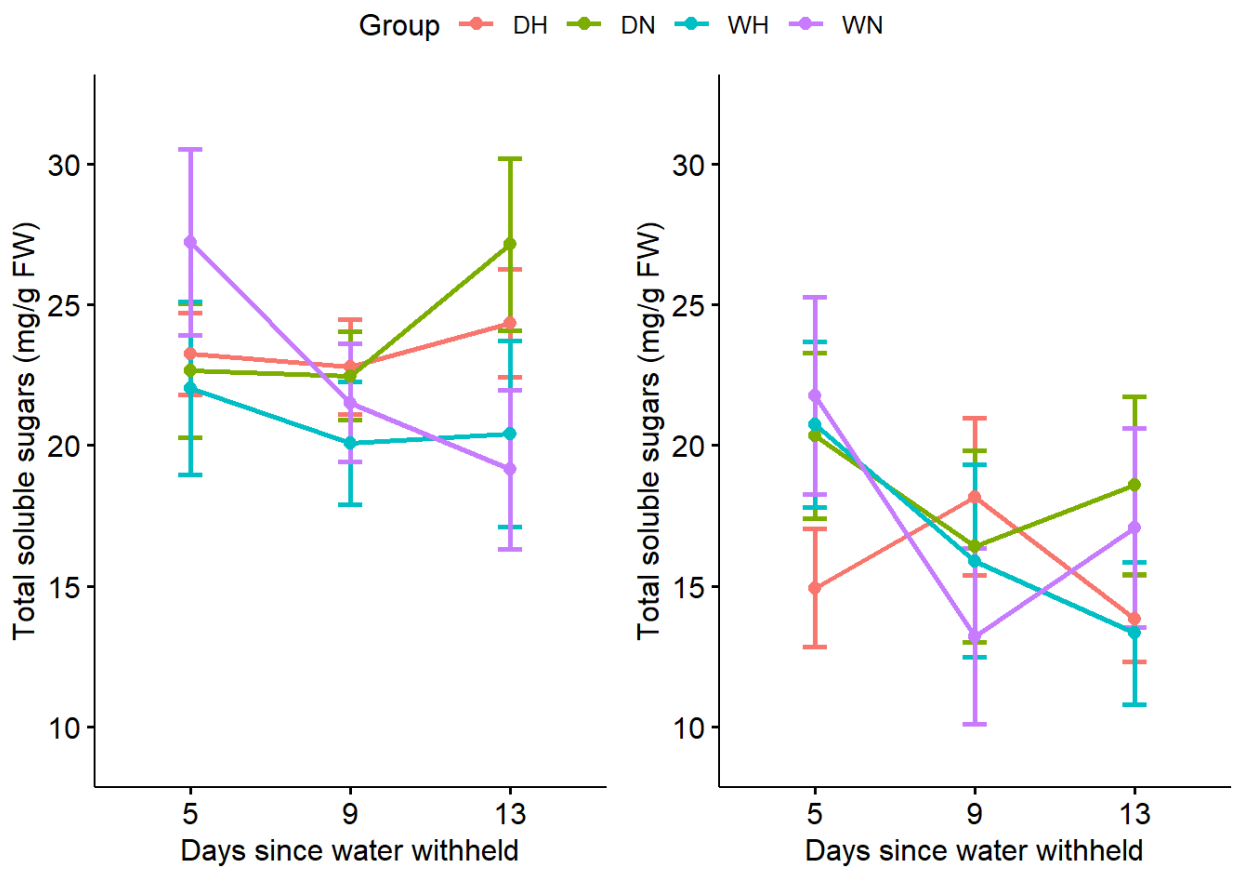


Table 4.1. Analysis of variance (ANOVA) table of the mean slope of daily log-transformed g_s.

Factor	**Df**	**Sum sq**	**Mean sq**	**F value**	**Pr(> F)**
Drought	1	1408.0	1407.95	8.0014	0.0134*
Error (main-plot)	14	2463.5	175.96		
Block	15	5119.6	341.31	9.8187	0.0001507*
Species	1	560.9	560.94	16.1371	0.0017083*
Species : drought	1	9.1	9.09	0.2615	0.6183585
Error	12	417.1	34.76		

Table 4.2. Analysis of variance (ANOVA) table of the mean intercept of daily log-transformed g_s.

Factor	**Df**	**Sum sq**	**Mean sq**	**F value**	**Pr(> F)**
Drought	1	12009	12008.7	1.2234	0.2873
Error (main-plot)	14	137426	9816.1		
Block	15	188827	12588	7.1974	0.0007127*
Species	1	151109	151109	86.3965	7.844e-07*
Species : drought	1	186	186	0.1062	0.7500877
Error (split-plot)	12	20988	1749		

Table 4.3. Analysis of variance (ANOVA) table of the regression on g_s against SWC log-transformed stomatal conductance on sampling days.

Factor	**Df**	**Sum sq**	**Mean sq**	**F value**	**Pr(> F)**
Drought	1	1.39741	1.39741	38.171	2.401e-05*
Error (main-plot)	14	0.51253	0.03661		
Heat shock	1	17.0342	17.0342	1112.0253	2.936e-15*
Block	15	12.0692	0.8046	5.2915	1.576e-06*
Species	1	4.0044	4.0044	26.3351	3.366e-06*
Day	2	1.4667	0.7334	4.8229	0.011482*
Drought : heat shock	1	1.9638	1.9638	12.9148	0.000666*
Day : drought	2	2.0458	1.0229	6.7270	0.002335*
Day : heat shock	2	0.4135	0.2067	1.3597	0.264674
Species : drought	1	0.0034	0.0034	0.0223	0.881672
Species : heat shock	1	0.2863	0.2863	1.8827	0.175225
Species : day	2	0.1423	0.0712	0.4681	0.628517
Day : drought : heat shock	2	0.5062	0.2531	1.6646	0.198029
Species : drought : heat shock	1	0.5349	0.5349	3.5176	0.065669
Species : day : drought	2	0.3421	0.1711	1.1250	0.331503
Speces : day : heat shock	2	0.3117	0.1558	1.0248	0.365172
Species : day : drought : heat shock	2	0.2149	0.1075	0.7067	0.497410
Error (split-plot)	59	8.9713	0.1521		

Table 4.4. Analysis of variance (ANOVA) table of Ψ_{MD} on sampling days.

Factor	**Df**	**Sum sq**	**Mean sq**	**F value**	**Pr(> F)**
Drought	1	8.7392	8.7392	5.1517	0.03955*
Error (main-plot)	14	23.7492	1.6964		
Heat shock	1	7.34	7.341	0.7939	0.376529
Block	15	187.21	12.480	1.3497	0.203269
Species	1	13.46	13.465	1.4562	0.232354
Day	2	260.73	130.364	14.0987	9.892e-06*
Drought : heat shock	1	7.34	7.341	0.7939	0.376529
Day : drought	2	202.06	101.029	10.9261	9.188e-05*
Day : heat shock	2	18.42	9.209	0.9960	0.375480
Species : drought	1	82.32	82.320	8.9028	0.004137*
Species : heat shock	1	16.42	16.424	1.7762	0.187738
Species : day	2	10.55	5.274	0.5704	0.568387
Day : drought : heat shock	2	20.77	10.387	1.1234	0.332035
Species : drought : heat shock	1	30.75	30.749	3.3255	0.073281
Species : day : drought	2	45.14	22.569	2.4408	0.095842
Speces : day : heat shock	2	3.74	1.870	0.2022	0.817486
Species : day : drought : heat shock	2	0.33	0.167	0.0181	0.982075
Error (split-plot)	59	545.55	9.247		

Table 4.5. Analysis of variance (ANOVA) table of LDMC on sampling days.

Factor	**Df**	**Sum sq**	**Mean sq**	**F value**	**Pr(> F)**
Drought	1	0.0002297	0.00022972	0.4279	0.5236
Error (main-plot)	14	0.0075165	0.00053689		
Heat shock	1	0.005667	0.0056668	1.8272	0.18162
Block	15	0.046698	0.0031132	1.0038	0.46386
Species	1	0.001281	0.0012807	0.4130	0.52296
Day	2	0.016766	0.0083830	2.7030	0.07530
Drought : heat shock	1	0.003203	0.0032033	1.0329	0.31363
Day : drought	2	0.012903	0.0064513	2.0802	0.13398
Day : heat shock	2	0.037019	0.0185094	5.9682	0.00436*
Species : drought	1	0.000067	0.0000667	0.0215	0.88395
Species : heat shock	1	0.002547	0.0025474	0.8214	0.36846
Species : day	2	0.003652	0.0018262	0.5888	0.55820
Day : drought : heat shock	2	0.006750	0.0033749	1.0882	0.32349
Species : drought : heat shock	1	0.000085	0.0000847	0.0273	0.86931
Species : day : drought	2	0.001544	0.0007722	0.2490	0.78040
Speces : day : heat shock	2	0.008646	0.0043229	1.3939	0.25616
Species : day : drought : heat shock	2	0.006402	0.0032012	1.0322	0.36257
Error (split-plot)	59	0.182980	0.0031014		

Table 4.6. Analysis of variance (ANOVA) table of the log-transformed SLA on sampling days.

Factor	**Df**	**Sum sq**	**Mean sq**	**F value**	**Pr(> F)**
Drought	1	0.004837	0.0048374	0.3514	0.5628
Error (main-plot)	14	0.192709	0.0137649		
Heat shock	1	0.0273	0.0273	0.3786	0.5407
Block	15	1.3237	0.0882	1.2230	0.2812
Species	1	5.8698	5.8698	81.3485	1.061e-12*
Day	2	0.2332	0.1166	1.6157	0.2074
Drought : heat shock	1	0.1062	0.1062	1.4719	0.2299
Day : drought	2	0.5980	0.2990	4.1439	0.0207*
Day : heat shock	2	0.2757	0.1379	1.9107	0.1570
Species : drought	1	0.0020	0.0020	0.0277	0.8683
Species : heat shock	1	0.0048	0.0048	0.0669	0.7969
Species : day	2	0.0013	0.0007	0.0090	0.9910
Day : drought : heat shock	2	0.1124	0.0562	0.7790	0.4635
Species : drought : heat shock	1	0.0078	0.0078	0.1087	0.7428
Species : day : drought	2	0.0344	0.0172	0.2386	0.7885
Speces : day : heat shock	2	0.1742	0.0871	1.2068	0.3064
Species : day : drought : heat shock	2	0.2163	0.1081	1.4987	0.2318
Error (split-plot)	59	4.2572	0.0722		

Table 4.7. Analysis of variance (ANOVA) table of the log-transformed *Ac* on sampling days.

Factor	**Df**	**Sum sq**	**Mean sq**	**F value**	**Pr(> F)**
Drought	1	0.00618	0.0061774	0.2272	0.641
Error (main-plot)	14	0.38069	0.0271921		
Heat shock	1	0.1606	0.16062	0.4258	0.51659
Block	15	2.4567	0.16378	0.4342	0.96200
Species	1	0.5139	0.51394	1.3625	0.24781
Day	2	1.5478	0.77390	2.0516	0.13760
Drought : heat shock	1	1.0394	1.03941	2.7555	0.10223
Day : drought	2	0.5054	0.25271	0.6699	0.51559
Day : heat shock	2	0.2370	0.11852	0.3142	0.73158
Species : drought	1	0.2658	0.26583	0.7047	0.40459
Species : heat shock	1	1.1777	1.17769	3.1221	0.08241
Species : day	2	0.5238	0.26192	0.6944	0.50343
Day : drought : heat shock	2	0.5747	0.28733	0.7617	0.47139
Species : drought : heat shock	1	0.0649	0.06492	0.1721	0.67975
Species : day : drought	2	0.8013	0.40064	1.0621	0.35225
Speces : day : heat shock	2	0.9065	0.45325	1.2016	0.30797
Species : day : drought : heat shock	2	0.3440	0.17198	0.4559	0.63609
Error (split-plot)	59	22.2557	0.37721		

Table 4.8. Analysis of variance (ANOVA) table of *Chl*$_a$ on sampling days.

Factor	**Df**	**Sum sq**	**Mean sq**	**F value**	**Pr(> F)**
Drought	1	0.10204	0.102040	5.8941	0.02927*
Error (main-plot)	14	0.24237	0.017312		
Heat shock	1	0.2891	0.28908	2.0159	0.160918
Block	15	1.9510	0.13007	0.9071	0.560724
Species	1	0.6372	0.63723	4.4439	0.039284*
Day	2	0.9462	0.47310	3.2993	0.038328*
Drought : heat shock	1	0.0729	0.07290	0.5084	0.478639
Day : drought	2	0.3011	0.15057	1.0501	0.356362
Day : heat shock	2	0.0783	0.03914	0.2730	0.762082
Species : drought	1	0.1421	0.14212	0.9911	0.323533
Species : heat shock	1	0.1297	0.12974	0.9048	0.345386
Species : day	2	0.7073	0.35367	2.4664	0.093604
Day : drought : heat shock	2	1.0368	0.51839	3.6151	0.033036*
Species : drought : heat shock	1	0.6732	0.67322	4.6949	0.034303*
Species : day : drought	2	2.3969	1.19847	8.3578	0.000637*
Speces : day : heat shock	2	0.5126	0.25629	1.7873	0.176362
Species : day : drought : heat shock	2	0.1205	0.06027	0.4203	0.658799
Error (split-plot)	59	8.4603	0.14340		

Table 4.9. Analysis of variance (ANOVA) table of the log-transformed Chl_b on sampling days.

Factor	**Df**	**Sum sq**	**Mean sq**	**F value**	**Pr(> F)**
Drought	1	0.000971	0.0009706	0.0751	0.788
Error (main-plot)	14	0.180877	0.0129198		
Heat shock	1	0.3697	0.36967	1.2272	0.27245
Block	15	1.2975	0.08650	0.2871	0.99499
Species	1	0.1694	0.16939	0.5623	0.45630
Day	2	0.4506	0.22528	0.7479	0.47781
Drought : heat shock	1	0.4629	0.46290	1.5367	0.22002
Day : drought	2	0.1703	0.08514	0.2826	0.75481
Day : heat shock	2	0.0394	0.01971	0.0654	0.93672
Species : drought	1	0.0223	0.02225	0.0739	0.78672
Species : heat shock	1	1.7441	1.74410	5.7899	0.01927*
Species : day	2	1.1308	0.56539	1.8769	0.16208
Day : drought : heat shock	2	0.4671	0.23353	0.7753	0.46521
Species : drought : heat shock	1	0.1913	0.19126	0.6349	0.42875
Species : day : drought	2	1.1827	0.59133	1.9631	0.14949
Speces : day : heat shock	2	0.1603	0.08013	0.2660	0.76735
Species : day : drought : heat shock	2	0.2171	0.10854	0.3603	0.69898
Error (split-plot)	59	17.7725	0.30123		

Table 4.10. Analysis of variance (ANOVA) table of the log-transformed carotenoids content on sampling days.

Factor	**Df**	**Sum sq**	**Mean sq**	**F value**	**Pr(> F)**
Drought	1	3.5244	3.5244	5.1451	0.03966*
Error (main-plot)	14	9.5900	0.6850		
Heat shock	1	3.960	3.960	0.7669	0.384730
Block	15	66.756	4.450	0.8618	0.607878
Species	1	0.450	0.450	0.0872	0.768777
Day	2	59.359	29.680	5.7475	0.005242*
Drought : heat shock	1	11.480	11.480	2.2231	0.141291
Day : drought	2	4.096	2.048	0.3966	0.674397
Day : heat shock	2	3.960	3.960	0.7669	0.384730
Species : drought	1	8.284	8.284	1.6043	0.210273
Species : heat shock	1	6.826	6.826	1.3218	0.254903
Species : day	2	7.207	3.603	0.6978	0.501732
Day : drought : heat shock	2	14.632	7.316	1.4168	0.250616
Species : drought : heat shock	1	7.855	7.855	1.5211	0.222340
Species : day : drought	2	64.485	32.243	6.2438	0.003470*
Speces : day : heat shock	2	27.682	13.841	2.6803	0.076886
Species : day : drought : heat shock	2	2.850	1.425	0.2759	0.759847
Error (split-plot)	59	304.670	5.164		

Table 4.11. Analysis of variance (ANOVA) table of TSS on sampling days.

Factor	**Df**	**Sum sq**	**Mean sq**	**F value**	**Pr(> F)**
Drought	1	10.579	10.5792	5.1289	0.03993*
Error (main-plot)	14	28.877	2.0627		
Heat shock	1	63.53	63.53	1.6632	0.20221
Block	15	392.78	26.19	0.6856	0.78804
Species	1	1034.19	1034.19	27.0757	2.586e-06*
Day	2	209.28	104.64	2.7395	0.07283
Drought : heat shock	1	2.79	2.79	0.0731	0.78777
Day : drought	2	160.79	80.39	2.1048	0.13093
Day : heat shock	2	61.55	30.77	0.8056	0.45165
Species : drought	1	14.54	14.54	0.3807	0.53961
Species : heat shock	1	0.00	0.00	0.0001	0.99303
Species : day	2	10.45	5.22	0.1367	0.87248
Day : drought : heat shock	2	34.94	17.47	0.4574	0.63515
Species : drought : heat shock	1	0.99	0.99	0.0259	0.87264
Species : day : drought	2	107.94	53.97	1.4130	0.25152
Speces : day : heat shock	2	13.18	6.59	0.1726	0.84191
Species : day : drought : heat shock	2	35.89	17.95	0.4698	0.62742
Error (split-plot)	59	2253.58	38.20		

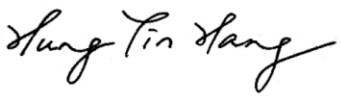
Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	The first chromosome-level genome assembly of Siamese rosewood *Dalbergia cochinchinensis* and comparative genomics of resistance genes
Publication Status	□Published □ Accepted for Publication □Submitted for Publication ☑Unpublished and unsubmitted work written in a manuscript style
Publication Details	Hung, T. H.*, Boshier, D. H. & MacKay, J. J.* (unpublished). The first chromosome-level genome assembly of Siamese rosewood *Dalbergia cochinchinensis* and comparative genomics of resistance genes.

Student Confirmation

Student Name:	Tin Hang Hung		
Contribution to the Paper	designed the study, coordinated the sample collection, conducted the sequencing experiment, conceived and conducted the bioinformatic analyses, drafted the manuscript		
Signature 		Date	8 MAR 2022

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: John MacKay, Wood Professor of Forest Science		
Supervisor comments		
Signature 	Date	8 MAR 2022

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 5. Landscape genomics of two threatened rosewood species *Dalbergia cochinchinensis* and *D. oliveri* in the Greater Mekong Subregion

Tin Hang Hung[1,*], Thea So[2], Bansa Thammavong[3], Hoa Thi Tran[4], Voradol Chamchumroon[5], Ida Theilade[6], Chhang Phourin[2], Somsanith Bouamanivong[7], Ida Hartvig[8], Hannes Gaisberger[9,10], Riina Jalonen[11], David H. Boshier[1], John J. MacKay[1,*]

1. Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, United Kingdom
2. Institute of Forest and Wildlife Research and Development, Phnom Penh, Cambodia
3. National Agriculture and Forestry Research Institute, Forestry Research Center, Vientiane, Laos
4. Forest Genetics and Conservation, Center for Biodiversity and Biosafety, Institute of Agricultural Genetics, Vietnam Academy of Agricultural Sciences, Hanoi, Vietnam
5. The Forest Herbarium, Department of National Park, Wildlife and Plant Conservation, Ministry of Natural Resources and Environment, Bangkok, Thailand
6. Faculty of Science, University of Copenhagen, Denmark
7. National Herbarium of Laos, Biotechnology and Ecology Institute, Ministry of Science and Technology, Vientiane, Laos
8. Forest Genetics and Diversity, Department of Geosciences and Natural Resource Management, University of Copenhagen, Denmark
9. The Alliance of Bioversity International and CIAT, Rome, Italy
10. Paris Lodron University, Salzburg, Austria
11. The Alliance of Bioversity International and CIAT, Serdang, Malaysia

5.1 Abstract

Rosewoods were the world's most trafficked wild product between 2005 and 2014. In particular, *Dalbergia cochinchinensis* and *D. oliveri* in the Greater Mekong Subregion are threatened by habitat degradation, overexploitation, and climate change. They become the priority species in both national and regional programmes for conservation and restoration. However, genetic risks can compromise the effectiveness of restoration programmes, for example, genetic bottleneck during seed sourcing and maladaptation in plantation. Although there has been a growing knowledge on their genetic diversity and population structure, their adaptation under changing environment remains unstudied. Here we present the first genomic resources and ecological genomic study on ~770 samples across their range and ~200,000 genetic markers. We used a gradient forest model on 34 environmental predictors to elucidate their gene-environment association and local adaptation. We predicted the future adaptation and vulnerability across their species ranges under a range of climate scenarios (SSP126, 245, 370, and 585) from 2041 to 2100. We developed a web application based on R Shiny, *seedeR*, which allows end-users to predict genomic similarity between current germplasm sources and future restoration site. The outcomes of these studies will ultimately enhance the germplasm management of *Dalbergia* for their sustainable production and long-term genetic conservation.

5.2 Introduction

Dalbergia cochinchinensis Pierre and *D. oliveri* Gamble ex Prain are two of the around 250 species in the pantropical genus *Dalbergia* L. f. (Fabaceae: Faboideae), many of which produce extremely valuable rosewood timber[1]. Rosewoods were the world's most trafficked wild product between 2005 and 2014, with *D. cochinchinensis* being the most sought-after species[2]. The growing demand for and the diminishing populations in both species is driving their exploitation to commercial extinction, meaning a constant supply can no longer be maintained in the market, but at the same time drastically escalating their economic value due to their rarity[3]. The two species were classified as vulnerable and endangered in the IUCN Red List in 1998, with their international trade regulated under CITES Appendix II since 2017.

These two species are endemic to the Greater Mekong Subregion (GMS) in Southeast Asia, consisting of Cambodia, southwest region of China, Laos, Myanmar, Thailand, and Vietnam. GMS is of high ecological and conservation concern as 84% of its area overlaps with the Indo-Burmese mega biodiversity hotspot[4]. The complex biogeographical and geological history may have contributed to the high species richness and endemism levels in GMS[5]. In particular, the ancient changes in the distribution of terrestrial and water bodies in the region have been associated with changes in vegetation types and cover[6]. The topography is highly varied, ranging from lowlands to mountains, with many freshwater bodies including the Mekong River and the Tonle Sap Lake[7]. As a result, the GMS has a very heterogeneous landscape and supports a wide range of habitats. *Dalbergia cochinchinensis* grows in open deciduous and semi-deciduous forests at elevation up to 400 metres, while *D. oliveri* grows in a wider range of forest types from deciduous to evergreen and can grow at elevations of up to 1,200 metres[8]. However, there exist only a few studies on these two valuable species and biological knowledge surrounding them and their populations are still relatively unknown.

Three main studies have examined the spatial genetic variation of the species[8–10]. Hartvig et al. (2018) reported high levels of genetic differentiation among populations in both *D. cochinchinensis* and *D. oliveri*. They found a link between population genetic structure with landscape features in the region, such that the Tonlé Sap, Mekong river and Cardamon Mountains effectively block gene flow among populations and could explain the observed genetic differentiation. A similar observation has been made on other animal and plant species that are also endemic to the region[11,12]. Although deforestation and forest degradation are severe in the region, the two *Dalbergia* species seem to maintain higher levels of genetic diversity. The two species display substantial among-population variations, compared to other tree species in the area, including *Afzelia xylocarpa*[13], *Prunus cerasoides*[14], *Shorea obtusa*[15], and *Xylia xylocarpa*[16], and no inbreeding was evident in any of the *Dalbergia* populations studied so far. It may be too early to conclude however that regional habitat fragmentation did not influence the genetic diversity because trees species such as *Dalbergia* have long generation time and any effect may appear in later generations[17].

The studies reported significant differences in the genetic variation of the species, which could be explained by the different ecological niches they occupy. *Dalbergia oliveri* had a higher level of gene flow and more evenly distributed genetic diversity among populations than *D. cochinchinensis*. A potential explanation is that *D. oliveri* has both a broader geographical distribution and wider ecological niche, while *D. cochinchinensis* is more pioneering and thus genetic bottlenecks have potentially affected individual populations. Previous studies used neutral genetic markers (e.g. AFLP and SSR) to understand the microevolutionary processes of gene flow and genetic drift in these two species as an evidence base to inform conservation and forest management. However, the role of environment in driving genetic divergence between populations remains largely unexplored,

and this lack of knowledge on adaptability may hinder conservation management in the long term.

Development of new and economic technologies open doors to using genomic approaches and incorporating them into biodiversity conservation and landscape restoration[18]. In particular, recent advances in high-throughput genotyping and sequencing offer the means to discover and genotype a much larger set of single-nucleotide polymorphisms (SNP)[19,20], that represent the genetic variation more comprehensively than a small number of near-neutral loci, thus leading to less biased estimates of population genetic parameters[21]. The universality of sequencing methods enables genomic studies in non-model and novel organisms[22]. At the same time, advances in remote sensing and geodata tools, such as satellite imaging and GIS, allow scientists to obtain and process environmental and landscape data at higher spatial and temporal resolutions reducing the effort of *in situ* measurement[23].

The gene-environment interaction (G x E) of trees has been demonstrated in many tree species, revealing adaptive variations in response to environmental differences[24–26]. However, many of these population genetic studies use neutral markers such as microsatellites and amplified fragment length polymorphism (AFLP)[27]. As a result, most of the genomic basis of the genotype-by-environment interactions that impact and connect the population fitness and adaptation remain unknown[28]. The use of genomic tools can significantly broaden the understanding of the relationships among environment, selection, gene expression and fitness.

In short, these technical and conceptual advances lead to an accelerating trend to incorporate genomic models into landscape research for conservation and restoration[29], and thus the emerging field of “landscape genomics”.

Landscape genomics is an analytical framework that attempts to unravel geographic patterns of genome-wide genetic variation[30]. It detects the association between loci and

environmental variables, and further dissects the relative contribution of these variations in explaining the difference between spatial structure of associated and neutral loci[31]. Most landscape genomics studies incorporate three analytical steps[32]. First, they examine the genetic variation among populations and its patterns and microevolutionary processes. Second, they quantify environmental heterogeneity from landscape data. Third, they incorporate statistical analyses to model the relationship between genetic and environmental data. An outlier test is commonly used to distinguish two components of genetic variation, neutral and adaptive, and their corresponding genetic structure and processes[33]. Neutral variation in that has been successfully and widely analysed using microsatellites or short DNA sequences reflects the environmental effects on functional connectivity and is mainly driven by genetic drift and gene flow. Adaptive variation is driven by natural selection, resulting in possible local adaptations, and is highly related to the local bioclimatic factors. Landscape genomics uses a suite of environmental association analyses (EAA) to correlate local adaptation to environmental heterogeneity[34], including logistic regressions, spatial analysis method (Joost et al., 2007), canonical correlation analysis[36], mixed effect models[37]. Some of the models were developed based on community-level frameworks, such as machine-learning based gradient forest (GF)[38] and generalized dissimilarity modelling (GDM)[39], which make them ideal to measure turnover in allele frequencies along environmental gradients. Different statistical methods have their own strengths and weaknesses, for example, only a few of them can be used to account for spatial autocorrelation (e.g. spatial analysis method, partial Mantel test, GDM)[40].

We can see that landscape genomics is still an emerging field and the statistical methods are developing rapidly to address analysis challenges such as the detection of false positives[41]. Although considerable progress has been made in the field throughout the past decade, the lack of a universal analytical method and a central hypothesis is still a core gap in

this young field. Research has shown the importance of the sampling design and the usefulness of complementing the methods with other lines of evidence such as common garden experiments for validation of adaptive genetics[30].

Landscape genomics receives considerable attention in the study of forest due to their important ecological and economical values. Understanding how forests are adapted to climate and landscape structure is one of the keys to sustainability of forests. Forest trees are also considered suitable model organisms for landscape genomics because of their high levels of genomic and phenotypic variations. Since the first mention of 'landscape genomics'[35], a wide range of applications on forest trees have emerged, but mainly on temperate and boreal species[42–45]. The scale of genomic coverage in landscape genomics studies also increased from a few hundred[46] to tens of thousands of SNPs[47]. It was only very recently that the first genomic assessment of local adaptation was done in the tropical plant on *Ipomoea cavalvantei*[29]. To the best of our knowledge, there are no landscape genomics studies published on any tropical forest trees.

This paper presents the first study to survey the genomic variations and their association with the environment across the present range of *D. cochinchinensis* and *D. oliveri*. First, we produce the first genome assembly of *D. oliveri* using long-read Nanopore sequencing and multi-contact chromosome conformation capture. Second, we genotype the range-wide individuals and differentiate neutral genetic structure and signatures of adaptation. Third, we project the current genotypes onto future climate scenarios and predict the potential maladaptation of populations. Finally, we deploy an interactive application to predict optimal seed sources for restoration under future climate.

.

5.3 Methods

5.3.1 Dalbergia oliveri de novo genome assembly

Dried seeds of *Dalbergia oliveri* were collected at the Institute of Forest and Wildlife Research and Development, Phnom Penh, Cambodia in 2018. We germinated the seeds in a greenhouse at 30°C and photoperiod 16L/8D. Leaf tissues were harvested from a selected 1-year-old individual and ground in liquid nitrogen with a mortar and pestle.

High-molecular-weight genomic DNA was extracted from the reference individual with Carlson lysis buffer (100 mM Tris-HCl, pH 9.5, 2% CTAB, 1.4 M NaCl, 1% PEG 8000, 20 mM EDTA) followed by purification using the QIAGEN Genomic-tip 500/G. The quantity and quality of genomic DNA were determined with NanoDrop 2000 (Thermo, Wilmington, United States) and Qubit 4 (Thermo Fisher Scientific, United States). DNA integrity was preliminary assessed with a 0.4% agarose gel against a NEB Quick-Load® 1 kb Extend DNA Ladder. DNA sample passed the quality check only when a single band could be mapped near a lambda DNA band (~ 48.5 kb).

5.3.2 Nanopore library preparation and sequencing

9 µg of extracted DNA was size-selected using the Circulomics Short Read Eliminator XL Kit (Maryland, United States) to deplete fragments < 40 Kbp. 3 µg of size-selected DNA was used in each library preparation with the Oxford Nanopore Technologies Ligation Sequencing Kit (SQK-LSK110), totalling to 3 libraries for the same sample. The libraries were sequenced on two R10.3 (FLO-109D) flow cells on a GridION sequencer for ~ 72 hours. Real-time basecalling was performed in MinKNOW release 19.10.1. Raw reads with Phred score lower than 8 were filtered.

5.3.3 Pore-C proximity ligation library preparation and sequencing

Pore-C library was prepared with the protocol and reagents described by Belaghzal et al.[48] with minor modifications. We harvested 2 g of fresh leaf from the same reference individual as for the Nanopore library and immediately cross-linked the finely chopped tissues in 1% formaldehyde for 20 minutes. The cross-linking was then quenched with 125 mM glycine for 20 minutes and the samples were then ground in liquid nitrogen with a mortar and a pestle. Cell nuclei were isolated with a buffer containing 10 mM Trizma, 80 mM KCl, 10 mM EDTA, 1 mM spermidine trihydrochloride, 1 mM spermine tetrahydrochloride, 500 mM sucrose, 1% (w/v) PVP-40, 0.5% (v/v) Triton X-100, and 0.25% (v/v) β-mercaptoethanol, and then passed through a 40 µm cell strainer. The suspension was centrifuged at 3,000 g, according to the estimated genome size of ~ 700 Mbp. Chromatin was denatured with the restriction enzyme NlaIII at a final concentration of 1 U/µL (New England Biolabs, United Kingdom) at 37°C for 18 hours. The enzyme was heat-denatured at 65°C for 20 minutes at 300 rpm rotation. Proximity ligation, protein degradation, decrosslinking, and DNA extraction were performed according to the original Belaghzal protocol. The Pore-C library was prepared with the Oxford Nanopore Technologies Ligation Sequencing Kit (SQK-LSK110), then sequenced on two R10.3 (FLO-109D) Nanopore flow cells on a GridION sequencer for ~ 72 hours. The flow cell was washed once every 24 hour with the Flow Cell Wash Kit (EXP-WSH003).

5.3.4 Assembly pipeline

Nanopore raw reads shorter than 500 bp were filtered. Due to the heterozygous nature of the wild individual, we assembled the assembly with Canu 2.1.1 using the options "`corOutCoverage=200 correctedErrorRate=0.16 batOptions=-dg 3 -db 3 -dr 1 -ca 500 -cp 50`". We then used purge_haplotigs v1.1.1 with the cutoff

`-a 98` to collapse the assembly by separating the primary assembly and haplotigs. Pore-C reads were mapped to the draft genome assembly and used to generate contact map with the Pore-C-Snakemake (https://github.com/nanoporetech/Pore-C-Snakemake) to produce a merged_nodups (.mnd) file, which contains a duplicate-free list of paired alignments from the Pore-C reads to the draft assembly. The draft assembly and the merged_nodups file were used for scaffolding in 3D-DNA (version 180419) to produce the final genome Daoli 0.2.

To validate the scaffold arrangement, Daoli 0.2 was aligned to that of *D. cochinchinensis* (Dacoc 1.2) using minimap2 and D-GENIES[49] (http://dgenies.toulouse.inra.fr/) to produce a dot plot for visualising similarity, repetitions, breaks, and inversions, with a minimum identity of 0.25.

Filtered mRNA-sequencing data (54.4 Gb) for *D. oliveri* from a previous project[50] (NCBI Bioproject: PRJNA593817) were aligned against the genome assembly using STAR v2.7.6 and assembled using the genome-guided mode of Trinity v2.13.2. Both *de novo* repeat library and gene models were then constructed following the same pipeline in Hung et al (unpublished). The quality of the gene models was assessed with two metrics: the annotation edit distance (AED) in MAKER 3.01.03[51] and the BUSCO score (v5.1.2)[52].

5.3.5 Population sampling

We obtained a collection of 435 and 331 foliage samples of *Dalbergia cochinchinensis* and *D. oliveri* from 35 and 28 localities across their native range (Gaisberger et al., under review). These samples were a combination of those collected in a previous study[8] and newly between 2019 and 2020. Genomic DNA was purified using a two-round modified CTAB protocol (2% CTAB, 1.4 M NaCl, 1% PVP-40, 100 mM Tris-Cl pH 8.0, 20 mM EDTA pH 8.0, 1% 2-mercaptoethanol) with sorbitol pre-wash (0.35 M Sorbitol, 1% PVP-40, 100 mM Tris-Cl pH 8.0, and 5 mM EDTA pH 8.0) as the samples were rich in

polyphenols and polysaccharides[53]. Genomic DNA was treated with 5 μL RNase (10 mg/mL). Quality and quantity of the genomic DNA were assessed using NanoDrop One (Thermo, Wilmington, United States) and Qubit dsDNA BR Assay kit on Qubit 4 (Thermo, Wilmington, United States) respectively.

5.3.6 Genotyping-by-sequencing (GbS)

DNA samples were normalised to 200 ng suspended in 10 μL water and sent to the Genomic Analysis Platform, Institute of Integrative Biology and Systems, Université Laval, Quebec, Canada for GbS library preparation. DNA was digested with a combination of restriction enzymes PstI/NsiI/MspI, ligated with barcoded adapter, and pooled to equimolarity. The pooled library was amplified by PCR and sequenced on a Illumina NovaSeq6000 S4 with paired-end reads of 150 bp at the Génome Québec, Canada.

5.3.7 Variant calling

Variant calling was done with the Fast-GBS v2.0 pipeline[54]: Illumina raw reads were demultiplexed with Sabre 1.0[55] and trimmed with Cutadapt 1.18[56] to remove the adaptors. Trimmed reads shorter than 50 bp were discarded. Reads were aligned against the Dacoc 1.0 genome (Hung et al., unpublished) and the Daoli 0.1 genome using BWA-MEM 0.7.17[57]. The SAM alignment files were converted to BAM format and indexed using SAMtools 1.9[58]. Variant calling was performed in Platypus[59] and variants were filtered with proportion of missing data of 0.2 and minimum allele frequency (MAF) of 0.01 using VCFtools 0.1.16[60]. Missing genotype was imputed using Beagle 5.2. Finally, linkage equilibrium among SNPs was detected using BCFtools 1.9[58], and one SNP was removed from all SNP pairs with $r^2 > 0.5$ in a genomic window of 5 Kbp.

5.3.8 Environmental heterogeneity characterisation

Environmental data were obtained at different sources (34 variables in total, see Table 5.1) and represented different measurers of temperature, precipitation, their seasonality, soil, elevation, and vegetation. We calculated a correlation matrix across the sampling localities and highly inter-correlated variables (|pairwise correlation coefficient| > 0.7) were detected. For each inter-correlated variable pair, the one variable with the largest mean absolute correlation across all variables was removed.

5.3.9 Population genetic structure and identification of putatively adaptive loci

Population genetic structure was assessed with sparse non-negative matrix factorisation (sNMF) to estimate the number of discrete genetic clusters (K)[61]. The sNMF was run for 10 repetitions for each value of K from 1 to 15 with a maximum iteration of 200. The optimal K was selected based on the lowest cross-entropy value from the sNMF run, or where the value began to plateau. Admixture plots were drawn for K = {2, 4, 8, optimal K}. Population structure-based outlier analysis was also conducted with sNMF, in which outlier SNPs that are significantly differentiated among populations, based on estimated F_{ST} values from the ancestry coefficients obtained from sNMF[62], were obtained and mapped on the 10 putative chromosomes for *D. cochinchinensis* or the 16 longest scaffolds for *D. oliveri* in a Manhattan plot.

We used latent factor mixed modelling (LFMM) to test for significant associations between environmental variables and SNP allele frequencies. The optimal K obtained from the sNMF was used in LFMM to correct for the neutral genetic structure. LFMM was run for 3 repetitions with a maximum iteration of 1,000 and 500 burn-ins. Z-scores were obtained for all repetitions for each environmental variable, and then the median was taken for each SNP. Next, the genomic inflation factor λ, defined as the observed median of Z-scores divided by

the expected median of the chi-squared distribution for each environmental association[63], was calculated to calibrate for *P*-values:

$$\lambda = \frac{median(Z^2)}{\chi_1^2(0.5)}$$, such that $$P_{adjusted} = \chi_1^2(\frac{Z^2}{\lambda})$$.

The calibration was then inspected on a histogram of *P*-values for each environmental association. Finally, multiple testing was corrected with the Benjamini and Hochberg method to obtain *Q*-values.

The sNMF and LFMM calculations were performed in R 4.1.0 using the packages LEA 3.4.0[64].

5.3.10 Gradient forest modelling

For all predictions in gradient forest models, resampling was necessary because not all environmental raster layers had the same resolution and extent. They were all cropped to the latest-updated modelled species distribution (Gaisberger, 2021) and reprojected to the WorldClim bioclimatic rasters, as they have the highest resolution, using bilinear interpolation or nearest neighbour method for continuous and categorical variables respectively.

To correct for the genetic structure, spatial variables were generated using the principal coordinates of neighbour matrices (PCNM) approach[65]. Only half of the positive PCNM values were kept. Gradient forest model was used to predict and rank the importance of environmental variables in genomic variation, as its machine learning algorithm worked best with minimal prior and confounding variables. All SNPs and putatively adaptive SNPs were used as the response variables and all the filtered environment variables and PCNM

variables were used as the predictor variables in the gradient forest model for 500 regression trees. The maximum number of splits to evaluate was determined as follows:

$$\text{Maximum number of splits} = \log_2 \frac{(0.368 \times \text{number of predictor variables})}{2}$$

The turnovers of allelic frequencies were then projected spatially across the latest-updated species distribution (Gaisberger, 2021) using the fitted gradient forest model and the environmental values across the range. Principal component analysis (PCA) was used to summarise the genomic variation across the distribution and the first three principal components (PC1, PC2, and PC3) were used for visualisation of genomic variation across the range.

The PCAs of turnovers of allelic frequencies between adaptive SNPs and all SNPs were compared using the Procrustes rotation, and its residuals were used to map where adaptive genomic variation deviates from neutral variation.

5.3.11 Prediction of genomic vulnerability

Future climate projection was obtained from five general circulation models (GCM) (MIROC6, BCC-CSM2-MR, IPSL-CM6A-LR, CNRM-ESM2-1, MRI-ESM2-0) participating in the World Climate Research Programme Coupled Model Intercomparison Project 6 (WCRP CMIP6) for four shared socio-economic pathways (SSPs) (126, 245, 370, and 585) over four 20-year periods (2021–2040, 2041–2060, 2061–2080, 2081–2100). The gradient forest model was used to predict the patterns of genetic variation and local adaption under future environmental scenarios. The allelic frequency turnover function was fitted on the future landscape and the genomic offset, defined as the required genomic change in a set of putatively adaptive loci to adapt to a future environment[66], was calculated in a grid-by-grid

basis using the following equation for Euclidean distance, where *p* is the number of environmental (predictor) variables:

$$\text{Genetic offset} = \sqrt{\sum_{n=1}^{p}(\text{Future allelic turnover} - \text{Current allelic turnover})^2}$$

The genetic offset was then scaled across all SSPs and time periods.

5.3.12 Prediction of genomic similarity between current germplasm source and future restoration site

It is of practical interest to forestry end-users to predict if a current germplasm source is a good match for future restoration sites, or where to source suitable germplasm for a future restoration site. We developed an interactive web application based on R Shiny and hosted the application on the shinyapps.io server. *seedeR* v 1.0 is open source and freely available from https://trainingidn.shinyapps.io/seeder/. The analysis workflow consists of the selection of species of interest, time period and scenario for future climate, and input of geographical coordinates for the restoration site (Figure 5.1).

The application mapped the predicted turnovers of allelic frequencies on a future hypothetical restoration site onto the current landscape in a grid-by-grid basis, then the genetic offset was calculated as described above. After scaling, the values were reversed on the 0-1 scale to represent the genomic similarity between current germplasm source and future restoration site.

5.4 Results

5.4.1 Genome assembly of Dalbergia oliveri

The genome was sequenced from a single *Dalbergia oliveri* seedling grown from a seed collected at the Institute of Forest and Wildlife Research and Development, Cambodia. Oxford Nanopore sequencing produced 15.13 Gbp (~ 22X) and 13.46 Gbp (~ 20X) of raw reads and Pore-C reads respectively, with a N50 of 44.83 Kbp and 5.79 Kbp. Canu produced a diploid-aware assembly of 814.69 Mbp with 3,249 contigs, achieving a N50 of 474.02 Kbp and the longest contig being 6.01 Mbp. Purging the haplotig and scaffolding with chromosome conformation Pore-C data reduced the assembly size by 15.40% and scaffold number by 8.37%, and increased the scaffold N50 by approximately 8 folds. The final assembly (Daoli_0.2) has a total length of 689.25 Mbp in 2,977 scaffolds with a N50 of 38.43 Mbp (Table 5.2 and Figure 5.2). The assembly has a BUSCO v5.1.2 completeness of 93.6% using the eudicots_odb10 reference dataset, with 86.7% being complete single-copy. Syntenic analysis between Daoli_0.2 against the 10 putative chromosomes of *D. cochinchinensis* in the assembly Dacoc_1.2 showed that the 16 largest scaffolds in Daoli_0.2 had 1-to-1 or 2-to-1 correspondences to Dacoc_1.2, implying that Daoli_0.2 was at chromosome arm length (Figure 5.3).

The *de novo* repeat library of Daoli_0.2 was constructed with RepeatModeler and LTRharvest. The final library contained 453 Mbp of repeat elements (65.71% of the genome), majority of which contained LTR elements (48.55%) such as Ty1/Copia (15.75%) and Gypsy/DIRS1 (31.96%) (Table 5.3 and Figure 5.4).

The MAKER pipeline predicted 33,558 gene models with a mean length of 3942.71 bp and 54,062 transcripts. 95.5% of the gene models had a AED score less than 0.5, which was considered a good gene model (Figure 5.5). 88.3% of the gene models were complete BUSCOs (Table 5.4).

5.4.2 Variant calling

Illumina raw reads were obtained for 435 and 331 samples of *D. cochinchinensis* and *D. oliveri* respectively. After demultiplexing, filtering and trimming low-quality reads, and variant calling, 1,832,629 and 3,377,855 SNPs were initially obtained. The Platypus step kept 246,225 and 250,954 SNPs. After pruning for linkage disequilibrium, the final variant pools contained 180,944 and 193,724 SNPs respectively for the two species (Table 5.5).

5.4.3 Population differentiation tests

sNMF was run for each value of K from 1 to 15 for all filtered SNPs of *D. cochinchinensis* and *D. oliveri*. The optimal K for *D. cochinchinensis* and *D. oliveri* were determined to be 13 and 14 respectively, based on the cross-entropy criterion (Figure 5.6). sNMF showed a highly-resolved hierarchical population genetic structure for both species (Figure 5.7). Population differentiation tests with sNMF reported 50,951 (28.1%) and 54,154 (28.0%) outlier SNPs for the two species respectively ($P < 0.05$) (Figure 5.8).

A subset of 13 and 12 environmental variables for *D. cochinchinensis* and *D. oliveri* were retained, which were related to temperature, precipitation, their seasonality, soil, elevation, and vegetation, after filtering the highly inter-correlated variables (|pairwise correlation coefficient| > 0.7). These environmental variables were then used for the LFMM analysis.

Genomic inflation factor λ for *D. cochinchinensis* ranged from 0.071 (evapotranspiration) to 0.25 (precipitation of driest quarter), with a mean of 0.13 and standard deviation of 0.049; that for *D. oliveri* ranged from 0.038 (evapotranspiration) to 0.081 (mean diurnal range), with a mean of 0.056 and a standard deviation of 0.016 (Table 5.6).

After correcting for the neutral structure (optimal K) and the genomic inflation (λ), LFMM predicted 20,373 (11.3%) and 6,953 (3.59%) of the SNPs to be adaptive (| *Z*-value | > 2 & *Q*-values < 0.01) for at least one of the environmental variables (Figure 5.9). The

frequency distribution of adjusted *P*-values of the SNPs is rather flat among the non-outliers ($P > 0.05$), suggesting that false discovery rate is controlled (Figure 5.10).

5.4.4 Gradient forest

10 and 8 PCNM values were kept for *D. cochinchinensis* and *D. oliveri* respectively as uncorrelated spatial variables to be incorporated into the GF model, together with the filtered environmental variables. GF model predicted that PCNMs were predominantly the most important variables that explained both neutral and adaptive genomic variation, but among non-spatial variables, wc2.1_30s_bio_3 (isothermality) was the main driver of both neutral and adaptive genomic variation in *D. cochinchinensis*, while wc2.1_30s_bio_16 (precipitation of wettest quarter) was that in *D. oliveri* (Figure 5.11 and Figure 5.12). Soil factors were among the lowest ranked variables for both species. The fitted GF model for adaptive genomic variation was mapped across the native range of both species (Figure 5.13). *D. cochinchinensis* and *D. oliveri* had a different geographic pattern of adaptive variation: *D. cochinchinensis* had a strong differentiation between North and South at around 16°N, that was mainly driven by wc2.1_30s_bio_3 (isothermality) as seen in the PCA loadings (Figure 5.14). On the other hand, *D. oliveri* had the major differentiation between coastal and inland areas, that was driven by both wc2.1_30s_bio_16 (precipitation of wettest quarter) and wc2.1_30s_bio_2 (mean diurnal range).

The allelic frequency turnover functions of the neutral and adaptive genomic variation were compared for each environmental predictor variable (Figure 5.15). Adaptive genomic variation, which was captured by only adaptive SNPs, were more strongly associated with environmental gradients than neutral variation, which was represented by all SNPs. There was only one exception, where s_AWCh1_sl5 (available soil water capacity at depth of 60 cm) had similar importance in explaining neutral and adaptive variation, regardless of the environmental gradient.

Procrustes residuals were used as a proxy of local adaptation to identify areas where adaptive genomic variation deviated from neutral variation (Figure 5.16). Coastal areas in east Vietnam and Cardamom Mountains in southwest Cambodia were common hotspots of local adaptation for both *D. cochinchinensis* and *D. oliveri*.

5.4.5 Prediction of genomic vulnerability

Genetic offset in form of Euclidean distance represented the mismatch of current and future gene-environment association, which was averaged over the five GCMs (MIROC6, BCC-CSM2-MR, IPSL-CM6A-LR, CNRM-ESM2-1, MRI-ESM2-0) under WCRP CMIP6. For both species, genetic offset generally increased over shared socioeconomic pathway, which implies increased carbon emission, and time. However, the geographic patterns of genetic offset differed between the two species: *D. cochinchinensis* had an increasing offset universally across the range, while the southeast part of the range of *D. oliveri* had a distinctly high offset (Figure 5.17).

5.4.6 Genomic similarity between current germplasm source and future restoration site

seedeR is open source and freely available from https://trainingidn.shinyapps.io/seeder/, where users could select the species (only *D. cochinchinensis* and *D. oliveri* for now), SSP, time period, and geographical coordinates of future restoration site. *seedeR* would predict the genomic similarity between current germplasm source and future restoration site from allelic frequency turnover functions and genetic offset and project onto the species range. We demonstrated the utility of *seedeR* by selecting a hypothetical future restoration site (106° N, 14° E) in northeast Cambodia for both *D. cochinchinensis* and *D. oliveri*, under the future climate scenario of SSP370 between 2081 and 2100. In both predictions, the genomic similarity was the highest at proximity and decreased when further away. Commonly, coastal regions in northeast Vietnam, which was

predicted to have the strongest local adaptation in both species, showed a low genomic similarity. An example of the maps as *seedeR* output was shown in Figure 5.18. The output of *seedeR* may be viewed directly online or saved as a PNG file in local instances. The output included a title showing both selected SSP and time period. The selected future restoration site was marked as a white asterisk on the map. The genomic similarity was colour coded between white (the most dissimilar) and dark green (the most similar).

5.5 Discussion

Global biodiversity is threatened by human footprints in modern times, including anthropogenic climate change[67], habitat degradation[68], and overexploitation[69].

Dalbergia cochinchinensis and *D. oliveri* are highly threatened and their distributions have been declining over the past decades due to escalating demand of their rosewood timber. They become the priority species for conservation and restoration in both national and regional programmes to establish management strategies, germplasm guidelines, networking, partnership, and capacity building[70–72]. Our study is the first to assess the association between environmental factors and genomic variation in the genus *Dalbergia*, and predict its change in future climate scenarios.

5.5.1 Genome assembly of D. oliveri

The high-quality, chromosome-arm-scale assembly of *D. oliveri* adds an important resource for genomic analysis of the valuable genus, as reflected by the contiguity, gene completeness, synteny with *D. cochinchinensis,*. For the genus currently, there are only two available genomes: *D. hupeana*[73] and *D. odorifera*[74], and seven transcriptomes[50,75]. Despite the recent advances in sequencing technologies, genomic resources remain scarce for such a big genus consisting of over 250 species[1]. New reference genomes will enable genome-wide analysis, including characterisation of functional genes, higher-resolution species delineation, association mapping and adaptation, genetic rescue, and genome editing[76], which address important conservation questions such as genetic monitoring of introduced and relocated populations, predicting population viability, disease resistance, synthetic alternatives, and de-extinction[77,78]. Specifically, previous studies in *Dalbergia* mainly focused on limited number of loci and microsatellites[8,79–84], which potentially hinders the broader understanding of their evolutionary and ecological processes and the application of genetic tools in conservation management.

5.5.2 Higher-resolution population differentiation

Our study assessed the population genetic structure using the sNMF method and obtained the optimal number of ancestral populations, K = 13 and K = 14, for *D. cochinchinensis* and *D. oliveri*, which were much higher than previous studies where K = 5 was reported for both species[8,10]. We believed the difference can be accounted by three main reasons. First, the enormous number of SNPs (> 180,000) gave a much higher resolution than the previous studies using only nine microsatellite markers and 19 SNPs. Second, there were more populations in this study (35 and 28) than the previous one (26 and 23). Third, sNMF is able to detect lower hierarchical levels of population substructure, while the ΔK method, as done in previous studies, can only detect higher ones or fail[85,86]. It was also reported that in large human datasets K did not stabilise for K ≤ 10 using the minimal cross-entropy criterion in sNMF[61], thus a higher number of K (> 10) is necessary to describe the population substructure. In summary, we provided a higher resolution of neutral population structure of *D. cochinchinensis* and *D. oliveri* across their species range, which could be useful in delineating conservation management units and assessing gene flows in relevance to landscape features.

5.5.3 Environmental drivers of adaptive variation

Our gradient forest model identifies a number of environmental variables related to precipitation, soil, spatial, temperature, and vegetation as significant drivers of genomic variations across the species range. Spatial variables, in the form of PCNMs, explain the highest variation, indicating strong spatial autocorrelation and isolation by distance, as previous studies already reported [8,10]. It confirms the importance of correcting for the spatial factor in analysing adaptive genomic variation. However, it sometimes can also imply important unmeasured environmental predictors[87], and thus have to be taken with caution. Although the two species have an overlapping range, they show different patterns in gene-

environment associations: *D. cochinchinensis* has the strongest association with isothermality, while *D. oliveri* with precipitation of wettest quarter. In particular, there is a steep turnover in allelic frequencies at around 55% for isothermality. Isothermality is generally a useful metrics in tropical environments[88], and has been shown to influence plant height growth[89]. In a controlled greenhouse experiment, it was found that *D. cochinchinensis* is more anisohydric than *D. oliveri*, which means that it optimises carbon assimilation and tolerates drought as a pioneering species with faster growth[90]. *D. oliveri* is also found often in moist areas and along streams and rivers[91], and the morphological characters of its seeds suggest that water dispersal is likely[8]. Nonetheless, temperature and precipitation, and their variability such as isothermality[92], have been reported as the most important drivers in other similar studies on tree species[93–95].

5.5.4 Genomic vulnerability and match

Genomic vulnerability, determined here as the mismatch between current and future gene-environment associations, can inform priority setting. Conservation resources are limited, particularly in lower-income countries, and thus need efficient and practical priority setting[96]. As expected, our study reports that genomic vulnerability increases under higher emission scenarios along with time. However, the two species show a different pattern: *D. cochinchinensis* has an increased vulnerability rather uniformly across the species range, while *D. oliveri* has distinct regions with higher vulnerability, such as inland at the borders among Laos, Cambodia, and Vietnam. It is reported that some areas of their range show limited gene flow that can be explained by geographical barriers, with *D. oliveri* generally having a higher level of gene flow than *D. cochinchinensis*[8]. Due to the longevity and long generation time of trees, they usually have a low rate of migration which can hardly keep track of the climate velocity in many scenarios. Forest trees usually experience a higher rate of climate change per generation than mobile and short-living organisms[97]. In a rapidly

changing environment, forest trees either persist through migration or phenotypic plasticity, or extirpate[93] when environmental change outpaces the adaptation. By developing a spatially explicit model of vulnerability, we provide a metric which quantifies the balance between *in situ* adaptation and migration, where populations with lower vulnerability are likely to persist through adaptation[98]. We thus also inform important decision making in conservation between *in situ* and *ex situ*, where *ex situ* is recommended for populations with higher vulnerability to preserve the genetic resources that may disappear.

On the other hand, our gradient model also allows prediction of matches between current germplasm source and future restoration site, on which the resilience and productivity of the species in the future greatly depend. In case of maladaptation and extirpation due to environmental change[99], when the classical preference of local germplasm for local provenance may no longer hold, deliberate transfer of germplasm along climate gradients may be necessary[100]. Examples of such strategies include assisted migration based on predictive provenancing to facilitate adaptation of the populations under climate change[101], which can be classified into assisted gene flow and assisted colonization.

5.5.5 Conservation implications for assisted migration

Assisted gene flow (AGF) refers to the managed translocation of individuals or germplasms among populations within the current range (i.e. *in situ*), as to facilitate more rapid adaptation to the predicted future climate, or to mitigate local maladaptation[99]. AGF in *D. cochinchinensis* and *D. oliveri* may target hotspots of local adaptation in the coastal Vietnam, which have higher priority to conserve their genetic diversity. Genetic materials from these regions can be moved to suitable regions using the *seedeR* prediction to supply populations with individuals of higher fitness, helping the maintenance of the unique genetic components of the population.

Assisted colonization (AC) refers to the managed translocation of individuals or germplasms outside contemporary species ranges (i.e. *ex situ*). They are particularly important when setting up seed production areas and plantations in the restoration of *Dalbergia* species to multiply genetic materials. In particular, hotspots of vulnerable populations such as those in northern Cambodia will need to find new suitable areas to conserve the genetic diversity. Also, the edge of species range, especially in western Thailand, can benefit from AC to expand the species range.

Our newly developed application *seedeR* helps visualising these spatially explicit predictive models of genomic vulnerability and match, which are most useful because the science that underpins the predictions can be of little use to frontline practitioners and managers[87]. It informs the conservation and management strategies as discussed above and narrows the 'conservation genomics gap' which is related to the challenge of disseminating genomic knowledge, by simplifying the analytical pipelines and presenting a user-friendly platform[19].

5.6 Conclusion

This study presents a genome assembly of *Dalbergia oliveri* and a study of genomic variation of *Dalbergia* species. Genome-wide analysis enables us to detect signatures of gene-environment association and local adaptation, which are useful in ecological genomic models to predict the variation both spatially (e.g. across the species range) and temporally (e.g. different future climate scenarios). When coupled with gradient forest model, the genomic technology offers a faster, cheaper, and more robust route to test gene-environment associations than large-scale controlled experiment such as common garden or reciprocal transplantation. It is particularly important for decision-making in conservation management under rapidly changing environment. The outcomes of this study encourage two future avenues: (1) the spatially explicit model enables genomic prediction of germplasm sources and restoration sites, as demonstrated by the vulnerability model and the *seedeR* application, which can benefit the planning of assisted migration; (2) the adaptive loci pool and predicted gene-environment association encourages further empirical study to understand the ecological adaptation of *Dalbergia* species, for example, a controlled experiment to study temperature-associated loci.

5.7 Acknowledgements

This work was supported by funding to T.H.H. from the Biotechnology and Biological Sciences Research Council (BBSRC) [grant number BB/M011224/1]; and to T.S., B.T., H. T. T., R. J., D.H.B., J.J.M from the UK Darwin Initiative (ref. 25-023).

5.8 References

1. Vatanparast, M. *et al.* First molecular phylogeny of the pantropical genus *Dalbergia*: Implications for infrageneric circumscription and biogeography. *South African J. Bot.* **89**, 143–149 (2013).
2. UNODC. *World Wildlife Crime Report: Trafficking in Protected Species*. (United Nations Publication, 2016).
3. EIA. *The Rosewood Racket: China's billion dollar illegal timber trade and the devastation of Nigeria's forests*. (2017).
4. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
5. Woodruff, D. S. Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. *Biodivers. Conserv.* **19**, 919–941 (2010).
6. Wurster, C. M. *et al.* Forest contraction in north equatorial Southeast Asia during the Last Glacial Period. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15508–15511 (2010).
7. Theilade, I., Schmidt, L., Chhang, P. & McDonald, J. A. Evergreen swamp forest in Cambodia: floristic composition, ecological characteristics, and conservation status. *Nord. J. Bot.* **29**, 71–80 (2011).
8. Hartvig, I. *et al.* Population genetic structure of the endemic rosewoods *Dalbergia cochinchinensis* and *D. oliveri* at a regional scale reflects the Indochinese landscape and life-history traits. *Ecol. Evol.* **8**, 530–545 (2018).
9. Moritsuka, E. *et al.* Genetic variation and population structure of a threatened timber tree *Dalbergia cochinchinensis* in Cambodia. *Tree Genet. Genomes* **13**, 115 (2017).
10. Hartvig, I. *et al.* Conservation genetics of the critically endangered Siamese rosewood (*Dalbergia cochinchinensis*): recommendations for management and sustainable use.

Conserv. Genet. 1–16 (2020) doi:10.1007/s10592-020-01279-1.

11. Geissler, P. *et al.* The Lower Mekong: an insurmountable barrier to amphibians in southern Indochina? *Biol. J. Linn. Soc.* **114**, 905–914 (2015).
12. Tantrawatpan, C. *et al.* Genetic differentiation among populations of *Brachytrupes portentosus* (Lichtenstein 1796) (Orthoptera: Gryllidae) in Thailand and the Lao PDR: the Mekong River as a biogeographic barrier. *Bull. Entomol. Res.* **101**, 687–696 (2011).
13. Pakkad, G., Kanetani, S. & Elliott, S. Genetic diversity and differentiation of an endangered tree species, *Afzelia xylocarpa* (Kurz) craib in Thailand revealed by nuclear microsatellite markers. *African J. Biotechnol.* **13**, 366–377 (2014).
14. Pakkad, G., James, C., Torre, F., Elliott, S. & Blakesley, D. Genetic variation of *Prunus cerasoides* D. Don, a framework tree species in northern Thailand. *New For.* **27**, 189–200 (2003).
15. Senakun, C., Changtragoon, S., Pramual, P. & Prathepha, P. Genetic structure and diversity of *Shorea obtusa* (Dipterocarpaceae) in Thailand. *J. Syst. Evol.* **49**, 120–125 (2011).
16. Wattanakulpakin, T., Iamtham, S., Grubbs, K. C. & Volkaert, H. A. Population genetic analysis of *Xylia xylocarpa* (Fabaceae—Mimosoideae) in Thailand. *Tree Genet. Genomes* **11**, 7 (2015).
17. Davies, S. J. *et al.* Pollen flow in fragmented landscapes maintains genetic diversity following stand-replacing disturbance in a neotropical pioneer tree, *Vochysia ferruginea* Mart. *Heredity (Edinb).* **115**, 125–9 (2015).
18. Holliday, J. A. *et al.* Advances in ecological genomics in forest trees and applications to genetic resources conservation and breeding. *Mol. Ecol.* **26**, 706–717 (2017).
19. Shafer, A. B. A. *et al.* Genomics and the challenging translation into conservation

practice. *Trends Ecol. Evol.* **30**, 78–87 (2015).

20. Galla, S. J. *et al.* Building strong relationships between conservation genetics and primary industry leads to mutually beneficial genomic advances. *Mol. Ecol.* **25**, 5267–5281 (2016).
21. Santure, A. W. *et al.* On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Mol. Ecol.* **19**, 1439–1451 (2010).
22. Storfer, A., Patton, A. & Fraik, A. K. Navigating the Interface Between Landscape Genetics and Landscape Genomics. *Front. Genet.* **9**, 68 (2018).
23. Ramos, B. *et al.* Landscape genomics: natural selection drives the evolution of mitogenome in penguins. *BMC Genomics* **19**, 53 (2018).
24. Hayatgheibi, H., Fries, A., Kroon, J. & Wu, H. X. Estimation of genetic parameters, provenance performances, and genotype by environment interactions for growth and stiffness in lodgepole pine (*Pinus contorta*). *Scand. J. For. Res.* **34**, 1–11 (2019).
25. O'brien, E. K., Mazanec, R. A., Krauss, S. L. & O'brien, E. K. Provenance variation of ecologically important traits of forest trees: implications for restoration. *J. Appl. Ecol.* **44**, 583–593 (2007).
26. Hufford, K. M., Veneklaas, E. J., Lambers, H. & Krauss, S. L. Genetic delineation of local provenance defines seed collection zones along a climate gradient. *AoB Plants* **8**, (2016).
27. Leimu, R., Mutikainen, P., Koricheva, J. & Fischer, M. How general are positive relationships between plant population size, fitness and genetic variation? *J. Ecol.* **94**, 942–952 (2006).
28. Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. (Kuke) & Hedrick, P. W. Conservation genetics in transition to conservation genomics. *Trends Genet.* **26**, 177–

187 (2010).

29. Lanes, É. C. *et al.* Landscape genomic conservation assessment of a narrow-endemic and a widespread morning glory from Amazonian savannas. *Front. Plant Sci.* **9**, 532 (2018).

30. Li, Y. *et al.* Ten years of landscape genomics: challenges and opportunities. *Front. Plant Sci.* **8**, 2136 (2017).

31. Sork, V. L. *et al.* Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genet. Genomes* **9**, 901–911 (2013).

32. Balkenhol, N., Cushman, S., Storfer, A. & Waits, L. *Landscape Genetics: Concepts, Methods, Applications*. (John Wiley & Sons, 2016).

33. Fischer, M. C. *et al.* Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Mol. Ecol.* **22**, 5594 (2013).

34. Joost, S. *et al.* Uncovering the genetic basis of adaptive change: on the intersection of landscape genomics and theoretical population genetics. *Mol. Ecol.* **22**, 3659–3665 (2013).

35. Joost, S. *et al.* A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol. Ecol.* **16**, 3955–3969 (2007).

36. Legendre, P. & Legendre, L. *Numerical ecology*. (Elsevier, 2012).

37. Frichot, E., Schoville, S. D., Bouchard, G. & François, O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol. Biol. Evol.* **30**, 1687–1699 (2013).

38. Ellis, N., Smith, S. J. & Pitcher, C. R. Gradient forests: calculating importance

gradients on physical predictors. *Ecology* **93**, 156–168 (2012).

39. Ferrier, S., Manion, G., Elith, J. & Richardson, K. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers. Distrib.* **13**, 252–264 (2007).

40. Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M. & Holderegger, R. A practical guide to environmental association analysis in landscape genomics. *Mol. Ecol.* **24**, 4348–4370 (2015).

41. De Kort, H. *et al.* Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Mol. Ecol.* **23**, 4709–4721 (2014).

42. Ćalić, I., Bussotti, F., Martínez-García, P. J. & Neale, D. B. Recent landscape genomics studies in forest trees—what can we believe? *Tree Genet. Genomes* **12**, 3 (2016).

43. Keller, S. R., Levsen, N., Olson, M. S. & Tiffin, P. Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. *Mol. Biol. Evol.* **29**, 3143–3152 (2012).

44. Namroud, M.-C., Beaulieu, J., Juge, N., Laroche, J. & Bousquet, J. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol. Ecol.* **17**, 3599–613 (2008).

45. Prunier, J., Laroche, J., Beaulieu, J. & Bousquet, J. Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Mol. Ecol.* **20**, 1702–1716 (2011).

46. Holliday, J. A., Wang, T. & Aitken, S. Predicting adaptive phenotypes from multilocus genotypes in Sitka spruce (*Picea sitchensis*) using random forest. *G3 (Bethesda).* **2**, 1085–93 (2012).

47. Brousseau, L. *et al.* Genomics of microgeographic adaptation in the hyperdominant Amazonian tree *Eperua falcata* Aubl. (Fabaceae). *bioRxiv* 312843 (2018) doi:10.1101/312843.

48. Belaghzal, H., Dekker, J. & Gibcus, J. H. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* **123**, 56 (2017).

49. Cabanettes, F. & Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958 (2018).

50. Hung, T. H. *et al.* Reference transcriptomes and comparative analyses of six species in the threatened rosewood genus Dalbergia. *Sci. Rep.* **10**, 17749 (2020).

51. Holt, C. & Yandell, M. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).

52. Manni, M., Berkeley, M. R., Seppey, M., Sim~ Ao, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).

53. Inglis, P. W., Pappas, M. de C. R., Resende, L. V. & Grattapaglia, D. Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One* **13**, e0206085 (2018).

54. Torkamaneh, D., Laroche, J. & Belzile, F. Fast-GBS v2.0: an analysis toolkit for genotyping-by-sequencing data. *https://doi.org/10.1139/gen-2020-0077 Genome* **63**, 577–581 (2020).

55. Joshi, N. A. sabre - A barcode demultiplexing and trimming tool for FastQ files.

(2013).

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

57. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **1303.3997**, (2013).

58. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, 1–4 (2021).

59. Rimmer, A. *et al.* Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet. 2014 468* **46**, 912–918 (2014).

60. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

61. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196**, 973–983 (2014).

62. Martins, H., Caye, K., Luu, K., Blum, M. G. B. & François, O. Identifying outlier loci in admixed and in continuous populations using ancestral population differentiation statistics. *Mol. Ecol.* **25**, 5029–5042 (2016).

63. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807 (2011).

64. Frichot, E. & François, O. LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929 (2015).

65. Borcard, D. & Legendre, P. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol. Modell.* **153**, 51–68 (2002).

66. Rellstab, C., Dauphin, B. & Exposito-Alonso, M. Prospects and limitations of genomic offset in conservation management. *Evol. Appl.* **14**, 1202–1212 (2021).

67. Scheffers, B. R. *et al.* The broad footprint of climate change from genes to biomes to people. *Science.* **354**, (2016).

68. Newbold, T. *et al.* Global effects of land use on local terrestrial biodiversity. *Nature* **520**, 45–50 (2015).

69. Butchart, S. H. M. *et al.* Global biodiversity: Indicators of recent declines. *Science.* **328**, 1164–1168 (2010).

70. APFORGEN. *Conserving Rosewood genetic resources for resilient livelihoods in the Mekong - Project Inception Workshop Report*. (2018).

71. Jalonen, R., Choo, K. Y., Hong, L. T. & Sim, H. C. *Forest genetic resources conservation and management: status in seven South and Southeast Asian countries*. (2009).

72. Choo, K. Y., Jalonen, R., Hong, L. T. & Sim, H. C. *Forest genetic resources conservation and management: national consultative workshops of seven South and Southeast Asian countries*. (2010).

73. Li, C., Zheng, Y., Liu, Y., Lin, F. & Huang, P. Development of genomic SSR for the subtropical hardwood tee *Dalbergia hupeana* and assessment of their transferability to other related species. *For. 2021, Vol. 12, Page 804* **12**, 804 (2021).

74. Hong, Z. *et al.* The chromosome-level draft genome of *Dalbergia odorifera*. *Gigascience* **9**, 1–8 (2020).

75. Liu, F.-M. *et al.* De novo transcriptome analysis of *Dalbergia odorifera* and transferability of SSR markers developed from the transcriptome. *Forests* **10**, 98 (2019).

76. Supple, M. A. & Shapiro, B. Conservation of biodiversity in the genomics era. *Genome Biol.* **19**, (2018).

77. Allendorf, F. W., Hohenlohe, P. A. & Luikart, G. Genomics and the future of

conservation genetics. *Nat. Rev. Genet.* **11**, 697–709 (2010).

78. Desalle, R. & Amato, G. Conservation genetics, precision conservation, and de-extinction. *Hastings Cent. Rep.* **47**, S18–S23 (2017).

79. Bhagwat, R. M., Dholakia, B. B., Kadoo, N. Y., Balasundaran, M. & Gupta, V. S. Two new potential barcodes to discriminate *Dalbergia* species. *PLoS One* **10**, 1–18 (2015).

80. Wattoo, J. I., Saleem, M. Z., Shahzad, M. S., Arif, A. & Hameed, A. DNA Barcoding: Amplification and sequence analysis of rbcL and matK genome regions in three divergent plant species. *Adv. Life Sci.* **4**, 03–07 (2016).

81. Phong, D. T., Tang, D. Van, Hien, V. T. T., Ton, N. D. & Van, H. N. Nucleotide diversity of a nuclear and four chloroplast DNA regions in rare tropical wood species of *Dalbergia* in Vietnam: a DNA barcode identifying utility. *Asian J. Appl. Sci.* **02**, 116–125 (2014).

82. Hartvig, I., Czako, M., Kjær, E. D., Nielsen, L. R. & Theilade, I. The use of DNA barcoding in identification and conservation of rosewood (*Dalbergia* spp.). *PLoS One* **10**, (2015).

83. RIBEIRO, R. A. *et al.* Development of polymorphic microsatellite markers for *Dalbergia nigra* (Papilionoideae), an endangered tree from the Brazilian Atlantic Forest. *Mol. Ecol. Resour.* **9**, 203–206 (2009).

84. Resende, L. C., Ribeiro, R. A. & Lovato, M. B. Diversity and genetic connectivity among populations of a threatened tree (*Dalbergia nigra*) in a recently fragmented landscape of the Brazilian Atlantic Forest. *Genetica* **139**, 1159–1168 (2011).

85. Kuismin, M. O., Ahlinder, J. & Sillanpää, M. J. CONE: Community oriented network estimation is a versatile framework for inferring population structure in large-scale sequencing data. *G3 Genes, Genomes, Genet.* **7**, 3359–3377 (2017).

86. Janes, J. K. *et al.* The $K = 2$ conundrum. *Mol. Ecol.* **26**, 3594–3602 (2017).

87. Martins, K. *et al.* Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of *Quercus rugosa*. *Evol. Appl.* **11**, 1842–1858 (2018).

88. Nix, H. A. A biogeographic analysis of Australian elapid snakes. in *Atlas of elapid snakes of Australia: Canberra, Australian Flora and Fauna Series 7* (ed. Longmore, R.) 4–15 (Australian Government Publishing Service, 1986).

89. Moles, A. T. *et al.* Global patterns in plant height. *J. Ecol.* **97**, 923–932 (2009).

90. Hung, T. H. *et al.* Physiological responses of rosewoods Dalbergia cochinchinensis and D. oliveri under drought and heat stresses. *Ecol. Evol.* (2020) doi:10.1002/ece3.6744.

91. Aerts, R. *et al.* Site requirements of the endangered rosewood *Dalbergia oliveri* in a tropical deciduous forest in northern Thailand. *For. Ecol. Manage.* **259**, 117–123 (2009).

92. Garnier-Géré, P. H. & Ades, P. K. Environmental surrogates for predicting and conserving adaptive genetic variability in tree species. *Conserv. Biol.* **15**, 1632–1644 (2001).

93. Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T. & Curtis-McLane, S. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol. Appl.* **1**, 95–111 (2008).

94. Supple, M. A. *et al.* Landscape genomic prediction for restoration of a *Eucalyptus* foundation species under climate change. *Elife* **7**, e31835 (2018).

95. Manel, S. *et al.* Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Mol. Ecol.* **21**, 3729–3738 (2012).

96. Game, E. T., Kareiva, P. & Possingham, H. P. Six common mistakes in conservation priority setting. *Conserv. Biol.* **27**, 480–485 (2013).

97. IPCC. *Climate Change 2014: Mitigation of Climate Change Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. (Cambridge University Press, 2014).

98. Gougherty, A. V., Keller, S. R. & Fitzpatrick, M. C. Maladaptation, migration and extirpation fuel climate change risk in a forest tree species. *Nat. Clim. Chang. 2021 112* **11**, 166–171 (2021).

99. Aitken, S. N. & Whitlock, M. C. Assisted gene flow to facilitate local adaptation to climate change. *Annu. Rev. Ecol. Evol. Syst.* **44**, 367–388 (2013).

100. Bozzano, M. *et al. Genetic Considerations in Ecosystem Restoration Using Native Tree Species*. (FAO and Bioversity International, 2014).

101. Chmura, D. J. *et al.* Forest responses to climate change in the northwestern United States: ecophysiological foundations for adaptive management. *For. Ecol. Manage.* **261**, 1121–1142 (2011).

5.9 Figures and Tables

Figure 5.1. The user interface of the "Predict" function of *seedeR* (https://trainingidn.shinyapps.io/seeder/). In the left panel "Inputs", the user can choose a species, a SSP, a time period and input the longitude and the latitude of the future restoration site. After clicking "Run", the process will begin and print the output In the right panel "Output".

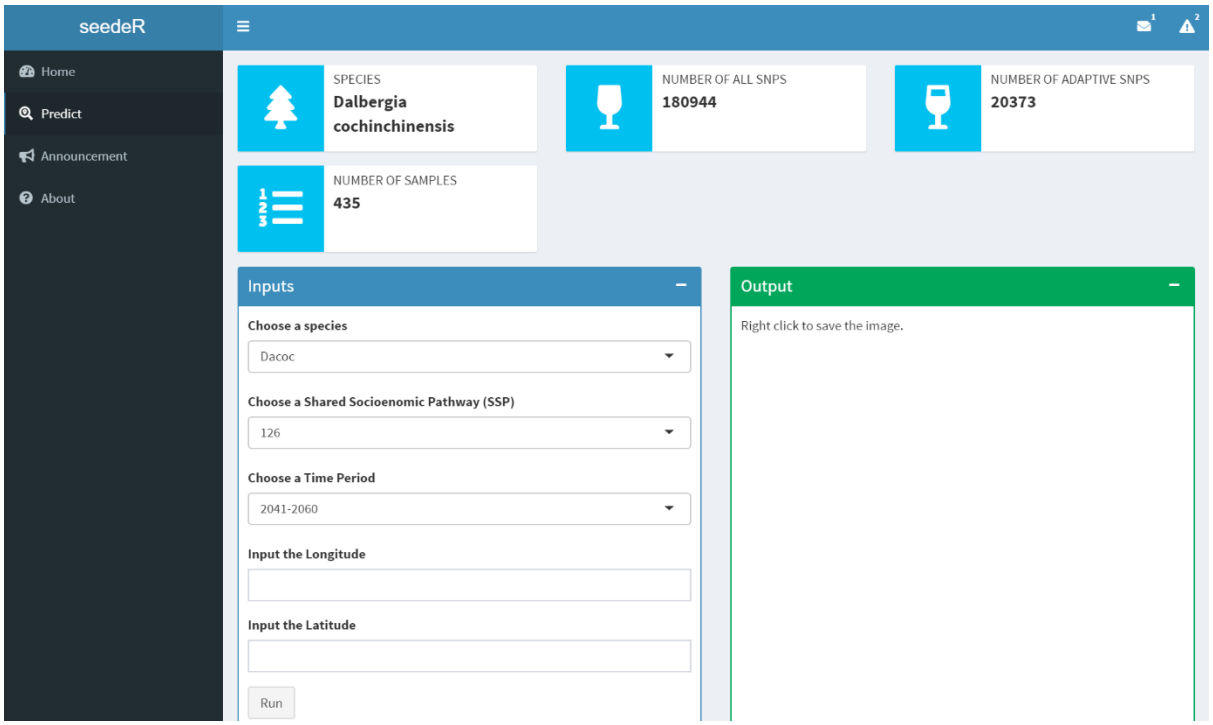


Figure 5.2. Snailplot of the genome assembly Daoli_0.2, showing scaffold statistics, BUSCO statistics, and assembly base composition.

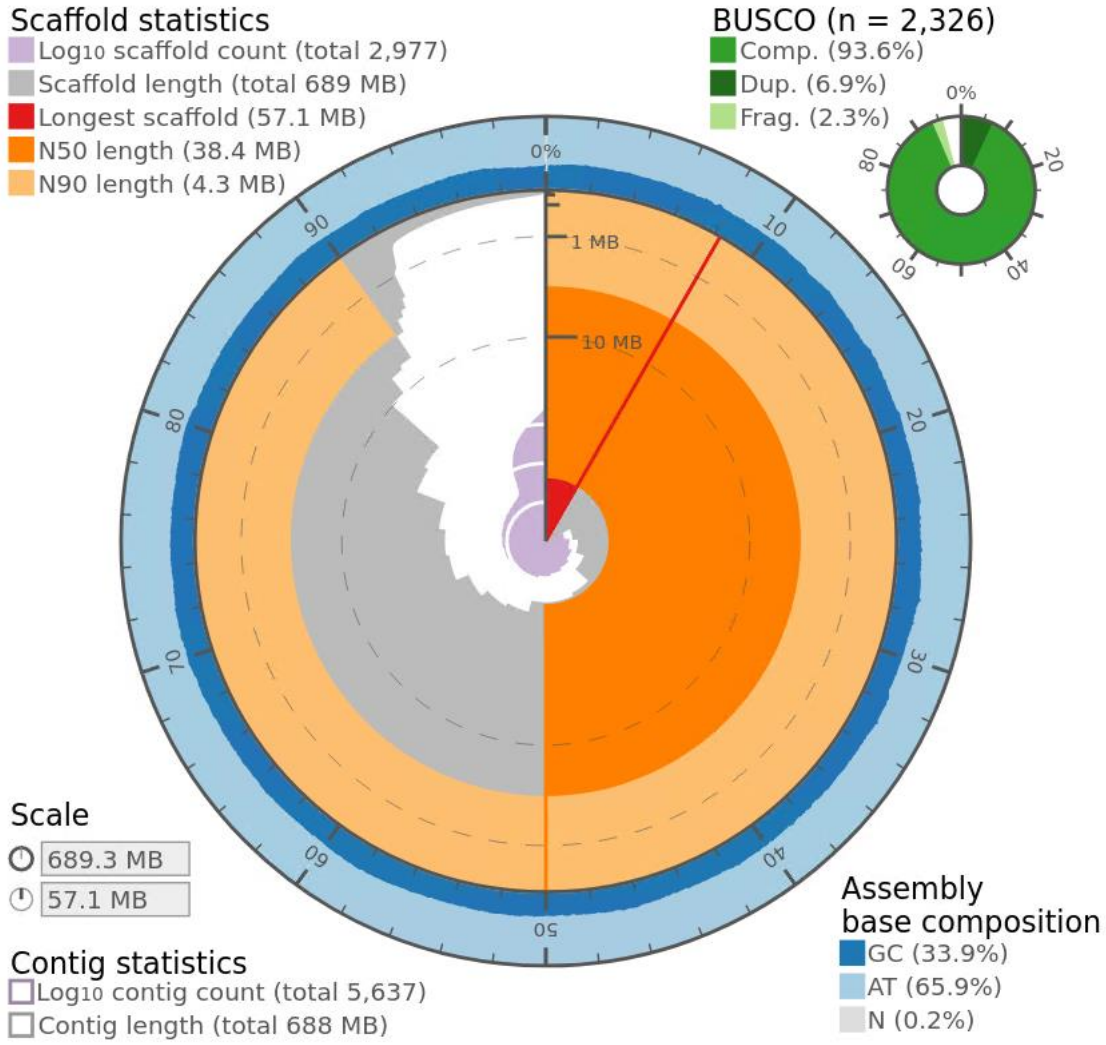


Figure 5.3. Syntenic dot plot of *D. oliveri* (Daoli_0.2) against *D. cochinchinensis* (Dacoc_1.2) with a minimum identity of 0.25.

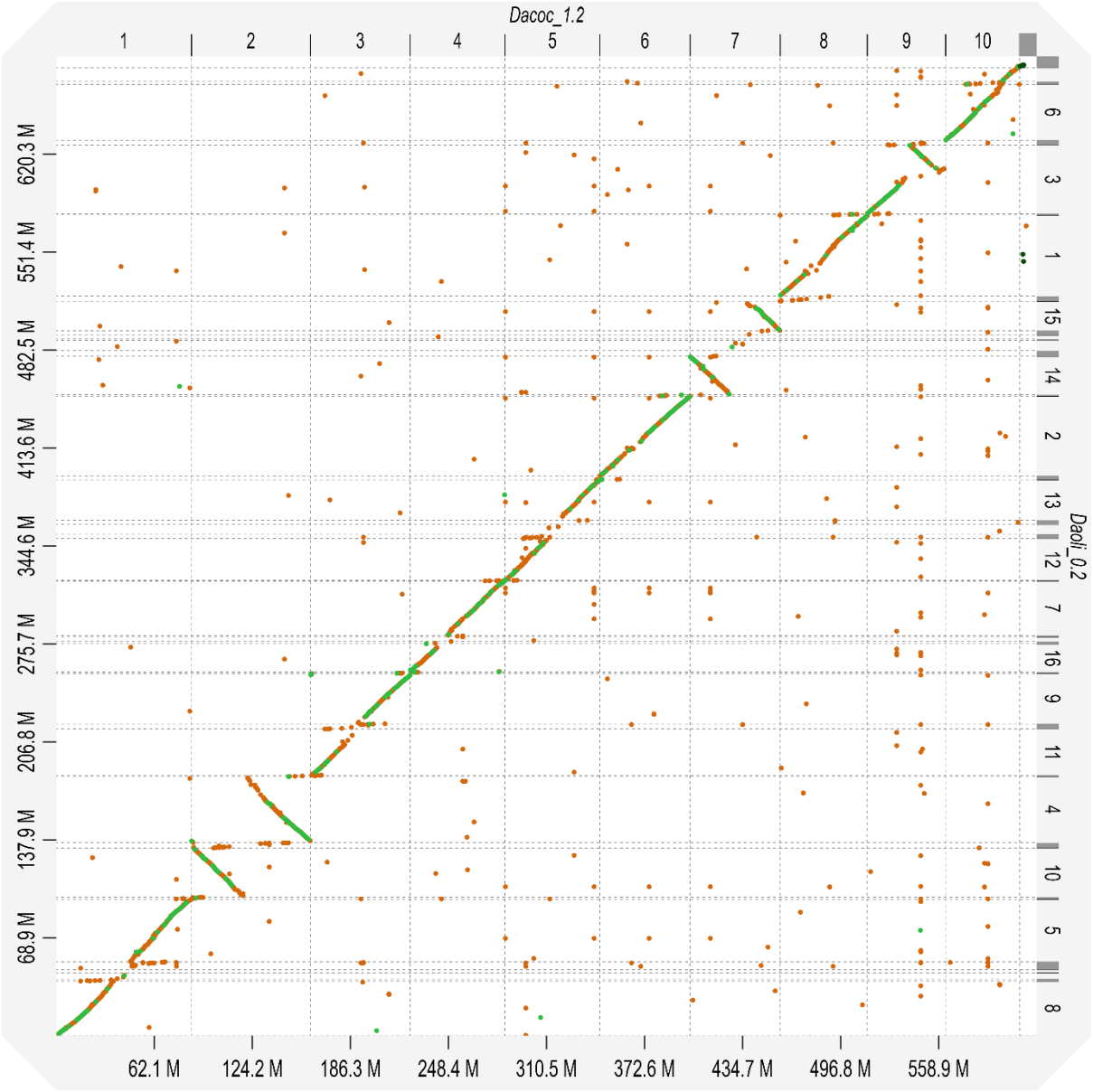


Figure 5.4. Percentage of repeat elements in the repeat library of Daoli_0.2.

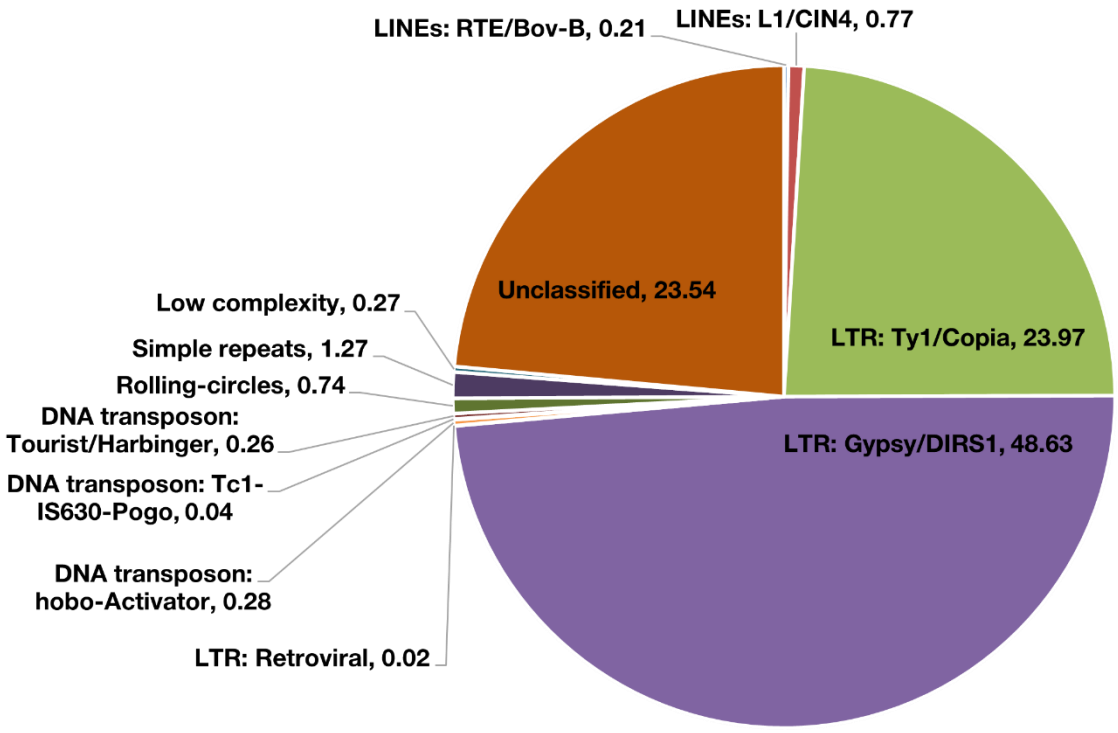


Figure 5.5. Cumulative frequency of gene models of Daoli_0.2 against annotation edit distance (AED).

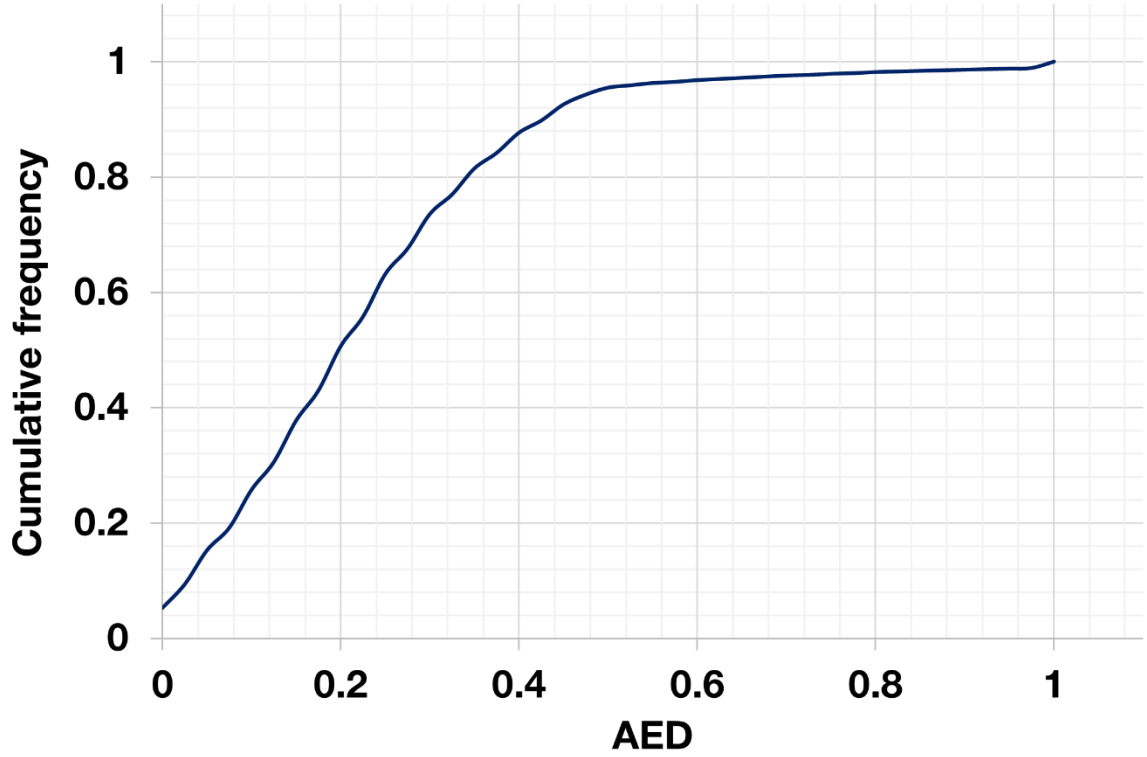
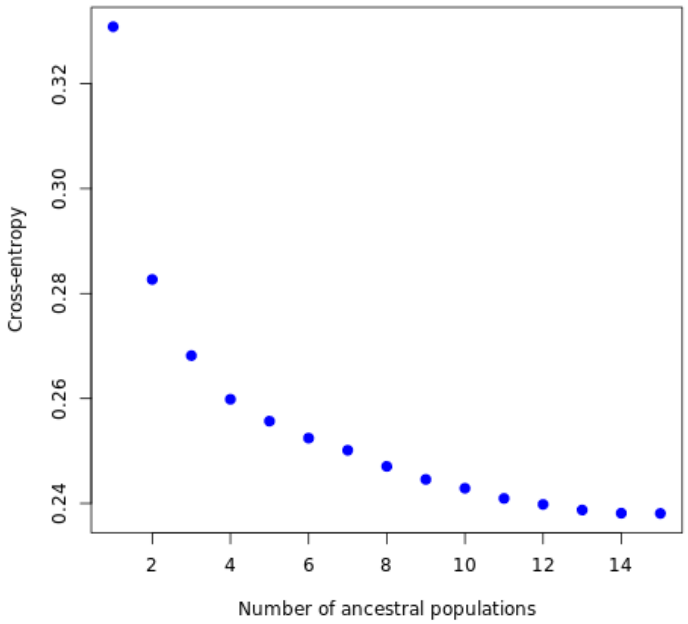


Figure 5.6. Cross-entropy criterion as a function of the number of factors in the sNMF analysis for **(a)** *D. cochinchinensis* and **(b)** *D. oliveri*.

(a)



(b)

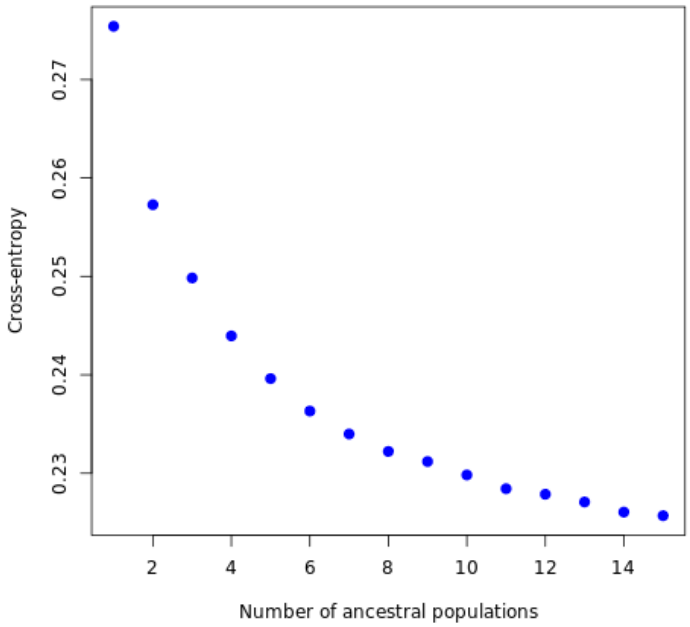
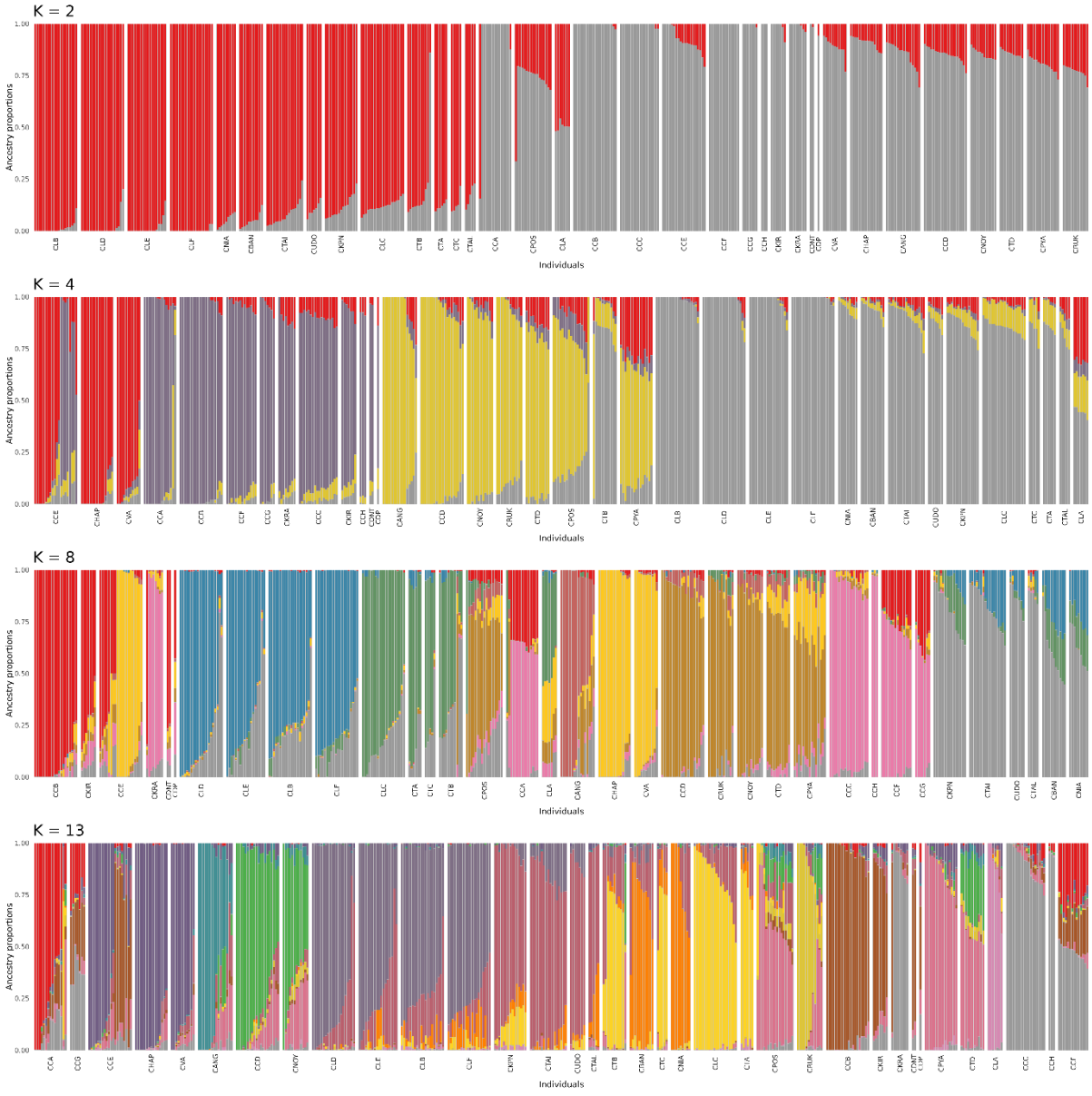


Figure 5.7. Admixture results of sNMF analysis showing hierarchical population substructure of **(a)** *D. cochinchinensis* and **(b)** *D. oliveri*.

(a)



(b)

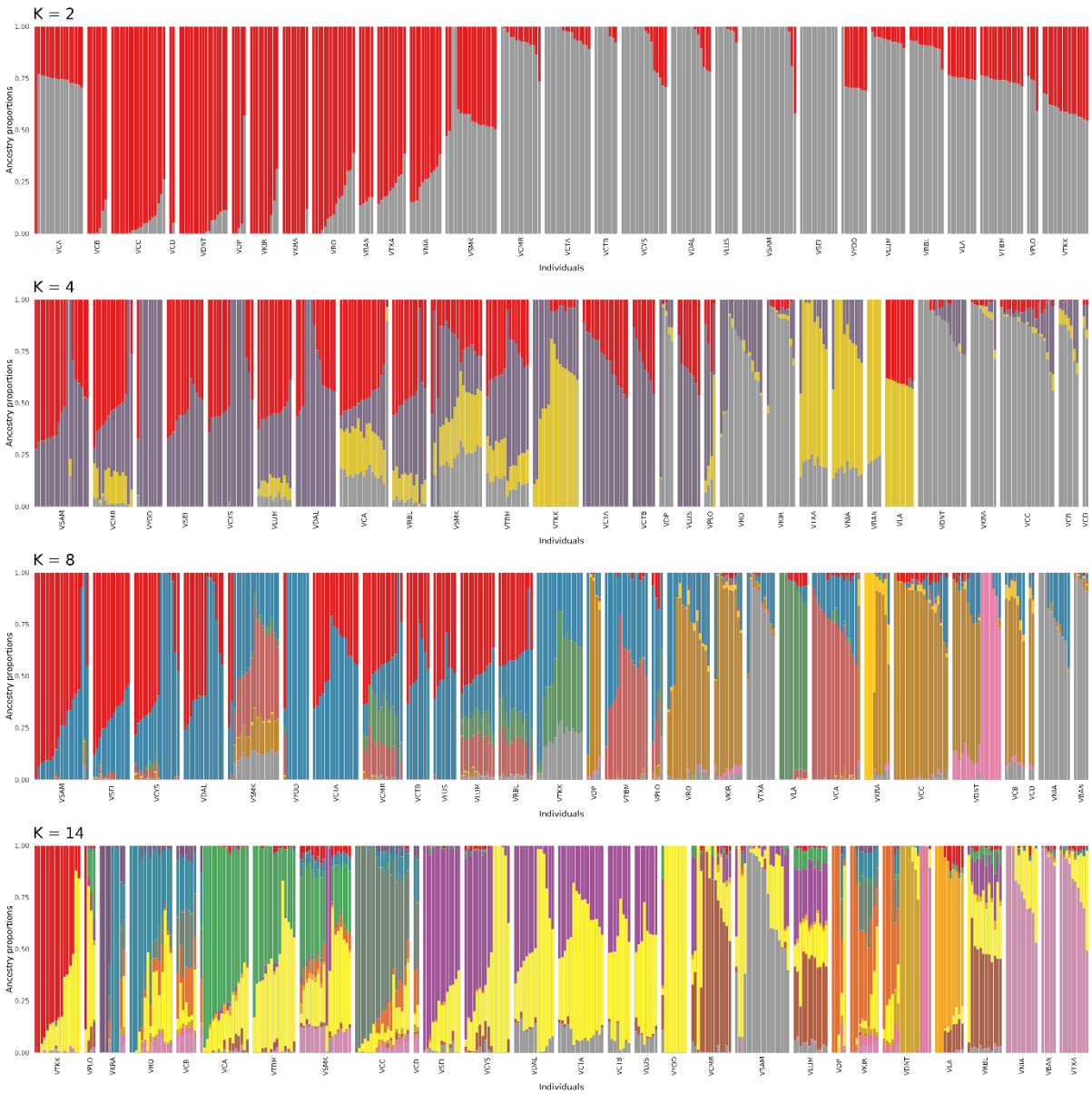
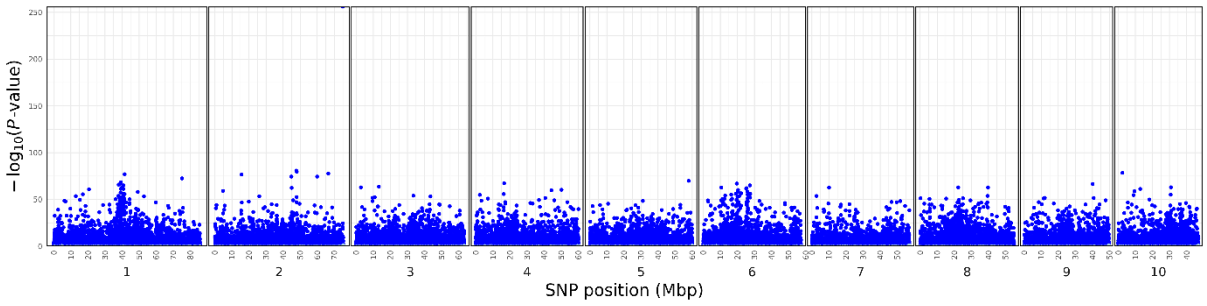


Figure 5.8. Manhattan plot of outlier loci predicted by population differentiation tests with sNMF for **(a)** *D. cochinchinensis* and **(b)** *D. oliveri*. Blue dots are outlier SNPs (*P*-value < 0.05), light-blue dots are non-outlier SNPs.

(a)



(b)

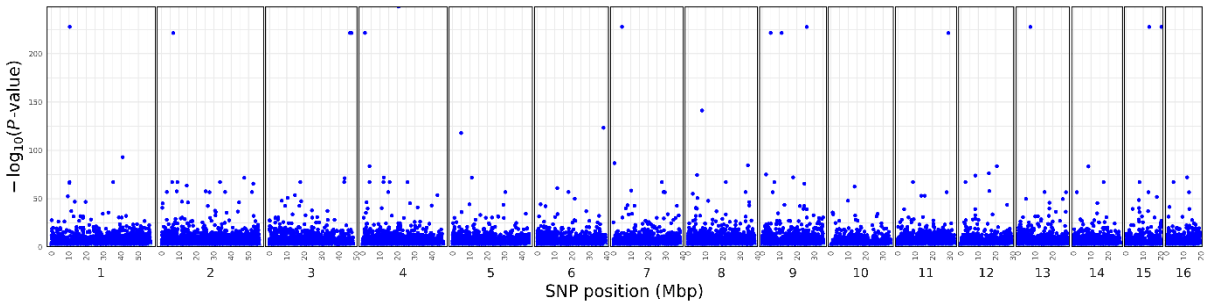
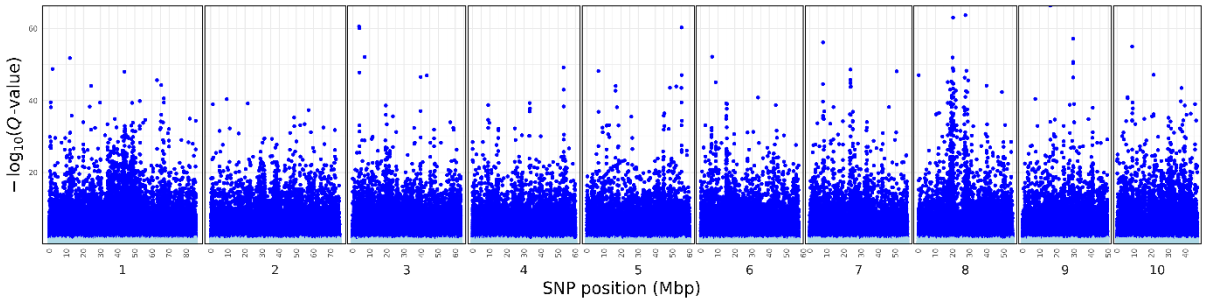


Figure 5.9. Manhattan plot of adaptive SNPs predicted by LFMM for **(a)** *D. cochinchinensis* and **(b)** *D. oliveri*. Blue dots are adaptive SNPs for at least one environmental variable (| *Z*-value | > 2 & *Q*-value < 0.01), light-blue dots are neutral SNPs.

(a)



(b)

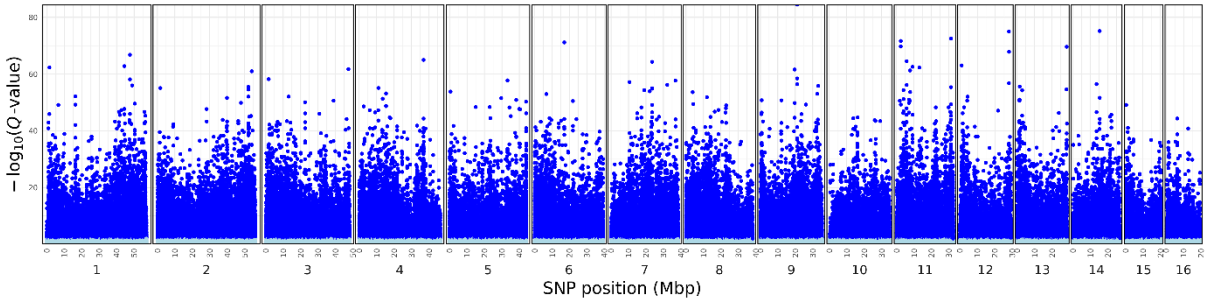
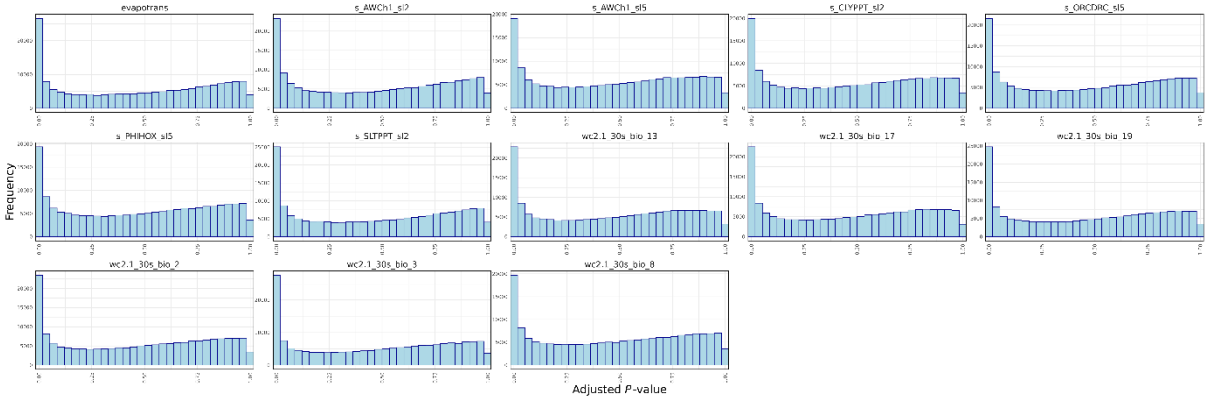


Figure 5.10. Frequency histogram of adjusted *P*-values for each environmental association of the LFMM analysis in **(a)** *D. cochinchinensis* and **(b)** *D. oliveri*.

(a)



(b)

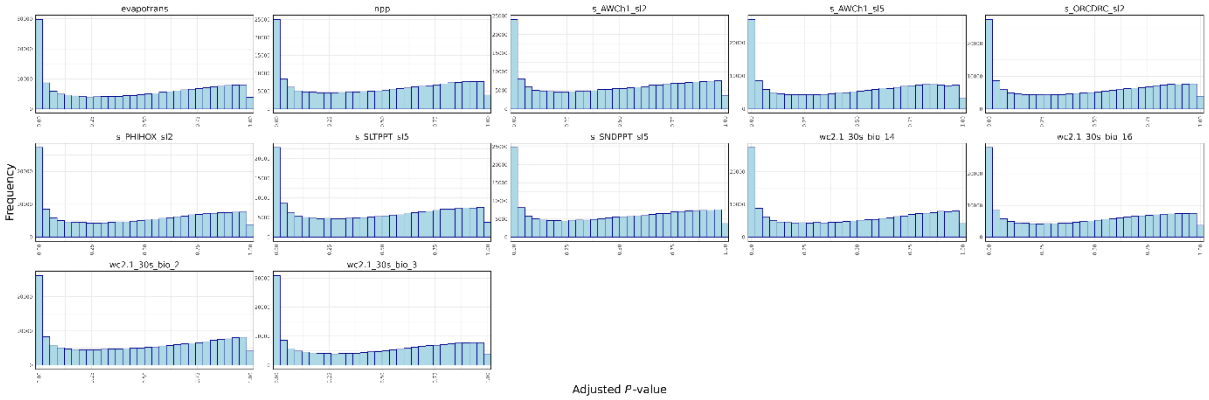
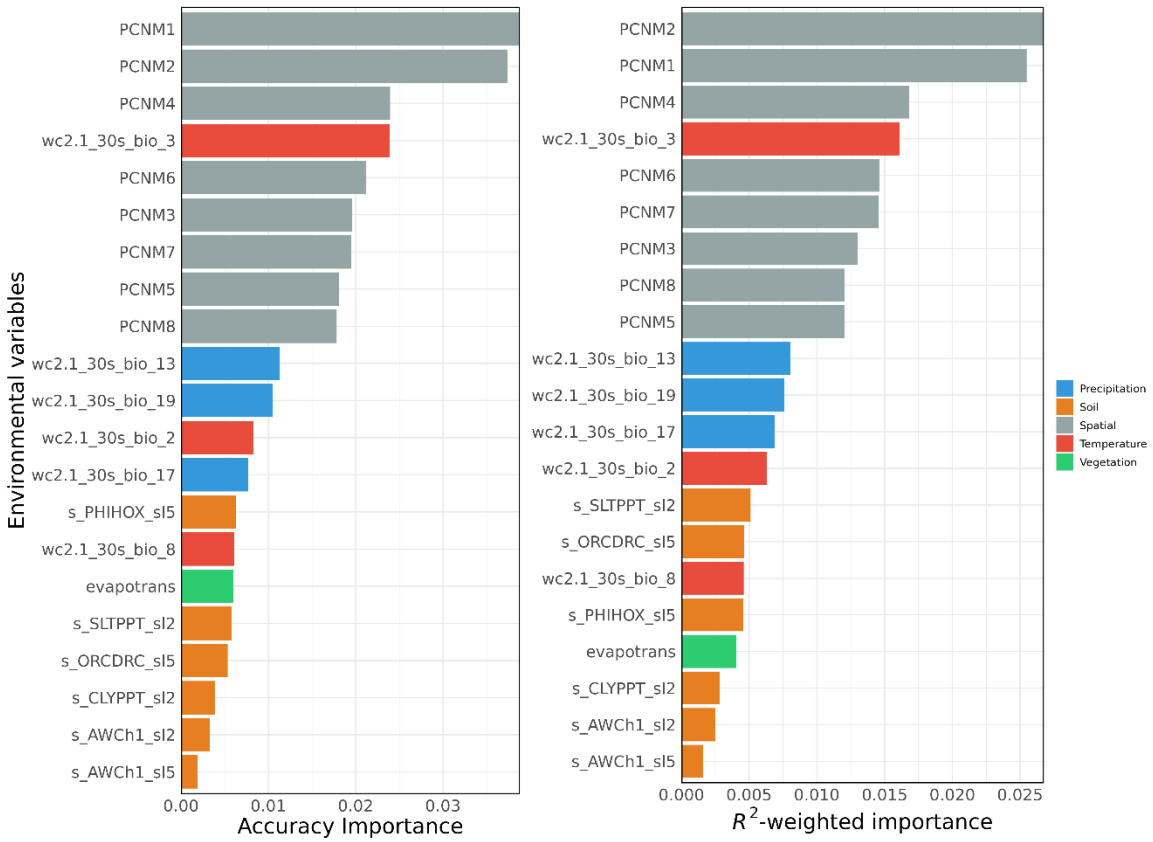


Figure 5.11. Accuracy and R^2-weighted importance for the environmental predictor variables which explained the neutral genomic variation (all SNPs) for **(a)** *D. cochinchinensis* and **(b)** *D. oliveri* predicted by the GF model.

(a)



(b)

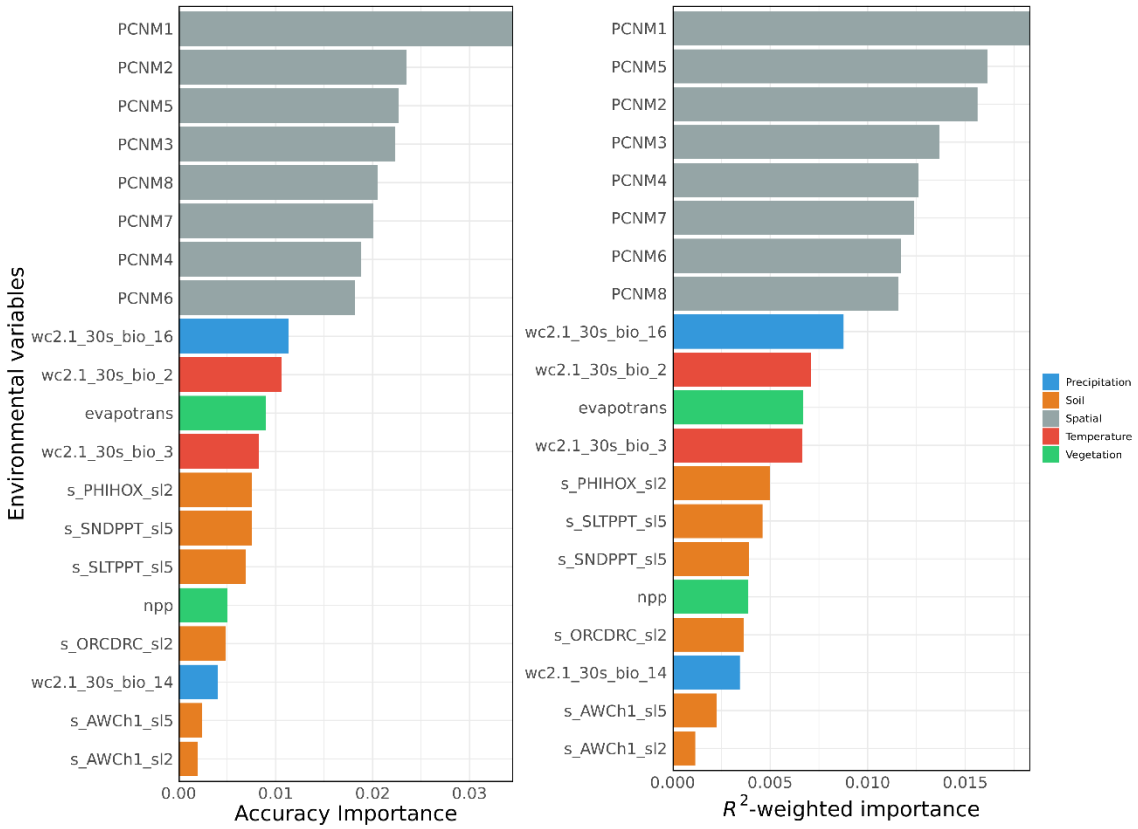
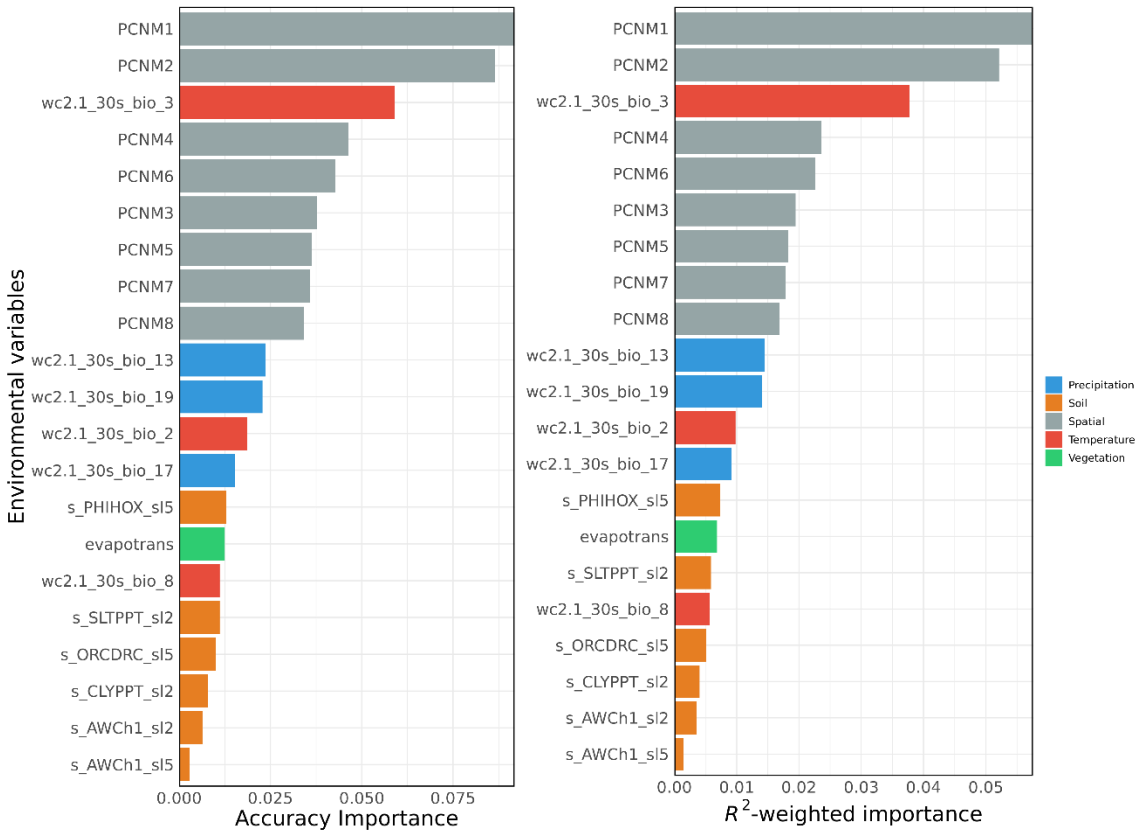

Environmental variables
PCNM1
PCNM2
PCNM5
PCNM3
PCNM8
PCNM7
PCNM4
PCNM6
wc2.1_30s_bio_16
wc2.1_30s_bio_2
evapotrans
wc2.1_30s_bio_3
s_PHIHOX_sl2
s_SNDPPT_sl5
s_SLTPPT_sl5
npp
s_ORCDRC_sl2
wc2.1_30s_bio_14
s_AWCh1_sl5
s_AWCh1_sl2
0.00
0.01
0.02
0.03
Accuracy Importance
PCNM1
PCNM5
PCNM2
PCNM3
PCNM4
PCNM7
PCNM6
PCNM8
wc2.1_30s_bio_16
wc2.1_30s_bio_2
evapotrans
wc2.1_30s_bio_3
s_PHIHOX_sl2
s_SLTPPT_sl5
s_SNDPPT_sl5
npp
s_ORCDRC_sl2
wc2.1_30s_bio_14
s_AWCh1_sl5
s_AWCh1_sl2
0.000
0.005
0.010
0.015
R^2-weighted importance
Precipitation
Soil
Spatial
Temperature
Vegetation

Figure 5.12. Accuracy and R^2-weighted importance for the environmental predictor variables which explained the adaptive genomic variation (adaptive SNPs) for **(a)** *D. cochinchinensis* and **(b)** *D. oliveri* predicted by the GF model.

(a)



(b)

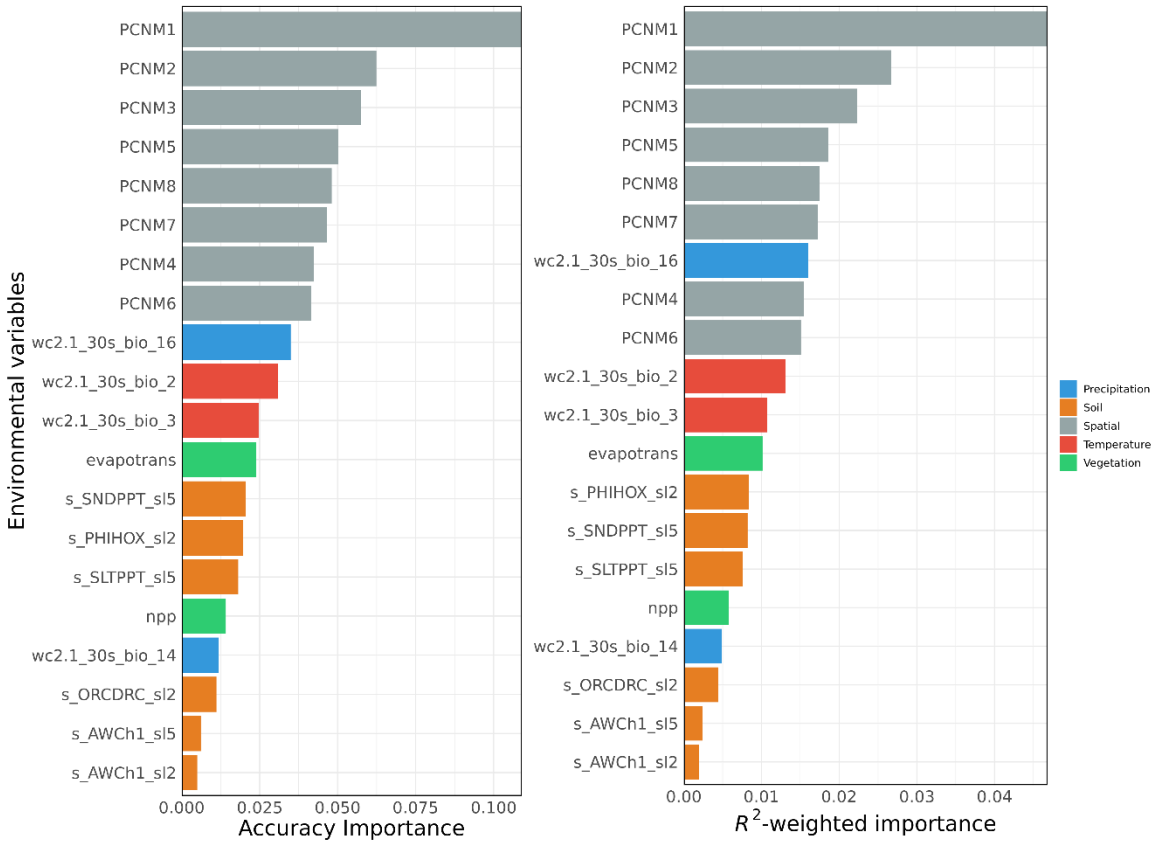
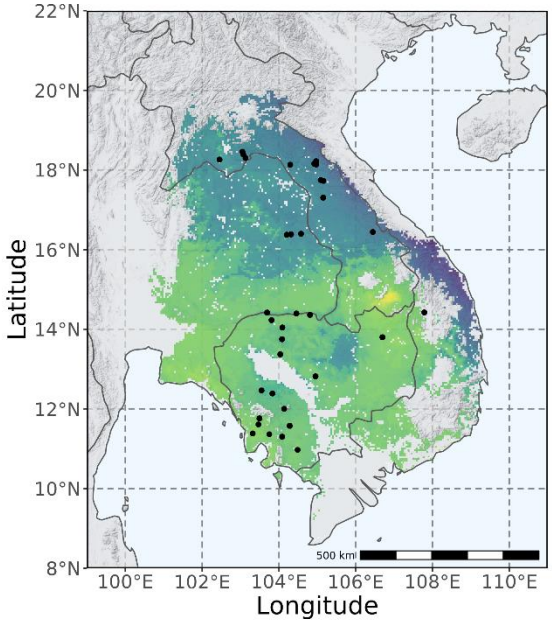

Environmental variables
PCNM1
PCNM2
PCNM3
PCNM5
PCNM8
PCNM7
PCNM4
PCNM6
wc2.1_30s_bio_16
wc2.1_30s_bio_2
wc2.1_30s_bio_3
evapotrans
s_SNDPPT_sl5
s_PHIHOX_sl2
s_SLTPPT_sl5
npp
wc2.1_30s_bio_14
s_ORCDRC_sl2
s_AWCh1_sl5
s_AWCh1_sl2
0.000
0.025
0.050
0.075
0.100
Accuracy Importance
0.00
0.01
0.02
0.03
0.04
R^2-weighted importance
Precipitation
Soil
Spatial
Temperature
Vegetation

Figure 5.13. Adaptive genomic variation across the species range predicted by GF model for **(a)** *D. cochinchinensis* and **(b)** *D. oliveri*, visualised using the first two principal axes from the PCA.

(a)



(b)

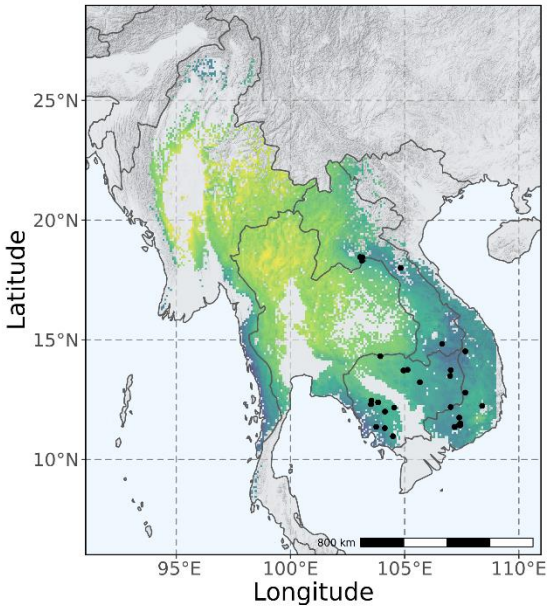
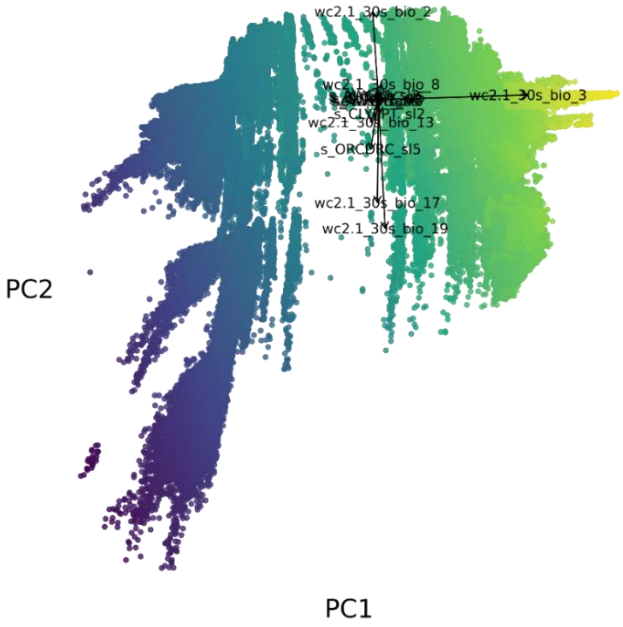


Figure 5.14. Principal component analysis (PCA) of the adaptive genomic variation predicted by the GF model across the species range for **(a)** *D. cochinchinensis* and **(b)** *D. oliveri.* The loadings are the environmental factors.

(a)



(b)

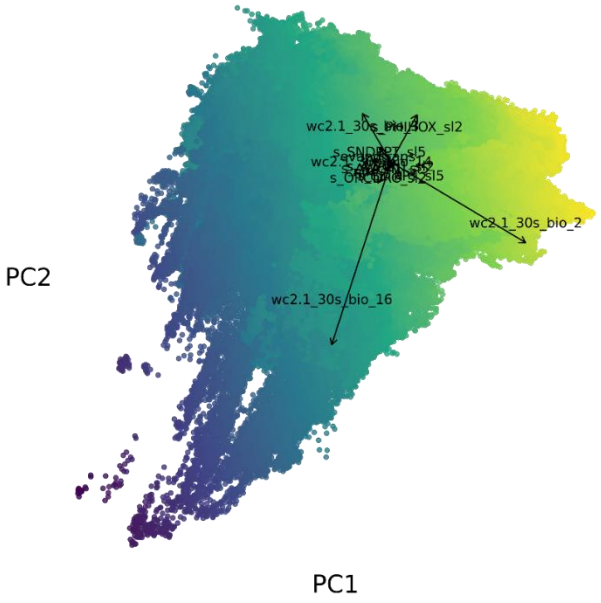
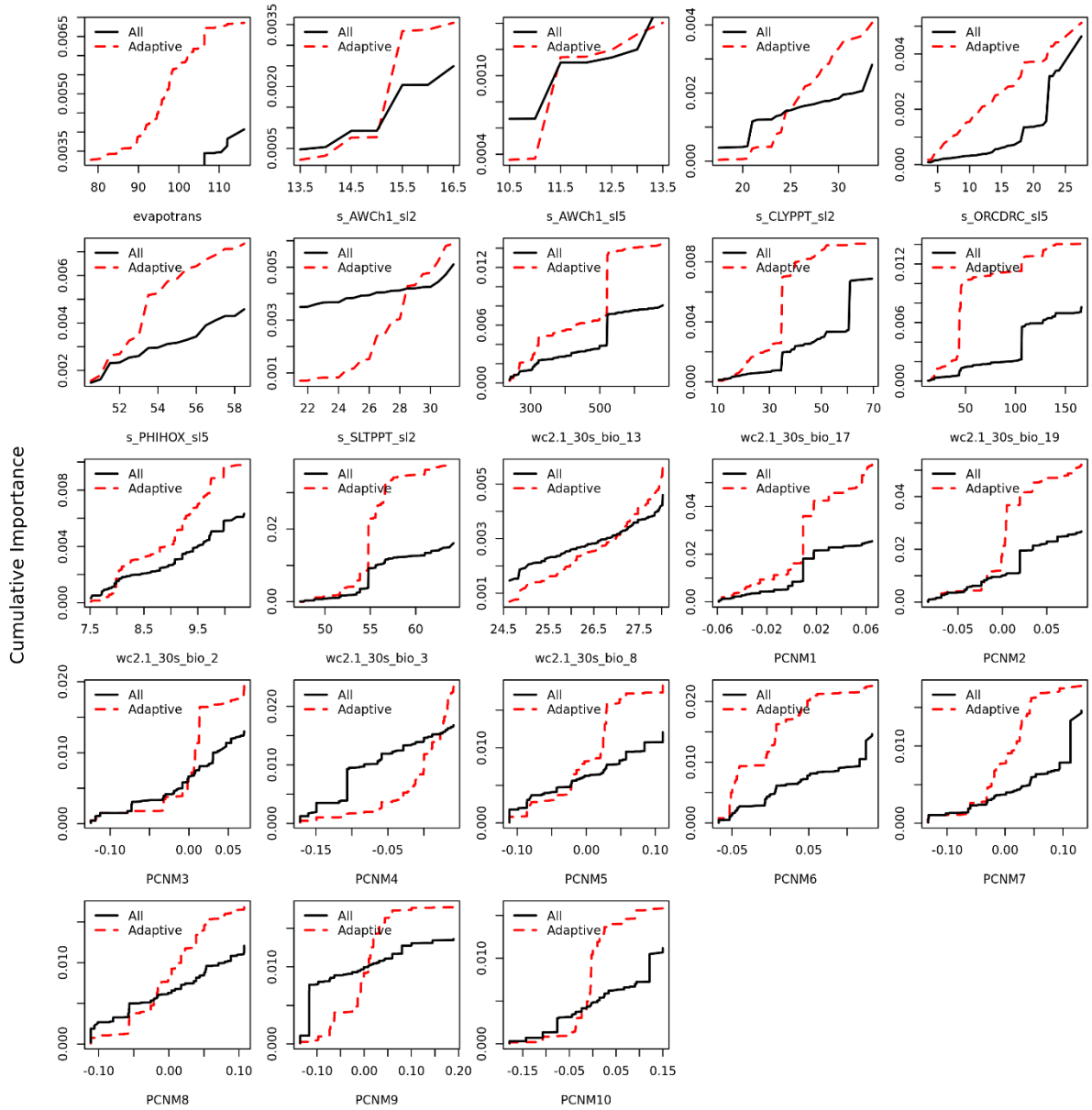


Figure 5.15. Allelic turnover functions of all and adaptive SNPs explaining the cumulative importance of each environmental predictor variable in **(a)** *D. cochinchinensis* and **(b)** *D. oliveri*.

(a)



(b)

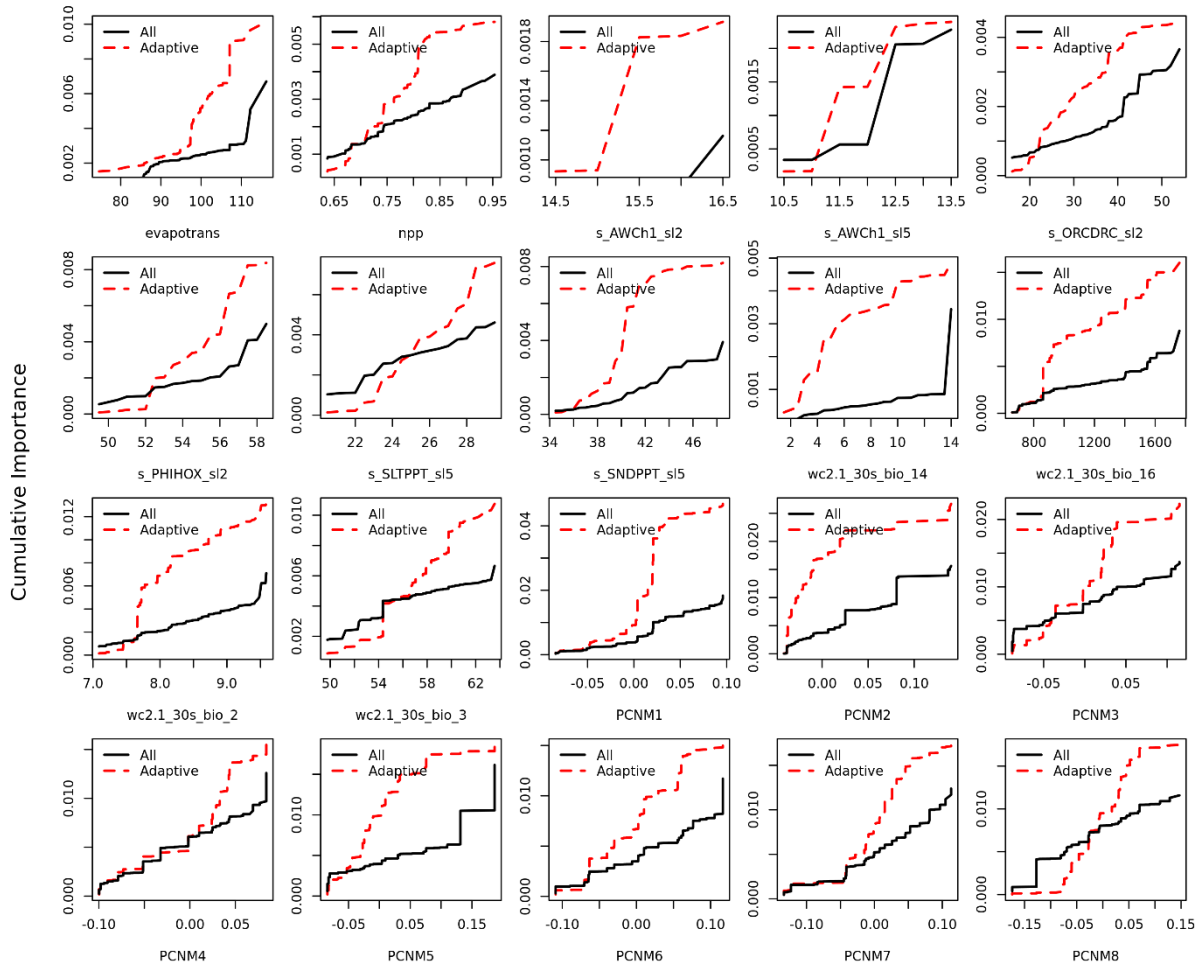
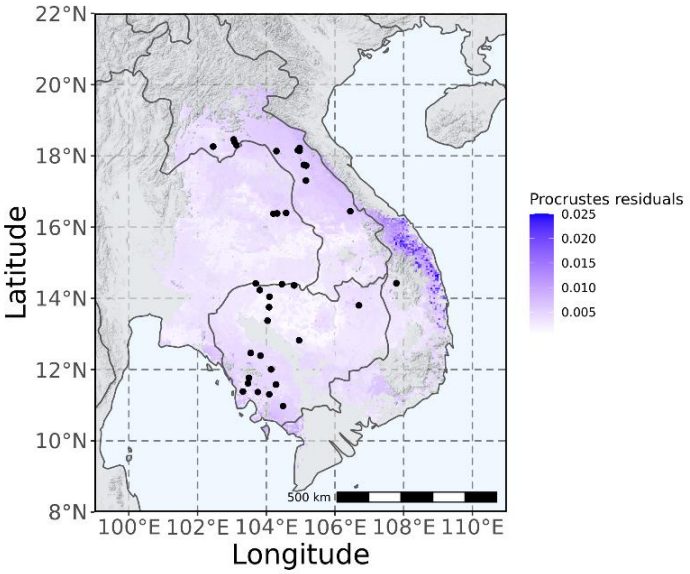

Cumulative Importance
All
Adaptive
evapotrans
npp
s_AWCh1_sl2
s_AWCh1_sl5
s_ORCDRC_sl2
s_PHIHOX_sl2
s_SLTPPT_sl5
s_SNDPPT_sl5
wc2.1_30s_bio_14
wc2.1_30s_bio_16
wc2.1_30s_bio_2
wc2.1_30s_bio_3
PCNM1
PCNM2
PCNM3
PCNM4
PCNM5
PCNM6
PCNM7
PCNM8

Figure 5.16. Procrustes residuals between neutral and adaptive gene-environmental associations as an indicator of strength of local adaptation for **(a)** *D. cochinchinensis* and **(b)** *D. oliveri.*

(a)



(b)

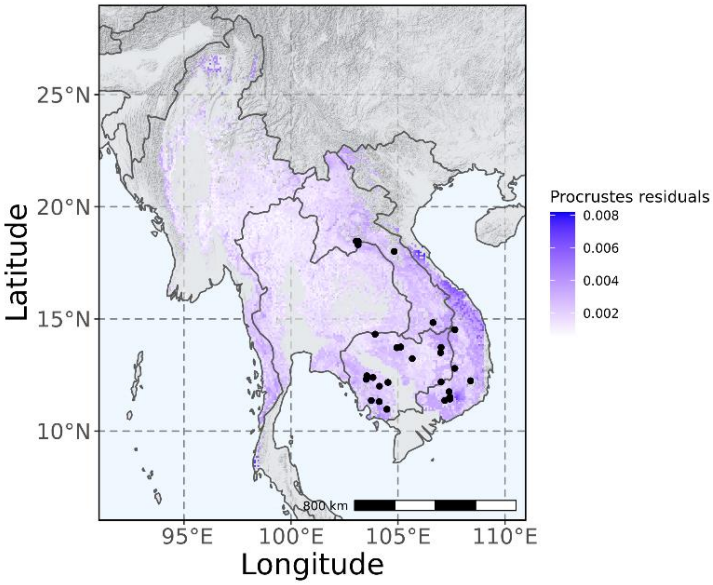
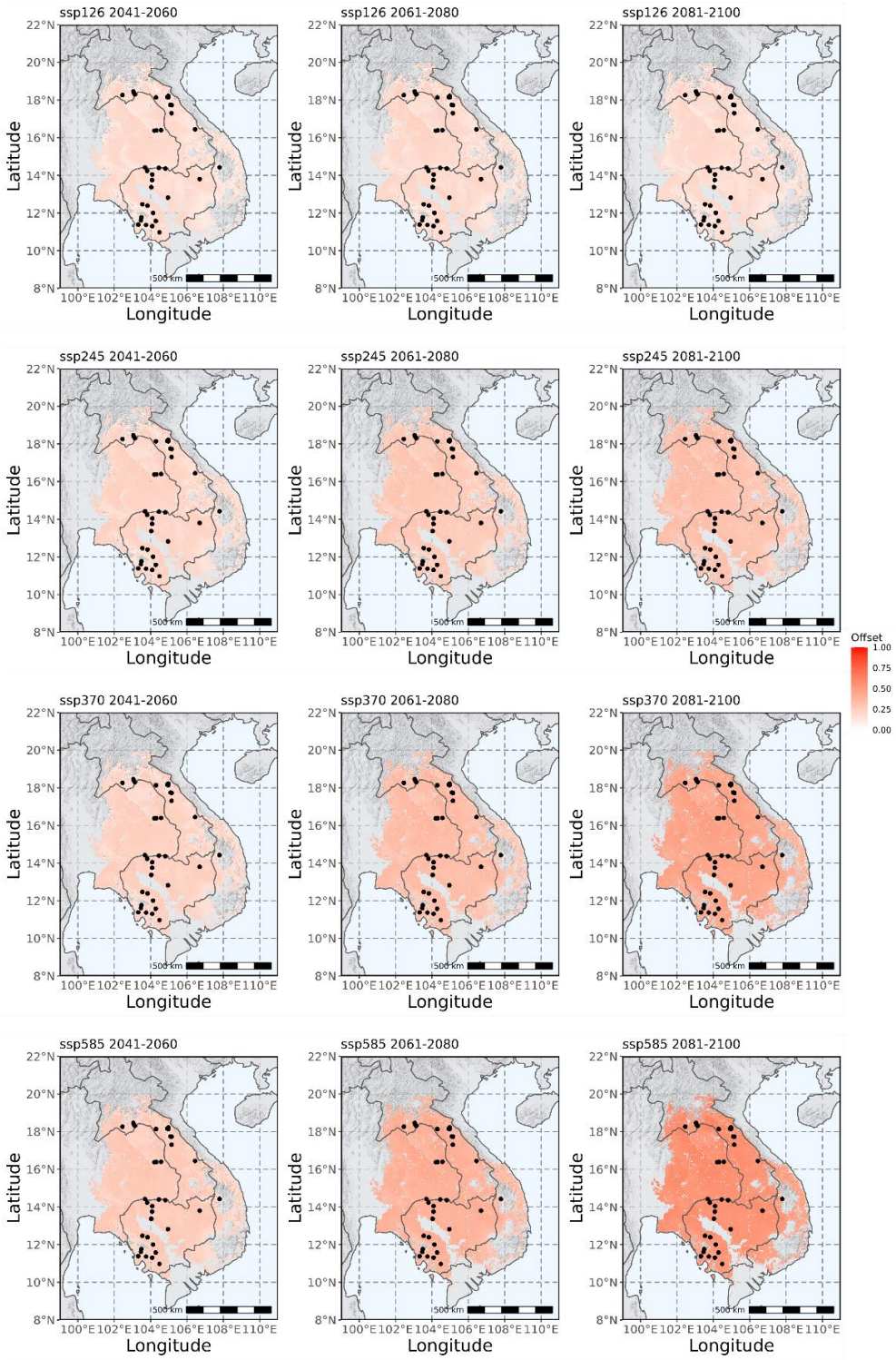


Figure 5.17. Genetic offset of gene-environmental association (scaled between 0, lowest offset, and 1, highest offset) across the range of **(a)** *D. cochinchinensis* and **(b)** *D. oliveri* in 4 SSPs (126, 245, 370, and 585) over three bidecades (2041–2060, 2061–2080, 2081–2100) averaged across five GCMs (BCC-CSM2-MR, CNRM-ESM2-1, IPSL-CM6A-LR, MIROC6, MRI-ESM2-0.

(a)


ssp126 2041-2060
ssp126 2061-2080
ssp126 2081-2100
ssp245 2041-2060
ssp245 2061-2080
ssp245 2081-2100
ssp370 2041-2060
ssp370 2061-2080
ssp370 2081-2100
ssp585 2041-2060
ssp585 2061-2080
ssp585 2081-2100
Latitude
Longitude
Offset
1.00
0.75
0.50
0.25
0.00

(b)

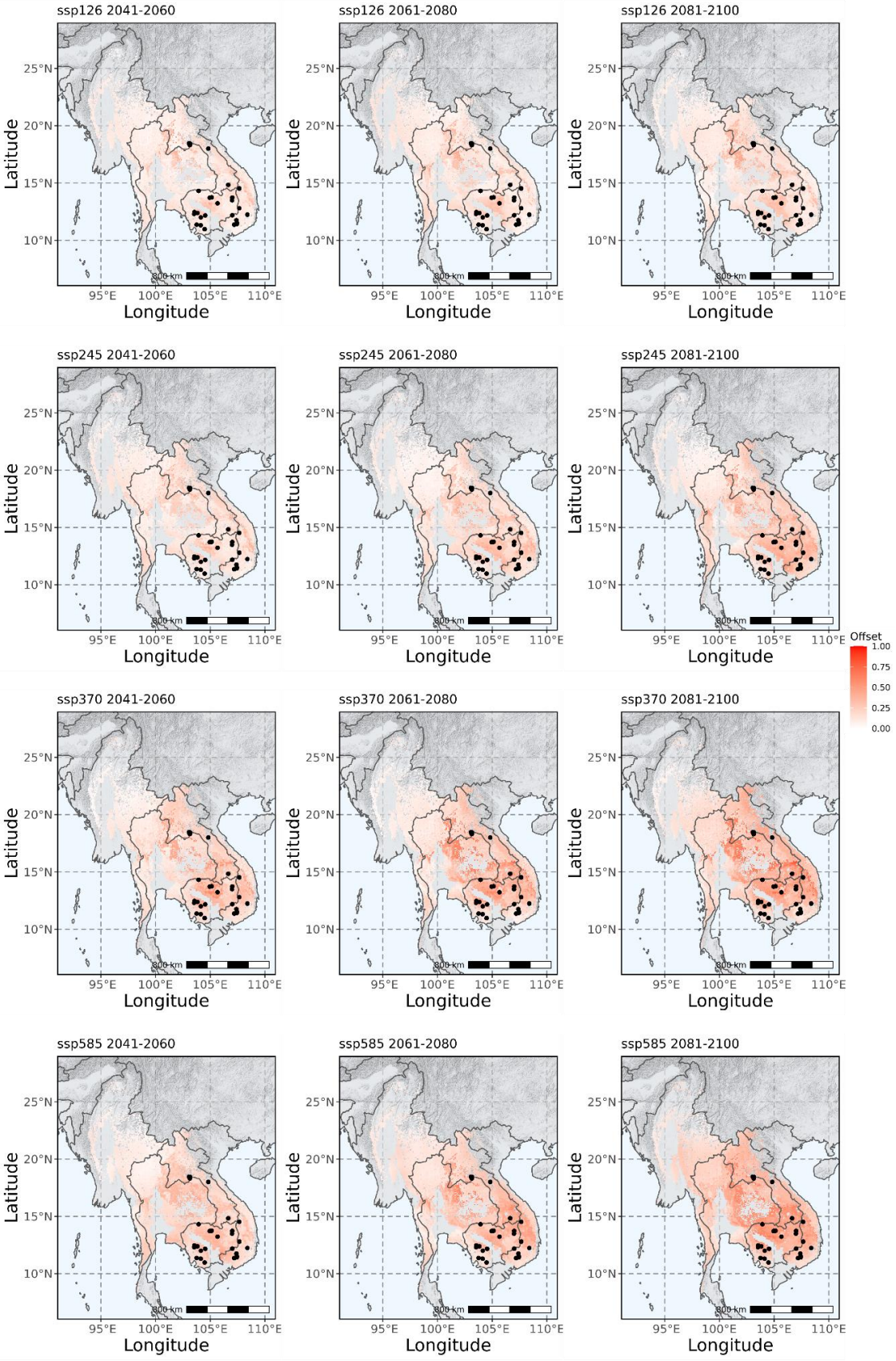
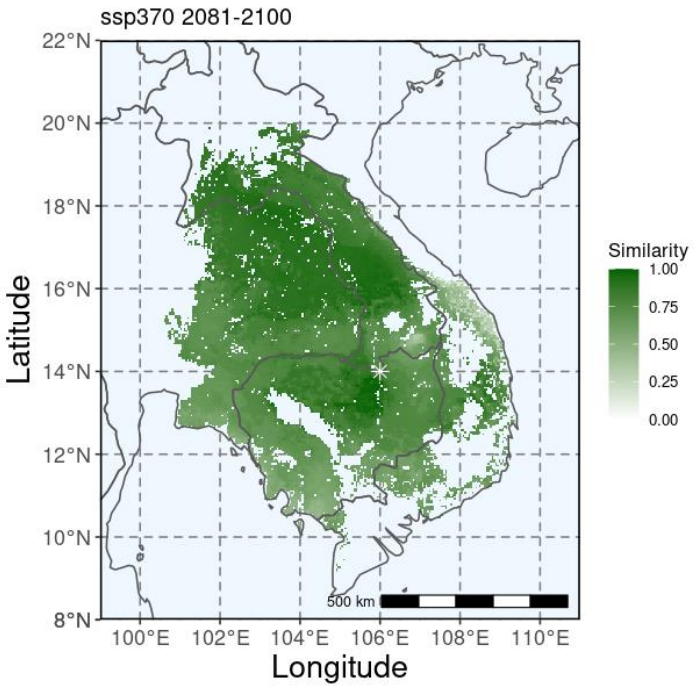

ssp126 2041-2060
ssp126 2061-2080
ssp126 2081-2100
ssp245 2041-2060
ssp245 2061-2080
ssp245 2081-2100
ssp370 2041-2060
ssp370 2061-2080
ssp370 2081-2100
ssp585 2041-2060
ssp585 2061-2080
ssp585 2081-2100
Latitude
Longitude
25°N
20°N
15°N
10°N
95°E
100°E
105°E
110°E
300 km
Offset
1.00
0.75
0.50
0.25
0.00

Figure 5.18. Genomic similarity (scaled between 0, most dissimilar, and 1, most similar) between a hypothetical future restoration site (106° N, 14° E) and the current potential germplasm sources under the future climate scenario of SSP370 between 2081 and 2100 for **(a)** *D. cochinchinensis* and **(b)** *D. oliveri* predicted on *seedeR* (https://trainingidn.shinyapps.io/seeder/).

(a)



(b)

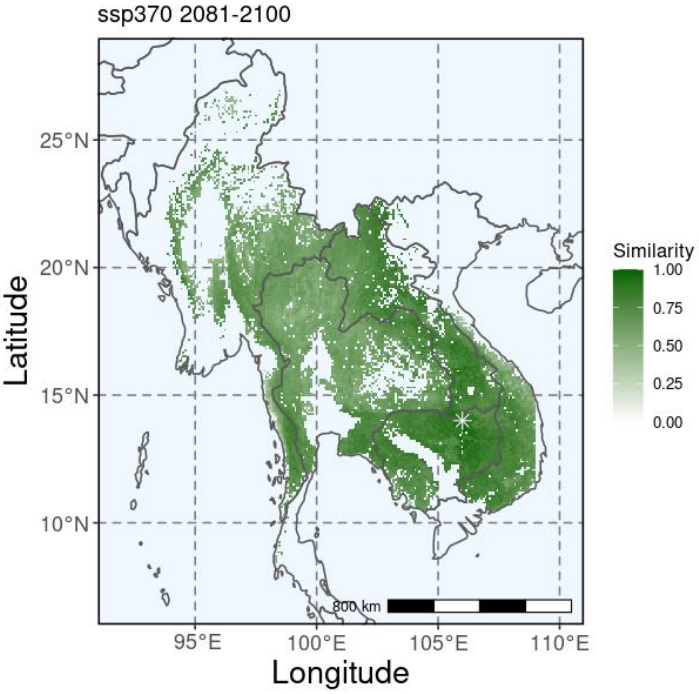

ssp370 2081-2100
25°N
20°N
15°N
10°N
Latitude
800 km
95°E
100°E
105°E
110°E
Longitude
Similarity
1.00
0.75
0.50
0.25
0.00

Table 5.1. List of environmental data used in the landscape genomics modelling of this study.

Symbol	Description	Category	Source	*D. cochinchinensis*	*D. oliveri*
wc2.1_30s_bio_1	Annual mean temperature	Temperature	WorldClim		
wc2.1_30s_bio_2	Mean diurnal range	Temperature	WorldClim	V	V
wc2.1_30s_bio_3	Isothermality	Temperature	WorldClim	V	V
wc2.1_30s_bio_4	Temperature seasonality	Temperature	WorldClim		
wc2.1_30s_bio_5	Maximum temperature of warmest month	Temperature	WorldClim		
wc2.1_30s_bio_6	Minimum temperature of coldest month	Temperature	WorldClim		
wc2.1_30s_bio_7	Temperature annual range	Temperature	WorldClim		
wc2.1_30s_bio_8	Mean temperature of wettest quarter	Temperature	WorldClim	V	
wc2.1_30s_bio_9	Mean temperature of driest quarter	Temperature	WorldClim		
wc2.1_30s_bio_10	Mean temperature of warmest quarter	Temperature	WorldClim		
wc2.1_30s_bio_11	Mean temperature of coldest quarter	Temperature	WorldClim		
wc2.1_30s_bio_12	Annual precipitation	Precipitation	WorldClim		
wc2.1_30s_bio_13	Precipitation of wettest month	Precipitation	WorldClim	V	
wc2.1_30s_bio_14	Precipitation of driest month	Precipitation	WorldClim		V
wc2.1_30s_bio_15	Precipitation seasonality	Precipitation	WorldClim		

wc2.1_30s_bio_16	Precipitation of wettest quarter	Precipitation	WorldClim		V
wc2.1_30s_bio_17	Precipitation of driest quarter	Precipitation	WorldClim	V	
wc2.1_30s_bio_18	Precipitation of warmest quarter	Precipitation	WorldClim		
wc2.1_30s_bio_19	Precipitation of coldest quarter	Precipitation	WorldClim	V	
wc2.1_30s_bio_elev	Digital elevation	Elevation	WorldClim		
s_AwCh1_sl2	Available soil water capacity (volumetric fraction) with FC = pF 2.0 at depth of 5 cm	Soil	SoilGrids	V	V
s_AwCh1_sl5	Available soil water capacity at depth of 60 cm	Soil	SoilGrids	V	V
s_CLYPPT_sl2	Clay content in percent at depth of 5 cm	Soil	SoilGrids	V	
s_CLYPPT_sl5	Clay content in percent at depth of 60 cm	Soil	SoilGrids		
s_ORCDRC_sl2	Soil organic carbon stock at depth of 5 cm	Soil	SoilGrids		V
s_ORCDRC_sl5	Soil organic carbon stock at depth of 60 cm	Soil	SoilGrids	V	
s_PHIHOX_sl2	Soil pH measured in water solution at depth of 5 cm	Soil	SoilGrids		V
s_PHIHOX_sl5	Soil pH measured in water solution at depth of 60 cm	Soil	SoilGrids	V	
s_SLTPPT_sl2	Silt content in percent at depth of 5 cm	Soil	SoilGrids	V	

s_SLTPPT_sl5	Silt content in percent at depth of 60 cm	Soil	SoilGrids		V
s_SNDPPT_sl2	Sand content in percent at depth of 5 cm	Soil	SoilGrids		
s_SNDPPT_sl5	Sand content in percent at depth of 60 cm	Soil	SoilGrids		V
npp	Net primary productivity	Vegetation	The Atlas of the Biosphere		V
evapotrans	Evapotranspiration	Vegetation	The Atlas of the Biosphere	V	V

Table 5.2. Assembly statistics of the *D. oliveri* genome (Daoli_0.2).

Assembly identifier	Daoli_0.2
Species	*Dalbergia oliveri* Gamble ex Prain
NCBI taxonomy ID	1030150

	Raw reads	***Pore-C reads***	***Contigs***	***Primary haplotigs***	***Scaffolds***
Total size (Mbp)	15,127.06	13,462.40	814.69	687.92	689.25
Number of sequences	959,359	3,098,972	3,249	2,239	2,977
Average length of sequences	15,767.9	4,344.2	250,750	307,244.3	231,525
N50	44,832	5,789	474,015	623,924	38,426,337

Table 5.3. Repeat elements in the repeat library of Daoli_0.2.

	Number of elements	Total length (bp)	Percentage of sequence
Retroelements	455479	339092238	49.2
SINEs	0	0	0
Penelope	0	0	0
LINEs	10758	4445290	0.64
CRE/SLACS	0	0	0
L2/CR1/Rex	0	0	0
R1/LOA/Jockey	0	0	0
R2/R4/NeSL	0	0	0
RTE/Bov-B	3783	973287	0.14
L1/CIN4	6975	3472003	0.5
LTR elements	444721	334646948	48.55
BEL/Pao	0	0	0
Ty1/Copia	229457	108551031	15.75
Gypsy/DIRS1	200968	220263371	31.96
Retroviral	181	91342	0.01
DNA transposons	16939	12338687	1.89
hobo-Activator	2978	1273955	0.18
Tc1-IS630-Pogo	1440	179489	0.03
En-Spm	0	0	0
MuDR-IS905	0	0	0
PiggyBac	0	0	0
Tourist/Harbinger	2277	1185567	0.17
Other (Mirage, P-element, Transib)	0	0	0
Rolling-circles	5146	3333791	0.48
Unclassified	376428	91174602	13.23
Total interspersed repeats	1757530	442605527	64.22

Simple repeats	143462	5735706	0.83
Low complexity	25174	1218960	0.18

Table 5.4. BUSCO statistics of gene models of Daoli_0.2 using eudicots_odb10 as the reference dataset (*n* = 2,326).

Complete BUSCOs	2,054 (88.3%)
Complete and single-copy BUSCOs	1,878 (80.7%)
Complete and duplicated BUSCOs	176 (7.6%)
Fragmented BUSCOs	107 (4.6%)
Missing BUSCOs	165 (7.1%)

Table 5.5. Statistics of SNPs for *D. cochinchinensis* and *D. oliveri* at each bioinformatic step.

	D. cochinchinensis	***D. oliveri***
# Samples	435	331
# Localities	35	28
# SNPs after variant calling	1,832,629	3,377,855
# SNPs after filtering for missing data and MAF	246,225	250,954
# SNPs after pruning for linkage disequilibrium	180,944	193,724
# Adaptive SNPs predicted by sNMF	50,951	54,154
# Adaptive SNPs predicted by LFMM	20,373	6,953

Table 5.6. Genomic inflation factor λ for the gene-environment associations of **(a)** *D. cochinchinensis* and **(b)** *D. oliveri*.

env	**lambda**
evapotrans	0.071053
s_AWCh1_sl2	0.1012
s_AWCh1_sl5	0.097883
s_CLYPPT_sl2	0.0773
s_ORCDRC_sl5	0.103751
s_PHIHOX_sl5	0.11886
s_SLTPPT_sl2	0.117395
wc2.1_30s_bio_13	0.166612
wc2.1_30s_bio_17	0.247199
wc2.1_30s_bio_19	0.136046
wc2.1_30s_bio_2	0.186022
wc2.1_30s_bio_3	0.144369
wc2.1_30s_bio_8	0.088284
mean	**0.127383**
sd	**0.049323**

env	**lambda**
evapotrans	0.038403
npp	0.046587
s_AWCh1_sl2	0.037444
s_AWCh1_sl5	0.057747
s_ORCDRC_sl2	0.05131
s_PHIHOX_sl2	0.05708
s_SLTPPT_sl5	0.050496
s_SNDPPT_sl5	0.032456
wc2.1_30s_bio_14	0.070037
wc2.1_30s_bio_16	0.079196
wc2.1_30s_bio_2	0.080899
wc2.1_30s_bio_3	0.069888
mean	**0.055962**
sd	**0.016241**

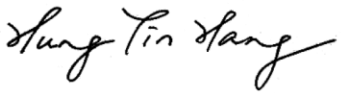
Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Landscape genomics of two threatened rosewood species *Dalbergia cochinchinensis* and *D. oliveri* in the Greater Mekong Subregion
Publication Status	□Published □ Accepted for Publication □Submitted for Publication ☑Unpublished and unsubmitted work written in a manuscript style
Publication Details	Hung, T. H.*, So, T., Thammavong, B., Train, H. T., Chamchumroon, V., Theilade, I., Phourin, C., Bouamanivong, S., Hartvig, I., Gaibserger, H., Jalonen, R., Boshier, D. H. & MacKay, J. J.* (unpublished). Landscape genomics of two threatened rosewood species *Dalbergia cochinchinensis* and *D. oliveri* in the Greater Mekong Subregion

Student Confirmation

Student Name:	Tin Hang Hung	
Contribution to the Paper	designed the study, coordinated the sample collection, conducted the sample preparation, conceived and conducted the bioinformatic analyses, drafted the manuscript	
Signature 	Date	8 MAR 2022

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: John MacKay, Wood Professor of Forest Science		
Supervisor comments		
Signature 	Date	8 MAR 2022

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 6. General Discussion

Conservation of *Dalbergia cochinchinensis* and *D. oliveri* depend on the sustainable multiplication of genetic materials of genetic and adaptive diversity, in order to reduce the existing pressure on the natural populations and increase their number. This thesis produces genomic resources and an evidence base of adaptability that will help narrow the existing knowledge gap and enable real-world conservation applications.

In this chapter, we will discuss (1) the major findings of this thesis and their implications in a broader context of adaptation in tropical forest trees, (2) the pathways to impact and potential applications of the findings in their conservation, and (3) future research directions that will complement with this thesis.

6.1 Adaptation: linking gene, environment, and physiology

Species range is dependent on the environmental heterogeneity and its shift is thus impacted by the ongoing environmental change. To avoid extinction or extirpation, species need to either adapt *in situ* or colonise newly suitable habitat[1]. Therefore, understanding of the adaptability and its natural-occurring variation remains a central goal of conservation biology studies. This thesis presented the first genomic study of adaptation on two endangered *Dalbergia* species in the Greater Mekong Subregion, *D. cochinchinensis* and *D. oliveri*.

A key objective of the thesis was to develop genomic resources of *D. cochinchinensis* and *D. oliveri*, which was met by developing reference genomes and reference transcriptomes, along with gene models, annotations, and comparative analyses. Green plants

are hugely diverse and are estimated to comprise of ~ 500,000 species[2]; however, genomic resources available for a relatively small fraction of them to date. Biodiversity, together with associated evolutionary, ecological, developmental, morphological, and physiological processes, is driven by fundamental genomic diversity. Genomes are markers of evolutionary history and drivers of evolutionary change[3]. There has been an exponential growth in genomic resources, set with ambitious objectives such as to sequence the genome of every species on Earth under the Earth BioGenome Project[4]. However, when compared to the animal counterpart (e.g. Vertebrate Genomes Project[5]), the progress is still lagging behind for plant genomes (e.g. 10KP Plants Project[6]). The development of genomic resources for *Dalbergia* open up the opportunities of ecological and evolutionary studies in forest trees, which offer unique insights into genetic basis of longevity and adaptation to the environment throughout their sessile but long lifespan Chromosome-scale assemblies of *Dalbergia* genome and transcriptomes thus meets major objectives of forest tree genomics research, including the evolution and structure of tree genomes, monitoring adaptive gene diversity, and developing genomics-based tools[7]. *Dalbergia* genomes thus fill two diversity gaps: progress towards angiosperm tree genomes, which were largely confined to temperate trees of *Populus*, *Eucalytpus*, *Salix*, and *Quercus*[8]; and progress towards legume genomes, which only received attention for their herbaceous crop species such as soybeans, peanuts, and peas[9]. Our comparative genomics studies revealed insightful differences between the expansion of certain gene families in tropical *Dalbergia* species and the significant difference in resistance gene numbers between perennial *Dalbergia* species and annual species.

One important dimension of adaptation is the interaction between genes and the bio-physical environment. Both *D. cochinchinensis* and *D. oliveri* co-occur in a highly heterogeneous environment and thus it is hypothesised that they exhibit substantial amounts of adaptive variation and have patterns of local adaptation. While there is virtually no *a priori*

knowledge on their adaptability, the landscape genomic study reported in this thesis provides a comprehensive scan of adaptive signals and models the range-wide gene-environment association. We found significant differences between the species when comparing the environmental determinants of their adaptive variation, where *D. cochinchinensis* is driven by temperature variability and *D. oliveri* by precipitation variability. Their contrasting patterns of adaptation may be related to their different ecological niches as a pioneering species and climax species, respectively. This leads to different predictions of patterns of genetic maladaptation in the future climate scenarios and therefore the optimal seed sourcing and restoration strategies are likely different for the two species. However, although landscape genomics is particularly attractive in studying forest trees, which harbour higher levels of genomic and phenotypic variation than their annual counterparts, there is a noticeably large gap in landscape genomic studies among both tropical and angiosperm trees. This would be the first study in the Greater Mekong Subregion to date. While many of other studies on temperate and subtropical tree species have identified temperature to be a major driver of adaptive variation instead of precipitation[10], the contrasting adaptive difference towards temperature and precipitation between *D. cochinchinensis* and *D. oliveri* may suggest the differential effects of environmental factors on tropical tree species. Numerous studies have already demonstrated the associations of tropical forest trees to a suite of environmental variables including elevation, edaphic content, temperature, and light regimes and the evidence of niche differentiation[11]. While a study on two *Dalbergia* species cannot draw a broad conclusion on tree adaptation in the tropics, this study will be the first step to fully understand pattern of adaptations in Greater Mekong Region, which harbours great environmental heterogeneity and biodiversity.

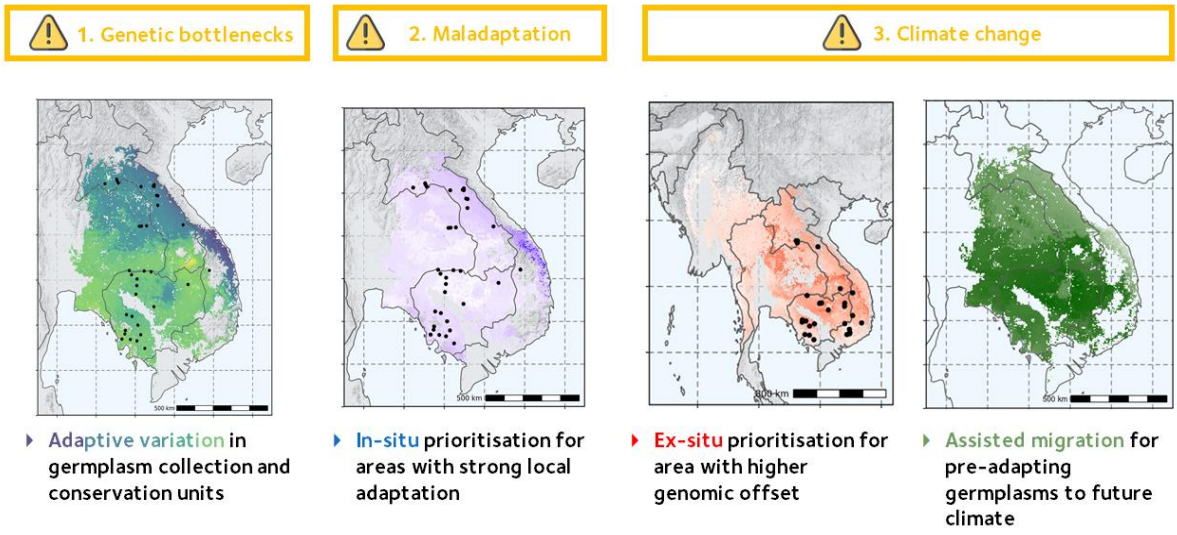
The controlled stress experiment presented in this thesis sheds light into physiological responses and provides further evidence of their contrasted adaptation pattern, where *D.*

cochinchinensis is found to be less conservative in terms of water relations during drought stress. Regulation of water relations is driven by the sustained stomatal opening and decreasing water potential. On the other hand, carbon assimilation is maximised and photoprotective pigments appear to equip *D. cochinchinensis* to tolerate drought more readily as indicated by its anisohydric behaviour. *D. cochinchinensis* is a pioneering species and thus may benefit from such a physiological adaptation to maximise productivity for more efficient colonisation in early ecological succession[12]. The trade-off between short-term growth and long-term survival is well aligned to the niche spectrum between pioneer and climax species[13], such as *D. cochinchinensis* and *D. oliveri* respectively. This experiment on stress response complements the major findings of contrasting adaptive differences between *D. cochinchinensis* and *D. oliveri* in the landscape genomic studies despite of their co-occurrence, as it provides a physiological explanation to the implied niche differentiation of gene-environment associations.

6.2 Conservation implications: enhancing germplasm diversity and adaptability

A central question of *Dalbergia* conservation is: “How do we ensure a sustainable supply of genetically and adaptively diverse germplasm for forest restoration?”. While its sustainable supply is mainly dependent on the establishment of a seed supply chain, we identified three potential genetic risks that may compromise the seed supply and restoration end-goal, namely genetic bottlenecks, maladaptation, and climate change.

Figure 6.1. Conservation implications of the knowledge of adaptation in *D. cochinchinensis* and *D. oliveri*.



The deliverable outputs of this thesis have direct implications on conservation (Figure 6.1). First, the knowledge of adaptive variation enables conservation practitioners to assess if any genetic materials are missed by the current germplasm production areas. This knowledge also allows the designation of conservation units based on the genetic uniqueness of populations. Second, resources to establish new conservation areas are limited and priority setting is important to balance cost and budget[14], therefore, *in situ* prioritisation should be given to areas with strong local adaptation as germplasm brought outside their native habitats are likely to have a reduced fitness. Third, while climate change is likely to impact the entire species range, populations in areas with higher genomic offset are more likely to undergo local extirpation and therefore should be prioritised for *ex situ* conservation. Assisted germplasm transfer will help pre-adapt germplasm to new areas and future climate.

The only genetic studies to date on *D. cochinchinensis* and *D. oliveri*, which served the current conservation practice, are based on neutral genetic markers, namely microsatellites[15,16]. Our study provided a higher resolution of population differentiation,

which can be informative in assigning new conservation areas. These, previous studies identified the main driver for population structure as the Indochinese landscape features. In contrast, our studies found that the environmental variation (isolation-by-environment) had a strong role in shaping the population structure, resulting in a different pattern of population genetic structure, although we also found that spatial factors contributed significantly to isolation-by-distance. The future designation of *in situ* and *ex situ* conservation units thus needs to take into account the environmental factors in addition to the spatial factors.

6.3 Narrowing the 'conservation genomics gap'

This thesis illustrates how genomic technologies are advantageous in studying adaptation and guiding conservation activities in *D. cochinchinensis* and *D. oliveri*, especially when our knowledge of their biology is limited. Although the benefits of genomics in conservation have been widely discussed and advocated, there are still significant barriers to translate fundamental genomic research into end-user conservation applications – this is known as 'conservation genetics gap'[17]. These barriers are hypothesised to include: extensive use of jargon in genomic research coupled with poor communication between geneticists and conservation practitioners[18], cost of sequencing and genotyping, inaccessibility to genetic knowledge, frontier research, and methodologies[19]. It has been argued that the scientific and conservation communities operate in largely separate circles, and the transition from conservation genetics to genomics is likely expanding the gap[20], where the increased number of genetic markers and more sophisticated models in conservation genomics translate to higher financial cost and a wider knowledge barrier.

It is hoped that this study will help to narrow the conservation genetics gap through a number of perspectives. First, this study benefited from the infrastructure of a region-wide

network and collaboration to obtain range-wide samples for a comprehensive analysis, within the framework of a Darwin Initiative project. Second, the main deliverable outputs, such as the prediction of seed sourcing and vulnerability, are made as user-friendly and use spatially explicit models. Third, through engaging policy makers and forest authorities in target countries, this study has identified the conservation opportunities and consolidated pathways to impact. The range-wide consorted efforts have set ambitious goals for the two *Dalbergia* species, such as ≥ 50% increase in number of *in situ* or *ex situ* conservation units and 10–25% improvement in species' survival and growth. Our genomic models enable visualisation of adaptive patterns and thus will accelerate the decision-making and map-drawing exercises to designate conservation units and favourable nurseries conditions for their growth.

Genomic prediction in conservation of threatened species is moving forward among different stakeholders in the practitioners' circle . While most genomic studies have a main focus on risk assessment, conservation management, and decision prioritisation[21], we also emphasise the importance of incorporating the aspects of value chains and sustainable production to generate incentives for conservation and restoration. In addition, critical gains from genomic analysis also pushes international parties such as the Convention on International Trade in Endangered Species (CITES) or the International Union for Conservation of Nature (IUCN) Red List to incorporate genomic tools in more rapid assessment of threatened species, as these parties continue to remain as the gold standard among the practitioners.

6.4 Two is better than one: future directions of research

The landscape genomic study revealed substantial adaptive variation in *D. cochinchinensis* and *D. oliveri*, while the physiological study identified a suite of

physiological traits proposed to contribute to the contrasting adaptive strategies between the two species. While landscape genomics is usually considered a 'bottom-up' approach where the first step is to identify the potential loci under selection without *a priori* knowledge, common garden experiments are a more quantitative 'top-down' approach to identify phenotypic variation of different provenances and populations[22]. Therefore, we see a promising avenue for a new study to combine genomic and common garden experiments to unravel the genetic and environmental control of complex traits such as morphological and physiological traits. It is hard to avoid false positives in genome scans for natural selection[23], therefore, if a strong signal of adaptation is detected in both genomic and phenotypic data from a combined common garden approach, it constitutes two pieces of evidence to test the local adaptation hypothesis[24].

Combining the two approaches is believed to better explain the complex and multigenic nature of plant adaptation[25], and could potentially delineate genotype-by-phenotype-by-environment interactions of local adaptation[26]. Although the combined approach is often advocated, it was not very commonly applied, as a previous literature survey revealed that only 11% of common garden studies incorporated a genetic component of any kind[27]. A few studies have been able to couple landscape genomics and common garden experiments to: detect linkage between phenotypic variation and outlier analysis[28], study local adaptation at fine spatial scales[29], and examine heritability of the associated traits[30].

At the same time, common garden experiments provide an opportunity to study associations between phenotypes and genotypes, which are impractical in their natural habitats as the phenotypic traits can be confounded with environmental effects. Common

gardens thus overcome the challenge and help isolate the genotypic effect on phenotypes by providing uniform environmental conditions for all genotypes[31].

Storfer et al.[32] distinguished four research frameworks to understand the relationship between genetic adaptability and environment: (a) correlative analyses based on outlier test and/or EAA; (b) phenotypic analyses on QTL or GWAS; (c) candidate gene analyses; and (d) exome or transcriptome analyses. Combining landscape genomics and common garden experiments may support all of these frameworks and increase the statistical power to link adaptive genetics and environment.

While regional conservation efforts are increasing the number of *in situ* and *ex situ* conservation units and the germplasm production capacity in *Dalbergia*, there is an expanding opportunity to obtain genetic materials and construct range-wide common garden experiments. We are planning to study the effects of drought and heat stresses on different ecotypes of *D. cochinchinensis* and *D. oliveri*. The objectives of this planned study is three-fold: (1) to detect their local adaptation towards water availability and temperature; (2) to predict their vulnerability and recovery towards drought and heat stresses under extreme climate; (3) to dissect the genetic underpinning of the phenotypic variation and adaptation towards water availability and temperature.

6.5 References

1. Christmas, M. J., Breed, M. F. & Lowe, A. J. Constraints to and conservation implications for climate change adaptation in plants. *Conserv. Genet. 2015 172* **17**, 305–320 (2015).
2. Lughadha, E. N. *et al.* Counting counts: revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. *Phytotaxa* **272**, 82–88–82–88 (2016).
3. Soltis, P. S. & Soltis, D. E. Plant genomes: Markers of evolutionary history and drivers of evolutionary change. *Plants, People, Planet* **3**, 74–82 (2021).
4. Lewin, H. A. *et al.* Earth BioGenome Project: Sequencing life for the future of life. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 4325–4333 (2018).
5. Rhie, A. *et al.* Towards complete and error-free genome assemblies of all vertebrate species. *Nat. 2021 5927856* **592**, 737–746 (2021).
6. Leebens-Mack, J. H. *et al.* One thousand plant transcriptomes and the phylogenomics of green plants. *Nat. 2019 5747780* **574**, 679–685 (2019).
7. Neale, D. B. & Kremer, A. Forest tree genomics: growing resources and applications. *Nat. Publ. Gr.* **12**, (2011).
8. Tuskan, G. A. *et al.* Hardwood tree genomics: Unlocking woody plant biology. *Front. Plant Sci.* **871**, 1799 (2018).
9. Schaefer, H. *et al.* Systematics, biogeography, and character evolution of the legume tribe Fabeae with special focus on the middle-Atlantic island lineages. *BMC Evol. Biol.* **12**, 1–19 (2012).
10. Martins, K. *et al.* Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of *Quercus rugosa*. *Evol. Appl.* **11**, 1842–1858 (2018).

11. Brown, C. *et al.* Multispecies coexistence of trees in tropical forests: spatial signals of topographic niche differentiation increase with environmental heterogeneity. *Proc. R. Soc. B Biol. Sci.* **280**, (2013).

12. So, N. V. The potential of local tree species to accelerate natural forest succession on marginal grasslands in Southern Vietnam. in *Proceedings of the Workshop on Forst Restoration for Wildlife Restoration for Wildlife Conservation* (eds. Elliot, S., Kerby, J. & Blakesly, K. et al.) (International Tropical Timber Organization and The Forest Restoration Unit, Chiang Mai University).

13. L, P. & F, B. Leaf traits are good predictors of plant performance across 53 rain forest species. *Ecology* **87**, (2006).

14. Armsworth, P. R. *et al.* Factoring economic costs into conservation planning may not improve agreement over priorities for protection. *Nat. Commun. 2017 81* **8**, 1–10 (2017).

15. Hartvig, I. *et al.* Population genetic structure of the endemic rosewoods *Dalbergia cochinchinensis* and *D. oliveri* at a regional scale reflects the Indochinese landscape and life-history traits. *Ecol. Evol.* **8**, 530–545 (2018).

16. Hartvig, I. *et al.* Conservation genetics of the critically endangered Siamese rosewood (*Dalbergia cochinchinensis*): recommendations for management and sustainable use. *Conserv. Genet.* 1–16 (2020) doi:10.1007/s10592-020-01279-1.

17. R. Taylor, H., Dussex, N. & van Heezik, Y. Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Glob. Ecol. Conserv.* **10**, 231–242 (2017).

18. Hoban, S. *et al.* Conservation Genetic Resources for Effective Species Survival (ConGRESS): Bridging the divide between conservation research and practice. *J. Nat. Conserv.* **21**, 433–437 (2013).

19. Cook, C. N. & Sgrò, C. M. Aligning science and policy to achieve evolutionarily enlightened conservation. *Conserv. Biol.* **31**, 501–512 (2017).

20. Shafer, A. B. A. *et al.* Genomics and the challenging translation into conservation practice. *Trends Ecol. Evol.* **30**, 78–87 (2015).

21. Supple, M. A. *et al.* Landscape genomic prediction for restoration of a *Eucalyptus* foundation species under climate change. *Elife* **7**, e31835 (2018).

22. Sork, V. L. *et al.* Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genet. Genomes* **9**, 901–911 (2013).

23. Mallick, S., Gnerre, S., Muller, P. & Reich, D. The difficulty of avoiding false positives in genome scans for natural selection. *Genome Res.* **19**, 922 (2009).

24. Holderegger, R. *et al.* Land ahead: using genome scans to identify molecular markers of adaptive relevance. *https://doi.org/10.1080/17550870802338420* **1**, 273–283 (2010).

25. Bourret, V., Dionne, M. & Bernatchez, L. Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: polygenic multilocus analysis surpasses genome scan. *Mol. Ecol.* **23**, 4444–4457 (2014).

26. Lepais, O. & Bacles, C. F. E. Two are better than one: combining landscape genomics and common gardens for detecting local adaptation in forest trees. *Mol. Ecol.* **23**, 4671–4673 (2014).

27. Johnson, L. C. *et al.* Reciprocal transplant gardens as gold standard to detect local adaptation in grassland species: New opportunities moving into the 21st century. *J. Ecol.* **110**, 1054–1071 (2022).

28. De Kort, H. *et al.* Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Mol.*

Ecol. **23**, 4709–4721 (2014).

29. Eckert, A. J. *et al.* Local adaptation at fine spatial scales: an example from sugar pine (*Pinus lambertiana*, Pinaceae). *Tree Genet. Genomes* **11**, 42 (2015).

30. McKown, A. D. *et al.* Geographical and environmental gradients shape phenotypic trait variation and genetic structure in *Populus trichocarpa*. *New Phytol.* **201**, 1263–1276 (2014).

31. Villemereuil, P. de *et al.* Common garden experiments in the genomic era: new perspectives and opportunities. *Heredity (Edinb).* **116**, (2016).

32. Balkenhol, N., Cushman, S., Storfer, A. & Waits, L. *Landscape Genetics: Concepts, Methods, Applications*. (John Wiley & Sons, 2016).