UNIVERSITY OF COPENHAGEN

Københavns Universitet

Range-wide differential adaptation and genomic offset in critically endangered Asian rosewoods

Hung, Tin Hang; So, Thea; Thammavong, Bansa; Camchumroon, Voradol ; Theilade, Ida; Phourin, Chhang; Bouamanivong, Somsanith ; Hartvig, Ida; Gaisberger, Hannes; Jalonen, Riina; Boshier, David H.; MacKay, John J.

Published in:
PNAS

DOI:
10.1073/pnas.2301603120

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Hung, T. H., So, T., Thammavong, B., Camchumroon, V., Theilade, I., Phourin, C., Bouamanivong, S., Hartvig, I., Gaisberger, H., Jalonen, R., Boshier, D. H., & MacKay, J. J. (2023). Range-wide differential adaptation and genomic offset in critically endangered Asian rosewoods. *PNAS*, *120*(33), Article e2301603120. https://doi.org/10.1073/pnas.2301603120

Download date: 25. Sep. 2026

PNAS RESEARCH ARTICLE | ECOLOGY ENVIRONMENTAL SCIENCES OPEN ACCESS

Range-wide differential adaptation and genomic offset in critically endangered Asian rosewoods

Tin Hang Hung[a,1], Thea So[b], Bansa Thammavong[c], Voradol Chamchumroon[d], Ida Theilade[e], Chhang Phourin[b], Somsanith Bouamanivong[f], Ida Hartvig[g,h], Hannes Gaisberger[i,j], Riina Jalonen[k], David H. Boshier[a], and John J. MacKay[a,1]

Edited by Donald Ort, University of Illinois at Urbana Champaign, Urbana, IL; received January 29, 2023; accepted June 15, 2023

In the billion-dollar global illegal wildlife trade, rosewoods have been the world's most trafficked wild product since 2005. *Dalbergia cochinchinensis* and *Dalbergia oliveri* are the most sought-after rosewoods in the Greater Mekong Subregion. They are exposed to significant genetic risks and the lack of knowledge on their adaptability limits the effectiveness of conservation efforts. Here, we present genome assemblies and range-wide genomic scans of adaptive variation, together with predictions of genomic offset to climate change. Adaptive genomic variation was differentially associated with temperature and precipitation-related variables between the species, although their natural ranges overlap. The findings are consistent with differences in pioneering ability and in drought tolerance. We predict their genomic offsets will increase over time and with increasing carbon emission pathway but at a faster pace in *D. cochinchinensis* than in *D. oliveri*. These results and the distinct gene–environment association in the eastern coastal edge of Vietnam suggest species-specific conservation actions: germplasm representation across the range in *D. cochinchinensis* and focused on hotspots of genomic offset in *D. oliveri*. We translated our genomic models into a seed source matching application, *seedeR*, to rapidly inform restoration efforts. Our ecological genomic research uncovering contrasting selection forces acting in sympatric rosewoods is of relevance to conserving tropical trees globally and combating risks from climate change.

rosewood | ecological genomics | climate vulnerability | adaptation | conservation

Rosewoods have been the world's most trafficked wild product since 2005, amounting to 30 to 40% of the global illegal wildlife trade (1), which is estimated at 7 to 23 billion USD annually (2). *Dalbergia cochinchinensis* Pierre and *Dalbergia oliveri* Gamble ex Prain are among the most sought-after and threatened rosewood species. Exploited for their extremely valuable timber (3), alongside many other valued and threatened tree species in Asia's tropical and subtropical forests (4), the growing demand and limited supply have driven prices as high as 50,000 USD per cubic metre (5). Both these *Dalbergia* species were classified as Vulnerable and Endangered in the 1998 IUCN Red List (6, 7). The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) has listed the entire *Dalbergia* genus in its Appendix II since 2017 to reduce sequential exploitation of other closely related species (8). In the IUCN's latest reassessment of their endangered status to Critically Endangered in 2022 (9, 10), it is suspected that the populations of both species have already experienced a decline of at least 80% over the last three generations, and the decline is likely to continue (11).

D. cochinchinensis and *D. oliveri* are sympatric species, endemic to the Greater Mekong Subregion (GMS) in Southeast Asia, an area of high ecological and conservation concern as 84% of the GMS overlaps with the Indo-Burmese mega biodiversity hotspot (12). The complex biogeographical and geological histories of the GMS have contributed to its high species richness, heterogeneous landscapes, and high endemism levels (13). Ancient changes in the distribution of terrestrial and water bodies have been associated with changes in vegetation types and cover (14). These forests contribute substantially to local livelihoods, economies, food security, and human health (15, 16), though overexploitation undermines their potentially central role to nature-based solutions and most of them are unprotected (2).

Species- and environment-specific conservation approaches represent an immediate need in response to declining populations (4). Conservation, collection, and use of genetically diverse germplasm are key to conserving diversity and restoring these rosewood populations. Genetic conservation actions were started in the early 2000s but were limited in scale, usually including fewer than 50 seed-producing trees per country (17–19). Newer capacity-building initiatives targeting tree nurseries and seed value chain development (20) may still carry genetic risks associated with the supply and use of germplasm, and

Significance

In the billion-dollar global illegal wildlife trade, rosewoods have been the world's most trafficked wild product since 2005, with *Dalbergia cochinchinensis* and *Dalbergia oliveri* being the most sought-after and endangered species in Southeast Asia. Emerging efforts for their restoration have lacked a suitable evidence base on adaptability and adaptive potential. We integrated range-wide genomic data and climate models to detect the differential adaptation between *D. cochinchinensis* and *D. oliveri* in relevance to temperature- and precipitation-related variables and projected their genomic offset until 2100. We highlighted the stronger local adaptation in the coastal edge of the species ranges suggesting conservation priority. We developed genomic resources including chromosome-level genome assemblies and a web-based application *seedeR* for genomic model-enabled assisted migration and restoration.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

Although PNAS asks authors to adhere to United Nations naming conventions for maps (https://www.un.org/geospatial/mapsgeo), our policy is to publish maps as provided by the authors.

[1]To whom correspondence may be addressed. Email: tin-hang.hung@biology.ox.ac.uk or john.mackay@biology.ox.ac.uk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2301603120/-/DCSupplemental.

Published August 7, 2023.

Downloaded from https://www.pnas.org by COPENHAGEN UNIVERSITY LIBRARY on August 8, 2023 from IP address 130.226.236.6.

may compound the effects of overexploitation. First, underrepresented genetic diversity during the sourcing of genetic materials can create a genetic bottleneck for the species and reduce the species' ability to adapt and evolve in a changing climate (21). Second, mismatch of habitat suitability can result in maladaptation, if populations have strong local adaptation (22). Third, climate change will likely impose new forces of selection on the current genetic diversity, thus reducing the species' adaptability, affecting population functioning (23, 24), and leading to increased risk of local extirpations and species' range collapse (25). If unaddressed, these risks will reduce both short- and long-term effectiveness of restoration projects. The genetic risks call for an understanding of adaptation and its genetic basis in *Dalbergia* species in the GMS to safeguard on-going conservation and restoration efforts. *Dalbergia* are high-value species that could be used sustainably and generate income for farmers in developing countries if well-adapted planting material is available (4). Planting for economic purposes and reducing risks to remaining natural populations of these species seem necessary, where ecological restoration alone is insufficient.

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 listed animals (26). There is a critical lack of genomic resources in threatened species and a disproportionate representation across taxa, in contrast with the rapid growth in genomic technologies. New reference genomes in threatened species will enable the analysis, of functional genes, higher resolution studies of species delineation, association mapping and adaptation, genetic rescue, and genome editing (27). These in turn will help to address important conservation and restoration questions such as genetic monitoring of introduced and relocated populations, predicting population viability, disease resistance, synthetic alternatives, and deextinction (28, 29).

This paper develops an unprecedented understanding of adaptation in critically endangered rosewoods, which integrates genomic analyses and a resource base to inform and expand ongoing conservation efforts. 1) We present genome assemblies of *D. cochinchinensis* and *D. oliveri* at the chromosomal and near-chromosomal scale, respectively. 2) We analyze range-wide patterns of adaptation by genotyping ~800 trees, and identify differential drivers of adaptive genetic diversity between the two species by using gene-by-environment association analyses. 3) We predict the potential maladaptation of populations by assessing the genomic offset in future climate scenarios. 4) We deploy an interactive application to predict optimal seed sources, based on our landscape genomic results, in *D. cochinchinensis* and *D. oliveri* for use in restoration under future climate scenarios. Our ecological genomic study in the GMS fills crucial knowledge gaps for genomic adaptation in tropical tree species which are highly underrepresented in the current research literature.

Chromosome-Scale Genome Characterization

The *D. cochinchinensis* reference genome assembly (Dacoc_1.4) was 621 Mbp in size comprised of 10 pseudochromosomes (Fig. 1*A* and *SI Appendix*, Fig. S1 and Table S1). Whole-genome sequencing of a single seedling of *D. cochinchinensis* produced 165 Gbp (~260 X) long-read data. A diploid-aware draft assembly of 1.3 Gbp with 6,443 contigs and a N50 of 1.35 Mbp was first obtained, with the longest contig between 33.2 Mb at chromosome-arm length. We purged the haplotig and scaffolded the draft genome with 54.97 Gbp (~88.52X) Hi-C chromosome conformation capture reads into 511 scaffolds with a N50 of 60.0 Mb (*SI Appendix*, Table S2).

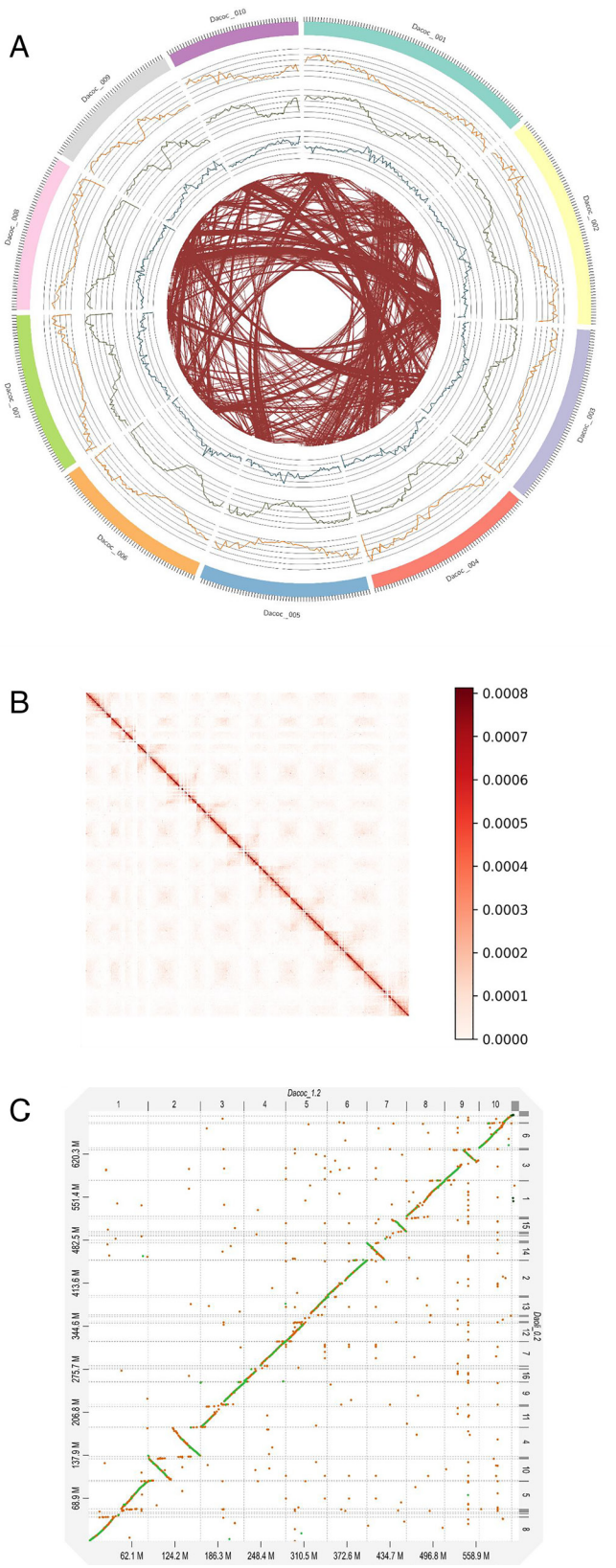

Fig. 1. (*A*) Genomic landscape of the 10 assembled pseudochromosomes of *D. cochinchinensis* (Dacoc_1.4), showing tick marks every 1 Mb, gene density (orange), repeat density (green), 5-mC density (blue), and interchromosomal syntenic arrangement (brown). The densities are calculated in 1-Mb sliding window. (*B*) High-resolution contact probability map of the final *D. cochinchinensis* genome assembly after scaffolding, revealing the 10 pseudochromosomes at 100-Kbp resolution. (*C*) Syntenic dot plot of assemblies of *D. oliveri* (Daoli_0.3) against *D. cochinchinensis* with a minimum identity of 0.25.

Downloaded from https://www.pnas.org by COPENHAGEN UNIVERSITY LIBRARY on August 8, 2023 from IP address 130.226.236.6.

The 10 longest scaffolds were considered pseudochromosomes, and 98.3% of the contigs were mapped onto them (Fig. 1*B*).

The *D. oliveri* draft genome assembly (Daoli_0.3) was 689.25 Mbp in size (*SI Appendix*, Fig. S1 and Table S3). Whole-genome sequencing of a single seedling of *D. oliveri* produced 15.13 Gbp (~22X) long-read data. We first obtained a diploid-aware draft assembly of 814.69 Mbp with 3,249 contigs and a N50 of 474.02 Kbp. We purged the haplotig and scaffolded the draft genome with 13.46 Gbp (~20X) Pore-C multicontact chromosome confirmation capture reads into 2,977 scaffolds with a N50 of 38.43 Mbp. Syntenic analysis of the *D. oliveri* assembly (Daoli_0.3) against the 10 pseudochromosomes obtained in *D. cochinchinensis* (Dacoc_1.4) showed that the 16 largest scaffolds in Daoli_0.3 had 1-to-1 or 2-to-1 correspondences to Dacoc_1.4, implying that Daoli_0.3 was at chromosome-arm length (Fig. 1*C*).

We constructed de novo repeat libraries of Dacoc_1.4 and Daoli_0.3, which contained 402 Mbp and 453 Mbp of repeat elements, respectively (64.80% and 65.71% of the genomes) (*SI Appendix*, Tables S4 and S5), the majority of which were annotated as containing LTR elements (46.63% and 48.55%) such as Ty1/Copia (15.25% and 15.75%) and Gypsy/DIRS1 (30.51% and 31.96%). The repeat content of the two genomes was significantly higher than the average among Fabids (~49%), which may be due to the near double amount of LTRs (~22%) (30).

We predicted and annotated 27,852 and 33,558 gene models in Dacoc_1.4 and Daoli_0.3 respectively, using previous RNA sequencing data (*SI Appendix*, Table S6) and protein homology of *Arabidopsis thaliana* and *Arachis ipaensis*. The gene models had a mean length of 4,284.20 and 3942.71 bp, respectively, of which 98.3% and 95.5% had an AED (annotation edit distance) score less than 0.5, considered as strong confidence (*SI Appendix*, Fig. S2). The gene models had a BUSCO v5.1.2 completeness of 96.2% and 88.3% using the eudicots_odb10 reference dataset, with 92.1% and 86.7% being both complete and single copy.

Range-Wide Genomic Scan for Adaptive Signals

We obtained initial pools of 1,832,629 and 3,377,855 SNPs from genotyping 435 and 331 individuals of *D. cochinchinensis* and *D. oliveri,* respectively, across their natural ranges (*SI Appendix*, Table S7), and final pools of 180,944 and 193,724 SNPs after filtering for missing data, minimum allele frequency (MAF), and linkage disequilibrium. The samples represented previous sampling work (31, 32) and new sampling that covered all known existing populations.

We employed the sparse nonnegative matrix factorization (sNMF) algorithm to determine the optimal number of ancestral populations (K) for *D. cochinchinensis* and *D. oliveri* as 13 and 14, respectively (*SI Appendix*, Figs. S3–S5). These results were much higher than the previous estimation of K = 5–9 for the same species using nine microsatellite markers and 19 SNPs (31, 32). The analysis revealed a highly resolved hierarchical genetic structure for both species and distinct population clusters around the Cardamon Mountains in southwest Cambodia and in northern Laos. Our calculation gave a larger genomic inflation factor (λ) in *D. cochinchinensis* (range from 0.071 (evapotranspiration) to 0.25 (precipitation of driest quarter), mean of 0.13, SD of 0.049) than that in *D. oliveri* [range from 0.038 (evapotranspiration) to 0.081 (mean diurnal range)], mean of 0.056, SD of 0.016 (*SI Appendix*, Table S8).

The numbers of SNPs found to be adaptive for at least one of the environmental variables were 20,373 (11.3%) and 6,953 (3.59%) in *D. cochinchinensis* and *D. oliveri* respectively ($|Z\text{-value}| > 2$ & $Q\text{-value} < 0.01$), after correcting for population structure (optimal K) and genomic inflation (*SI Appendix*, Figs. S6 and S7 and Dataset S1). Relatively few SNPs were associated with all or many environmental variables; 4 SNPs were associated with 11 out of 13 variables tested in *D. cochinchinensis*, and 46 SNPs were associated with all 12 variables in *D. oliveri*. These findings revealed the complex and polygenic nature of environmental adaptation, where multiple forces of natural selection can act together via different environmental cues and affect overlapping loci.

In *D. cochinchinensis*, "precipitation in the driest quarter" was the environmental variable (wc2.1_30s_bio_17) that had the strongest gene–environmental association with a SNP on chromosome 3 at position 36,345,659 (LFMM Z = 6.07237, Q = 4.77e-29). The SNP was located within the gene Dacoc08834, a homologue of the Ubiquitin-like specific protease 1B AtULP1B in *A. thaliana*. In particular, an allele of this SNP were found to cluster in the southwest of Cambodia and the northeast of Cambodia, both with the highest precipitation of the driest quarter (Fig. 2). Very little has been studied on AtULP1B, except that it is part of a highly transposon-mediated expanded family of ubiquitin-like specific proteases (ULP) (33). ULPs are responsible in mediating the maturation and deconjugation of a small ubiquitin-like modifier (SUMO) from target proteins as part of posttranslational modification (34), which regulates stress responses including to drought, heat, salinity, and pathogens (35–37) and timing of flower initiation (38). In an analysis of transcriptomes from 6 *Dalbergia* species, ubiquitin-related proteins were found to be overrepresented compared to other legumes (26). These observations warrant further studies into the function and potential divergence of ubiquitin-related proteins in *Dalbergia* adaptation to water availability.

By contrast, the strongest association in *D. oliveri* was between precipitation of the wettest quarter (wc2.1_30s_bio_16) and a SNP on the scaffold Daoli_0035 at the position 107,725 (LFMM Z = 6.1895, Q = 6.36e-102). The locus was 3,254 bp upstream of a predicted gene model Daoli32516 and 5,010 bp downstream of the gene Daoli32517, a homologue of tatC-like protein YMF16.

Differential Adaptation Related to Temperature and Precipitation

Isothermality (wc2.1_30s_bio_3) was identified as the most important overall driver of both neutral and adaptive genomic variation among nonspatial environmental variables in *D. cochinchinensis*, based on our gradient forest (GF) model (Fig. 3 and *SI Appendix*, Fig. S8*A*), in contrast to "precipitation of the wettest quarter" (wc2.1_30s_bio_16) in *D. oliveri* (Fig. 4 and *SI Appendix*, Fig. S8*B*). Spatial variables, as principal coordinates of a neighborhood matrix (PCNM), were the most important variables that explained both neutral and adaptive genomic variation, which was unsurprising given strong isolation by distance was known in these species (31) and environmental adaptation only affects a small portion of the genome (39). Soil factors were among the lowest ranked variables for gene–environment associations for both species. We observed different patterns of geographic variation in *D. cochinchinensis* and *D. oliveri* when fitting the GF models across their native ranges. *D. cochinchinensis* had strong differentiation between North and South populations at around 16°N, that was mainly driven by isothermality (wc2.1_30s_bio_3) as seen in the principal component analysis (PCA) loadings. On the other hand, *D. oliveri's* major differentiation was between coastal and inland areas, driven by both precipitation of the wettest quarter (wc2.1_30s_bio_16) and mean diurnal range (wc2.1_30s_bio_2). The eastern coastal areas in Vietnam showed particularly strong

Downloaded from https://www.pnas.org by COPENHAGEN UNIVERSITY LIBRARY on August 8, 2023 from IP address 130.226.236.6.

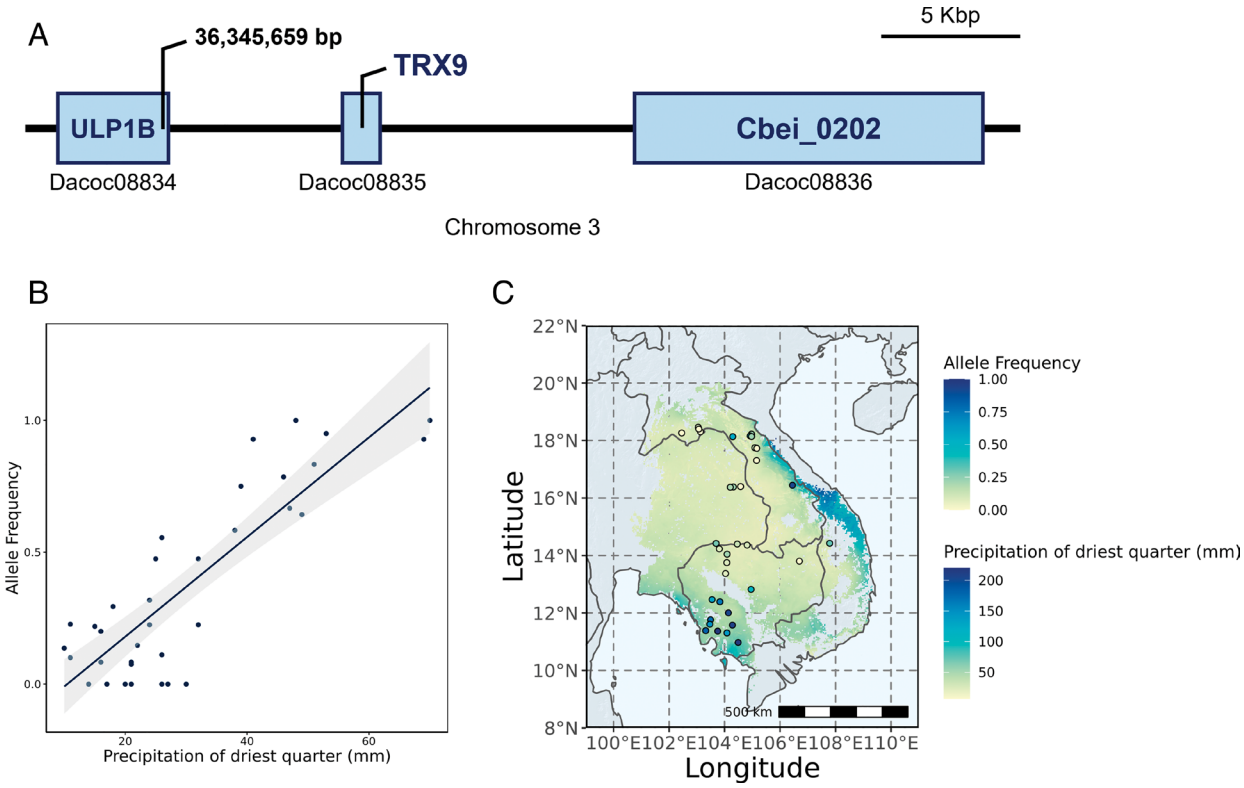

Fig. 2. (*A*) The most significant gene–environment association at 36,346,659 bp on chromosome 3, within the Dacoc08834 gene and upstream of Dacoc08835 and Dacoc08836 genes, which are homologues of ULP1B, TRX9, and Cbei_0202, respectively. (*B*) Correlation between allele frequency and wc2.1_30s_bio_17 (Precipitation of driest quarter) for this locus. In (*C*), the distribution is color-coded with the precipitation of driest quarter and the individual locality is color-coded with the allele frequency.

differences in environmental associations with adaptive variation and neutral variation for both *D. cochinchinensis* and *D. oliveri* (Fig. 5).

We compared the allelic frequency turnover functions of the neutral and adaptive genomic variation for each environmental predictor variable. Adaptive genomic variation was significantly more strongly associated with environmental gradients than neutral variation (*SI Appendix*, Fig. S9). There was only one exception, where available soil water capacity at a depth of 60 cm (s_AWCh1_sl5) was near-zero but of similar importance in explaining neutral and adaptive variation, regardless of the environmental gradient.

When exposed to drought stress under controlled conditions, *D. cochinchinensis* was more anisohydric than *D. oliveri*, which means that *D. cochinchinensis,* as a pioneering species with faster growth, optimizes carbon assimilation and better tolerates reduced water availability (40). *D. oliveri* is often found in moist areas and along streams and rivers (41), and the morphological characteristics of its seeds suggest that secondary dispersal by water is likely (31). This could explain how isothermality, which has been an important predictor in environmental niche models in tropical environments (42) and shown to influence plant height growth (43), had a dominant effect in the adaptive variation only in *D. cochinchinensis*. Pioneering species maximize height growth in early successional habitats to meet their light requirements (44), consistent with the observation of higher photosynthetic pigment levels in *D. cochinchinensis* (40). On the other hand, the effect of precipitation of the wettest quarter could act on selection in seed dispersal and survival in *D. oliveri* in the wet season. Temperature and precipitation, and their variability such as isothermality (45) have been widely reported as the most important drivers shaping patterns of productivity and adaptation in tree species across the world (46–48).

To fill the current gaps in existing conservation actions, populations that are underrepresented but display distinct adaptive variation should be prioritized to avoid the potential loss of unique genetic diversity. Populations at the edge of the species ranges should be prioritized based on our findings on adaptive variation showing their distinct allelic frequencies and adaptation; however, they are currently underrepresented in conservation efforts and existing protected area networks. Importantly, hotspots of differential adaptive variation near the edges of species ranges are shared between *D. cochinchinensis* and *D. oliveri*. This observation reinforces the role of marginal populations in preserving evolutionary potential for range expansion and persistence due to their adaptation to distinct environmental conditions (49).

Genomic Offset under Different Climate Change Scenarios

Genetic offset in the form of Euclidean distance represented the mismatch between current and future gene–environment association, which was modeled over five general circulation models (GCMs), namely MIROC6, BCC-CSM2-MR, IPSL-CM6A-LR, CNRM-ESM2-1, MRI-ESM2-0, under WCRP CMIP6 (*SI Appendix*, Fig. S10). For both *Dalbergia* species, genetic offset generally increased over time (P = 2.71e–10) and shared socioeconomic pathway (P = 4.54e–14), which are associated with increased carbon emission (Fig. 6*A* and

Downloaded from https://www.pnas.org by COPENHAGEN UNIVERSITY LIBRARY on August 8, 2023 from IP address 130.226.236.6.

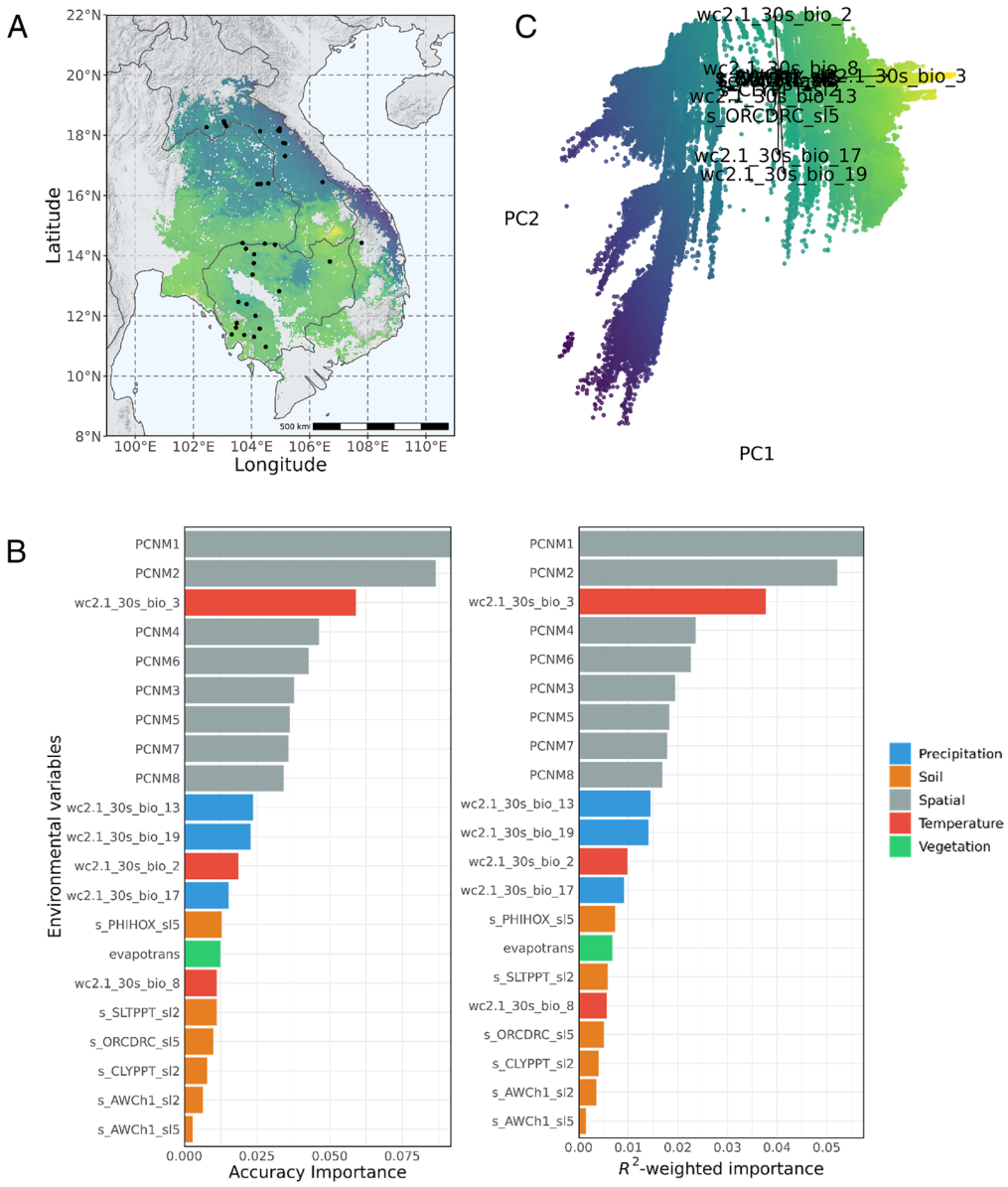

Fig. 3. (*A*) Adaptive genomic variation across the species range predicted by GF model for *D. cochinchinensis*, visualized using the first two principal axes from the PCA. (*B*) Accuracy and R^2-weighted importance for environmental predictor variables which explained adaptive genomic variation (adaptive SNPs) by the GF model. (*C*) PCA of the adaptive genomic variation predicted by the GF model across the species range. Loadings are the environmental factors.

SI Appendix, Table S9). However, *D. cochinchinensis* shows a significantly larger increase in genetic offset over time compared to *D. oliveri* (P = 0.025), suggesting that *D. cochinchinensis* is more susceptible to any mismatch of current genotypes and future climate. The geographic patterns of genetic offset also differed between the two species: *D. cochinchinensis* had an increasing offset across the entire range, while *D. oliveri* had a distinctly high offset in the southeast part of the range (Fig. 6 *B* and *C*). The variation in genomic offset between two species was mainly driven by the strong association with isothermality (wc2.1_30s_bio_3) in *D. cochinchinensis,* as demonstrated in the GF model, as it contributed to ~75% of the genomic offset on average (Fig. 6*D*). Isothermality had a smaller effect (~35%) in *D. oliveri* (Fig. 6*E*).

Our prediction contrasts with a separate sensitivity-and-exposure modeling study, which predicted that *D. oliveri* is likely to be slightly more vulnerable to climate change by 2055 (2041 to 2070 period) than *D. cochinchinensis* (11). It used growth rate and seed weight as proxy traits, predicting that both species have equally high sensitivity to climate change, but that *D. oliveri* is more exposed to the threat. Our findings predict that the dominant environment factor of isothermality could give more weight to the species' genomic offset. As discussed, isothermality is likely to affect the productivity and growth in pioneering species like *D. cochinchinensis* more than later successional species like *D. oliveri*. Our work supports that isothermality and other temperature variation factors will serve as more reliable indicators to predict the climate response of *D. cochinchinensis* and encourages further studies of this response, such as greenhouse or common garden experiments to validate the prediction with empirical data.

The different geographical patterns of genomic offset support species-specific recommendations in conservation and restoration. While climate change is likely to affect *D. cochinchinensis* evenly across its range, greater attention is needed on the representation of adaptive variation in germplasm collection and conservation units; sampling should target edge populations in particular as they show potential signals of local adaptation, where the environmental associations between adaptive and neutral variation are the greatest. By contrast, we recommend

Downloaded from https://www.pnas.org by COPENHAGEN UNIVERSITY LIBRARY on August 8, 2023 from IP address 130.226.236.6.

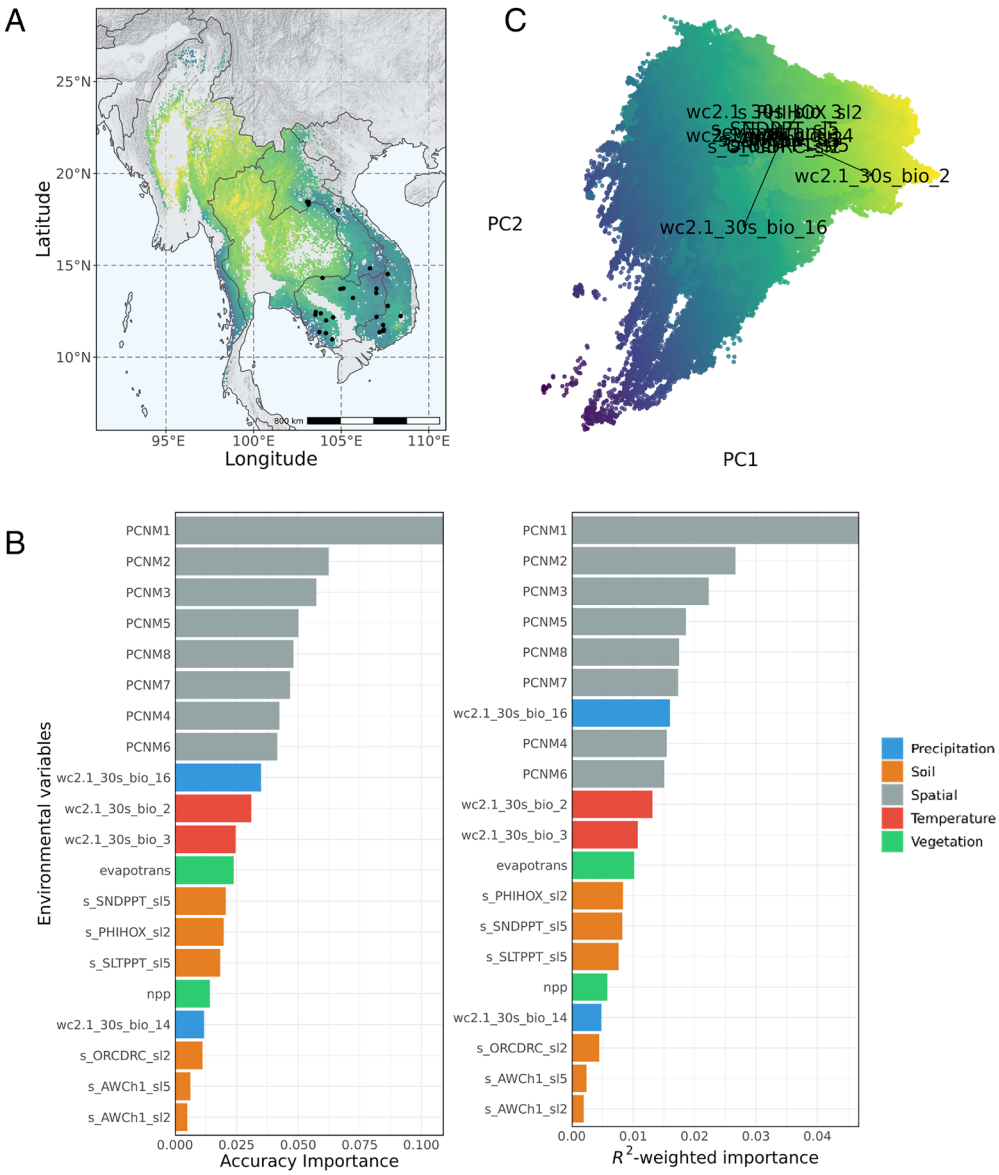

Fig. 4. (*A*) Adaptive genomic variation across the species range predicted by GF model for *D. oliveri*, visualized using the first two principal axes from the PCA. (*B*) Accuracy and R^2-weighted importance for the environmental predictor variables which explained the adaptive genomic variation (adaptive SNPs) by the GF model. (*C*) PCA of the adaptive genomic variation predicted by the GF model across the species range. The loadings are the environmental factors.

targeting hotspots of genomic offset in *D. oliveri*, especially around the borders between Cambodia, Laos, Vietnam, and Thailand, to improve conservation efforts.

In a rapidly changing environment, forest trees either persist through migration or phenotypic plasticity, or will extirpate (46) when environmental change outpaces adaptation potential. The spatially explicit model of genomic offset helps to develop conservation decisions balancing between in situ adaptation and assisted migration, as populations with lower genomic offset are likely to persist through adaptation (50).

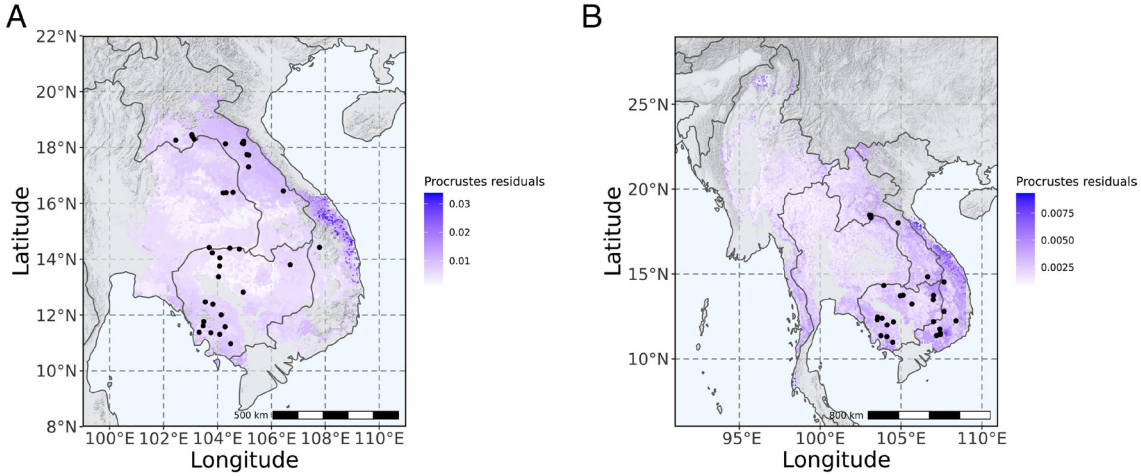

Fig. 5. Procrustes residuals between neutral and adaptive gene–environmental associations for (*A*) *D. cochinchinensis* and (*B*) *D. oliveri*.

Downloaded from https://www.pnas.org by COPENHAGEN UNIVERSITY LIBRARY on August 8, 2023 from IP address 130.226.236.6.

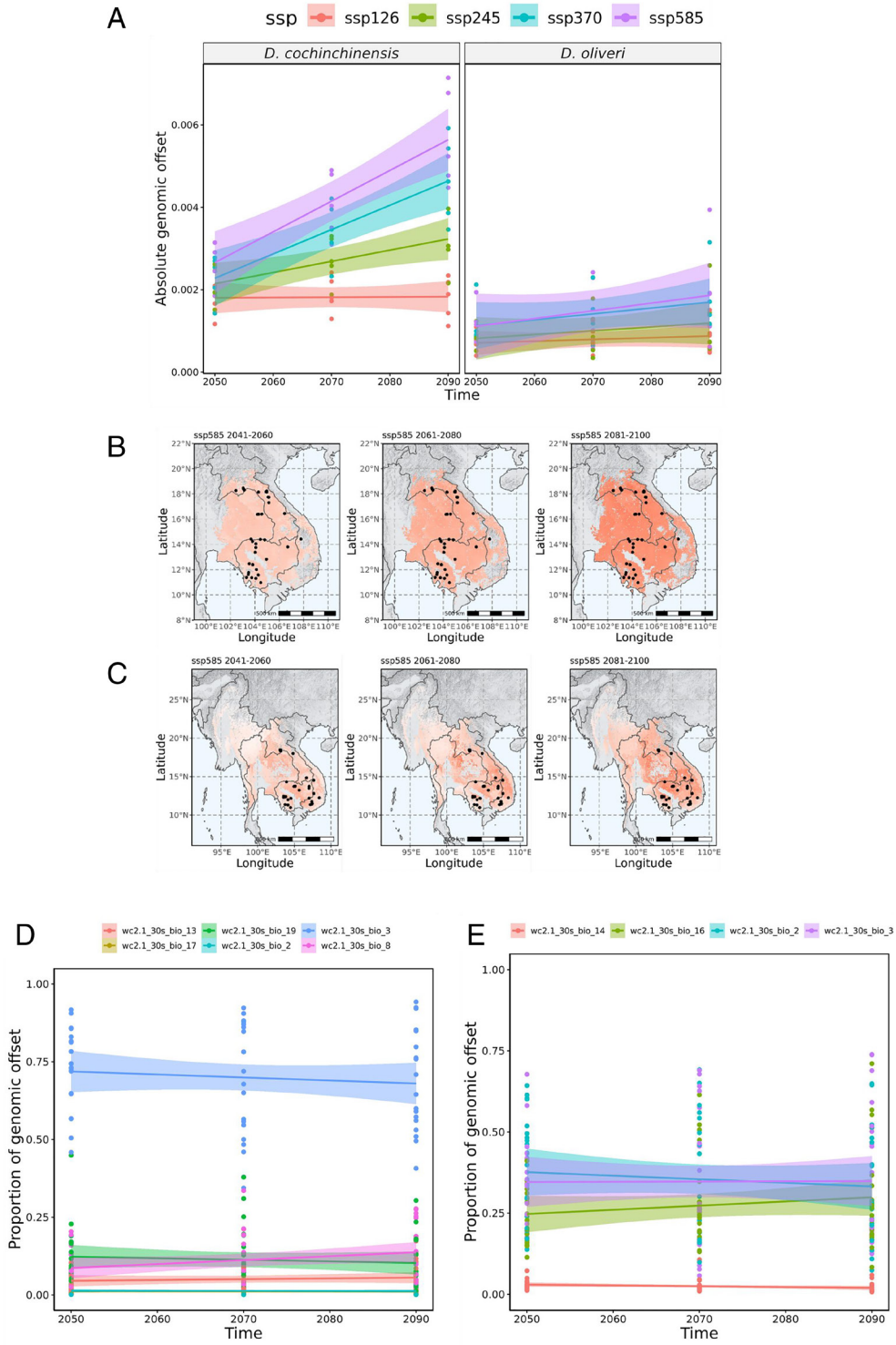

Fig. 6. (*A*) Absolute genomic offset of gene–environment association, quantified as the Euclidean distance, of *D. cochinchinensis* and *D. oliveri* in 4 SSPs (126, 245, 370, and 585) over three bidecades (2041 to 2060, 2061 to 2080, 2081 to 2100) averaged across five GCMs (BCC-CSM2-MR, CNRM-ESM2-1, IPSL-CM6A-LR, MIROC6, MRI-ESM2-0). Scaled genomic offset across the range of (*B*) *D. cochinchinensis* and (*C*) *D. oliveri*, using SSP585 between 2041 and 2060 as an example. Proportion of genomic variation explained by environmental variables in (*D*) *D. cochinchinensis* and (*E*) *D. oliveri*.

Genomic Model-Enabled Assisted Migration and Restoration

We developed *seedeR*, an open-source web application that is freely available from https://www.github.com/hung-th/seedeR, where users can input the species (*D. cochinchinensis* or *D. oliveri*), shared socioeconomic pathways (SSP), time period, and geographical coordinates of the target restoration or planting site. With these inputs, *seedeR* predicts the genomic similarity between a current germplasm source and target site from allelic frequency turnover functions and genetic offset and projects them onto the species range. We demonstrate the utility of *seedeR* for a hypothetical target 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 (Fig. 7). In both predictions, the genomic similarity was the highest at proximity to several hundreds of kilometers and decreased when further away. Commonly, coastal regions in northeast Vietnam, which were predicted to have the strongest local adaptation in both species, showed a lower genomic similarity. The geographical scale of suitable seed sources has an important implication as too many forest landscape projects collect seeds from very close (a few kilometers) to restoration sites to feed

Downloaded from https://www.pnas.org by COPENHAGEN UNIVERSITY LIBRARY on August 8, 2023 from IP address 130.226.236.6.

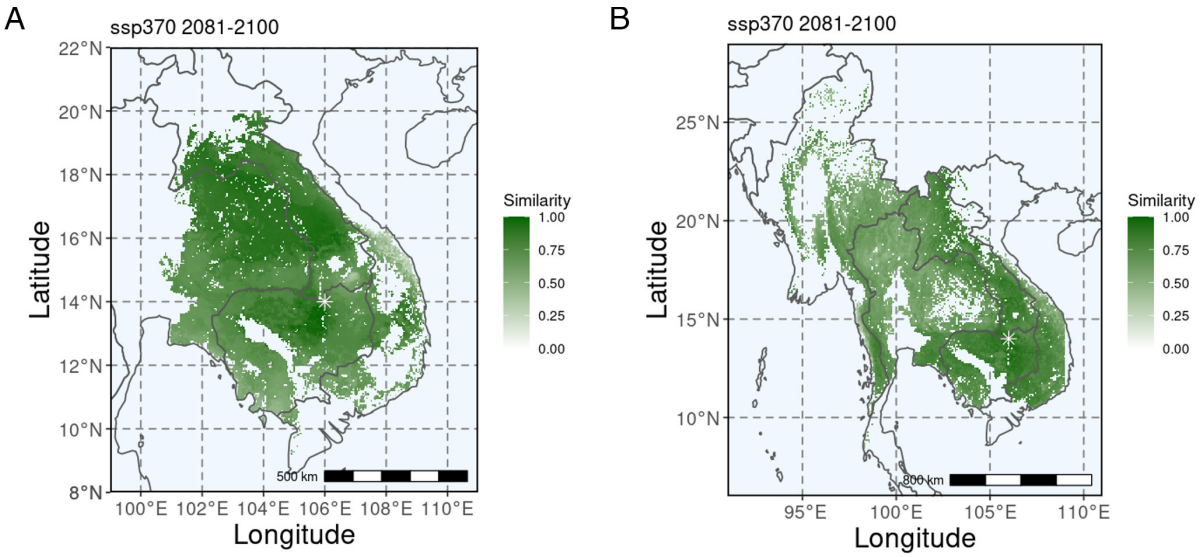

Fig. 7. Genomic similarity (scaled between 0, most dissimilar, and 1, most similar) between a hypothetical future restoration site (106°N, 14°E) and the current potential germplasm sources under the future climate scenario of SSP370 between 2081 and 2100 for (*A*) *D. cochinchinensis* and (*B*) *D. oliveri* predicted on seedeR (https://www.github.com/hung-th/seedeR).

the "local is best" paradigm (51), while our predictions showed otherwise. It is also important to note that local tree populations in landscapes in need of restoration are often degraded and have low genetic diversity. Genetic quality of seed should be ensured by collecting seed from large populations and many unrelated trees, even if this means collecting from trees at distances much further from the target restoration site.

Matching seed sources and restoration sites remains one of the keys for effective conservation and restoration (52), in line with the importance of adaptive variation and potential in genetic materials. Our genome-enabled prediction tool considers the future climate of restoration sites, which in turn will greatly influence the future resilience and productivity of these species. In the case of maladaptation and extirpation due to environmental change (53), when the classical preference for local provenance may no longer hold, deliberate transfer of germplasm along climate gradients may be necessary (54). Especially in the case of *Dalbergia*, when many local populations have extirpated or are very small in size, and large environmental association was predicted, assisted migration based on admixture and predictive provenancing are deemed more appropriate for the species to facilitate adaptation of the populations under climate change (55). Genetic materials from regions with strong adaptive genomic variation, such as coastal Vietnam, can be moved to suitable regions using the *seedeR* prediction to facilitate gene flow and maintain unique genetic components of the population by admixture (54). Hotspots of vulnerable populations such as those in northern Cambodia are suitable to be moved to new suitable areas to prevent loss of genetic diversity.

The *seedeR* application helps to visualize these spatially explicit predictive models of genomic offset and match, which are most useful to frontline practitioners and managers (56). Not only can it inform conservation and management strategies, but by simplifying the analytical pipelines through a user-friendly platform, it will also directly reduce the gap between conservation and genomics; a challenge faced for dissemination of genomic knowledge (57).

Narrowing the Gap between Conservation and Genomics

Our study characterizes range-wide gene–environment association in two sympatric endangered species, *D. cochinchinensis* and *D. oliveri,* for which there was virtually no prior knowledge on adaptability. Building on previous understanding of their different physiologies, we demonstrate their differential adaptive characteristics, which point to species-specific implications for their conservation. These findings on differential genomic adaptation between sympatric species sheds understanding on tropical forests, which harbor many threatened species at risk from threats associated with climate change (4).

We show how genomic technologies can directly support rapid decision-making and conservation activities. The separation between scientific and conservation communities represents a long-standing challenge, such that advances in scientific research and specifically genomic technologies are often inaccessible to the conservation side, which hinders translational science (57, 58). Through engagement with diverse stakeholders and conservation activities, we were strongly motivated to deliver the results of this study in a user-friendly (e.g., *seedeR*) and spatially explicit manner that can be integrated with ongoing conservation work.

Methods

Plant Materials and Sample Preparation for Genome Assemblies. Dried seeds of *D. cochinchinensis* and *D. oliveri* were collected from the Bolikhamxay, Khamkend, Laos, and Phnom Penh, Cambodia in 2018 by their forestry authorities, respectively. We germinated the seeds in a greenhouse at 30 °C with 16L/8D photoperiod. Leaf tissues were harvested from a selected 1-y-old individual for each species 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 Kingdom). DNA integrity was preliminary assessed with a 0.4% agarose gel against a NEB Quick-Load® 1 kb Extend DNA Ladder. A DNA sample passed the quality check only when a single band could be mapped near a lambda DNA band (~ 48.5 kb).

Genomic Sequencing and Assembly of *D. cochinchinensis*. For Oxford Nanopore 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. Three libarires were prepared each starting from 3 μg of size-selected DNA was used in each library preparation with the Oxford Nanopore

Downloaded from https://www.pnas.org by COPENHAGEN UNIVERSITY LIBRARY on August 8, 2023 from IP address 130.226.236.6.

Technologies Ligation Sequencing Kit (SQK-LSK110). The libraries were sequenced on two R10.3 (FLO-109D) flow cells on a GridION sequencer for ~ 72 h. Real-time basecalling was performed in MinKNOW release 19.10.1. Raw reads with Phred score lower than 8 were filtered.

For PacBio sequencing, DNA samples were sent to the Genomics & Cell Characterization Core Facility at the University of Oregon for DNA library preparation and sequencing. Throughout the sample preparation, the quality of DNA was assessed using Fragment Analyzer 1.2.0.11 (Agilent, United States). Then, 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 mode with a movie time of 30 h.

For Hi-C 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 min. The cross-linking was then quenched with glycine (125 mM). The cross-linked samples were 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.

Genomic Sequencing of *D. oliveri*. For Nanopore sequencing, the same protocol and procedure were used as for *D. cochinchinensis* (see above).

For Pore-C sequencing, the library was prepared with the protocol and reagents described by Belaghzal et al. (59) 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 min. The cross-linking was quenched with 125 mM glycine for 20 min and then the samples were 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 h. The enzyme was heat-denatured at 65 °C for 20 min at 300 rpm rotation in a thermomixer. 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 h. The flow cell was washed once every 24 h with the Flow Cell Wash Kit (EXP-WSH003).

Assembly Pipelines. Raw reads shorter than 500 bp were filtered. Due to the heterozygous nature of the wild individual, we assembled the sequences with Canu 2.1.1 (60) 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 (61) to collapse the assembly by separating the primary assembly and haplotigs.

Hi-C reads (for *D. cochinchinensis*) were mapped to the draft genome assembly using hicstuff 2.3.2 (62) to generate the contact matrix, which was then used to scaffold and polish the assembly using instaGRAAL 0.1.2 (63) with default options to produce the final assembly Dacoc_1.4 after removing contamination.

Pore-C reads (for *D. oliveri*) 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) and 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) and produce the final genome Daoli_0.3.

To validate the scaffold arrangement, Daoli_0.3 was aligned to that of *D. cochinchinensis* (Dacoc_1.4) using minimap2 and D-GENIES (64) to produce a dot plot for visualizing similarity, repetitions, breaks, and inversions, with a minimum identity of 0.25.

De Novo Repeat Library. A de novo repeat library was constructed using RepeatModeler 2.0.1 (65), which incorporated RECON 1.08 (66), RepeatScout 1.0.6 (67), and TRF 4.0.9 (68) for identification and classification of repeat families. We then used RepeatMasker 4.1.1 (69) 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 (70) and annotated with the GyDB database and profile HMMs using LTRdigest (71) 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 (72) ("-aS 80 -c 0.8 -g 1 -G 0 -A 80") to produce the final repeat library.

Gene Models and Annotation. Filtered mRNA-sequencing data for *D. cochinchinensis* (50.5 Gbp) and *D. oliveri* (54.4 Gbp) from a previous project (26) (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. Protein sequences were obtained from *A. thaliana* (Araport11) (73) and *A. ipaensis* (Araip1.1) (74). After soft-masking the genome with the de novo repeat library using RepeatMasker (Dfam libraries 3.2), the transcript and protein evidence were used to produce gene models using MAKER 3.01.03 (75). 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) (76) and AUGUSTUS 3.3.3 (77), which were used in the MAKER pipeline. tRNA genes were predicted with tRNAscan-SE 1.3.1 (78). The quality of the gene models was assessed with two metrics: the AED in MAKER 3.01.03 (75) and the BUSCO score (v5.1.2) (79).

Population Sampling. We obtained a collection of 435 and 331 foliage samples of *D. cochinchinensis* and *D. oliveri* from 35 and 28 localities across their native range (Dataset S2). These samples were a combination of those collected in a previous study (31) 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 prewash (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 (80). 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 the Qubit dsDNA BR Assay kit on Qubit 4 (Thermo, Wilmington, United States), respectively.

Genotyping-by-Sequencing (GbS). DNA samples were normalizsed to 200 ng suspended in 10 μL water and sent to the Genomic Analysis Platform, Institute of Integrative and Systems Biology, 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 an Illumina NovaSeq6000 S4 with paired-end reads of 150 bp at the Génome Québec Innovation Centre (Montreal, Canada).

Variant Calling. DNA sequence variant calling was done with the Fast-GBS v2.0 pipeline (81): Illumina raw reads were demultiplexed with Sabre 1.0 (82) and trimmed with Cutadapt 1.18 (83) to remove the adaptors. Trimmed reads shorter than 50 bp were discarded. Reads were aligned against the Dacoc_1.4 genome and the Daoli_0.3 genome using BWA-MEM 0.7.17 (84). The SAM alignment files were converted to BAM format and indexed using SAMtools 1.9 (85). Variant calling was performed in Platypus (86) and variants were filtered with proportion of missing data of 0.2 and MAF of 0.01 using VCFtools 0.1.16 (87). Missing genotypes were imputed using Beagle 5.2, which estimated the haplotype clusters based on the available data since a reference panel was not available. Finally, linkage equilibrium among SNPs was estimated using BCFtools 1.9 (85), and one SNP was removed from all SNP pairs with $r^2 > 0.5$ in a genomic window of 5 Kbp.

Environmental Heterogeneity Characterization. Environmental data were obtained from different sources (34 variables in total, *SI Appendix*, Table S10) and represented different measurers of temperature, precipitation, their seasonality, soil, elevation, and vegetation. We calculated a correlation matrix across the sampling localities and highly intercorrelated variables (pairwise correlation coefficient| > 0.7) were detected. For each intercorrelated variable pair,

Downloaded from https://www.pnas.org by COPENHAGEN UNIVERSITY LIBRARY on August 8, 2023 from IP address 130.226.236.6.

the one variable with the largest mean absolute correlation across all variables was removed.

Population Genetic Structure and Identification of Putatively Adaptive Loci. Population genetic structure was assessed with sNMF, which is more statistically robust to departures from population genetic model assumptions and computationally efficient than likelihood-based approaches such as STRUCTURE and ADMIXTURE (88), to estimate the number of discrete genetic clusters (K) (89). 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 (90), 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 modeling (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 (91), was calculated to calibrate for *P*-values:

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

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 (89).

GF Modeling. For all predictions in GF models, resampling was necessary because not all environmental raster layers had the same resolution and extent. They were all cropped to the latest-updated modeled and expert-validated species distribution (11) and reprojected to new rasters based on the resolution of the WorldClim bioclimatic rasters, using bilinear interpolation or nearest neighbor method for continuous and categorical variables respectively.

To correct for the genetic structure, spatial variables were generated using the PCNM approach (92). Only half of the positive PCNM values were kept. GF model (93, 94) was used to assess and rank the importance of environmental variables in genomic variation, as it has the ability to handle nonlinear and complex relationships between genetic and environmental variables (95) and noisy data (96). Putatively neutral 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 GF model for 500 regression trees. GF model implemented a conditional permutation approach (97) to correct for correlated environmental predictors, the maximum number of splits (*K*, not to be confused with the number of ancestral populations above) for the trees was determined as follows, to ensure at least two points per partition (see the appendix of ref. 93 for the discussion):

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

The turnovers of allelic frequencies were then projected spatially across the latest-updated predicted species distribution ranges (11) using the fitted GF model and the environmental values across the range. PCA was used to summarize the genomic variation across the distribution and the first three principal components (PC1, PC2, and PC3) were used for visualization of genomic variation across the range.

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

Prediction of Genomic Offset. Future climate projections were obtained from five 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 SSPs (126, 245, 370, and 585) over four 20-y periods (2021 to 2040, 2041 to 2060, 2061 to 2080, 2081 to 2100). The GF model was used to predict 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 (94, 95), 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 mix-max scaled across all SSPs and time periods to allow visual comparisons among different future environmental scenarios.

Prediction of Genomic Similarity between Current Germplasm Source and Future Restoration Site. It is of practical interest to a range of forestry stakeholders to predict if a current germplasm source is a good match for future restoration sites, or where to source suitable germplasm for a proposed 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://www.github.com/hung-th/seedeR. The analysis workflow consists of the selection of species of interest, time period and future climate scenario, and the restoration site's geographical coordinates (*SI Appendix*, Fig. S11).

The application maps the predicted turnover of allelic frequencies at a hypothetical future restoration site onto the current landscape on a grid-by-grid basis, with the genetic offset calculated as described above. After scaling, the values are reversed on a 0 to 1 scale to represent the genomic similarity between the current germplasm source and future restoration site.

Data, Materials, and Software Availability. The research materials supporting this publication, including genomic assemblies, raw reads, and annotations, can be publicly accessed either in *SI Appendix* or in NCBI GenBank under the BioProjects PRJNA841235 [Genome assembly of *D. cochinchinensis* (Dacoc_1.4)] (98), PRJNA841689 [Genome assembly of *D. oliveri* (Daoli_0.3)] (99), and PRJNA962334 (Genotyping by sequencing of *D. cochinchinensis* and *D. oliveri*) (100).

ACKNOWLEDGMENTS. The genomic work was supported by funding to T.H.H. from the Biotechnology and Biological Sciences Research Council (grant number BB/M011224/1) and to T.H.H., J.J.M. from the Google Cloud Academic Grant. The sampling work was supported by funding to T.S., B.T., I.T., R.J., D.H.B., J.J.M from the UK Darwin Initiative (ref. 25-023). The work of H.G. and R.J. was supported by the CGIAR Fund Donors (https://www.cgiar.org/funders/) through the CGIAR Research Programme on Forests, Trees and Agroforestry. T.H.H. wishes to thank Andrew Eckert and Stephen Harris as the examiners of his doctoral thesis, who provided very constructive feedback that improved this paper.

Author affiliations: [a]Department of Biology, University of Oxford, Oxford OX1 3RB, United Kingdom; [b]Institute of Forest and Wildlife Research and Development, Phnom Penh, Cambodia; [c]National Agriculture and Forestry Research Institute, Forestry Research Center, Vientiane, Laos; [d]The Forest Herbarium, Department of National Park, Wildlife and Plant Conservation, Ministry of Natural Resources and Environment, Bangkok 10900, Thailand; [e]Department of Food and Resource Economics, Faculty of Science, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark; [f]National Herbarium of Laos, Biotechnology and Ecology Institute, Ministry of Science and Technology, Vientiane, Laos; [g]Forest Genetics and Diversity, Department of Geosciences and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, 1958 Frederiksberg C, Denmark; [h]Center for Evolutionary Hologenomics, Globe Institute, University of Copenhagen, Øster Farimagsgade 5, 1353 Copenhagen K, Denmark; [i]Bioversity International, I-00057 Rome, Italy; [j]Department of Geoinformatics, Paris Lodron University, 5020 Salzburg, Austria; and [k]Bioversity International, 43400 UPM Serdang, Malaysia

Author contributions: T.H.H., D.H.B., and J.J.M. designed research; T.H.H., T.S., B.T., V.C., I.T., C.P., S.B., I.H., H.G., and R.J. performed research; T.H.H., T.S., B.T., V.C., I.T., C.P., S.B., I.H., H.G., and R.J. contributed new reagents/analytic tools; T.H.H. analyzed data; and T.H.H. wrote the paper.

Downloaded from https://www.pnas.org by COPENHAGEN UNIVERSITY LIBRARY on August 8, 2023 from IP address 130.226.236.6.

1. UNODC, *World Wildlife Crime Report: Trafficking in Protected Species* (United Nations Office on Drugs and Crime, 2020).
2. United Nations Environment Programme, *The Rise of Environmental Crime: A Growing Threat to Natural Resources Peace, Development and Security* (United Nations Environment Programme, 2016).
3. UNODC, *World Wildlife Crime Report: Trafficking in Protected Species* (United Nations Publication, 2016).
4. H. Gaisberger *et al.*, Tropical and subtropical Asia's valued tree species under threat. *Conserv. Biol.* **36**, e13873 (2022).
5. K. Winfield, M. Scott, C. Graysn, "Global status of Dalbergia and Pterocarpus rosewood producing species in trade" in *Convention on International Trade in Endangered Species 17th Conference of Parties* (Convention on International Trade in Endangered Species, Johannesburg, 2016).
6. Asian Regional Workshop (Conservation & Sustainable Management of Trees Viet Nam), Dalbergia cochinchinensis. The IUCN red list of threatened species (e.T32625A9719096, 1998), 10.2305/IUCN.UK.1998.RLTS.T32625A9719096.en.
7. N. H. Nghia, Dalbergia oliveri. The IUCN red list of threatened species 1998 (e.T32306A9693932, 1998), 10.2305/IUCN.UK.1998.RLTS.T32306A9693932.en.
8. 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).
9. M. Barstow *et al.*, Dalbergia cochinchinensis. The IUCN red list of threatened species 2022 (2022).
10. M. Barstow *et al.*, Dalbergia oliveri. The IUCN red list of threatened species 2022 (2022).
11. H. Gaisberger *et al.*, Range-wide priority setting for the conservation and restoration of Asian rosewood species accounting for multiple threats and ecogeographic diversity. *Biol. Conserv.* **270**, 109560 (2022).
12. N. Myers, R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, J. Kent, Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
13. D. S. Woodruff, 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).
14. C. M. Wurster *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).
15. M. Jansen *et al.*, Food for thought: The underutilized potential of tropical tree-sourced foods for 21st century sustainable food systems. *People Nat.* **2**, 1006–1020 (2020).
16. J. A. Oldekop *et al.*, Forest-linked livelihoods in a globalized world. *Nat. Plants* **6**, 1400–1407 (2020).
17. Centre for Forest, Landscape and Planning, D., Cambodia Tree Seed Project & Forestry Administration, C., Conservation of valuable and endangered tree species in Cambodia 2001–2006–A case study (Forest & Landscape: Development and Environment Series 3, 2006).
18. D. Version, *Conservation of Valuable and Endangered Tree Species in Cambodia 2001–2006 Moestrup, Søren; Sloth, Arvid; Burgess, Sarah* (2017).
19. E. V. Maningo, S. Thea, *Regional Project for Prootion of Forest Rehabilitation in Cambodia and Vietnam through Demonstration Models and Improvement of Seed Supply System: Lesson Learned*.
20. APFORGEN, *Conserving Rosewood Genetic Resources for Resilient Livelihoods in the Mekong–Project Inception Workshop Report* (2018).
21. R. Frankham *et al.*, "Loss of genetic diversity reduces ability to adapt" in *Genetic Management of Fragmented Animal and Plant Populations*, R. Frankham *et al.*, Eds. (Oxford University Press, 2017), 10.1093/OSO/9780198783398.003.0004.
22. O. Savolainen, M. Lascoux, J. Merilä, Ecological genomics of local adaptation. *Nat. Rev. Genet.* **14**, 807–820 (2013).
23. P. J. Verkerk *et al.*, Climate-smart forestry: The missing link. *For. Policy Econ.* **115**, 102164 (2020).
24. C. Petit-Cailleux *et al.*, Tree mortality risks under climate change in Europe: Assessment of silviculture practices and genetic conservation networks. *Front. Ecol. Evol.* **9**, 582 (2021).
25. M. Lindner *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).
26. T. H. Hung *et al.*, Reference transcriptomes and comparative analyses of six species in the threatened rosewood genus Dalbergia. *Sci. Rep.* **10**, 17749 (2020).
27. M. A. Supple, B. Shapiro, Conservation of biodiversity in the genomics era. *Genome Biol.* **19**, 131 (2018).
28. F. W. Allendorf, P. A. Hohenlohe, G. Luikart, Genomics and the future of conservation genetics. *Nat. Rev. Genet.* **11**, 697–709 (2010).
29. R. Desalle, G. Amato, Conservation genetics, precision conservation, and de-extinction. *Hastings Center Rep.* **47**, S18–S23 (2017).
30. X. Luo, S. Chen, Y. Zhang, PlantRep: A database of plant repetitive elements. *Plant Cell Rep.* **41**, 1163–1166 (2022).
31. I. Hartvig *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).
32. I. Hartvig *et al.*, Conservation genetics of the critically endangered Siamese rosewood (*Dalbergia cochinchinensis*): Recommendations for management and sustainable use. *Conserv. Genet.* **21**, 677–692 (2020), 10.1007/s10592-020-01279-1.
33. D. R. Hoen *et al.*, Transposon-mediated expansion and diversification of a family of ULP-like genes. *Mol. Biol. Evol.* **23**, 1254–1268 (2006).
34. D. Roy, A. Sadanandom, SUMO mediated regulation of transcription factors as a mechanism for transducing environmental cues into cellular signaling in plants. *Cell. Mol. Life Sci.* **78**, 2641–2664 (2021).
35. J. Lee *et al.*, Salicylic acid-mediated innate immunity in Arabidopsis is regulated by SIZ1 SUMO E3 ligase. *Plant J.* **49**, 79–90 (2007).
36. R. Catala *et al.*, The Arabidopsis E3 SUMO ligase SIZ1 regulates plant growth and drought responses. *Plant Cell* **19**, 2952–2966 (2007).
37. C. Y. Yoo *et al.*, SIZ1 small ubiquitin-like modifier E3 ligase facilitates basal thermotolerance in Arabidopsis independent of salicylic acid. *Plant. Physiol.* **142**, 1548–1558 (2006).
38. J. B. Jin *et al.*, The SUMO E3 ligase, AtSIZ1, regulates flowering by controlling a salicylic acid-mediated floral promotion pathway and through affects on FLC chromatin structure. *Plant J.* **53**, 530–540 (2008).
39. R. A. Bay *et al.*, Predicting responses to contemporary environmental change using evolutionary response architectures. *Am. Nat.* **189**, 463–473 (2017).
40. T. H. Hung *et al.*, Physiological responses of rosewoods Dalbergia cochinchinensis and D. oliveri under drought and heat stresses. *Ecol. Evol.* **10**, 10872–10885 (2020), 10.1002/ece3.6744.
41. R. Aerts *et al.*, Site requirements of the endangered rosewood Dalbergiaoliveri in a tropical deciduous forest in northern Thailand. *For. Ecol. Manage.* **259**, 117–123 (2009).
42. H. A. Nix, "A biogeographic analysis of Australian elapid snakes" in *Atlas of elapid snakes of Australia: Canberra, Australian Flora and Fauna Series 7*, R. Longmore, Ed. (Australian Government Publishing Service, 1986), pp. 4–15.
43. A. T. Moles *et al.*, Global patterns in plant height. *J. Ecol.* **97**, 923–932 (2009).
44. D. Woodcock, A. Shier, Wood specific gravity and its radial variations: The many ways to make a tree. *Trees* **16**, 437–443 (2002).
45. P. H. Garnier-Géré, P. K. Ades, Environmental surrogates for predicting and conserving adaptive genetic variability in tree species. *Conserv. Biol.* **15**, 1632–1644 (2001).
46. S. N. Aitken, S. Yeaman, J. A. Holliday, T. Wang, S. Curtis-McLane, Adaptation, migration or extirpation: Climate change outcomes for tree populations. *Evol. Appl.* **1**, 95–111 (2008).
47. M. A. Supple *et al.*, Landscape genomic prediction for restoration of a Eucalyptus foundation species under climate change. *Elife* **7**, e31835 (2018).
48. S. Manel *et al.*, Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Mol. Ecol.* **21**, 3729–3738 (2012).
49. J. B. Ledoux *et al.*, Potential for adaptive evolution at species range margins: Contrasting interactions between red coral populations and their environment in a changing ocean. *Ecol. Evol.* **5**, 1178 (2015).
50. A. V. Gougherty, S. R. Keller, M. C. Fitzpatrick, Maladaptation, migration and extirpation fuel climate change risk in a forest tree species. *Nat. Clim. Change* **11**, 166–171 (2021).
51. R. Jalonen, M. Valette, D. Boshier, J. Duminil, E. Thomas, 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).
52. T. Fremout *et al.*, Diversity for Restoration (D4R): Guiding the selection of tree species and seed sources for climate-resilient restoration of tropical forest landscapes. *J. Appl. Ecol.* **59**, 664–679 (2022).
53. S. N. Aitken, M. C. Whitlock, Assisted gene flow to facilitate local adaptation to climate change. *Annu. Rev. Ecol. Evol. Syst.* **44**, 367–388 (2013).
54. M. Bozzano *et al.*, *Genetic Considerations in Ecosystem Restoration Using Native Tree Species* (FAO and Bioversity International, 2014).
55. M. F. Breed, M. G. Stead, K. M. Ottewell, M. G. Gardner, A. J. Lowe, Which provenance and where? Seed sourcing strategies for revegetation in a changing environment *Conserv. Genet.* **14**, 1–10 (2013).
56. K. Martins *et al.*, Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of *Quercus rugosa*. *Evol. Appl.* **11**, 1842–1858 (2018).
57. A. B. A. Shafer *et al.*, Genomics and the challenging translation into conservation practice. *Trends Ecol. Evol.* **30**, 78–87 (2015).
58. R. Taylor, H. Dussex, N. Y. van Heezik, Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Glob. Ecol. Conserv.* **10**, 231–242 (2017).
59. H. Belaghzal, J. Dekker, J. H. Gibcus, Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* **123**, 56 (2017).
60. K. Sergey *et al.*, Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
61. M. J. Roach, S. A. Schmidt, A. R. Borneman, Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics.* **19**, 460 (2018).
62. C. Matthey-Doret *et al.*, koszullab/hicstuff: Use miniconda layer for docker and improved P(s) normalisation (2020), 10.5281/ZENODO.4066363.
63. L. Baudry *et al.*, InstaGRAAL: Chromosome-level quality scaffolding of genomes using a proximity ligation-based scaffolder. *Genome Biol.* **21**, 1–22 (2020).
64. F. Cabanettes, C. Klopp, D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958 (2018).
65. J. M. Flynn *et al.*, RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9451–9457 (2020).
66. Z. Bao, S. R. Eddy, Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* **12**, 1269 (2002).
67. A. L. Price, N. C. Jones, P. A. Pevzner, De novo identification of repeat families in large genomes. *Bioinformatics* **21** (suppl. 1), i351–i358 (2005).
68. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
69. M. Tarailo-Graovac, N. Chen, Using repeatmasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **25**, 4.10.1–4.10.14 (2009).
70. D. Ellinghaus, S. Kurtz, U. Willhoeft, LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 1–14 (2008).
71. S. Steinbiss, U. Willhoeft, G. Gremme, S. Kurtz, Fine-grained annotation and classification of de novo predicted LTR retrotransposons. *Nucleic Acids Res.* **37**, 7002–7013 (2009).
72. W. Li, A. Godzik, Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
73. C.-Y. Cheng *et al.*, Araport11: A complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* **89**, 789–804 (2017).
74. D. J. Bertioli *et al.*, The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. *Nat. Genet.* **48**, 438–446 (2016).
75. C. Holt, M. Yandell, MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
76. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 1–9 (2004).
77. M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
78. P. P. Chan, T. M. Lowe, tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* **1962**, 1 (2019).
79. M. Manni, M. R. Berkeley, M. Seppey, F. A. Simão, E. M. Zdobnov, 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).
80. P. W. Inglis, M. de Castro R Pappas, L. V. Resende, D. Grattapaglia, 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).

Downloaded from https://www.pnas.org by COPENHAGEN UNIVERSITY LIBRARY on August 8, 2023 from IP address 130.226.236.6.

81. D. Torkamaneh, J. Laroche, F. Belzile, Fast-GBS v2.0: An analysis toolkit for genotyping-by-sequencing data. *Genome* **63**, 577–581 (2020), 10.1139/gen-2020-0077.
82. N. A. Joshi, sabre–A barcode demultiplexing and trimming tool for FastQ files. Preprint (2013). https://github.com/najoshi/sabre (Accessed 29 January 2023).
83. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
84. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [Preprint] (2013). https://doi.org/10.48550/arXiv.1303.3997 (Accessed 29 January 2023).
85. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, 1–4 (2021).
86. A. Rimmer *et al.*, Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **46**, 912–918 (2014).
87. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
88. E. Frichot, F. Mathieu, T. Trouillon, G. Bouchard, O. François, Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196**, 973–983 (2014).
89. E. Frichot, O. François, LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929 (2015).
90. H. Martins, K. Caye, K. Luu, M. G. B. Blum, O. François, Identifying outlier loci in admixed and in continuous populations using ancestral population differentiation statistics. *Mol. Ecol.* **25**, 5029–5042 (2016).
91. J. Yang *et al.*, Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807 (2011).
92. D. Borcard, P. Legendre, All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol. Model.* **153**, 51–68 (2002).
93. N. Ellis, S. J. Smith, C. R. Pitcher, Gradient forests: Calculating importance gradients on physical predictors. *Ecology* **93**, 156–168 (2012).
94. M. C. Fitzpatrick, S. R. Keller, Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecol. Lett.* **18**, 1–16 (2015).
95. C. Rellstab, B. Dauphin, M. Exposito-Alonso, Prospects and limitations of genomic offset in conservation management. *Evol. Appl.* **14**, 1202–1212 (2021).
96. R. Genuer, J. M. Poggi, C. Tuleau-Malot, Variable selection using random forests. *Pattern. Recognit. Lett.* **31**, 2225–2236 (2010).
97. C. Strobl, A. L. Boulesteix, T. Kneib, T. Augustin, A. Zeileis, Conditional variable importance for random forests. *BMC Bioinformatics* **9**, 1–11 (2008).
98. T. H. Hung, D. H. Boshier, J. MacKay, Genome assembly of *D. cochinchinensis* (Oxon_Dacoc_1.4). NCBI BioProject. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA841235. Accessed 21 May 2022.
99. T. H. Hung, D. H. Boshier, J. MacKay, Genome assembly of *D. oliveri* (Daoli_0.3). NCBI BioProject. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA841689. Accessed 12 July 2023.
100. T. H. Hung, D. H. Boshier, J. MacKay, Range-wide genotyping by sequencing of *Dalbergia cochinchinensis* and *Dalbergia oliveri*. NCBI BioProject. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA962334. Accessed 27 April 2023.

Downloaded from https://www.pnas.org by COPENHAGEN UNIVERSITY LIBRARY on August 8, 2023 from IP address 130.226.236.6.